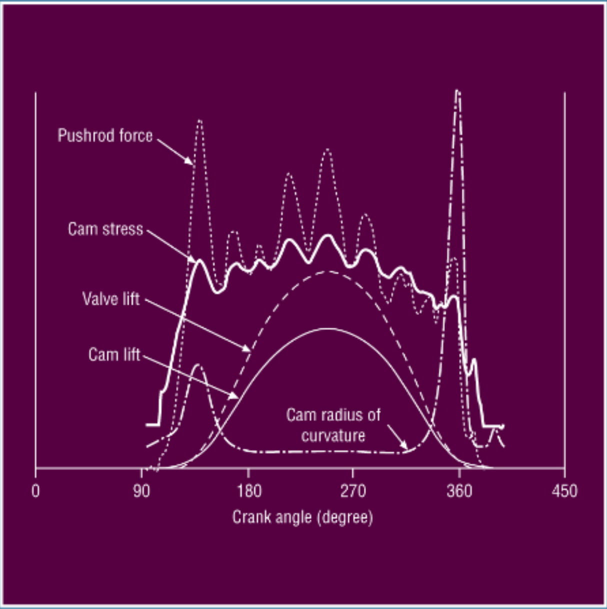
WOODHEAD PUBLISHING IN MECHANICAL ENGINEERING

Diesel engine system design

Qianfan Xin



WOODHEAD PUBLISHING

Diesel engine system design

Related titles:

Advanced direct injection combustion engine technologies and development Volume 1: Gasoline and gas engines
(ISBN 978-1-84569-389-3)
Direct injection enables precise control of the fuel/air mixture so that engines can be tuned for improved power and fuel economy, but ongoing research challenges remain in improving the technology for commercial applications. As fuel prices escalate, DI engines are expected to gain in popularity for automotive applications. This important book, in two volumes, reviews the science and technology of different types of DI combustion engines and their fuels. Volume 1 deals with direct injection gasoline and CNG engines, including history and essential principles, approaches to improved fuel economy, design, optimisation, optical techniques and their application.

Advanced direct injection combustion engine technologies and development Volume 2: Diesel engines
(ISBN 978-1-84569-744-0)
Volume 2 of the two-volume set *Advanced direct injection combustion engine technologies and development* investigates diesel DI combustion engines which, despite their commercial success, are facing ever more stringent emission legislation worldwide. Direct injection diesel engines are generally more efficient and cleaner than indirect injection engines and, as fuel prices continue to rise, DI engines are expected to gain in popularity for automotive applications. Two exclusive sections examine light-duty and heavy-duty diesel engines. Fuel injection systems and after treatment systems for DI diesel engines are discussed. The final section addresses exhaust emission control strategies, including combustion diagnostics and modelling, drawing on reputable diesel combustion system research and development.

Tribology and dynamics of engine and powertrain: Fundamentals, applications and future trends
(ISBN 978-1-84569-361-9)
Tribology is one element of many interacting within a vehicle engine and powertrain. In adopting a detailed, theoretical, component approach to solving tribological problems, the minutiae can be overwhelmingly complex and practical solutions become elusive and uneconomic. The system perspective generally adopted in industry, however, can lead to shortcuts and oversimplifications, industrial projects are subject to *ad hoc* trial and error, and subsequent 'fire-fighting' activity is required. This book seeks to bridge this divide, using a multi-physics approach to provide sufficient fundamental grounding and understanding of both detailed and approximate analyses – thereby making 'first time right' design solutions possible. Tribological issues and solutions in piston systems, valve train systems, engine bearings and drivetrain systems are addressed. New developments in materials, micro-engineering, nano-technology and MEMS are also included.

Details of these and other Woodhead Publishing books can be obtained by:

- visiting our web site at www.woodheadpublishing.com
- contacting Customer Services (e-mail: sales@woodheadpublishing.com; fax: +44 (0) 1223 832819; tel.: +44 (0) 1223 499140 ext. 130; address: Woodhead Publishing Limited, 80 High Street, Sawston, Cambridge CB22 3HJ, UK)

If you would like to receive information on forthcoming titles, please send your address details to: Francis Dodds (address, tel. and fax as above; e-mail: francis.dodds@woodheadpublishing.com). Please confirm which subject areas you are interested in.

Diesel engine system design

Qianfan Xin

WOODHEAD PUBLISHING

Oxford Cambridge Philadelphia New Delhi

Published by Woodhead Publishing Limited,
80 High Street, Sawston, Cambridge CB22 3HJ, UK
www.woodheadpublishing.com

Woodhead Publishing, 1518 Walnut Street, Suite 1100, Philadelphia, PA 19102-3406, USA

Woodhead Publishing India Private Limited, G-2, Vardaan House, 7/28 Ansari Road,
Daryaganj, New Delhi – 110002, India
www.woodheadpublishingindia.com

First published 2011, Woodhead Publishing Limited

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

ISBN 978-1-84569-715-0 (print)
ISBN 978-0-85709-083-6 (online)

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp which is processed using acid-free and elemental chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

Typeset by Replika Press Pvt Ltd, India
Printed by TJI Digital, Padstow, Cornwall, UK

Contents

Part IV Heat rejection, air system, engine controls, and system integration in diesel engine system design

Nomenclature

A	area
A_c	contact area
A_{EGR}	the theoretical effective flow area of the EGR valve opening
A_{ex}	the instantaneous flow area of the engine exhaust valve
A_h	heat transfer area
A_{in}	the instantaneous flow area of the engine intake valve
A_s	the availability or exergy in the second law of thermodynamics
A_T	turbine effective cross-sectional area
A_V	vehicle frontal area
$A_{VAL,cyl}$	the engine valve head area exposed to the cylinder side
$A_{VAL\ eff}$	engine valve effective flow area
$A_{VAL,port}$	the engine valve back area exposed to the port side
a	acceleration or deceleration
$\dot{a}$	jerk
a_{amp}	acceleration vibration amplitude
a_{CAM}	cam acceleration
a_e	the long half-axis length of the elliptic contact area
a_V	vehicle acceleration
a_{VACL}	vibration acceleration level in dB
a_{VAL}	valve acceleration
B	the percentage of fuel energy lost to the base engine coolant heat rejection
B_E	engine cylinder bore diameter
b	width
b_e	half width of Hertzian elliptic contact area
C, C_i	coefficient (i = 1, 2, 3, …), different in each formula
C_c	cost
C_{cav}	the cavitation factor accounting for the effect of cavitation or oil film rupture on friction reduction
C_{cf}	the cost to functionality if tolerance is exceeded
C_{cl}	confidence level

C_{CMD}	Taguchi's mean square deviation
C_d	the coefficient of flow restriction of an air system valve or a device/system
$C_{d,EGR}$	EGR valve flow restriction coefficient, or the flow restriction coefficient of the entire EGR circuit (when especially mentioned)
$C_{d,exh}$	total exhaust flow restriction coefficient, including all the components downstream from the turbine outlet
$C_{d,int}$	total intake flow restriction coefficient, including all the components upstream to the compressor inlet
$C_{d,IT}$	intake throttle valve flow restriction coefficient, or the coefficient of flow restriction of both the intake throttle valve and the CAC (when especially mentioned)
C_E	engine cylinder-to-cylinder centerline distance
C_f	the coefficient of valve and port flow discharge
C_{ij}, $C_{i,j}$	coefficient (i = 1, 2, 3, …; j = 1, 2, 3, …), different in each formula
C_q	quality loss coefficient
C_s	sound coefficient
C_{sr}	the speed ratio of torque converter
C_{T0}	the theoretical gas flow velocity in the turbine under the isentropic condition
C_{tr}	the torque ratio of torque converter
c	clearance or lash
c_B	bearing radial clearance
c_E	engine capacity factor
c_P	piston skirt clearance
c_p	constant-pressure specific heat
c_{SC}	the clearance of the spring between coils
c_{TV}	the input capacity factor of torque converter
c_{VT}	valvetrain lash
c_v	constant-volume specific heat
D	damage
D_s	destruction in the second law of thermodynamics
d	diameter
d_{SC}	spring coil diameter
d_{SP}	spring mean diameter
d_T	turbine wheel average diameter
d_{VAL}	valve diameter
$d_{VAL,ref}$	valve reference diameter
E	energy
E_a	activation energy
E_k	kinetic energy

$E_{k,slap}$	the kinetic energy of piston slap
$E_{k,V}$	vehicle kinetic energy
E_p	potential energy
$E_{p,V}$	vehicle potential energy
$\dot{E}$	energy rate or power
e	eccentricity or offset
e_1	the lateral distance from cylinder bore centerline to crankshaft axis (positive value means offset toward the anti-thrust side of the piston)
e_2	the lateral distance from piston centerline to piston pin (positive value means offset toward the anti-thrust side of the piston)
F	force
$F(\ldots)$	the constrained single-objective function in multi-objective optimization
F_a	vehicle aerodynamic drag resistance force
F_{acc}	the resistance force acting on the vehicle wheels caused by vehicle accessory loads
F_{amp}	force vibration amplitude
F_{br}	the force of the service brakes (wheel brakes) acting on the vehicle wheels
F_{df}	the drivetrain friction force acting on the vehicle wheels
F_{dr}	the resistance force of drivetrain retarders acting on the vehicle
F_{er}	the engine brake retarding force acting on the vehicle wheels
F_f	friction force
$F_{f,stem}$	valve stem friction force
$F_{f,v}$	viscous friction force
F_{gas}	gas loading (force)
$F_{gas,VAL}$	the net gas loading acting on the valve
F_{gl}	the gravity force on a gradient acting on the vehicle wheels along the longitudinal direction (i.e., along the road direction)
F_i	the vehicle inertia force along the longitudinal direction
$F_{lub,an}$	the lubricant force acting on the anti-thrust side of the piston skirt in the normal direction
$F_{lub,th}$	the lubricant force acting on the thrust side of the piston skirt in the normal direction
F_n	the force or load acting in the normal direction
F_{pre}	valve spring preload force
F_{rf}	vehicle tire–road rolling friction resistance force
F_{rt}	piston ring tension
F_{SP}	spring force (load)
F_t	the vehicle tractive force from engine firing acting on the wheels

$\widetilde{F}_{groove}$	the lateral friction force between the ring and the ring groove per unit length of the piston ring
$\widetilde{F}_{lub}$	the lubricating oil film force per unit length of the component
$\widetilde{F}_{lub,ring}$	the lubricating oil film force per unit length of the piston ring
$\widetilde{F}_{n}$	the normal loading force per unit length
$\widetilde{F}_{tension}$	the tension force per unit length of the piston ring
$f(\ldots)$	function or objective function
f^R	a percentile of the function $f(...)$
f_a	the coefficient of aerodynamic resistance
f_{adp}	the coefficient of adaptability
f_{adh}	the coefficient of road adhesion
$f_{A/F}$	engine air–fuel ratio
$f_{A/F,stoi}$	stoichiometric air–fuel ratio
f_b	the coefficient of engine torque backup
$f_{C\text{-}C}$	the ratio of connecting rod length to crank radius (i.e., conrod–crank ratio)
$f_{CAM,ac}$	the coefficient of cam acceleration pulse width
$f_{CAM,lub}$	the coefficient of cam lubrication characteristics
f_{CDF}	cumulative distribution function
f_{cm}	cooling capability multiplier for flow rate
f_{cor}	the coefficient of restitution
f_{cov}	the coefficient of variation
$f_{c\varepsilon}$	cooling capability multiplier for effectiveness
f_{dd}	the dynamic deflection factor of the valvetrain
f_{EGR}	EGR rate
f_f	feasibility of increasing a component's reliability
f_{fail}	failure rate
f_{fri}	the coefficient of friction
f_{hu}	humidity
f_{iEGR}	internal residue fraction or internal EGR rate
f_{lift}	the coefficient of the roundness of the valve lift
f_{NTU}	the number of transfer units in heat exchangers
f_n	natural frequency
$f_{n,SP}$	valve spring natural frequency (as distributed mass)
f_{O_2}	oxygen mass fraction
$f_{O_2,air}$	the mass fraction of O_2 (oxygen) in the ambient air
$f_{O_2,IM}$	the mass fraction of O_2 (oxygen) in the intake manifold charge
f_{ova}	the degree of ovality of the piston skirt
f_{PDF}	probability density function
f_{pr}	pressure ratio

f_{ql}	cyclic loading frequency
f_q	frequency
f_r	fraction
f_{RA}	rocker arm ratio
f_{rf}	the coefficient of rolling friction resistance
f_{ri}	the coefficient of rated intensity
f_{rpr}	the rate of cylinder pressure rise
$f_{RMSE\%}$	the percentage of root-mean-square error
f_{sd}	the static deflection factor of the valvetrain
f_{sf}	safety factor
f_{slip}	tire slip
f_{sr}	the coefficient of speed reserve
$f_{S/N}$	signal-to-noise ratio
f_{swr}	swirl ratio
$f_v\,(...)$	the variance of objective function $f\,(...)$
f_w	wear coefficient
f_{Wahl}	the Wahl stress correction factor for springs
f_x	mole fraction
G	the percentage of fuel energy lost to miscellaneous heat losses
G_G	Gibbs free enthalpy
G_r	road grade
g	the acceleration due to gravity
$g\,(\ldots)$	constraint function in optimization
H	enthalpy
$\dot{H}_{exh,firing}$	the exhaust enthalpy rate in the firing operation
$\dot{H}_{exh,retarding}$	the exhaust enthalpy rate in the retarding operation
h	specific enthalpy, or in-cylinder gas specific enthalpy
h_o	lubricating oil film thickness
h_r	hazard function in reliability engineering
h_w	the penetration hardness of surface in wear
$h1, h2, h3, h4$	the clearance or oil film thickness at the four corners of the piston skirt in the piston thrust plane
I	moment of inertia
I_{drive}	vehicle driveline moment of inertia
I_E	engine moment of inertia
I_{irr}	irreversibility in the second law of thermodynamics
I_R	reliability importance
I_{RA}	the moment of inertia of the rocker arm
I_s	sound intensity
I_{SIL}	sound intensity level in dB
I_{TC}	the moment of inertia of the turbocharger
i	index number

i_{ax}	axle ratio
i_F	the indicator of front geartrain or rear geartrain
i_{gr}	transmission gear ratio
i_{grn}	transmission gear number
J	torque
J_E	engine brake torque
$J_{E,r}$	the engine retarding brake torque acting on the crankshaft
J_f	friction torque
j	index number
K	reaction constant, or other constant
K_h	overall heat transfer coefficient
K_s	the stiffness of a component or physical system
$K_{s,SP}$	spring rate (stiffness)
$K_{s,VAL}$	valve or valve stem stiffness
$K_{s,VT}$	overall valvetrain stiffness
k	the total number of DoE factors
k_c	thermal conductivity
k_r	reaction rate
k_v	viscous friction coefficient
L	length (of pipe or component, etc.), or the width of the component in the direction of motion
L_B	bearing length
$L_{CG,P}$	the vertical distance from the piston center of gravity to the top of the piston skirt (positive value means the piston center of gravity is located below the top of the piston skirt, and a larger positive value means a lower position of the piston center of gravity)
L_{clear}	the clearance height in the cylinder and the combustion chamber
L_q	quality loss
l	displacement, travel distance, or lift
l_a	altitude level
$l_{a,V}$	the altitude level for the vehicle
l_{CAM}	cam lift
$l_{CAM,ramp}$	cam ramp height as a function of cam angle
l_{RA}	rocker arm length
$l_{r,V}$	vehicle braking distance
l_{SP}	spring length
l_V	vehicle travel distance
l_{VAL}	valve lift
$l_{VAL,design}$	the valve lift at the "design speed" of the valvetrain
$l_{VAL,max}$	maximum valve lift
l_{VDL}	vibration displacement level in dB

$l_{w,sl}$	sliding distance in wear
M	moment
M_a	Mach number
m	mass, or in-cylinder gas mass, or the total number of index
$m\ (\ldots)$	equality constraint functions in optimization
$m_{E,mov}$	the engine assembly mass of the moving parts
m_{fuel}	the consumed fuel mass
m_{fuelB}	the fuel mass injected into the cylinder
m_{fuelC}	the total injected fuel mass per engine cycle
m_J	journal mass
m_P	piston mass
m_V	total vehicle mass
m_{VAL}	valve mass
m_{VT}	valvetrain equivalent mass
$\dot{m}$	mass flow rate
$\dot{m}_{air}$	engine fresh air mass flow rate
$\dot{m}_{EGR}$	EGR mass flow rate
$\dot{m}_{ex}$	the engine exhaust gas mass flow rate at the turbine inlet or in the exhaust manifold or port (Note: This parameter is affected by EGR pickup.)
$\dot{m}_{exh}$	the engine exhaust gas mass flow rate at the turbine outlet
$\dot{m}_{fuel}$	fuel mass flow rate
$\dot{m}_T$	actual turbine mass flow rate
$\dot{m}_{WG}$	turbine wastegate mass flow rate
$\tilde{m}_{AT}$	precious metal loading in aftertreatment
$\tilde{m}_{ring}$	the mass of the piston ring per unit length of the ring
N	speed
N_B	journal bearing rotational speed
N_C	compressor shaft speed
N_{design}	valvetrain design speed
N_E	engine crankshaft rotational speed
N_f	the number of loading cycle to reach failure
N_T	turbine shaft speed
N_{TC}	turbocharger speed
N_u	the Nusselt number
N_V	vehicle speed
N_{Vw}	the vehicle speed relative to the wind
n	the total number of index, or the number of DoE runs, or sample size (number of samples), or the number of grid points, etc.
n_c	combustion index for noise
n_E	the number of engine cylinders
n_G	the number of gear meshes between the crankshaft and the fuel system

n_l	the number of loading cycle
n_{SC}	the number of spring coils
n_s	the number of crankshaft revolutions in one engine cycle
n_{sm}	smoke number
n_{VAL}	the number of engine valves
O	valve, throttle, wastegate, or vane opening; or pedal position
O_a	opening area
O_{ang}	opening angle
O_{dia}	opening diameter
P	probability
P_{fail}	the probability of failure
P_r	the Prandtl number
p	pressure, or in-cylinder gas pressure, or the number of regression coefficients in the emulator model in DoE
$p_{asperity}$	the asperity contact pressure in the mixed or boundary lubrication
p_{comp}	in-cylinder compression pressure
p_{cyl}	cylinder pressure
p_{EM}, p_3	exhaust manifold pressure
p_{ex}	exhaust port pressure
p_{IM}, p_{2a}	intake manifold boost pressure
p_{in}	intake port pressure
p_{inj}	fuel injection pressure
p_l	the loading pressure acting on the sliding component, or the mean Hertzian contact pressure
p_{lub}	lubricating oil film pressure
p_{max}	peak cylinder gas pressure
p_{port}	port pressure
p_s	sound pressure
p_{SPL}	sound pressure level in dB
$p_{SPL,E}$	overall engine noise sound pressure level
Q	heat, or the heat exchanged through the system boundary
$\dot{Q}$	heat transfer rate (i.e., heat rejection)
Q_{fuel}	fuel energy
Q_{wall}	the heat transfer through the walls of the cylinder head, the piston, and the liner
$\dot{Q}_{base\text{-}coolant}$	the heat rejection from the base engine to the coolant
$\dot{Q}_{hrj,firing}$	the engine coolant heat rejection in the firing operation
$\dot{Q}_{hrj,retarding}$	the unrecoverable heat dissipation to the cooling system in the retarding operation
$\dot{Q}_{miscellaneous}$	miscellaneous heat losses
q	heat per unit of mass, or the total number of objective functions in multi-objective optimization

q_{LHV}	the lower heating value of the fuel
$\tilde{q}$	heat flux
R	reliability
R_{adj}^2	the adjusted coefficient of determination
R_d^2	the coefficient of determination
$R_{prediction}^2$	the coefficient of predicted variation
R_e	the Reynolds number
R_{ex}	the specific gas constant of exhaust gas
R_{gas}	specific gas constant
R_i	the reliability of component i or subsystem i
R_{in}	the specific gas constant of intake charge
R_N	the number of repairs
R_S	system reliability
r	radius, or index number
r_B	bearing radius
r_{ROC}	the radius of curvature
r_{tire}	tire dynamic radius
S	material strength
S_E	engine stroke
S_e	entropy
S_E/B_E	engine stroke-to-bore ratio
S_{lub}	lubrication duty parameter
$S_{lub,sr}$	lubrication parameter
S_{ODE}	the stiffness ratio of ODE system
S_{SE}, S_{SR}, S_{ST}	statistical functions in response surface methodology (RSM)
s	stress
s_e	specific entropy
s_{fl}	fatigue limit stress
s_{SP}	the torsional stress of the spring
s_u	ultimate stress
T	temperature, or in-cylinder gas temperature
T_{CACout}	the air or gas temperature at the CAC outlet (before EGR mixing)
T_{ch}	characteristic temperature
T_{EGR}	the EGR gas temperature at the EGR cooler outlet
T_{ex}	exhaust port gas temperature
T_f	the required time to reach creep failure
T_{flash}	flash temperature
T_{in}	intake port gas temperature
T_{sink}	cooling medium inlet temperature, i.e., the sink temperature
T_{wall}	the component metal temperature at the wall
$T_{1,ROA}$	compressor inlet temperature including the rise-over-ambient (ROA) effect

T_{2a}	intake manifold gas temperature
T_3	exhaust manifold gas temperature
t	time
t_l	loading time
U	internal energy, or the internal energy of the in-cylinder gas
u	specific internal energy, or index number
V	volume, or instantaneous in-cylinder volume
$V_{act,E}$	the total engine displacement of the activated cylinders
V_{cyl}	cylinder displacement
V_E	engine displacement
V_o	battery voltage
V_w	wear volume
$\dot{V}$	volume flow rate
$\tilde{V}_{EV}$	the engine volume per displacement volume
$\tilde{V}_{EW}$	engine specific volume
v	velocity, or relative velocity
v_a	acoustic velocity
v_{air}	the velocity of the local oscillating motion of the air particles
v_{CAM}	cam velocity
v_{im}	impact velocity
v_{mp}	mean piston speed
v_P	piston sliding velocity
v_{sound}	the speed of sound
v_{sp}	space velocity
v_T	turbine wheel average tip speed
v_{VVL}	vibration velocity level in dB
W	work
$\dot{W}$	power
$\dot{W}_C$	compressor power
$\dot{W}_E$	engine brake power
$\dot{W}_{E_{acc}}$	engine accessory power
$\dot{W}_f$	friction power
$\dot{W}_{f,E}$	engine friction power
$\dot{W}_{f,TC}$	turbocharger friction power
$\dot{W}_{ind}$	engine indicated power
$\dot{W}_r$	retarding power
$\dot{W}_s$	sound power
$\dot{W}_{SWL}$	sound power level in dB
$\dot{W}_T$	turbine power
$\tilde{\dot{W}}_E$	engine specific brake power (i.e., power per displacement volume)
$\tilde{\dot{W}}_{EA}$	the engine brake power per piston area

w	weight, or weighting factor
w_c	canonical variables in RSM
$\bar{w}_{C_xH_y}$	the molecular weight of C_xH_y
$\tilde{w}_{EV}$	the engine weight per displacement volume
$\tilde{w}_{EW}$	engine specific weight (i.e., engine weight divided by power)
X	factor
X_d	deterministic control factors (deterministic design variables)
X_{fuel}	the instantaneous fraction of the fuel burnt within an engine cycle
X_p	nondeterministic random noise factors
X_r	nondeterministic random control factors (random design variables)
$\bar{X}$	the average or mean value of X
$\widehat{X}$	the normalized value of X
x	any dummy parameter used in integral or differential
x_{ring}	piston ring diametrical (lateral) displacement
x, y, z	the motion displacement in x, y, and z directions
$\dot{x}$, $\dot{y}$, $\dot{z}$	the velocity in x, y, and z directions
$\ddot{x}$, $\ddot{y}$, $\ddot{z}$	the acceleration in x, y, and z directions
x_2	piston pin lateral displacement
Y	response, or functional performance parameter
Y^*	the transformation of the response Y
$\bar{Y}$	the average or mean value of Y
$\widehat{Y}$	the normalized value of Y
$\hat{Y}$	the response Y approximated or predicted by regression model
$y(\phi)$	piston vertical displacement as a function of crank angle
Z	the composite of functional performance or response
Z_r	resistance
Z_s	acoustic impedance
Z_{sm}	mechanical impedance
Z_{sr}	radiation impedance

Greeks

Δ	change, or difference
$\Delta\varepsilon$	strain range
$\Delta\phi$	crank angle step
$\Delta\phi_{CAM,ac}$	the width of cam positive acceleration hump
Δp	pressure differential
$\Delta p_{1\text{-}2}$	the pressure drop from 1 to 2
Δs	stress range, or stress amplitude
Δt	time step, or time interval

ΔT	temperature drop
$\Delta TROA$	rise-over-ambient temperature increase
Φ	the cooling capability of a cooler or cooling system
Λ	impulse
Θ	kurtosis
Ω	engine compression ratio
Ψ	skewness
α	the location parameter in statistical distributions
α_g	the in-cylinder heat transfer coefficient from gas to inner cylinder wall
β	the scale parameter in statistical distributions
β_2	the piston tilting angle in piston dynamics
δ	the shape factor of combustion heat release rate, or thickness, or deformation
ε	strain
ε_B	the dimensionless eccentricity of bearing
ε_{cooler}	cooler effectiveness
ϕ	angular displacement or engine crank angle
ϕ_{CAM}	cam timing
ϕ_{VAL}	valve timing
φ	the circumferential direction of the piston or the bearing
φ_B	bearing attitude angle
γ	the shape parameter in statistical distributions
η	efficiency, usually in the first law of thermodynamics
η_{BSFC}	brake specific fuel consumption
η_{cyl}	the energy transfer efficiency from in-cylinder to the turbine inlet
η_C	compressor efficiency
η_{FC}	fuel consumption
$\eta_{E,mech}$	engine mechanical efficiency
$\eta_{r,CR}$	compression-release brake retarding process efficiency
$\eta_{r,exh}$	retarder available exhaust energy ratio
$\eta_{r,heat}$	retarder heat dissipation ratio
η_{regen}	DPF regeneration efficiency
η_T	turbine efficiency
η_{TC}	turbocharger efficiency
$\eta_{TC,mech}$	turbocharger mechanical efficiency
η_t	drivetrain efficiency
η_{th}	engine thermal efficiency
η_{trax}	the mechanical efficiency of transaxles
η_{ts}	the efficiency of turbocharging system
η_{vol}	engine volumetric efficiency
λ	the 'lambda' ratio, i.e., the ratio of lubricating oil film thickness to surface roughness

$\lambda_{ODE,max}$	the largest eigenvalue in stiff ODE
μ	mean
μ_v	lubricant dynamic viscosity
ω	angular velocity
ϖ_{IMEP}	indicated mean effective pressure
ϖ_{BMEP}	brake mean effective pressure
θ	road slope angle, or other angles
ϑ_e	emissivity
ϑ_m	the modulus of elasticity, or shear modulus
ϑ_P	the Poisson's ratio
ϑ_{SB}	the Stefan-Boltzmann constant
ρ	air or gas density
σ	standard deviation
σ^2	variance
σ_{sr}	surface roughness
σ^2_{SV}	sample variance
ς	excess air–fuel ratio, or the reciprocal of the equivalence ratio
τ	shear stress
υ	error
υ_{RMSE}	root-mean-square error
ξ	vehicle rotational mass coefficient
ψ	brake specific emissions
ζ	damping coefficient

Subscripts

Notes: The subscripts are usually used in the following three scenarios.

1. the name of a system, a component, identity, or acronym (using upper-case letters);
2. working fluids, physical processes, or phenomena (using lower-case letters);
3. index such as 1, 2, 3, …; x, y, z ... .

The hyphen sign (-) represents from one to another, or between the two (e.g., *h-m*). The / sign represents one or the other (e.g., *T/C*). Descriptive or self-explanatory subscripts are often used, provided that space is available in the equation, e.g., *intake*; *base-coolant*; *EGRcoolerGasOut*; *CAM,ramp*.

AMB	ambient
a	aerodynamic drag resistance
ac	acceleration
acc	accessory
act	activated or active cylinders
active	the active coils of the spring

air	fresh air
amp	the amplitude of variation
ave	average
B	bearing
b	boundary lubrication
bc	the beginning of compression stroke
br	service brake
C	compressor
CAC	charge air cooler
CACcooling	the cooling medium (sink) of the CAC
CACout	CAC outlet charge air or gas
CAM	cam
c	creep
cal	model calibration or tuning
cc	creep in tension and creep in compression
com	combustion
comp	compressor
corr	corrected
cp	creep in tension and plastic in compression
cyl	cylinder
D	damage
d	deterministic control factors
design	the "design speed" of valvetrain
df	drivetrain friction
dr	drivetrain retarder
drive	drivetrain
E	engine
EGRC	EGR cooler
EM	exhaust manifold
e	elastic
ec	the end of compression stroke
eff	effective
eq	equivalent
er	engine retarding (engine braking)
ex	the exhaust event or exhaust gas before the turbine inlet
exh	the exhaust system from the turbine outlet to the ambient; or the exhaust gas flow at or after the turbine outlet
exhaust	exhaust stroke
f	friction
fc	fatigue and creep
fire	engine firing
g	gravity
gas	gas or gas loading

gr	transmission gear
grade	road grade or gradient
h	hydrodynamic lubrication
hrj	heat rejection
IM	intake manifold
i	inertia force
ia	inactive
i, j, k	index number
id	ignition delay
im	impact
in	intake event or intake air/gas
ind	indicated
inj	fuel injection
inl	the inlet of the control volume
int	the intake system from the ambient to the compressor inlet
intake	intake stroke
J	journal
LHV	lower heating value
LubeOilCons	lube oil consumption
l	loading
level	level ground
lower	lower bound
lub	lubrication
m	mixed lubrication
max	maximum
mech	mechanical friction
min	minimum
mixture	the mixture gas of fresh air and EGR
mot	motoring
n	normal direction
O_2	oxygen
o	the total or stagnation state of the gas flow
obs	observed
out	the outlet of the control volume
P	piston
p	plastic, or nondeterministic random noise factors
pc	plastic in tension and creep in compression
pcyl	cylinder pressure
pp	plastic strain in tension and plastic in compression
pre	spring preload
q	frequency
R	reliability
RA	rocker arm

RAD	radiator
r	retarding, or nondeterministic random control factors
recir	recirculation
ref	reference state or reference object
rf	vehicle tire–road rolling friction resistance
ring	piston ring
S	system
SC	spring wire coil
Sh	strength
SOC	the start of combustion
SP	spring
SPL	sound pressure level
s	sound, or the static state of the gas flow
sf	safety
sink	cooling medium inlet or the sink environment
sl	sliding
slap	piston slap
sr	speed ratio
ss	stress
stoi	stoichiometric
T	turbine
TC	turbocharger
T/C	turbine or compressor
Texh	exhaust manifold gas temperature
TV	torque converter
t	tractive force
tan	tangential direction
tr	torque ratio
turb	turbine
upper	upper bound
V	vehicle
VAL	valve
VT	valvetrain
v	viscous
w	wear
wall	metal wall of the component
x	horizontal, lateral or the x direction
y, z	vertical or the y or z direction
0	reference state, or initial state, or the dead state in the second law of thermodynamics
1, 2	energy balance methods 1 and 2
1, 2, 3	the locations in engine gas flow network, or the index number of coefficients or DoE cases, etc.

1	compressor inlet
2	compressor outlet
2*a*	the location after location 2 in engine system layout or engine gas flow network (i.e., intake manifold)
3	turbine inlet or exhaust manifold
4	turbine outlet

Superscripts

g	global
L	lower bound
l	local
limit	limit
max	the maximum in optimization
min	the minimum in optimization
n	nominal
opt	optimal
p	peak
T	transposed
U	upper bound
v	valley
target	design target value
$(1, 2, 3, \ldots, i)$	the number of iterations
’	transposed vector

Special symbols

d	ordinary differential mathematical operator
e	the Euler’s number (≈ 2.718), the base of the natural logarithm
exp	the exponential function, $\exp(x) = \mathrm{e}^x$
log, ln	logarithmic functions
sin, cos, tan	trigonometric functions (sine, cosine, tangent)
π	a mathematical constant (≈ 3.14159265) whose value is the ratio of any circle’s circumference to its diameter in Euclidean space.
∞	infinity
°	degree
$\int$	integral mathematical operator
∂	partial differential mathematical operator
Σ	summation
Π	multiplication
$\lvert x \rvert$	the absolute value of x

Units

absolute	absolute pressure
Btu/min	British thermal unit per minute, a unit of power
ft	foot, a unit of length
ft.lb	foot pound-force (lbf), a unit of torque
gauge	gauge pressure
hp	horsepower, a unit of power
in	inch, a unit of length
inHg	inch Hg or inches of mercury, in Hg, or ″Hg, a unit of measurement for pressure
K	Kelvin, a unit of temperature
L	liter, a unit of volume
lb	pound-force, lb_F or lbf, a unit of force
lb/(hp.hr)	pound-mass (lbm) per horsepower per hour, a unit of specific fuel consumption rate
lb/hr	pound-mass (lbm) per hour, a unit of mass flow rate
lb/min	pound-mass (lbm) per minute, a unit of mass flow rate
lbs	pounds, the plural form of lbm or pound-mass, which is a unit of mass
m	meter, a unit of length
ppm	parts per million, a unit of concentration
psi	pound-force per square inch, or lbf/in^2, a unit of pressure or stress
psia	psi, absolute pressure
s	second, a unit of time
°C	degree Celsius, a unit of temperature
°F	degree Fahrenheit, a unit of temperature

List of abbreviations and acronyms

AAMA	The American Automotive Manufacturers Association
ABT	(emission) averaging, banking, and trading
ACEA	Association des Constructeurs Européens d'Automobiles
AECD	auxiliary emissions control device
A/F	air-to-fuel ratio or air–fuel ratio
ALT	accelerated life testing
ANOVA	analysis of variance
API	American Petroleum Institute
APQP	advanced product quality planning
APU	auxiliary power unit
A/R	turbine ratio of area to distance
ASQ	American Society for Quality
ASTM	American Society for Testing and Materials
ATDC	after top dead center
B5	a fuel blend of 5% biodiesel and 95% diesel
BB	Box–Behnken
BDC	bottom dead center
BEM	boundary element method
BGR	braking gas recirculation
BMEP	brake mean effective pressure
BOM	bill of materials
BS	brake specific
BSFC	brake specific fuel consumption
BSN	Bosch smoke number
BTDC	before top dead center
BVO	brake valve opening (timing)
C	carbon atom, or criticality
CA50	crank angle for 50% burn in heat release analysis
CAC	charge air cooler
CAFÉ	corporate average fuel economy
CAI	controlled auto-ignition
CARB	California Air Resources Board

CBSFC	cycle brake specific fuel consumption
CCC	central composite circumference
CCF	central composite faced
CCV	crankcase ventilation
CDA	cylinder deactivation
CDF	cumulative distribution function
CDPF	catalyzed diesel particulate filter
CFD	computational fluid dynamics
CFR	United States Code of Federal Regulations
CG	center of gravity
CGI	compacted graphite iron
CI	compression ignition
CNG	compressed natural gas
CO	carbon monoxide
CO_2	carbon dioxide
conrod	connecting rod
CR	compression ratio
CVT	continuously variable transmission
D	detection of failure modes
DESD	diesel engine system design
DF	deterioration factor
DI	direct injection
DIT	dynamic injection timing
DME	dimethyl ether
DOC	diesel oxidation catalyst
DoE	design of experiments
DOF	degrees of freedom
DPF	diesel particulate filter
E	energy transfer
EBP	exhaust back pressure
ECA	emission control area (along shorelines for marine engines)
ECU	engine control unit
EGR	exhaust gas recirculation
EHD	elastohydrodynamic lubrication
EHL	elastohydrodynamic lubrication
EMA	Engine Manufacturers Association
EOI	end of injection
EOTD	engine outlet coolant temperature difference over ambient
EPA	US Environmental Protection Agency
EPI	exhaust-pulse-induced (compression brake)
EPSI	engine performance and system integration
ESC	European Stationary Cycle
EVC	exhaust valve closing

EVO	exhaust valve opening
FE	fuel economy
FEA	finite element analysis
FEAD	front end accessory devices
FEL	family emissions level
FMEA	failure mode effect analysis
FMEP	friction mean effective pressure
FMETA	failure mode and effect tree analysis
FSN	filter smoke number
F-T	Fischer–Tropsch
FTP	Federal Test Procedure
FTP-75	Federal Test Procedure drive cycle for light-duty vehicle emissions
GA	genetic algorithm
GAWR	gross axle weight rating
GCVW	gross combined vehicle weight
GDI	gasoline direct injection (engine)
GVW	gross vehicle weight
GVWR	gross vehicle weight rating
GTL	gas-to-liquid
H	hydrogen atom
HC	hydrocarbons
HCCI	homogeneous charge compression ignition
HCF	high cycle fatigue
HD	heavy duty
HD-UDDS	heavy-duty Urban Dynamometer Driving Schedule
HEV	hybrid electric vehicle
HHV	hybrid hydraulic vehicle
HIL	hardware-in-the-loop
HLA	hydraulic lash adjuster
H_2O	water
HP	high pressure (stage)
HPL	high pressure loop
HRR	heat release rate
HSDI	high speed direct injection
HT	high temperature (stage)
HTHSV	high-temperature high-shear viscosity
HTR	high temperature radiator
HWFET	highway fuel economy test
HWY	highway
I	information exchange
I4	inline four-cylinder engine
I6	inline six-cylinder engine

ICP	injection command pressure
ICE	internal combustion engine
IDI	indirect injection
IMEP	indicated mean effective pressure
IMO	International Maritime Organization
IMT	intake manifold gas temperature
IMTD	intake manifold temperature difference (the difference between the charge air cooler outlet fresh air temperature and the ambient temperature)
INCOSE	International Council on Systems Engineering
ISC	inter-stage cooler
ISFC	indicated specific fuel consumption
ISO	International Organization for Standardization
IT	intake throttle
IVC	intake valve closing
IVO	intake valve opening
JFO	Jakobsson–Floberg–Olsson (JFO) cavitation boundary condition
LCF	low cycle fatigue
LD	light duty
LNG	liquefied natural gas
LNT	lean NO_x trap
LP	low pressure (stage)
LPG	liquefied petroleum gas
LPL	low pressure loop
LSD	low sulfur diesel, a diesel fuel which contains less than 500 ppm of sulfur
LSL	lower specification limit
LT	low temperature (stage)
LTR	low temperature radiator
M	material exchange
MAF	mass air flow
MAP	manifold air pressure
MCS	Monte Carlo simulation
MIMO	multi-input multi-output
MLA	mechanical lash adjuster
MOGA	multi-objective genetic algorithm
mpg	miles per gallon (of fuel)
MTBF	mean time between failures
MTTF	mean time to failure
MV	mean value (model)
N	nitrogen atom
N_2	nitrogen

NEDC	New European Driving Cycle
NMHC	non-methane hydrocarbons
NN	neural network
NO	nitric oxide
NO_2	nitrogen dioxide
NO_x	oxides of nitrogen
NRSC	non-road stationary cycle
NRTC	non-road transient composite test cycle (for EPA Tier 4 emissions certification)
NTE	Not-to-Exceed
NTU	number of transfer unit
NVH	noise, vibration, and harshness
O	oxygen atom, or occurrence of the effects
O_2	oxygen
OBD	on-board diagnostics
ODE	ordinary differential equation
OEM	original equipment manufacturer
OHC	overhead cam
OHV	overhead valve
OPOC	opposed-piston opposed-cylinder (engine)
OTAQ	EPA Office of Transportation and Air Quality
P	physical connection
PCP	peak cylinder pressure
PD	proportional-differential
PDF	probability density function
PFQI	Pareto front quality index
PI	proportional-integral
PID	proportional-integral-differential
PM	particulate matter
PMEP	pumping mean effective pressure
QALT	quantitative accelerated life test
RAR	rear axle ratio
RBDO	reliability-based design optimization
rev	revolution
RMS	root-mean-square
RMSE	root-mean-square error
ROA	rise over ambient
ROC	radius of curvature
RPN	risk priority number
rpm	revolution per minute
RSM	response surface methodology
S	sulfur, or severity of the effects
SAE	Society of Automotive Engineers

SASR	solid ammonia storage and release
S/B	stroke-to-bore ratio
SCR	selective catalytic reduction
SEA	statistical energy analysis
SECA	SO_x Emission Control Area
SET	Supplemental Emissions Test
SFC	specific fuel consumption
SI	spark ignition
SIL	software-in-the-loop
S/N	signal-to-noise ratio
SO_x	sulfur oxides
SOC	state of charge (for battery)
SOC	start of combustion (for engine)
SOF	soluble organic fraction
SOI	start of injection
SPL	sound pressure level
SUV	sport utility vehicle
TBN	total base number (of engine oil)
TC	turbocharging, turbocharged, or turbocharger
TDC	top dead center
TDI	turbocharged direct injection
TM	thermo-mechanical
TS–SRP	total strain–strain range partitioning
turbo	turbocharger
ULSD	ultra-low sulfur diesel, a diesel fuel which contains less than 15 ppm of sulfur
US06	Supplemental Federal Test Procedure (SFTP) driving cycle to complement the FTP-75 test cycle and represent aggressive, high speed and/or high acceleration driving behavior and rapid speed fluctuations for light-duty vehicle emissions
USCAR	United States Council for Automotive Research
USL	upper specification limit
V6	Vee-bank six-cylinder engine
V8	Vee-bank eight-cylinder engine
VAT	variable area turbine
VCR	variable compression ratio
VGC	variable geometry compressor
VGT	variable geometry turbine
VNT	variable nozzle turbine
vol. effi.	volumetric efficiency
VVA	variable valve actuation
VVL	variable valve lift
VVT	variable valve timing

WE-VVA	wastegating elimination variable valve actuation
WG	wastegate or wastegate opening
WHR	waste heat recovery
WR-VVA	wastegating reduction variable valve actuation
ZWB	zero-wheel-braking
1-D	one-dimensional
2-D	two-dimensional
3-D	three-dimensional
°CA	degrees of crank angle

Dedicated to and in memory of
my dear father Xianxi Xin

About the author

Dr Qianfan Xin (also known as Harry Xin) received his B. Sc. degree in Thermal Engineering in 1991 from Tongji University in China and his M. Sc. (1997) and D. Sc. (1999) degrees in Mechanical Engineering from the Washington University in St Louis, USA. Dr Xin has been working at Navistar, Inc. since 1999, and is a Product Manager in the area of advanced simulation analysis on diesel engine performance and system integration. He specializes in diesel engine system design.

Contact details

Dr Qianfan (Harry) Xin
Product Manager of Engine Performance and System Integration
Advanced Analysis and Simulation Department
Navistar, Inc.
10400 West North Avenue
Melrose Park, IL 60160
USA
E-mail: Harry.Xin@navistar.com; QianfanHarryXin@yahoo.com
Phone: +1 708 865 3555

Preface

The diesel engine is recognized as the most promising powertrain in the foreseeable future due to its superior thermal efficiency and reliability. The diesel engine has been widely used in commercial vehicles, industrial applications and today's passenger cars and light-duty trucks. Modern emissions standards and customer demands are driving diesel engineering to become a fast growing applied engineering discipline in order to meet the requirements of designing optimum diesel engines. The engineering population in diesel engine design is growing fast. The need for advanced design theories and professional reference books has become pressing and obvious. This book presents my own experience and findings in many interrelated areas of diesel engine performance analysis and system design. The book also intends to establish an emerging area of diesel engine system design for the diesel industry.

Diesel engine design is very complex. It involves many people and companies from OEM (original equipment manufacturer) to suppliers. A system design approach to set up correct engine performance specifications is essential in order to streamline the process. While working at Navistar, a global leading manufacturer of diesel engines and trucks, I noticed several common challenges existing in today's engine industry:

1. Certain disconnections between academic education and industrial design practice (i.e., sometimes engineers do not know how to apply their classroom engine knowledge to analysis or design in their daily work).
2. The lack of comprehensive explanatory references or textbooks on the design approaches related to diesel engine performance and system integration. People have to rely on scattered literatures of published technical papers, unpublished company reports, and the oral accounts of engineers working in each related field. Unfortunately, very often this scattered information does not provide the direct answer to the needs of a practicing engineer for his daily design challenges.
3. The absence of a unified and systematic theory about diesel engine

system design. People from different areas or organizations use different or sometimes wrong approaches, resulting in much confusion and inconsistency. This has made collaboration and communication within the diesel engine industry itself difficult.

4. Insufficient guidance for research in applied engineering so that it supports the needs of diesel engine system design.

The diesel engine is a sophisticated electro-mechanical machine. The system attributes of a diesel engine can be classified into four categories, namely performance, durability, packaging, and cost. Among them, performance (or function) is the leading attribute. Both static (steady state) design and dynamic (transient) design are important to satisfy the performance requirements of the diesel engine. As an industrial commercial product, the diesel engine needs to be designed not only for nominal target performance, but also for taking into account statistical variability and reliability. Engine performance and system design is such a broad area that it affects almost everyone in the engine design business. It involves different areas and job functions, such as system/subsystem/component designs, emissions testing and calibration, vehicle and aftertreatment integration, engine electronic controls, durability testing, etc. A few years ago, I began to receive requests and suggestions from working engineers to write a book to address the above-mentioned needs related to system design. I believe the book should serve as an effective tool in training new engineers. It should supply comprehensive yet easy-to-use references. It should provide a standard approach to conduct system design and analysis. It should also provide a vision for future research needs to improve the quality of diesel engine system design.

By witnessing the challenges faced by engineers sometimes lacking the knowledge of how an engine system is designed, and by seeing some problems applying the academic fundamental knowledge to industry design practice, I am convinced that a bridge linking the following three closely inter-related areas is needed: textbooks teaching the fundamental principles; advanced research; and the design practice in the real world of diesel engineers.

System design is very important for the integration of simultaneous engineering processes for diesel engine product development, ranging from high-level product strategy planning to detailed production designs. Diesel engine system design is a performance-based emerging technical area for modern engines. The design theory proposed in this book is led by systems engineering principles, and based on advanced optimization theories to achieve precise and probabilistic system designs to integrate a wide range of attributes (performance, durability, packaging, and cost) from the system level to the component level. The system design is conducted by comparing engine configurations and producing system performance specifications with analytical tools in various areas (e.g., engine cycle simulation, vehicle and powertrain dynamics). The system design covers a range of core technical specialties including vehicle–engine–aftertreatment integration, thermodynamic cycle performance, engine air system design and turbocharger matching, system friction, NVH (noise, vibration and harshness) synthesis, and electronic controls.

The importance of diesel engine system design (DESD) has been recognized in the design and development process. The concept of DESD was presented in a book chapter, 'Heavy-duty diesel engine system design' written by me in 2008 as a part of the book *Advanced Direct Injection Combustion Engine Technologies and Development* (*Volume 2*: *Diesel engines*) edited by Professor Hua Zhao from Brunel University, UK. In the 2009 SAE Commercial Vehicle Engineering Congress held in Chicago on October 6, I had the opportunity to present the concept of diesel engine system design to an audience of about 70 people from the industry.

My original interest relating to diesel engine system design dated back to 1995–99 when I studied for my Doctor of Science (D. Sc.) degree at Washington University in St. Louis, Missouri, USA. My research related to engine piston-assembly lubrication dynamics. That experience inspired my interest in engine friction, dynamics and the relationship between subsystem design and overall engine performance. Since I joined Navistar in 1999, I have been working on advanced simulation analysis of engine performance and system integration. I had the opportunity to assemble a unique functional area – diesel engine system design, and became the Product Manager of the engine system design group in 2003.

This book aims to establish the theory of diesel engine system design, including the approaches used in its modeling, design, and advanced research. The central theme of the book is to design a good engine system with performance specifications in the early stage of the product development cycle. Every component that has a major impact on the quality of system design is considered in the theory.

The book tries to link everything diesel engineer's need to know about engine performance and system design in order for them to quickly master all the important topics. The book consists of four parts. Part I addresses the fundamental concepts of diesel engine system design and durability. Parts II-IV focus on engine performance and system integration (EPSI). Due to the limits of space, it is impossible to cover in detail every single subject in one book. The book therefore focuses on less well-developed areas of research. These include: Chapter 1 (the concepts of engine system design), Chapter 2 (engine system durability and reliability), Chapter 3 (optimization approaches used in system design), Chapter 4 (mathematical fundamentals of engine air system theory), Chapter 5 (vehicle performance), Chapter 6 (engine brake), Chapter 7 (from combustion to system design and calibration), Chapter 9 (valvetrain system design), Chapter 10 (system friction), Chapter 11 (system NVH), Chapter 12 (heat rejection), Chapter 13 (pumping loss and air system design theory) and Chapter 15 (system specification design and subsystem interaction optimization). For the topics that are briefly discussed, extensive organized learning materials are provided in the references and bibliography to direct the reader to find the most important advances in the area.

Chapter 2 (durability and reliability) is the most difficult but an extremely important area for system design in the long run. Engine performance specifications are directly tied to durability constraints, and subject to design for reliability as an ultimate design goal.

The first part of Chapter 5 (vehicle performance and engine–vehicle matching) provides a detailed summary to facilitate the reader with a streamlined analysis approach although this area has a long standing history. The second part of Chapter 5 (hybrid powertrains) is briefly summarized to introduce the role of engine system design in hybrid powertrain development. The reader is referred to the references and bibliography provided for more detailed discussion on hybrid powertrains.

Chapter 11 (system NVH) addresses a very challenging area in system-level modeling. System NVH is a very important research direction for diesel engine system design. NVH is extremely complex to analyze. The summary in Chapter 11 puts together all the pieces and tries to set up a foundation for system engineers. NVH attributes that are difficult to analyze usually do not impede the generation of engine system design specifications. Without a good upfront system design to start with, it will be more costly to fix the NVH problems in a later design stage at the component level. Many NVH analysis methods are available at the component and subsystem levels, and the system engineer needs to integrate them.

Chapter 8 (aftertreatment integration) and Chapter 14 (system dynamics, transient performance, and engine controls) are two major areas in system design for improving the competitiveness of engine products. They are discussed in a less detailed manner compared to other chapters in this book due to limited space. A comprehensive list of carefully selected references and bibliography is provided for these areas to facilitate the reader. In particular, diesel engine transient performance and electronic controls are not addressed in great detail because these topics have been thoroughly covered in two recently published books by other authors (*Diesel Engine Transient Operation – Principles of Operation and Simulation Analysis*, by Rakopoulos and Giakoumis and *Introduction to Modeling and Control of Internal Combustion Engine Systems*, by Guzzella and Onder in 2004).

The essence of diesel engine system design can be comprehended briefly as:

- four major driving goals (emissions, power density, fuel economy, and reliability)
- four attributes (performance, durability, packaging, and cost)

- four cornerstones of analysis (static or steady-state design, dynamic design or system dynamics, first law of thermodynamics, and the second law of thermodynamics).

Engine thermodynamics is the foundation of static system design with state variables, while system dynamics forms the basis of dynamic system design and engine controls that handle the transient processes. This book tries to provide abundant information to illustrate those points.

The focus of the book is to introduce advanced analysis methods to solve practical design problems rather than conduct pure theoretical research or teach engine fundamentals. A large amount of engine performance simulation analysis results are illustrated. The main readership is design and development engineers rather than software developers. Therefore, the book tries to limit the contents of complex theoretical equations that are usually the foundation of many widely-used commercial software packages. The discussions about such equations can be found from other published books or literatures.

I hope the book can benefit a broad range of engine professionals in different disciplines who work on the design and development of modern low-emissions diesel engines equipped with turbocharging and EGR (exhaust gas recirculation). The book tries to provide engineers with a systematic understanding of how engine system design specifications are generated. The book enables system design engineers to directly apply the methods and working knowledge to their daily design and research. The book also introduces engine design knowledge to academic researchers in order to broaden their vision so that they may better support the practical and critical needs of the diesel engine industry. Finally, the book may also help senior-level undergraduate or postgraduate students understand how industrial design problems are handled at a system level.

Writing such a book on top of my busy working schedule has been a great challenge. Almost all the evening time, weekends and holidays during 2009 were devoted to this book. I am very grateful to my wife, Katty, for her understanding and support. Without her support, this book would not have been possible. I want to thank Professor Hua Zhao at Brunel University for providing helpful suggestions. I greatly appreciate the valuable discussions I have had with my brilliant colleagues at Navistar. I would also like to thank Sheril Leich of Woodhead Publishing for commissioning this book and thank Sheril and Cathryn Freear at Woodhead for their professional support in preparing the book.

Qianfan Xin
Chicago, USA

1
The analytical design process and diesel engine system design

Abstract: Diesel engine system design (DESD) is an important and leading function in the design and development of modern low-emissions EGR diesel engines. It creates a paradigm shift in how engine design is carried out. It leads and integrates the designs from the system level to the component level by producing high-quality system design specifications with advanced analytical simulation tools. This chapter introduces the fundamental concepts in diesel engine system design and provides an overview on the theory and approaches in this emerging technical field. The central theme is how to design a good engine system performance specification at an early stage of the product development cycle. The chapter employs a systems engineering approach and applies the concepts of reliability and robust engineering to diesel engine system design to address the optimization topics encountered in design for target, design for variability, and design for reliability. An attribute-driven system design process is developed for advanced analytical engine design from the system level to the subsystem/component level in order to coordinate different design attributes and subsystems. Four system design attributes – performance, durability, packaging, and cost – are elaborated. The chapter also addresses competitive benchmarking analysis. By focusing on engine performance and system integration (EPSI), the technical areas, theoretical foundation, and tools in diesel engine system design are introduced.

Key words: diesel engine system design (DESD), engine performance and system integration (EPSI), systems engineering, robust engineering, reliability, variability, design target, design attributes, performance, durability, packaging, cost, competitive benchmarking, engine cycle simulation, exhaust gas recirculation (EGR).

1.1 Characteristics and challenges of automotive diesel engine design

1.1.1 Classification of diesel engines

A diesel engine is a type of compression-ignition engine using diesel fuel. Diesel engines can be classified into various categories (Table 1.1). Understanding the differences and the unique characteristics of each category of diesel engines is important for diesel engine system design. According to the number of crankshaft revolutions per working cycle, diesel engines are classified as four-stroke engines (two revolutions per cycle) and two-stroke

Table 1.1 Diesel engine classification

Classification criteria	Variant	Variant	Variant
Number of strokes	Four-stroke	Two-stroke	
Emissions standard	On-road	Off-road	
Application	On-road (trucks, buses, automobiles)	Off-road (marine, industrial, construction equipment, agricultural, locomotive)	Stationary
Emissions certification method for vehicles	Heavy heavy-duty, medium heavy-duty, light heavy-duty	Light duty	
Vehicle weight	Heavy duty	Medium duty	Light duty
Crankshaft rated speed	High speed (N_E > 1000 rpm or v_{mp} > 9 m/s)	Medium speed (N_E = 300 – 1000 rpm or v_{mp} = 6 – 9 m/s)	Low speed (N_E < 300 rpm or v_{mp} < 6 m/s)
Fuel injection	Direct injection	Indirect injection	
Air charging	Turbocharged (with or without after-cooling)	Mechanically supercharged (with or without after-cooling)	Naturally aspirated
Cooling medium	Water cooled	Air cooled	
In-cylinder NO_x emissions control	EGR	Non-EGR	
NO_x aftertreatment control	Non-SCR	SCR	
Number of cylinders	Single-cylinder	Multi-cylinder	
Design feature or configuration	OHC vs. OHV, etc.	Four-valve head vs. two-valve head	Inline, Vee, opposed piston, etc.
Fuel utilized	Light-liquid fueled	Heavy-liquid fueled	Multi-fueled (e.g., biodiesel, dual fuel – natural gas and diesel)

engines (one revolution per cycle). According to emissions standards and engine applications, diesel engines are classified as on-road, off-road and stationary. The on-road applications include trucks, buses and automobiles. The off-road applications include marine, industrial (e.g., diesel for compressors), construction equipment, agricultural (e.g., tractors), and locomotive. For emissions standards, the reader is referred to the websites of the US EPA at www.epa.gov, DieselNet at www.DieselNet.com, and Appendix A in Rakopoulos and Giakoumis (2009). According to emissions certification methods for on-road applications, diesel engines are classified as heavy-duty and light-duty (Table 1.2). The American Automotive Manufacturers Association (AAMA) classifies trucks into three categories (heavy-duty, medium-duty, and light-duty) and eight classes by vehicle weight (Table 1.3). Heavy-duty applications include trucks, buses, certain off-road vehicles, marine engines, generator sets (gensets), industrial power plants, etc. Light-duty applications include passenger cars, sport-utility vehicles and light trucks (e.g., some pickup trucks).

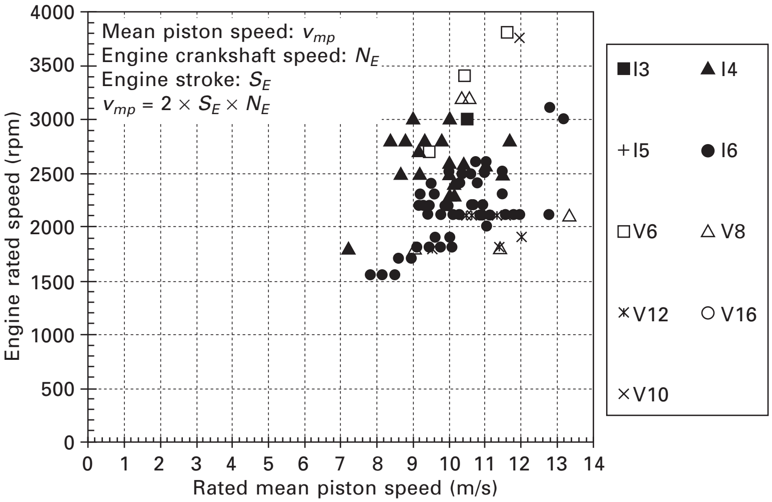
Diesel engines are also classified as low-speed (rated speed of crankshaft rotation slower than approximately 300 rpm), medium-speed (300–1000 rpm) and high-speed (faster than 1000 rpm). They can be also classified by using the mean piston speed at the rated speed. For example, a mean piston speed slower than 6 m/s is regarded as low-speed, 6–9 m/s as medium-speed, and faster than 9 m/s as high-speed. The engine crankshaft rotational speed versus mean piston speed of many high-speed diesel engines in North America is shown in Fig. 1.1. It is observed that there is a large variability in the data due to the differences in engine stroke. There is no clear design trend in the

Table 1.2 EPA vehicle weight classification

Vehicle weight	Classification	Emissions certification
<8500 lbs GVW	Light-duty (LD)	Vehicle chassis
>8500 lbs GVW	Light heavy-duty: 110,000 miles of useful life for compliance with the legislated emissions standards	Engine dynamometer
>8500 lbs GVW	Medium heavy-duty: 185,000 miles of useful life	Engine dynamometer
>8500 lbs GVW	Heavy heavy-duty: 290,000 miles of useful life	Engine dynamometer

Table 1.3 AAMA vehicle weight classification

	Vehicle weight	Classification
Class 1	<6000 lbs GVW	Light-duty (LD) trucks
Class 2	6001–10,000 lbs GVW	Light-duty (LD) trucks
Class 3	10,001–14,000 lbs GVW	Light-duty (LD) trucks
Class 4	14,001–16,000 lbs GVW	Medium-duty (MD) trucks
Class 5	16,001–19,500 lbs GVW	Medium-duty (MD) trucks
Class 6	19,501–26,000 lbs GVW	Medium-duty (MD) trucks
Class 7	26,001–33,000 lbs GVW	Heavy-duty (HD) trucks
Class 8	>33,000 lbs GVW	Heavy-duty (HD) trucks



1.1 North American HD and LD high-speed diesel engines.

correlation between these two speeds. The above classifications according to the speeds are only approximate.

According to different fuel injection methods, diesel engines are classified as indirect injection and direct injection. By air charging method, diesel engines are classified as naturally aspirated, mechanically supercharged and aerodynamically turbocharged engines. By cooling medium, diesel engines are divided into water-cooled and air-cooled. Most modern diesel engines are direct-injection, turbocharged (with inter-cooling or also known as after-cooling or charge air cooling) and water-cooled engines. According to in-cylinder NO_x emissions control and NO_x aftertreatment technology, diesel engines are classified into EGR (exhaust gas recirculation) engine, non-EGR engine, SCR (selective catalytic reduction) engine and non-SCR engine. Moreover, by different fuels utilized or fuel compatibility, diesel engines are classified as light-liquid-fueled, heavy-liquid-fueled or multi-fueled engines. The main body of this book focuses on four-stroke on-road heavy-duty turbocharged EGR diesel engines for automotive vehicle applications.

1.1.2 Comparison between diesel engines and gasoline engines

Understanding the fundamental characteristics of diesel engines is very important for engine system design and powertrain technology assessment. Compared to gasoline engines, diesel engines have the following advantages:

- *Low fuel consumption and low CO_2 emissions*. The high compression ratio used in diesel engines generally results in high thermodynamic cycle efficiency although mechanical friction may increase with peak cylinder pressure. Diesel engines usually use unthrottled operation so that the pumping loss can be lower.

- *High power*. Diesel combustion does not have the severe limitation of auto-ignition as seen in gasoline engines so that diesel engines can use a large cylinder diameter and tolerate a high level of turbocharging in order to produce high power.
- *High torque at low speeds and better drivability*. Diesel combustion can tolerate a high level of turbocharging so that they can burn more fuel to match the available charge air to produce higher torque than gasoline engines.
- *Low carbon monoxide (CO) and hydrocarbons (HC)* due to the high air–fuel ratio employed in diesel combustion.

It should be noted that in modern diesel engines with high EGR for NO_x control, the traditional advantage of unthrottled operation may be lost due to the need to drive EGR flow by closing an intake throttle valve at certain operating conditions, if the system is not carefully designed. Moreover, high EGR rate leads to high peak cylinder pressure due to the increased intake manifold boost pressure required. Such a high cylinder pressure may result in a reduction in the allowable air–fuel ratio or engine compression ratio. The superior fuel economy, high power, and high low-end (low speed) torque are the primary reasons for the dominance of diesel engines as medium- and heavy-duty power plants.

However, there are several design challenges for diesel engines compared with their gasoline counterparts as follows:

- Higher engine-out particulate matter (PM) and smoke due to the combustion with heterogeneous mixtures in the engine cylinder.
- Lower air utilization due to the heterogeneous combustion.
- More difficult control in tailpipe outlet NO_x. The three-way catalyst used for NO_x control on gasoline engines cannot be used in diesel engines because diesel engines are operated with lean air–fuel ratio. Diesel engine emissions control is detailed in Majewski and Khair, 2006)
- Lower exhaust temperature caused by lean burn combustion. This can make diesel particulate filter (DPF) regeneration difficult.
- Higher noise from fuel injection, combustion, and mechanical impact.
- Heavier engine weight: diesel engines need to use heavy structure to endure the high peak cylinder pressure produced by high compression ratio.
- Higher cost, primarily due to the sophisticated and expensive fuel injection equipment and the diesel particulate filter used in diesel engines.
- Lower engine rated speed, due to the limitation of slow combustion speed in the heterogeneous combustion in diesel engines. Instead of having

rated speed at 6000–7000 rpm like in gasoline engines, the rated speed of automotive diesel engines is usually limited to 2000–4000 rpm.
- Lower power density (i.e., lower specific power per volume of engine displacement), which is due to the limitation of rated speed and hence rated power.
- More difficult in cold start.

The above challenges are important system characteristics of diesel engines although they are inherently unfavorable. In diesel engine system design, it is paramount to further enhance the advantages of diesel engines and minimize, or at least not to increase, the disadvantages.

1.1.3 The history, characteristics, and challenges of diesel engines

The history of diesel engines extends back to Rudolf Diesel's pioneering work towards the end of the 19th century. Cummins Jr. (1993) and Grosser (1978) provide a detailed early history of diesel engines and a biography of Rudolf Diesel. The diesel engine is the most efficient liquid-fuel-burning prime mover and has a wide range of applications, from land and marine transportation propulsion to stationary power. The history of the diesel engine is accompanied by its characteristics and continued challenges.

The diesel engine could tolerate a wide range of diesel fuels of different quality due to its compression ignition mechanism. Therefore, before the First World War the diesel engine found its early wide application in ship propulsion and stationary power units where the inexpensive low-quality diesel fuels were usually used. Owing to its fuel injection and combustion mechanisms, the diesel engine runs at lower speeds than the gasoline engine, resulting in lower frictional losses. The diesel engine injects fuel into the cylinder to mix with the compressed hot air near the end of the compression stroke. It does not use an intake throttle valve to regulate the air charge as in the gasoline engine for the purpose of matching a stoichiometric air–fuel ratio. The diesel engine uses higher compression ratio than the gasoline engine, hence possesses a higher indicated thermal efficiency inherently. All the above characteristics (i.e., lower speed, less intake throttle pumping loss, leaner mixture of air and fuel, higher compression ratio) make the diesel engine run at a higher brake thermal efficiency than its gasoline engine counterpart. Therefore, during the first half of the 20th century, the diesel engine experienced rapid development in land transport and became firmly established as the most efficient powerplant for commercial vehicles (trucks and buses). The diesel engine dominated the market of heavy-duty truck, locomotive and marine applications where high efficiency, superior durability and reliability or high low-speed torque are required.

The development of diesel engines has been constantly driven by the requirement of higher power output. Due to the limitation of fuel injection systems and the required time duration for the slower diffusion combustion rate of heterogeneous combustion, the diesel engine cannot run at a very high speed to produce the high power density that is offered by a higher-speed gasoline engine. In the early years without supercharging, this resulted in a large gap in power density between the diesel and gasoline engines. The advent of supercharging technology changed that situation. The low-octane characteristic of the diesel fuel allows the diesel engine to be supercharged to withstand very high cylinder pressure and temperature without the risk of auto-ignition or 'knocking' that is encountered in the gasoline engine. The turbocharging and inter-cooling technology widely adopted during the second half of the 20th century greatly increased the power density of the diesel engine. According to Hikosaka (1997), by applying turbocharging the maximum BMEP (brake mean effective pressure) of heavy-duty diesel engines has increased exponentially from approximately 5 bar in the 1920s to in excess of 20 bar in the 1990s. In the passenger car market, turbocharging has triggered a sharp rise in power density of light-duty diesel engines since the 1980s to a level close to that of gasoline engines (Hikosaka, 1997). The diesel engine has become more compact, more powerful, and less noisy. The higher BMEP and exhaust energy recovery resulting from turbocharging has further enhanced the thermal efficiency of the diesel engine. On the other hand, it should be noted that the higher-BMEP engine demands a more durable structure (e.g., crankshaft, piston assembly) in order to sustain a higher peak cylinder pressure.

The oil crisis in the 1970s gave a considerable impetus to develop fuel efficient powertrains from the perspective of 'well-to-wheel' efficiency, which is defined as a combination of the efficiencies from 'well-to-tank' and 'tank-to-wheel'. Despite the disadvantages of higher noise (most apparent at idle) and higher cost, the diesel penetration into the light-vehicle and passenger-car markets has continued growing rapidly, especially in Europe where fuel prices are high and fuel taxation policies favor diesel.

The requirement of lower NVH (noise, vibration and harshness) has become increasingly important for personal transportation powertrains. Indirect injection (IDI) diesel engines produce lower combustion noise than direct injection (DI) engines at the expense of fuel consumption. Until the 1990s, indirect injection technology had a dominant market share in light-duty vehicles. With the advances in fuel injection systems to tackle the combustion noise problem and the improvement in structural design to reduce NVH, DI diesel engines have gradually replaced IDI engines in the light-duty market since the late 1980s. The commercial vehicle market has not been as sensitive to the noise issue as the light-duty market, and hence direct injection technology has been widely applied. Direct injection diesel engines have approximately 10–15% better fuel economy than indirect injection diesel engines, and 30–40% better than the port-injected gasoline engines. The reasons for DI engines' better fuel economy mainly include the following. First, DI engines do not use a divided combustion chamber (pre-chamber) and hence do not have its associated high pumping loss due to the flow restriction at the throat in the pre-chamber or swirl chamber. Secondly, DI engines have less heat rejection lost to the coolant due to less heat transfer area of the combustion chamber. A comparison of the fuel consumption between different engines was given by Hikosaka (1997).

Numerous advances have been made continuously to improve diesel engine designs during the past several decades, resulting in the reduction in fuel consumption, weight, NVH, and cost, as well as in the enhancement of transient performance, durability, and reliability. Engine friction and lube oil consumption have been reduced by improved piston-assembly designs. Pumping loss has been reduced by both carefully matching the turbocharger and improving the volumetric efficiency of the engine (e.g., by using the four-valve-head design). Mixed-flow and variable-geometry turbochargers have been used to reduce engine pumping loss and improve transient response. Variable-valve actuation and variable swirl have been researched to find their applications in diesel engines. Various combustion chamber designs including both quiescent and swirl-supported have been proposed in order to improve combustion efficiency and reduce heat losses. More compact and less restrictive heat exchangers have become available. Engine accessory parasitic losses have been greatly reduced due to the improvement made by suppliers.

Waste heat recovery has received attention since the 1980s. The potential of turbocompounding was investigated to try to recover some of the exhaust heat at the turbocharger's turbine outlet to drive another turbine in order to deliver some mechanical power that can be added back to the engine crankshaft. Moreover, it is worth noting the research on the low-heat-rejection engine (or the so-called 'adiabatic' engine) during the 1980–1990s. The research was initiated with the motivation of using large recovery of heat losses by insulating the power cylinder surfaces (e.g., piston, cylinder head, and liner). Soon after, it was discovered that the fuel economy benefit of the adiabatic engine was far less than initially imagined. Many other design challenges were also encountered, such as in-cylinder tribological issues and excessively high NO_x emissions. Since the more stringent emissions regulations were stipulated, the interest in the adiabatic engine gradually died down, at least for the vehicle categories that are not exempted from the stringent emissions regulations. After all, the development of the adiabatic diesel engine stimulated the interest in applying analysis of the second law of thermodynamics to internal combustion engines. The development also accumulated valuable experience and lessons learned for in-cylinder heat rejection control, which is still an important topic for modern low-NO_x engine design.

The worldwide enactment of emissions regulations (mainly in the US, Europe and Japan) for on-road and followed by off-road applications since the late 1990s gave another and the latest major impetus for the technological development of modern diesel engines, cleaner fuels and lubricant oils. In the US, heavy-duty on-road engines are certified according to the FTP (federal test procedure), SET (supplemental emissions test) and NTE (not-to-exceed) emissions regulations. Light-duty on-road engines are certified according to the FTP-75 and US06 regulations. Compared to its gasoline engine counterpart equipped with the three-way catalyst, the diesel engine without aftertreatment is characterized by lower unburned hydrocarbons (HC), negligible carbon monoxide (CO), and higher particulate matter (PM) and nitrogen oxides (NO_x, primarily NO and NO_2 with a NO_2/NO ratio approximately 5–25%) at the tailpipe. The reason for the lower HC and CO emissions of the diesel engine is that it operates at very lean air–fuel ratios, especially at part load. The diesel engine has near zero evaporative emissions due to lower volatility of the diesel fuel. The diesel engine also has lower cold-start emissions than the gasoline engine. The emissions challenge for the diesel engine is mainly NO_x and soot control. The diesel engine inherently possesses several trade-offs in emissions and performance, namely, the NO_x–soot trade-off, the NO_x–HC trade-off, and the NO_x–BSFC trade-off. The trade-off means improving one parameter is impossible without sacrificing the other when a design or calibration factor is changed. A typical example of the trade-off is the fuel injection timing effect. More retarded fuel injection timing results in a decrease in NO_x, but an increase in soot and BSFC (brake specific fuel consumption). In order to break the trade-off, another design factor needs to be improved. For instance, when fuel injection pressure is increased the soot emission and BSFC can be reduced. The design challenge has always been a simultaneous reduction of exhaust emissions and fuel consumption. Emissions compliance, power density, structural strength, and reliability are inter-related. Lower emissions require higher air or EGR flow and hence higher peak cylinder pressure, and this competes with the need for higher power density. The increase in power density of the engine has been slowing down since the emissions regulations became more stringent. Advanced high-pressure fuel system, electronic controls, aftertreatment (e.g., DOC or diesel oxidation catalyst, DPF) and low-sulfur fuel were the four major areas emerging since the 1980s for diesel emissions control.

Another important feature for modern diesel engines to meet the US 2004–2010 emissions regulations is EGR. EGR has been used as a very effective technology to reduce NO_x, especially in the US. In certain European applications, liquid urea-based selective catalytic reduction (SCR) was used instead of EGR to control NO_x. However, there are concerns about urea-based

SCR's capital and operating costs, weight, packaging space, increased exhaust restriction, complexity of electronic controls, ammonia slip, compatibility in very cold weather, readiness of infrastructure, maintenance burdens on vehicle customers, in-use emissions compliance, etc.

EGR reduces NO_x through the mechanism of reducing the in-cylinder oxygen concentration and combustion temperature. Turbocharging with inter-cooling has been used to increase power density and reduce PM emissions. EGR is used in modern engines mainly for NO_x control. Compared to the turbocharged non-EGR engines in the old era, EGR engines have very different design considerations from system to component. For instance, the usual practice in non-EGR engine designs was to match the turbocharger to make the intake manifold pressure higher than the exhaust manifold pressure because there was no need to drive EGR flow. In this case, a negative value of the pressure differential or negative 'engine delta P' was created. The engine delta P here is defined as exhaust manifold pressure minus intake manifold pressure. Such a negative engine delta P not only resulted in a net gain in BSFC due to a positive pumping work rather than a negative pumping loss, but also improved cylinder gas scavenging by using a large valve overlap to reduce the thermal load acting on the cylinder head, the exhaust valve/manifold, and the turbine. The valve overlap here refers to the timing difference between exhaust valve closing and intake valve opening. During the valve overlap period, both the exhaust valve and the intake valve are open. Moreover, the air–fuel ratio (A/F ratio) in non-EGR engines could be designed very high without the problem of exceeding the maximum cylinder pressure limit. The high A/F ratio increases combustion efficiency and decreases soot. However, in EGR engines those advantages disappear because the need to drive EGR flow requires a positive value of engine delta P (i.e., exhaust manifold pressure higher than intake manifold pressure). The increased exhaust restriction due to the addition of PM and NO_x aftertreatment devices further complicates the turbocharger matching. Moreover, the fluctuations of engine gas flow and temperature caused by EGR-on and EGR-off operations during transients or aftertreatment regeneration demand a careful system design approach in order to optimize all the subsystems involved. Other design challenges related to EGR include controlling intake condensate, coolant heat rejection and engine component wear.

In addition to EGR, modern diesel engines are also characterized by several other emissions control technologies. Innovative combustion concepts (e.g., low temperature combustion, HCCI) are being investigated intensively. Ultra-high injection pressures are demanded for fuel systems, along with the requirement of achieving high injection pressures at both low and high engine speeds (e.g., high-pressure common rail fuel system). Retarding fuel injection timing has become necessary in order to meet the most stringent NO_x standard and this results in a penalty on BSFC. With the retarded injection timing, the combustion pressure is no longer higher than the compression pressure at the firing TDC (top dead center) at many speed and load modes including full load. The compression pressure usually becomes the peak cylinder pressure. Multiple injection methods (pilot, main, and post injections) and injection rate shaping have also become necessary. The direction for clean and efficient combustion was illustrated by Hikosaka (1997) in heat release rate analysis. Intake charge cooling for the air–EGR mixture has been demonstrated as an effective solution for NO_x control, although there are challenges in controlling intake condensate and heat rejection. Moreover, DPF has become necessary to meet the stringent US 2007 emissions regulation. Advanced EGR, air management, fuel injection, combustion, aftertreatment, and electronic controls are the six key technology enablers for modern diesel engines to meet emissions standards.

Advanced fuel formulation (lower aromatics, ultra-low sulfur) and improvement in lubricant additives (for ash control and aftetreatment compatibility to avoid poisoning or fouling) are also required to meet the stringent emissions regulations. For example, the US EPA 2007 regulation is supported by the 2006 standard of ultra-low-sulfur diesel fuel. The sulfur content has been reduced from 500 ppm to 15 ppm starting in mid-2006 for the diesel fuels used in on-road vehicles. Such ultra-low-sulfur fuel will be required for most off-road vehicles/applications as well.

Meeting the coming fuel economy regulations and continuously improving engine fuel consumption is the next major impetus for diesel engine technology development. The diesel engine will remain as the most popular powerplant for the next several decades before fossil fuel resources become depleted. Almost the entire heavy-duty segment and much of the medium-duty segment run on diesel power today (Boesel *et al.*, 2003). Off-highway vehicles are also dominated by diesel power. According to the data from the US EPA, 23% of the energy used in US highway transportation was consumed by heavy trucks, 1% by buses, 32% by light trucks, and 44% by automobiles. Currently, meeting emissions regulations is the driving force for developing better engines. In the future, it will be fuel economy, along with greenhouse gases, NVH, reliability and cost. The regulations on CO_2 emission or fuel economy impose another challenge on engine manufacturers. There is a direct correlation between CO_2 emission and fuel economy (Menne and Rechs, 2002; Steinberg and Goblau, 2004). As pointed out by Dopson *et al.* (1995), the approach to fuel economy legislation has differed between the two main regulation bodies, the US and Europe. The US method was determined by the CAFE (corporate average fuel economy) standard which allows manufacturers to mix their fleet production and sales. The European method for fuel economy took the CO_2-only regulation. Some well-known projects of fuel economy development in the past included the 55% thermal efficiency heavy-duty engines in the US, the 3 L/100 km (equivalent to 78.4

mpg) fuel consumption passenger cars in Europe, and the PNGV (Partnership for a New Generation of Vehicles) program of 80 mpg cars in the US.

On the light-duty side, the PNGV program was formed by the US federal government and the United States Council for Automotive Research (USCAR), which was created by several major US automobile OEMs including Ford Motor Company, General Motors Corporation and Chrysler Corporation. The PNGV fuel economy target and the necessity of reducing vehicle weight were discussed by Boggs *et al*. (1997). The high-speed direct injection (HSDI) diesel engine has been identified as the most promising and fuel-efficient power plant (for conventional powertrains) or primary power source (for hybrid powertrains) in the program. Engine downsizing in displacement has become a trend in order to achieve high thermal efficiency in future passenger cars or hybrid vehicles. However, downsizing poses significant challenges in combustion, structural design, and NVH (e.g., converting a four-cylinder engine to three-cylinder, Ecker *et al*., 2000). The smaller engine displacement has a deteriorated combustion chamber surface-to-volume ratio and reduced proportion of air available for combustion. Improving fuel consumption while meeting the most stringent US Tier-2 Bin-5 light-duty emissions regulation with cost-effective engine-aftertreatment design is a major challenge for light-duty diesel engines.

On the heavy-duty side, research into exploring thermal efficiency improvement and better alternative fuels has been very active. The focus of heavy vehicle technology development has centered on reducing emissions, improving fuel economy, and using fuels from alternative non-petroleum feedstock (Eberhardt, 1999). Heavy-duty diesel engines running on conventional petroleum-based fuel have achieved more than 40% thermal efficiency, compared to the commonly reported 30% efficiency of gasoline engines. Recent investigations indicated that a thermal efficiency around 55% is achievable (Eberhardt, 1999). There are four main areas in future diesel engine development to achieve emissions compliance and high thermal efficiency while reducing the overall engine–vehicle cost:

1. in-cylinder combustion development;
2. exhaust aftertreatment;
3. hybrid powertrain engineering; and
4. fuel quality and alternative fuels.

Advanced clean diesel engine powertrain will still dominate the majority of the market share of heavy-duty sectors. However, hybrid-diesel powertrains are likely to gain high popularity as the most advanced powertrain technology in heavy-duty vehicles over the next few decades. Significant emissions reduction and fuel savings can be achieved with hybrid-electric or hybrid-hydraulic powertrain due to the benefits of regenerative braking energy recovery (especially for the urban stop-and-go duty cycles such as in delivery and transit transportations), engine idle-off, engine downsizing, and optimized engine operating conditions for driving.

Natural gas engines (e.g., CNG – compressed natural gas, LNG – liquefied natural gas) and dual-fuel natural-gas diesel engines will also increase their share in heavy-duty vehicles, especially in transit buses and urban delivery/ refuse trucks. Biodiesel, GTL (gas-to-liquid) or synthetic diesel fuels will be seen in more applications. Fuel cell as a primary propulsion power (versus as an APU – auxiliary power unit) still faces significant technical and cost challenges for commercialization in the foreseeable future.

Competition over fuel economy or fluid economy (e.g., including both fuel and aftertreatment fluid) will be fiercer in the market for customer satisfaction. Not only will competition happen among diesel engine manufacturers, it will also occur against the challenges from other advanced powertrain concepts such as the gasoline direct injection (GDI) engine (Asmus, 1999). Future diesel engines will be characterized by the advanced design features related to fuel economy improvement such as waste heat recovery, cylinder deactivation, variable valve actuation, hybrid electric/hydraulic integration, and advanced intelligent controls. All these challenges require an optimized system approach in diesel engine design (Xin, 2010) and vehicle powertrain integration (SAE PT-87, 2003; Seger *et al*., 2010). Applying a system design approach can also support technology innovation in each subsystem and offer new solutions to the traditional mechanical design issues such as friction and NVH reduction as well. The goal of advanced engine system design is to achieve a superior product with low cost and fast pace during the development cycle.

1.2 The concept of systems engineering in diesel engine system design

1.2.1 Principles of systems engineering

Definition of systems engineering

Systems engineering has been successfully employed for decades in aerospace and defense product development. Its sporadic application in the automotive industry emerged in the 1990s. There was no theory in the past on how to apply the principles of systems engineering to diesel engine system design. According to the definition of Kossiakoff and Sweet (2003) in their classic textbook on systems engineering, the function of systems engineering is ‘to guide the engineering of complex systems’. The system is defined as ‘a set of interrelated components working together toward some common objective’. In fact, there is no unanimity whether it should be spelled ‘systems engineering’ or ‘system engineering’. In diesel engine system design, there is only one system – the engine. Complex systems have a hierarchical structure that consists of a number of interacting elements called subsystems. A subsystem is composed of simpler functional entities such as subassemblies, components, subcomponents, and parts. A system may also become a subsystem if the hierarchical chain is expanded to a higher level. As pointed out by Kossiakoff and Sweet (2003), 'since the systems engineering function is that of guidance, authority is exercised by establishing goals (requirements and specifications), formulating task assignments, conducting evaluations (design reviews, analysis, and test), and controlling the configuration.'

From the INCOSE (International Council on Systems Engineering) definition quoted by Kanefsky *et al.* (1999), Armstrong (2002) and Austin (2007),

> Systems Engineering is an interdisciplinary approach and means to enable the realization of successful systems. It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, and then proceeding with design synthesis and system validation while considering the following complete problem: performance, cost & schedule, test, manufacturing, training & support, operations, and disposal. Systems engineering integrates these disciplines and specialty groups into a team effort forming a structured development process that proceeds from concept to production and to operation. Systems engineering considers both the business and the technical needs of all customers with the goal of providing a quality product that meets user needs.

Another definition given by Austin (2007) states that 'systems engineering is a formal process for the development of a complex system, driven by a set of established requirements, derived from the intended mission of the system throughout its life cycle.'

Jackson *et al.* (1991) attempted to introduce systems engineering to the automotive industry and applied the principles to automotive transmissions design. They explained the procedures and the tools of systems engineering used to plan, coordinate and execute in product development. Kanefsky *et al.* (1999) provided an overview on using a systems engineering approach in engine cooling design, focused on requirements analysis, functional analysis and target setting. Armstrong (2002) summarized the roles of systems engineering in different product development stages, with examples of electronic controls integration, in his SAE Buckendale Lecture. Austin (2007) provided an excellent introduction to systems engineering methodology for automotive engineering. The most comprehensive coverage on the general theory of systems engineering has been provided by Kossiakoff and Sweet (2003). In the following sections, key principles of systems engineering are examined and summarized first. Then, the disadvantages of certain traditional viewpoints of systems engineering are pointed out, and a new theory of systems engineering suitable for diesel engine system design is developed.

Interfaces – the focus of systems engineering

The performance of a system is affected by its external environment (e.g., ambient temperature and altitude) or interacts with other parallel systems (e.g., an engine vs. vehicle drivetrain). The definition and control of these effects or interactions at the external interfaces of the system is a particular responsibility of a system engineer. Inside the system, there are interactions occurring at various internal boundaries or interfaces of the subsystems. Again, the definition of the internal interfaces and their control for compatibility and reliability is also a primary responsibility of the system engineer. The control is applied to the elements that connect, isolate, or convert interactions. Experience has shown that a large fraction of system failures occurs at the interfaces. Strict control of the interface design between system elements is a critical topic in systems engineering. Interface requirements need to be included or reflected in system design specifications. Moreover, interface requirements validation needs to be included in the system validation plan.

Phases and roles of systems engineering

The systems engineering process is usually explained by a 'V' diagram (Austin, 2007). The left leg of the 'V shape' contains a top-down cascaded process from the system, subsystem to component level about product definition, requirements analysis, as well as development and allocation of design specifications. The bottom of the V shape has the product design and prototype build. The right leg of the V shape contains a bottom-up integration process from the component, subsystem to system level about product verification and testing validation. The left and right legs of the V shape are connected or communicated by passing the validation result.

The activities in systems engineering consist of three stages, namely concept development, engineering development, and post-development, in a system life cycle model. According to Kossiakoff and Sweet (2003), the concept development includes the following: establishing the system need; exploring feasible concepts to define functional performance requirements; selecting preferred system concept based on performance, cost, schedule, and risk; and defining system functional specifications. The engineering development includes the following: validating new technologies in advanced development; identifying and reducing risks; transforming concept into hardware and software designs according to design specifications; and building and testing the integrated production prototypes. The post-development includes producing and deploying the system as well as supporting system operation and maintenance.

Systems engineering is a part of project and program management. Systems engineering is a unified process of both engineering and management. As pointed out by Armstrong (2002), the essence of program management is the trade-offs between technical parameters, schedule, cost, market forces, and available resources to complete the total program. The tasks in systems engineering include the following:

- Assisting project and program management to manage statements of work and risks.
- System concept/architecture design or selection.
- System requirements analysis based on customer agreements (musts and wants) and constraints.
- System functional analysis to translate requirements from the customer domain to the engineering domain.
- Target setting.
- Identifying and managing key interfaces between system elements.
- System specification design with synthesis, optimization, and balance.
- Allocation of cascaded design specifications from the system level to the subsystem and component levels for both hardware and software systems.
- Technical coordination and guidance for subsystems with disciplined decision-making through system design reviews.
- System validation test to verify the requirements and the product through integration with good documentation.

Detailed descriptions of the above tasks are provided in Kanefsky *et al.* (1999), Armstrong (2002), and Kossiakoff and Sweet (2003). Systems engineering provides an orderly process for complex system development, and such a process is able to identify and resolve key problems in system design prior to subsystem designs. The role of systems engineering in the entire program is heavily biased toward the early stages (i.e., 'front' loading of work and resources). Systems engineering emphasizes the up-front analysis of the functional objectives, requirements, and constraints to define a logical product with customer input. Many experiences have shown that the later a requirements problem is identified, the more it costs to fix. In fact, it is not a linear proportionality, but usually exponential (Austin, 2007). One of the key aspects in systems engineering process is that the design must be strictly led by the complete analysis of functional requirements and system specifications.

Functional requirements are the basis of design specifications. Functional requirements are the reason that the item is included in the design. Requirements management is as important as concept/configuration management. Changing system requirements as a moving target makes system design difficult and unstable. However, very often changes required by the customer are inevitable. And sometimes it is difficult to accurately define the engineering target for the requirements, for example, accurately predicting the air–fuel ratio and EGR rate required to meet a certain level of NO_x emissions at the early stage of an engine program. It is necessary to document requirements traceability by recording the requirements and scoping the changes through system, subsystem, and component specifications.

Systems engineering allocates and balances the requirements to decide how much technology development risk each element should undertake, and decide the best trade-offs among all design attributes (e.g., performance, weight, size, cost) and elements. During the detailed design and testing phases, the role of systems engineering is to maintain the revised system design specifications and ensure that the system development follows the up-to-date specifications with good traceability from the original requirements to the final design. In addition to cascading system requirements and specifications at the early stages of the program, the system engineer should integrate the designs from the components and subsystems to the system, and manage their interfaces during the later system validation phases.

The system engineer also has the responsibility to take the initiative to establish a structured communication network between the program requirements and all subsystems. The system engineer plays an active role in program management meetings to review design status and resolve trade-offs. The systems engineering process has been proven to be able to reduce the amount and severity of the problems discovered in the late stages of the development program, and meanwhile to reduce the cost.

Risk management

Risk refers to the 'likelihood' of failing to meet requirements and the 'severity' or seriousness of the impact of such a failure to the success of the program. The methodology that is employed to identify and minimize risk in system development is called risk management. A more quantitative risk assessment is to consider the issue's probability of occurrence and its consequence should it occur (Austin, 2007). Weighting factors can be used to prioritize the risks. At the beginning of the program, the risk is high due to unpredictable adverse events. As the development progresses, risks are systematically reduced and eliminated by analysis and testing. Risk needs to be reduced largely in the concept development stage when the advanced technology development and system design occur. Advanced development matures new technologies or concepts and turns them into production-viable technologies. System design ensures trade-offs are analyzed and risks are assessed, balanced, and mitigated. The methods of risk mitigation include relief of excessive requirements and preparation of fallback alternatives, and so on. More details on system risks, design risks and risk reduction techniques are introduced in the later section of robust engineering principles.

Examples of the risks for engine development programs include increased peak cylinder pressure for structure strength, high heat rejection, conflicting packaging space (e.g., complicated turbochargers, cooling modules, aftertreatment devices), engine weight, product cost, etc.

Trade-offs and balanced design decisions

The principal viewpoint of systems engineering is to achieve the best balance among critical attributes or among the design trade-offs of various subsystems for the success of the system as a whole. Different system attributes should be appropriately weighted. An example of the best balance is the trade-off between performance (i.e., function) and cost. A curve of performance-to-cost ratio (or function-to-cost ratio, value ratio, cost-effectiveness) can be established as a function of cost so that the performance of the best overall system can be chosen close to the peak of the performance-to-cost ratio from the curve, as long as the peak point also satisfies the minimum acceptable performance. The principle of balanced system ensures that in system design decisions no system attribute is allowed to grow or become over-designed at the expense of other equally important attributes. It is important to conduct such value-oriented designs in engine system development (i.e., using a ratio of performance to cost). Another example of a balanced decision is the trade-off between technology advance and risk. New advanced technologies may contain many uncertainties (risks) and opportunities. Existing production technologies contain minimum risks but they may become obsolete very quickly. Making the right recommendation on technology path with affordable and balanced risk is a primary responsibility of a system engineer. The system engineer needs to have in-depth technical expertise and broad vision for overall project management in order to be able to make the right balanced decisions for the program. The system engineer is the ultimate authority on how the goals of system performance and affordability are achieved simultaneously.

Modularization and integration

Another principle of systems engineering is component modularization and integration in order to ease technical management. System integration focuses on the interfaces between system elements. Modulated design is easier to manage and assemble. Vehicle cooling module and powertrain control module are two examples in hardware design and electronic controls, respectively. Some of the modules are designed, tested, and provided by full-service system suppliers, while others are integrated by the OEM. Modularizing the system and integrating the subsystems at the interfaces is one of the central tasks in systems engineering.

Qualifications of system engineers

The need for system engineers is most apparent in large complex systems. The systems engineering function needs to adapt to the pre-existing organizational structure. In essence, the system engineer is undoubtedly the most influential leader for the entire product design. The subsystem specialist works downstream of the system engineer to realize the complete design of the given subsystem. If every subsystem specialist does a good job according to the coordination of the system engineer, the whole system will come out in good shape as a summation of all the subsystems with a high degree of integration. This process requires the system engineer to possess the following qualifications:

- have solid background in one or more traditional engineering disciplines;
- have demonstrated technical achievements and authority in one or more specialty areas;
- enjoy the interdisciplinary learning and the challenges from a wide range of technical disciplines;
- be familiar with all functional disciplines of the organization and knowledgeable in products and processes, especially system-level integration;
- have the courage to break the technical language barriers to bridge the gaps between different specialists;
- be proficient in handling large-scale optimization analysis;
- be open minded about new ideas and processes;
- have good communication skills and process management skills to lead a project and maintain effective communications among the many interacting individuals and groups.

1.2.2 Challenges of diesel engine system design in systems engineering

Academic background of engine system engineers

The traditional theory of systems engineering (Kossiakoff and Sweet, 2003) believed that systems engineering bridges the conventional engineering disciplines by guiding and coordinating the design of each individual element to assure that the interactions and interfaces between system elements are compatible and mutually supporting. The theory admits that systems engineering as a profession defined as such has not been widely recognized because systems engineering does not correspond to the conventional academic engineering disciplines such as mechanical or electrical engineering. The absence of a quantitative knowledge or specific technical expertise has inhibited traditional systems engineering from being recognized as a unique discipline. In fact, most technical people resist becoming generalists for fear that they will lose the recognition from their specialized profession.

Because the traditional theory of systems engineering considers that a job function of integration without getting into the design details is probably sufficient for system engineers, it is believed that graduates majoring in mathematics or physics without specialized engineering training are capable of the job. Unfortunately, this is usually not the case in diesel engine system design. A job function of sole coordination without producing system-level technical design data has been proven to be ineffective in engine companies. The engine system engineer needs to possess specialized skills in the conventional engineering disciplines in order to work on both system integration and guide component/subsystem design details if needed. Past experience has shown that in diesel engine system design, the most successful system engineers come out of the following engineering specialties: thermal/fluid science, combustion, dynamics, or electronic controls.

Technical breadth and depth

The traditional view of systems engineering requires a three-dimensional expertise in the following: great technical breadth, moderate technical depth, and moderate management skills. However, the term 'breadth' has not been clearly defined. In diesel engine system design, breadth means both the subsystems (e.g., valvetrain, turbocharger) and the attributes (i.e., performance, durability, packaging, and cost). The depth means the level of understanding of the details of a given subject, such as the understanding of the relevant influential factors, theoretical and experimental knowledge, and the interactions between different subjects. The subject here can refer to either a subsystem or an attribute.

Job responsibilities

The traditional view of systems engineering (Armstrong, 2002) believes that the systems engineering group within an organization needs to consist of an interdisciplinary team of people working at the system level and representing all the related functional areas. In essence, the systems engineers are believed to be technical project managers, working very closely with the design and development departments. In diesel engine system design, it is desirable for the system engineer to possess in-depth knowledge on certain subsystems and components. This is largely due to the requirements of system engineers working closely with component engineers to conduct component design that critically affects the system-level performance.

The traditional theory of systems engineering has been challenged by some negative voices since its inception, such as 'what is the value of "adding" a system engineer on top of specialists?' Such voices were raised because the complexity of the system design was not clearly comprehended. With the attribute-driven system design approach, the benefits of system engineers who are required to possess in-depth understanding of component designs are obvious. The approach does pose two challenges for the engineering staff. For system engineers, the expertise in a given attribute for all subsystems is highly desirable. For subsystem specialists, the capability of coordinating different attributes for a given subsystem is strongly recommended.

Design tools for system engineers

In diesel engine system design, a system engineer is required to conduct accurate and sophisticated calculations at a system optimization level by using advanced simulation tools. The approximate calculations for the complex engine processes are often inadequate. The high quality system design specifications generated by a system engineer ensures the successful design of subsystems and components.

1.2.3 Systems engineering in diesel engine system design – an attribute-driven design process

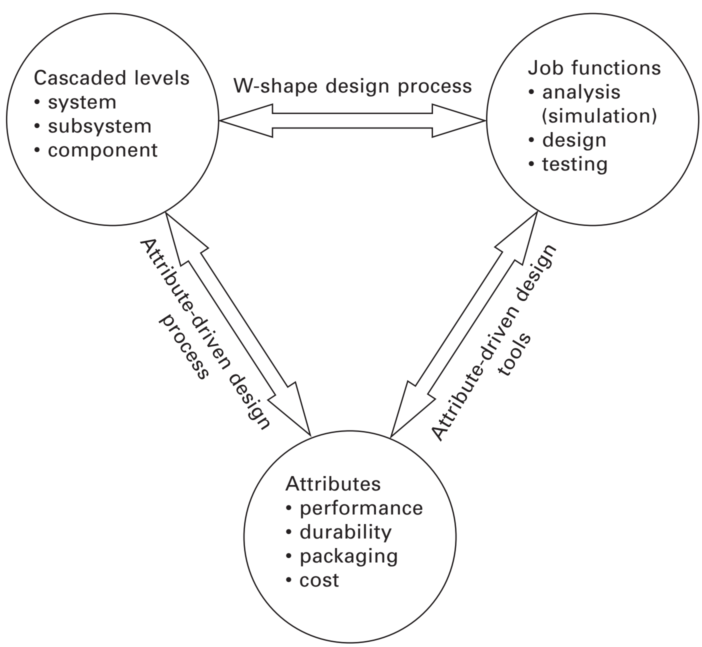
Attribute-driven system design process

The traditional theory of systems engineering (Armstrong, 2002) divided design activities into two phases: a preliminary design followed by a detailed design. The preliminary design follows two earlier phases in the systems engineering process, namely system requirements analysis and functional analysis. 'The functional analysis identifies the basic system functions and translates the previously defined requirements into specific system performance characteristics and constraints' (Armstrong, 1996). The detailed design phase is followed by the phases of prototype building, testing, integration, production, and support. Generating system specifications was regarded as a part of preliminary design. It was believed that each engineering specialty or subsystem specialty team conducts the detailed design by using the system specification as a guideline.

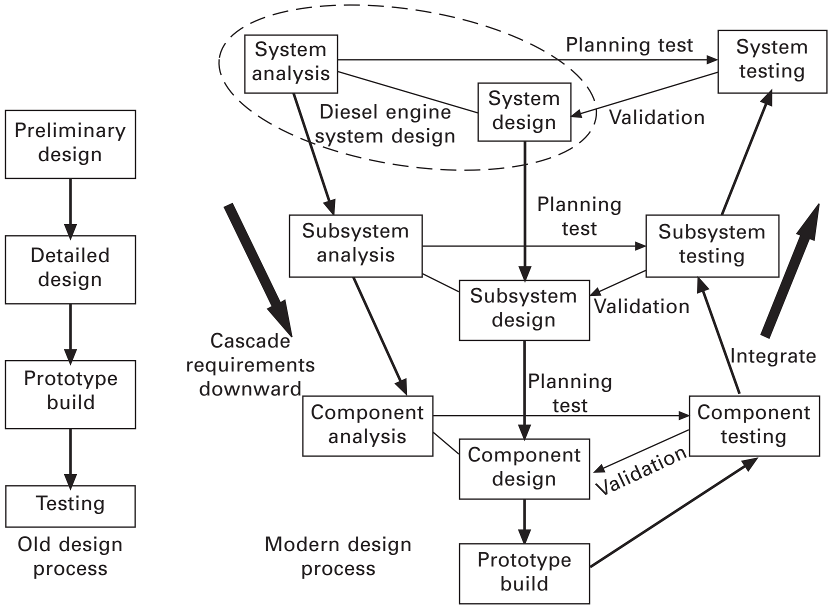
In diesel engine system design, because of the need for precise system specifications rather than a rough preliminary estimate, some of the detailed design work needs to be pulled up to the system level. For example, the detailed turbocharger matching needs to be conducted at the system specification level rather than a later detailed design stage. Therefore, it is not appropriate to differentiate the engine design between preliminary and detailed designs. Instead, the design work should be classified into system design, subsystem design and component design. For each design level, the work can be further classified into four attributes (Fig. 1.2):

- performance
- durability
- packaging
- cost.

The design work is shared properly by the system engineer and the subsystem specialist. In that sense, the performance attribute related work may be considered as the 'preliminary design' defined by the traditional theory (although its nature of design is already detailed and not preliminary at all), and the durability and packaging related design work can be regarded as the 'detailed design' in the traditional theory. The theory of diesel engine system design eliminates the 'preliminary design' stage by using an attribute-driven design approach woven with a three-layer cascaded process (Fig. 1.3). In addition, the system-level design can directly finalize the details of component level design if that component has critical interaction with others to affect the performance of the whole engine system. One example of this system-oriented design process is cam profile design. The overall camshaft design itself belongs to the component level, but the cam lobe profile has a dominant



1.2 Design space and elements of systems engineering for diesel engine system design.



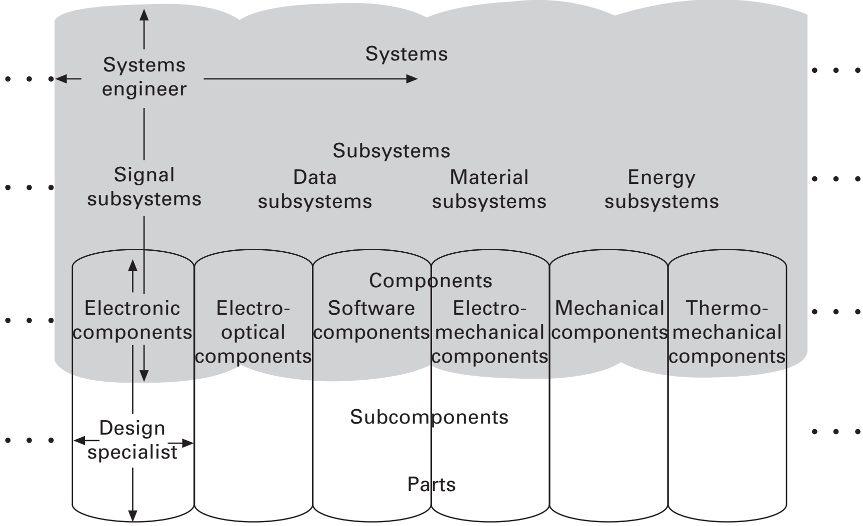
1.3 The 'W-shaped' systems engineering process and functions of engine system design.

impact on overall engine system performance, from volumetric efficiency to valvetrain dynamics. Therefore, the cam profile design is finalized at the system design level by the system engineer.

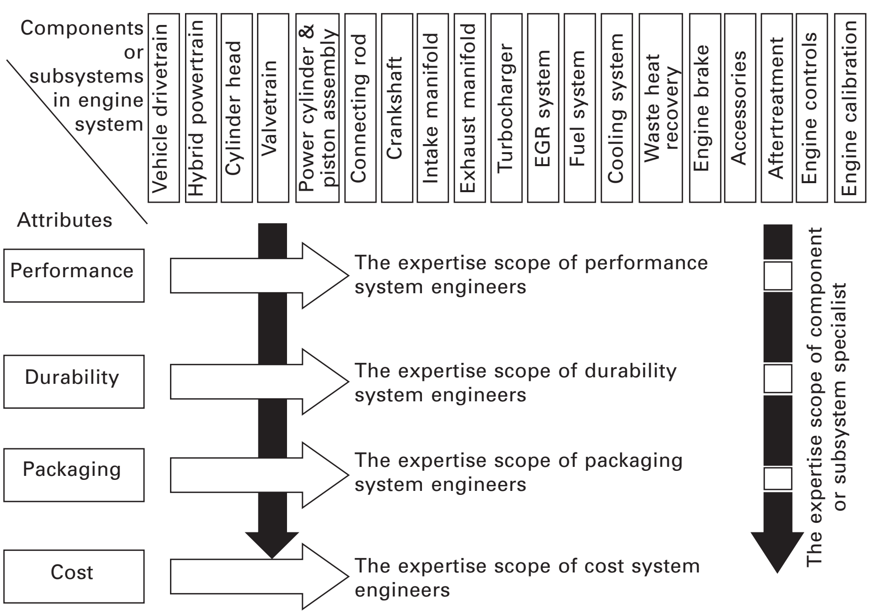
Attribute versus component domain and work scope

The systems engineering theory believes that the system engineers use their technical knowledge of the whole system to guide the system development (Fig. 1.4). For complex engine design, it is impossible for one person to know all the related areas. Therefore, it is necessary to define the work scope of the system engineers. In engine development, the design area can be described as a two-dimensional space of attribute versus component or subsystem (Fig. 1.5). The components or subsystems refer to the physical entities that a design engineer claims accountability for (e.g., cylinder head, turbocharger). The attributes refer to the factors to be considered for a given component, i.e., performance (function), durability, packaging (weight, geometry, appearance, and assembly), and cost.

Different activities and tools or job functions are required to achieve each attribute for each component. For example, performance development requires the activities in performance simulation analysis, solid modeling, design prototyping, and functional testing. Durability development requires



1.4 Knowledge domain of system engineer and design specialist (from Kossiakoff and Sweet, 2003).



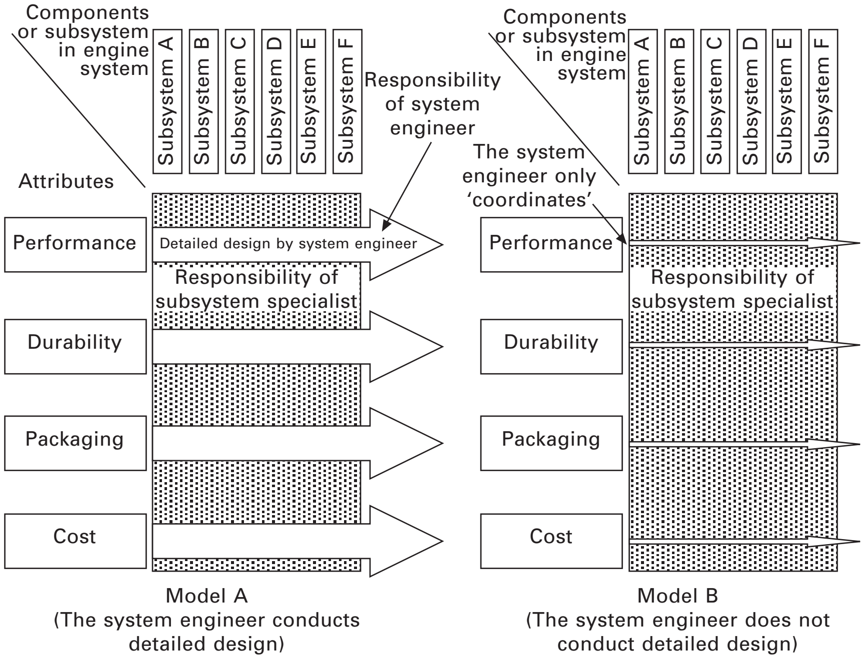
1.5 Subsystem areas and attributes in engine design in the attribute-driven design process.

activities in structural analysis, design prototyping, and durability testing. The packaging attribute requires solid modeling on the assemblies to check geometrical interference, drawing design, as well as performance and durability analyses to check functional interference. Packaging is usually not related to testing activity. The cost attribute requires mainly the activity of analysis rather than design and testing. As shown in Fig. 1.5, one person usually cannot become proficient in everything in the entire two-dimensional design space. Dividing the work scope along the horizontal and vertical directions makes the scope much more manageable and in good order.

Coordination between engineers

Figure 1.5 shows that there are two types of engineers: system engineers and subsystem (or component) specialists. In diesel engine system design, four types of system engineers are defined, namely, performance system engineer, durability system engineer, packaging system engineer, and cost system engineer. Usually the performance system engineer takes the lead in the entire system design because of the importance of system function in driving the design in the other three attributes. The expertise of each type of system engineer spans horizontally across all the subsystems or components (Fig. 1.5). The necessity of such a horizontal arrangement (or technical breadth) is the fact that the purpose of system design is to resolve the conflicts among all subsystems and integrate them as a whole. On the other hand, the vertical arrangement (or technical depth) of expertise within a given component (or subsystem) entails the career development of a component engineer or specialist. The necessity of such a vertical arrangement is the fact that the component engineer has the final ownership and accountability to reconcile all the conflicts among the four attributes for a given component. Moreover, as shown in Fig. 1.5, the subsystem design specialist has the responsibility to integrate the design from the bottom level parts, subcomponents and components all the way up to the subsystem level. Although the four types of system engineers can meet together to lay out a system-level compromise among the four attributes, it is eventually the subsystem specialist who has the design authority and ownership to finish all the details of the subsystem and component design.

The required complete engineering knowledge is reflected by the two-dimensional design space shown in Fig. 1.5, attributes versus subsystems, including their interactions. An engine program chief engineer ideally should be a person who masters all this knowledge. Moreover, it should be noted that ideally the total supplied expertise of the two types of engineers (system engineers and subsystem specialists) should be equal to the total required knowledge in the two-dimensional design domain (e.g., Model A in Fig. 1.6). A significant under-supply of the expertise from the total of the two types of engineers means a gap or deficiency for the product design. On the other hand, a significant over-supply of the expertise means a surplus of manpower, waste due to repeating, and an inefficient organization (i.e.,



1.6 Two models of system design in the attribute-driven design process.

an abused version of Model A in Fig. 1.6). The organizational efficiency decreases from Model A to B. But the degree of difficulty to train a system engineer is higher for Model A. Model B reflects the situation that the effectiveness of system engineers could be diminished when they only serve as 'messengers' between subsystems.

Without the guidance from a competent system engineer, subsystem specialists could easily lose the vision for the whole system if they only narrowly focus on their own design details. The abused version of Model A refers to the situation where both the system engineer and subsystem specialist unnecessarily duplicate their work on the same thing. In diesel engine system design, the role of system engineers in Model B is superficial and ineffective. Model A is a better systems engineering model for engine design that ensures the critical system design parameters are directly controlled by the system engineers in order to maximize design efficiency. It should be noted that Model B may be justified for extremely large and complex systems where many subsystems exist in parallel and there are many subsystem layers in the hierarchical chain (e.g., airplane design).

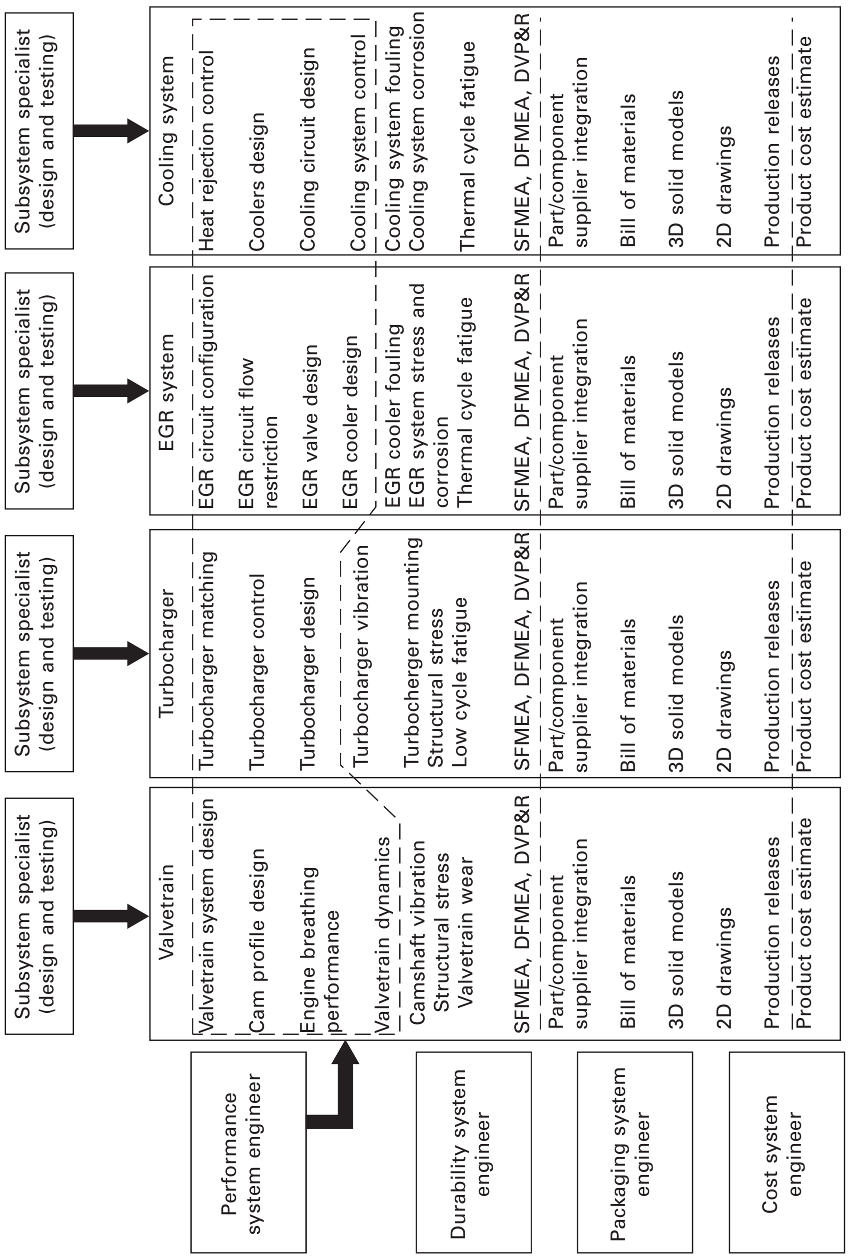
From Fig. 1.5, it is observed that there may be an overlap in the work area between a performance system engineer and a subsystem specialist, for example, in aftertreatment performance. With properly defined job responsibilities in coordination, such an overlap can be easily avoided. The general guideline is, within a given attribute, the item that ultimately affects the whole system integration needs to be designed or closely monitored by the system engineer. The remaining items local to the subsystem/component for all four attributes need to be designed by the subsystem specialist. Moreover, according to Fig. 1.5, the system engineer has the responsibility to integrate along the horizontal direction (i.e., across subsystems), and the subsystem specialist is responsible for integrating along the vertical direction (i.e., across attributes). Their work needs to be cross-reviewed in order to best utilize the expertise as a group to 'check-and-balance' in the two-dimensional design domain of 'attribute vs. component'.

The system engineer is desired to have in-depth knowledge on subsystem attribute design in order to conduct system integration seamlessly. Certainly, the degree of involvement of a system engineer on the component design details depends on the nature of the design attributes. For example, the performance attribute has more impact on the system-level integrity than the packaging attribute. This requires a performance system engineer to pay more attention to component design details than a packaging system engineer. Although the performance system engineer may alleviate the work load of the subsystem specialist from the performance attribute, the large amount of packaging and design details still needs to be carried out by the subsystem specialist. Therefore, both system engineers and component engineers are equally important in order to mature the design in a cross-functional team.

Figure 1.7 shows an example of the roles and responsibilities of system engineers and subsystem specialists. The system engineer defines system design specifications, and the subsystem specialist implements those specifications to realize them. In summary, Figs 1.5–1.7 present a powerful attribute-driven system integration approach for the engine system.

Work functions for different attributes

Among the four attributes (performance, durability, packaging, and cost), performance is the only one dominated by advanced analytical contents. Durability is largely handled by both structural modeling and experimental validation testing. Packaging is a mix of art and science, and is primarily in the realm of empirical design. Cost analysis has not grown to an advanced level in diesel engine design, although this area can be extremely complex and interesting. Table 1.4 summarizes the possible options to match the system design attributes with job functions. An analysis engineer is usually the best candidate for the position of a system engineer. A design or testing engineer is usually a better candidate for subsystem specialist because the majority of the subsystem/component work consists of design packaging and



1.7 Illustration of responsibility domain in engine system design.

Table 1.4 Design attributes and work functions

Attributes/work functions	Analysis engineer	Design engineer	Testing engineer
Performance system engineer	Y	N	N
Durability system engineer	Y	Y	Y
Packaging system engineer	N	Y	N
Cost system engineer	Y	Y	N

(Y: yes or possible; N: no or not likely)

durability issues. The focus of this book is on the performance attribute of diesel engine system design across all subsystems.

The theory of systems engineering is important in diesel engine system design. It helps set up a design process or mechanism that is more robust and less sensitive to human errors in system design. Better application of systems engineering principles can increase design quality and process efficiency, resulting in better products for customers.

1.2.4 Tools and methods of systems engineering

The concept of modification freedom

Provided by Menne and Rechs (2002), modification freedom is defined as the option to make design changes during the development phase that consists of a concept selection phase, a design phase, and a validation phase. The modification freedom is relatively large in the early concept selection phase. As the program moves forward to a later stage, less freedom is available for decision making, and more design data are finalized and released. There is no freedom at all immediately prior to the start of production.

Cause and effect diagram

The cause and effect diagram is also known as the 'fishbone' diagram, the Ishikawa diagram, or the 'feather' diagram. Moreover, a Pareto chart is another useful tool for breakdown or cause distribution analysis.

Decision tree

A decision-making process can be illustrated by a 'decision tree'. Decisions can be good decisions or bad decisions. Making the right decision at the early stage is very important. The disadvantage resulting from a single bad decision in an early stage may not be overcome entirely at the end, even by several good decisions later. This means it is immensely important to achieve the right engine system design specifications at the early stage of the development program.

Decision-making matrix for value-oriented design analysis

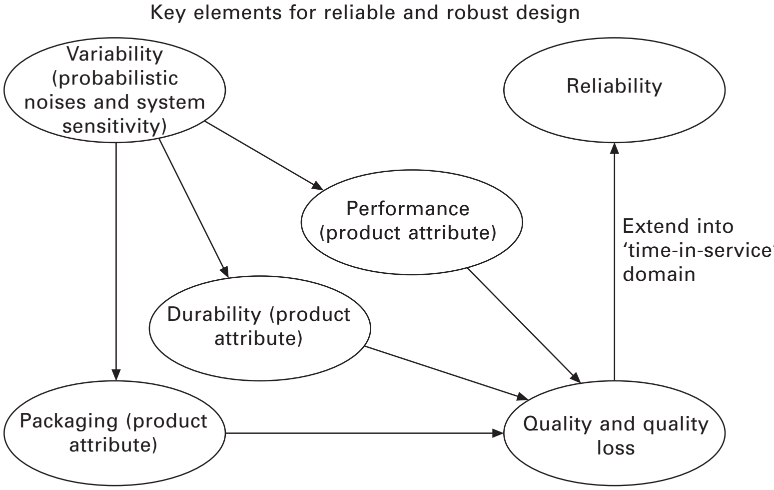
The decision-making process is an important part of system design. It provides the targets and criteria to judge the quality of technical concept selection between different alternatives. The targets can be classified into mandatory targets and ideal (or negotiable) targets. Whether or not the mandatory targets are met will result in a 'go' or 'no-go' rating or decision. The performance ratings for the negotiable targets need to be quantified and then weighted by multiplying weighting factors to balance the relative importance of the targets for a given application. Conflicting targets need to be negotiated by making trade-offs. Different design concepts need to be compared on the basis of total weighted performance rating, cost and the performance-to-cost ratio (or function-to-cost ratio, i.e., the value ratio).

1.3 The concepts of reliability and robust engineering in diesel engine system design

1.3.1 Key elements in reliability and robust engineering

The concepts of systems engineering introduced in the last section lay out the design processes of diesel engine system design. Four attributes of the engine product have been introduced: performance, durability, packaging, and cost. The concepts of reliability and robust engineering address the goal and the method of the design. Reliability engineering and robust engineering are about the quality, reliability, and the robustness of a product. They involve variability and probability. The ultimate goal of diesel engine system design is to design for reliability. There are six key elements in reliability and robust engineering for diesel engines, namely variability, performance, durability, packaging, quality, and reliability, as shown in Fig. 1.8. Each concept is discussed in the following sections.

The engine system attributes have a probabilistic nature caused by variability. Briefly, variability represents the probabilistic noise factors exposed to the system. Performance refers to engine functions (e.g., power, heat rejection, emissions). Durability refers to the structural capability of the engine (e.g., strength and stress, fatigue, and wear). Packaging refers to engine geometry, weight, relative positions, etc. Quality refers to the fulfillment of product specifications (assuming the specifications are good and meet customer requirements) before the engine is shipped to the customer. Reliability is



1.8 Concepts of reliability and robust engineering for engine system design.

basically a measure of the quality degradation in the time domain after the engine passes the quality inspection and is put in use by the customer. Quality and reliability are affected by both engineering design (i.e., controlling specifications to satisfy customer requirements) and manufacturing (i.e., realizing the design specifications). This book only addresses the aspects affected by engineering design.

For an engine product, the quality of performance and durability attributes are the most important for the customers. In fact, many quality problems due to packaging issues or errors (e.g., weight, clearance, size, shape) are reflected through performance and durability issues. A pure packaging problem (e.g., appearance and color) basically has little impact on engine reliability. Therefore, packaging is not elaborated in the following sections.

As shown by the arrows in Fig. 1.8, the process of diesel engine system design needs to go through a series of steps from assessing input variability to designing product attributes, then to achieving the specified quality, and finally arriving at a long period of assurance of quality (i.e., reliability) as the ultimate design goal. The process of reliability and robust engineering involves nondeterministic probabilistic design. It is more advanced than the fundamental deterministic design approach.

It should be noted that the terms 'quality', 'reliability', and 'durability' have often been misused in the literature without a clear distinction between them, including some of the references quoted in this chapter. Each important concept and its associated design methods are elaborated below.

1.3.2 The concept of variability

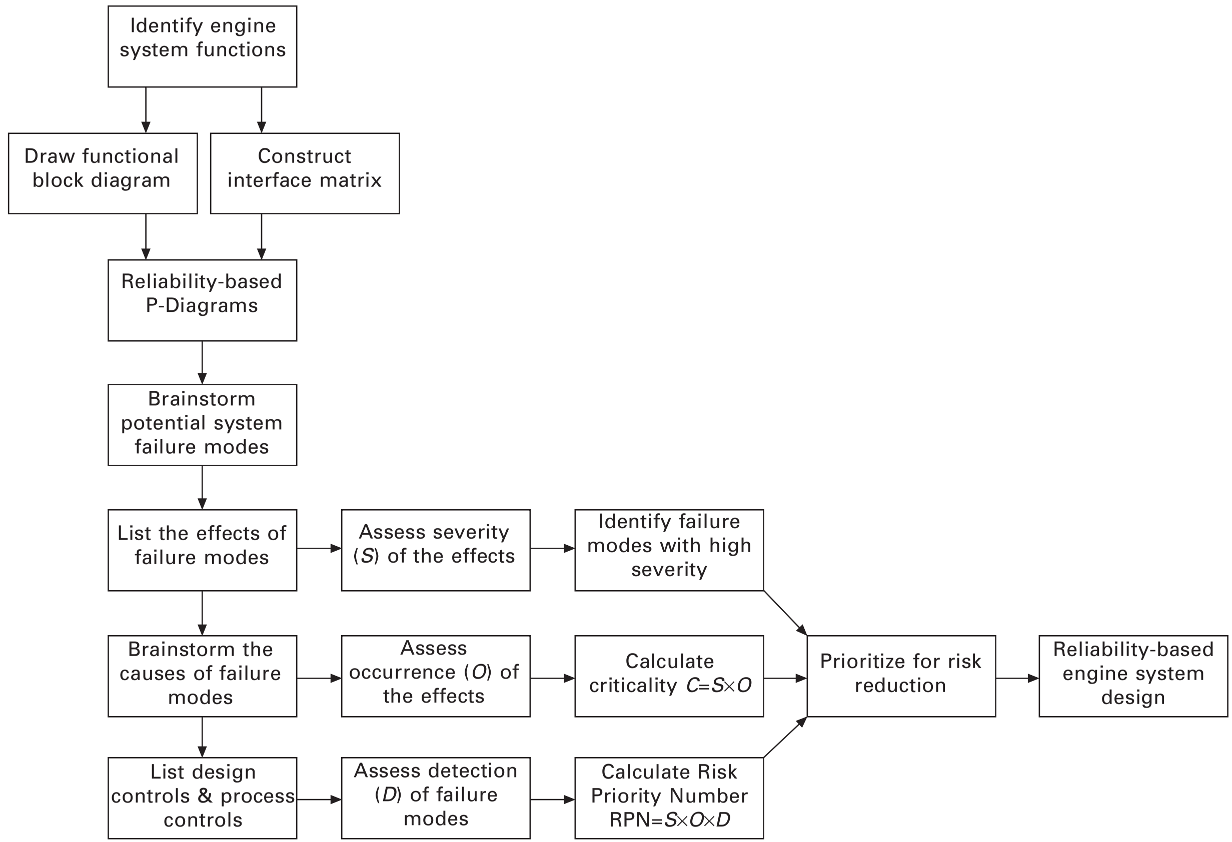
The concept of variability represents uncertain or uncontrolled noise factors and their impact on system sensitivity. It is a concept of nondeterministic design. Uncertainty introduces risk. A system sensitive to noise factors exhibits unstable performance (i.e., not robust). Traditionally, diesel engines were designed under deterministic conditions, namely under standard controlled laboratory testing conditions or certification conditions. Design quality could be guaranteed by controlling the product within the design specifications. However, when the diesel engine is put into real world usage, many problems caused by various noise factors surface in off-design conditions. The system may be over-designed for all the possible operating conditions (e.g., excessive margins on emissions, peak cylinder pressure, or coolant heat rejection), and consequently the product cost is too high. Or, the system could be under-designed for certain noise conditions and consequently it fails. In diesel engine system design robust engineering is needed in order to avoid those two scenarios. Design for variability (or design for probability) is required in order to achieve a cost-effective design for the majority of the product population. Moreover, design for variability is needed in order to detect at an early stage system failure modes caused by noise factors and to ensure reliability.

The risk associated with uncertainty or variability can be classified into system risk and design risk for an engine product. The risks are managed through FMEA (failure mode effect analysis). A detailed explanation of FMEA was provided by Stamatis (2003). A system/design FMEA approach usually consists of the following three steps:

1. Assess risk by identifying potential failure modes, the likelihood of occurrence, and the severity of their effects.
2. Establish priorities by ranking the failure modes with the risk priority number (RPN).
3. Take action to implement design/process changes to minimize the risk.

A system FMEA process for diesel engine system design is illustrated in Fig. 1.9. The FMEA is initiated by a cross-functional team of engineering, manufacturing, and reliability after the concept design is finished. The FMEA needs to be updated continuously prior to design release. Arcidiacono and Campatelli (2004) introduced an approach of failure mode and effect tree analysis (FMETA) for the reliability design process.

Tolerances are defined as the limit at which some economically measurable action is taken (Fowlkes and Creveling, 1995). Tolerances establish acceptable limits in specifications. In the system/design FMEA, it is always assumed that the failure modes result from design deficiencies. It is important to make



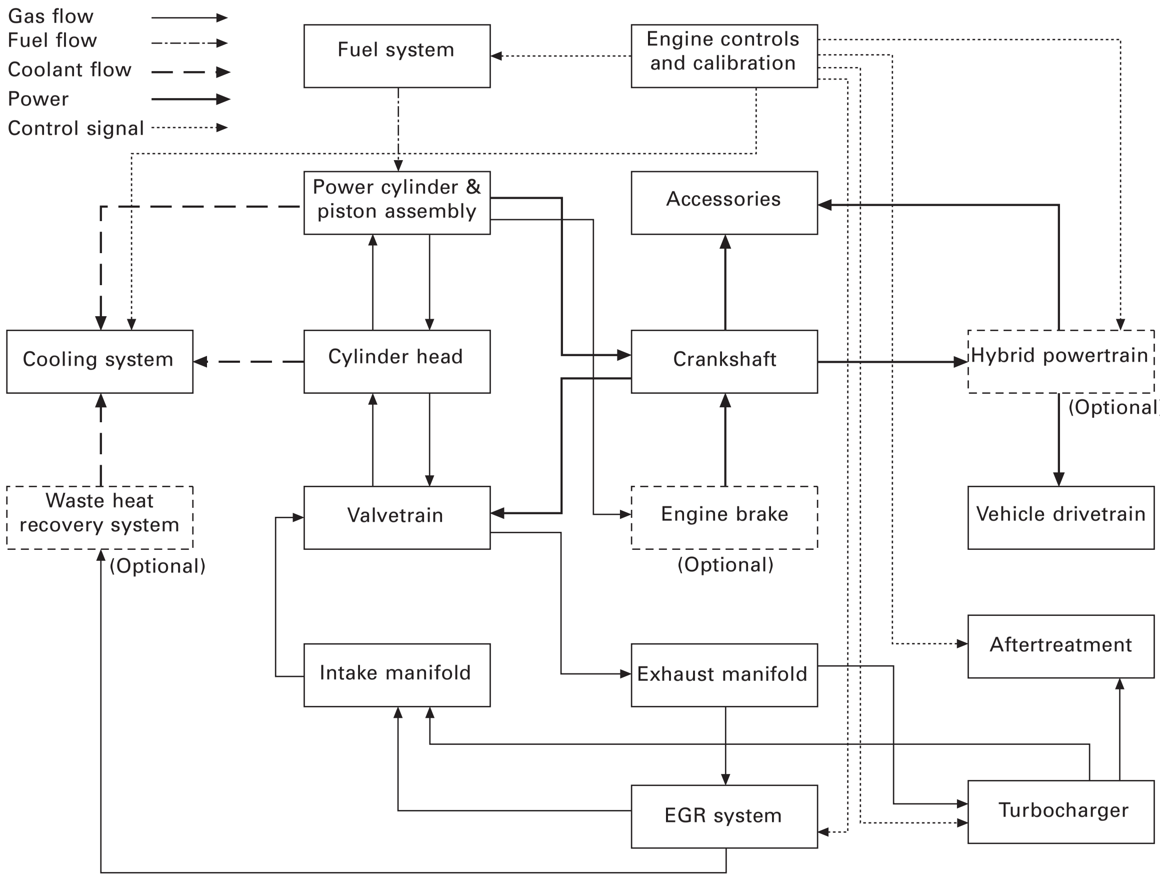
1.9 System FMEA process in engine system design.

the design specifications wide enough in tolerance (i.e., not over-constrained) so that the manufacturing process variation has little or no impact on the system/product performance.

The causes of failure modes may come from system design, component design, material selection, prototype build, etc. The failure modes can be classified into two categories: noise related and non-noise related. Each category is handled by a specific robust engineering tool. The functional block diagram and interface matrix for system FMEA development can be used to identify the non-noise-related failure modes. The purpose of a functional block diagram (Fig. 1.10) is to identify all the inputs and outputs of the functional blocks as well as their interfaces in the forms of physical or force connection, energy transfer, material exchange, and information or signal exchange. The system interface matrix (Fig. 1.11) quantifies the interfaces in terms of strength and importance and their potential effects. It is a useful tool for managing interfaces and the potential causes of failure resulting from subsystem interactions in diesel engine system design. Interactions present if the behavior of one subsystem depends on that of another subsystem.

The parameter diagram (i.e., P-Diagram, Fig. 1.12) is used to identify the noise-related failure modes. The P-Diagram identifies intended inputs and outputs, noise factors, control factors, and error states. Noise factors are unintended interfaces or sources of disturbing influences that may cause a deviation, disruption, or failure of the function during the mission time of the component or the engine. They are the causes of failure modes. Noise factors are uncontrollable (i.e., impossible, impractical or expensive to control) in product design or in-use operation. Generally, noise factors include the following five sources:

- Piece-to-piece or unit-to-unit variation (e.g., manufacturing variability in component geometry or material properties, variations in product-controlled parameters such as engine compression ratio, valve timing, turbocharger variance, fuel injection timing, injector variance or drift, tolerances in the controllable variables).
- Internal environment noise (also called system interaction noise or proximity noise, i.e., the unwanted effect of one subsystem on another, or subsystem interactions due to the variation of the input from neighboring subsystems or from the in-vehicle system operating environment, for example, variability in sensor signals, variance and drift of the air flow sensor, variability in the exhaust gas temperature or engine-out emissions composition).
- External environment noise (e.g., ambient temperature, humidity, altitude, road surface condition).
- Customer usage (e.g., accidental or foreseeable misuse and abuse of the product, real world usage duty cycle or load, different types of fuel or



1.10 Robustness tool – functional block diagram of diesel engine system.

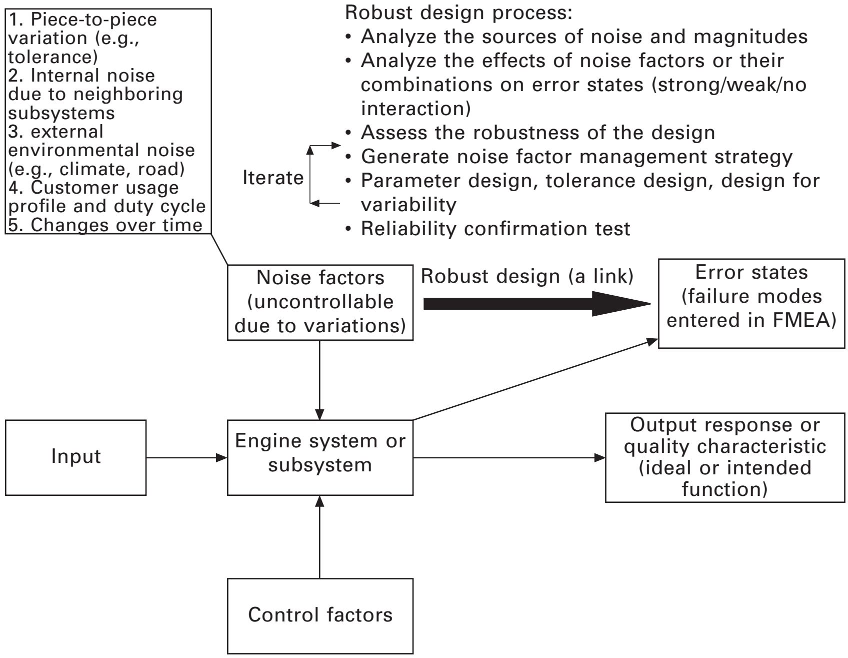
Causes (from) \ Effects (to) — Diesel engine system	Vehicle drive-train	Hybrid power-train	Cylinder head	Valve-train	Power cylinder & piston	Connect-ing rod	Crank-shaft	Intake manifold	Exhaust manifold	Turbo-charger	EGR system	Fuel system	Cooling system	Waste heat recovery	Engine brake	Acce-ssories	After-treat-ment	Engine controls	Engine cali-bration
Vehicle drive-train																			
Hybrid power-train																			
Cylinder head																			
Valve-train																			
Power cylinder & piston																			
Connect-ing rod																			
Crank-shaft																			
Intake manifold																			
Exhaust manifold																			
Turbo-charger																			
EGR system																			
Fuel system																			
Cooling system																			
Waste heat recovery																			
Engine brake																			
Acce-ssories																			
After-treat-ment																			
Engine controls																			
Engine cali-bration																			

System FMEA interface matrix

P: Physical connection
E: Energy transfer
M: Material exchange
I: Information exchange

2 = Necessary for function (i.e., required)
1 = Beneficial but not absolutely necessary for functionality (i.e., desired)
0 = Does not affect functionality (i.e., indifferent)
–1 = Causes negative effects but does not prevent functionality (i.e., undesired)
–2 = Must be prevented to achieve functionality (i.e., detrimental)

1.11 Robustness tool – interface matrix and subsystem interaction of diesel engine system.



1.12 Robustness tool – parameter diagram (P-Diagram) and robust design process.

lubricant or coolant used in an engine, accumulation of soot in the DPF and the associated increase in the back pressure).

- Changes over time or vehicle mileage (i.e., time-dependent deterioration or aging, e.g., component wear, corrosion, fatigue, component strength reduction, catalyst's physical or chemical degradation in conversion efficiency, build-up of impurities in recycled materials such as soot in engine oil, accumulation of ash in the DPF and the associated increase in the back pressure, EGR cooler fouling).

In system design or reliability verification, several noise factors may be combined to form a worse or the worst case noise level so that the number of tests can be reduced. When noise sources exist the pieces or units that marginally meet the minimum acceptable functional performance will suffer from a loss of function that may cause a failure. For example, a borderline component with maximum fuel injection quantity may fail at an extreme ambient condition. It should be noted that for a multi-cylinder engine, the cylinder-to-cylinder variation (for example, variations in EGR distribution, peak cylinder pressure, temperature, heat flux, exhaust runner pressure and temperature) is not a noise factor. The cylinder-to-cylinder variation is a controlled variable which can be improved by design. Usually, the cylinder's worst parameter should be selected as the characteristic parameter to represent the whole engine in probabilistic design. Moreover, the malfunction or catastrophic failure of a component or system is usually a noise factor. For example, when the EGR valve malfunctions to fully close at rated power, all the exhaust gas flows to the turbine and the compressor may over-speed. Although the turbocharger matching in engine system design should be targeted for normal operation rather than such a malfunctioning failure mode, the effect of such a failure mode needs to be checked as part of a reliable and robust design.

The control factors are the measures that can be changed in design to affect the mean response of the component or system to reduce variability. The response refers to the output of the component or system in the form of force, energy, material, signal, etc. For example, the set point of VGT vane opening is a control factor, but the tolerance of the VGT opening due to turbine actuator errors is a noise factor. Changing control factors can make the system function more robust (i.e., less sensitive to the influence of noises). The error states identify the failure modes. The error states reflect the deviation of the intended function, and they are potential failure modes. There are seven types of failure modes, namely omission of action, excessive action, incomplete action, erratic action, uneven action, action too slow, and action too fast. For example, the error states for a diesel aftertreatment system may include excessive tailpipe emissions, excessive back pressure, increased BSFC, decreased time between DPF regeneration cycles, etc.

At the end, a robustness checklist can be developed to manage the noise factors and failure modes. Each failure event is assessed by the frequency of occurrence of the cause (*O*), the severity of the effect (*S*) and the ability of detection (*D*, Fig. 1.9). To quantify the risks, each failure mode may have a risk priority number, calculated as the product of occurrence, severity, and detection. Design changes are usually required in order to reduce the occurrence and the severity and consequently the risk.

1.3.3 The concept of performance

Performance is the most important attribute for diesel engine systems. It represents the functionality of the engine. As shown in Fig. 1.8, performance is a major composition of the product quality. Engine performance includes the following six categories of characteristics:

1. displacement, velocity and acceleration (e.g., engine transient response of speed, vehicle acceleration, drivability)
2. force and torque (e.g., engine firing torque, vibration)
3. energy and energy rate (e.g., engine power, coolant heat rejection, exhaust gas exergy)

4. fluid flow (e.g., the pressure, temperature and flow rate of ambient air flow, engine gas flows, fuel flow and coolant flow, as well as the flow rate and concentration of emissions species)
5. noise (e.g., combustion noise, exhaust noise, piston slap noise)
6. electronic control signals.

It should be noted that the derived parameters by using the above parameters usually also belong to the performance attribute, for example, fuel economy, or BSFC (defined as the ratio of fuel flow rate to engine power). Both quality and reliability of the engine are measured by performance, durability, and packaging. For example, if the engine has a problem of power loss in service, it is a reliability problem in performance, rather than a quality problem due to durability.

Specific engine performance parameters can usually be found from engine program functional objectives, for example:

- ambient operating conditions (e.g., altitude range, temperature range, humidity range)
- engine speed range (e.g., maximum rated speed, maximum governed speed, high idle speed, maximum over-speed)
- engine power and torque output and speed (usually defined by full-load torque curve)
- engine friction and parasitic power losses (e.g., motoring friction power, power consumption of FEAD accessories)
- transient driving cycle emissions (e.g., HC, CO, NO_x or NMHC + NO_x, NMHC, particulate matter, formaldehyde, smoke during acceleration or at full load)
- white smoke level
- exhaust odor
- fuel economy (e.g., BSFC at rated power, peak torque and typical part-load modes)
- NVH (e.g., full-load noise, no-load noise, engine idle noise, turbocharger noise, 1/3 octave band center frequencies, maximum displacement of the powerplant mount, engine balance, resonances)
- engine startability (e.g., crank-to-start time at different cold ambient temperatures, hot restart time, glow plug's wait-to-start time, the minimum cranking speed for unaided cold start, cold cranking torque requirement)
- engine idle quality (e.g., hot curb idle speed and its maximum variation, misfire)
- fuel system (e.g., sulfur level, specifications of the diesel fuel or the alternative fuel, maximum fuel flow delivery capacity, fuel pressure, maximum fuel pump resistance, maximum return fuel temperature)
- air induction system (e.g., intake restriction, maximum air flow rate,

maximum rise-over-ambient temperature at the compressor inlet, maximum compressor outlet air temperature, charge air cooler restriction)
- exhaust system (e.g., maximum exhaust manifold gas temperature, maximum EGR flow rate, maximum exhaust flow rate, exhaust restriction)
- aftertreatment system (e.g., engine torque variations during DPF regeneration and LNT lean-rich modulations)
- cooling system (e.g., coolant type, heat rejections, water pump flow, maximum engine inlet and outlet coolant temperatures, maximum engine oil temperature, thermostat control set points)
- cab heater (e.g., heater core flow).

In the functional objectives the following are also stipulated but they are not performance attributes:

- engine durability life (e.g., B10 life in terms of the number of miles traveled)
- maintenance intervals (e.g., oil and filter change intervals, DPF service interval, valve lash adjustment; these are durability attributes)
- engine and aftertreatment weight (these are packaging attributes)
- engine architecture (e.g., displacement, cylinder configuration, bore, stroke, bore spacing, firing order, rotation direction, compression ratio, mean piston speed, package size in terms of length, width and height, configurations of valvetrain, combustion system and turbocharging).

It is observed from the above that the engine functional requirements consist primarily of performance parameters. Therefore, the core of this book is focused on engine performance in system design.

1.3.4 The concept of durability

The definition of durability

Engine durability or endurance is the other major attribute that affects the quality and reliability of the engine. It is usually confused with reliability in the literature. According to the definition given in the APQP (advanced product quality planning) manual developed jointly by Chrysler Corporation, Ford Motor Company and General Motors Corporation, durability is 'the probability that an item will continue to function at customer expectation levels, at the useful life without requiring overhaul or rebuild due to wear-out'. The APQP manual also defined reliability as 'the probability that an item will continue to function at customer expectation levels at a measurement point, under specified environmental and duty cycle conditions'. Another definition was given by O'Connor (2002), 'durability is a particular aspect of reliability, related to the ability of an item to withstand the effects of time (or of distance travelled, operating cycles, etc.) dependent mechanisms such as fatigue, wear, corrosion, electrical parameter change, etc. Durability is usually expressed as a minimum time before the occurrence of wear-out'.

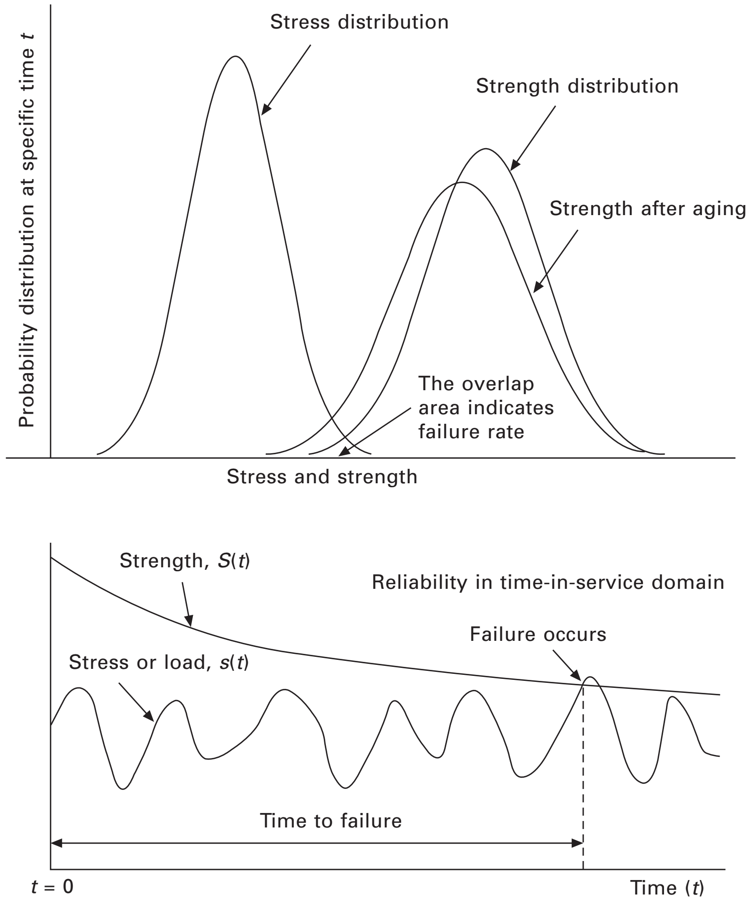
The above definitions of durability are somewhat ambiguous. Engine durability is better defined as the probability related to the hardware structural ability of an item to withstand the effects of time-dependent or non-time-dependent thermal, mechanical or chemical mechanisms such as fracture, fatigue, wear, corrosion, creep, deformation, fouling, plugging, electrical parameter change, etc. The time-dependent in the above refers to the accumulated service time. The non-time-dependent refers to the situations that are irrelevant to the accumulated time such as an abrupt rupture caused by over-load.

Durability is a part of quality before the engine product is released to the customer. It evolves to a part of reliability after the product is in use. In other words, a structural failure occurring in the engine test cell is called a durability problem rather than a reliability problem; and after the engine is released in service, a structural failure is called a reliability problem associated with a durability attribute. Such a definition clearly distinguishes between durability (as a design attribute) and reliability. The latter is a characteristic of the 'overall quality' extended into the time-in-service domain.

It should be noted that the concept of reliability covers the failures from all three attributes: performance, durability, and packaging. Therefore, it is not appropriate to assume equivalency between durability and reliability. Durability is usually expressed as a minimum time or vehicle mileage before the occurrence of any major type of structural failures (e.g., wear-out). For example, a B10 durability life is the expected life (e.g., 20,000 hours or one million miles) at which 10% of the population fails. A B50 durability life is the expected life at which 50% of the population fails.

Stress–strength interference model

The concept of structural durability is illustrated in the stress–strength interference model shown in Fig. 1.13. The figure shows a random probability distribution diagram of component strength and stress. The stress represents the load, and the strength represents the component's structural capability to resist the load. The stress and strength are random parameters and have probabilistic distributions corresponding to the variability in noise input factors. For example, the peak cylinder pressure load acting on the cylinder head can be regarded as a type of 'stress'. It has a probabilistic distribution due to the manufacturing tolerance in engine compression ratio, the variation in intake manifold boost pressure caused by the tolerance of turbocharger wastegate controls, the variation of exhaust restriction due to the change in DPF soot loading, the variation in ambient temperature, etc. The noise



1.13 Stress–strength interference model in durability and reliability.

factors affecting the structural strength of the cylinder head may include the manufacturing tolerances in cylinder head deck waviness, flatness and surface finish, the material properties of the head and gasket, and so on. A failure occurs when the stress is higher than the strength. The overlapping area between the stress and strength distribution curves in Fig. 1.13 indicates the probability of failure. The stress–strength model is elaborated in Chapter 2 for in-depth discussions on durability and reliability.

The role of durability in engine system design

Durability and performance are inter-related. Durability limits are used as design constraints in engine system design in order to determine the maximum achievable performance or appropriate hardware sizing. Figure

1.13 shows that in order to control the failure rate, either the strength needs to be increased (i.e., move the strength probability curve to the right or reduce its distribution range) or the stress has to be reduced. In the example of peak cylinder pressure, in order to reduce the pressure for better durability, the control factors such as engine compression ratio, intake manifold boost pressure or fuel injection timing need to be modified, and this affects engine performance and emissions. The durability analysis on the stress and strength distributions helps determine the maximum design limit and the nominal design/calibration target of design parameters (e.g., peak cylinder pressure and exhaust manifold gas temperature) that can be used for a durable system design. These limits or nominal targets can ensure the engine will not be overloaded and the structural strength is designed sufficiently strong.

1.3.5 The concepts of quality, robustness, and quality loss function

Quality is probably the most commonly known but ambiguous term in engineering system design. In order to define the role of quality in diesel engine system design, we need to review several important concepts in quality engineering. The engineering method of quality improvement is referred to as 'quality engineering' in Japan (due to the pioneering work by Dr Genichi Taguchi) and as 'robust design' in the West.

Quality has usually been defined as the degree to which performance meets expectations. The American Society for Quality (ASQ) provides a definition of quality as: 'Quality denotes an excellence in goods and services, especially to the degree they conform to requirements and satisfy customers.' ISO8402 standard defines quality as 'the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs'. Quality has been also sometimes loosely defined as conformance to specifications. Note that time dependency is not included in these definitions. For example, a non-time-dependent inspector's view of quality may be that a product is assessed against a specification. The product either passes the inspection or fails. When the product passes, it is delivered to the customer. The customer knows that it might fail at some future time and accepts such a 'reliability' risk over time. This approach provides no measure of quality over a period of time.

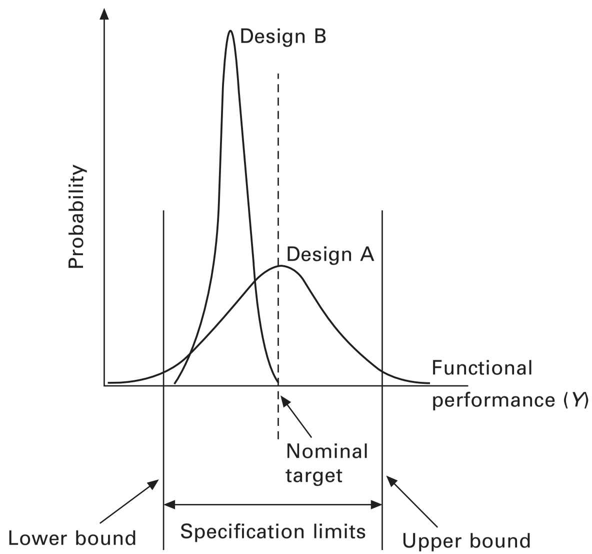
Time-based concept of quality was discussed by O'Connor (2002). Also, as pointed out by Rausand and Hoyland (2004), 'the quality of a product is characterized not only by its conformity to specifications at the time supplied to the user, but also by its ability to meet these specifications over its entire lifetime.' However, to avoid a controversy between the concepts of reliability and quality, Rausand and Hoyland (2004) again advocated that 'according to common usage, quality denotes the conformity of the product to its specification as manufactured, while reliability denotes its ability to continue to comply with its specification over its useful life. Reliability is therefore an extension of quality into the time domain'. Such a definition of quality is suitable for diesel engine system design. More details about quality and reliability engineering are provided by Chandrupatla (2009).

The quality of an engine product, as designed and manufactured, refers mainly to its consistency of conformance to the customer's requirements in performance, durability, and packaging attributes. As illustrated in Fig. 1.14, quality is an intermediate design goal of engine system design, and it assembles and measures all three design attributes. Quality ultimately extends or evolves to reliability in the time-in-service domain. During the course of diesel engine system design, the measurement and assessment of quality in performance, durability, packaging and their synthesis should not be overlooked. The quality target can be used as an objective in design optimization.

The demand for a product is usually affected by attributes and price (Hazelrigg, 1998). The quality and reliability of engine performance, durability and packaging features directly affect the brand image and product demand of diesel engines. It should be noted that sometimes people think a reliability problem is such a serious one that the product cannot function, while a quality problem is often regarded as a less serious one that the product can still function but with nuisance. That is a misunderstanding and wrong perception of the concepts of reliability and quality from an engineering sense. In fact, the severity of a problem can be characterized by a cost function of quality loss. A small loss of quality may cause a small cost to the customer without affecting the basic use of the product, while a large loss of quality at a moment of time in service during the lifetime of the product may cause a catastrophic failure of the function. Both scenarios present a reliability problem of different severities.

Before Dr Genichi Taguchi developed the methodology of continuous quality loss function, in the traditional discrete 'pass–fail' quality theory where samples were viewed as either pass or fail, a design within the range of specification was viewed as equally good, while any design outside the specification was viewed as equally bad (Fig. 1.14). However, the quality perception by the customer is not so simple. A product that is barely within the specification is certainly not as good as a product that is perfectly on the target. Therefore, quality loss, or functional performance loss, should be a continuous and gradual event with respect to the functional performance rather than a discrete or step event depicted in the traditional quality theory.

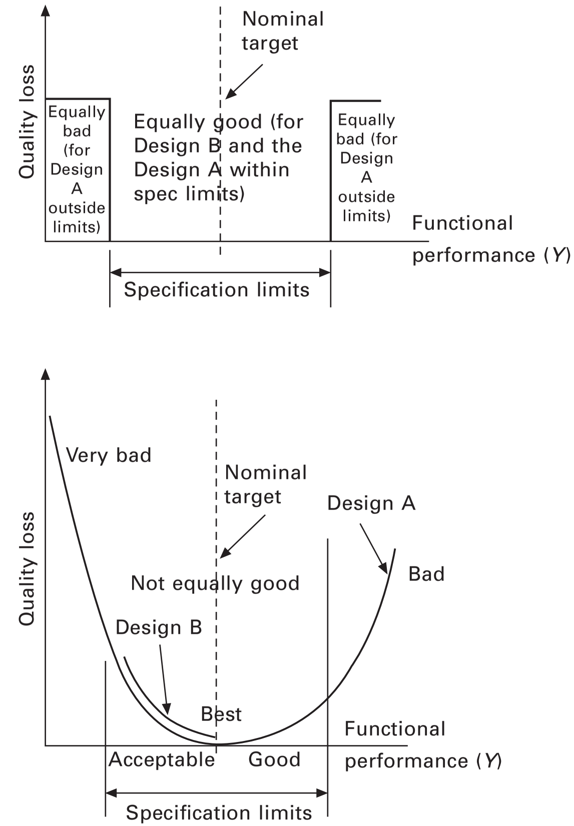
Quality engineering or robust design is a process to obtain the response that is insensitive to noise factors. Robust design requires a quantitative definition of quality. As mentioned earlier, the loss of quality is actually continuous rather than discrete and sudden when the product quality deviates from the



Quality loss function:

$$L_q(Y) = \frac{C_{cf\ upper}}{\tau_f^2}(Y - \overline{Y})^2 + \frac{C_{cf\ lower}}{\tau_f^2}(Y - \overline{Y})^2$$

τ_f: functional tolerance
C_{cf}: cost to functionality if tolerance is exceeded



1.14 Quality loss function.

nominal target due to various noise factors. The product quality is commonly defined by using a quality loss function (Fowlkes and Creveling, 1995). As shown in Fig. 1.14, the two designs (A and B) have different probabilistic distribution curves for a particular functional performance parameter. The specification is defined by a nominal target and a tolerance range for acceptable limits from a lower specification limit to an upper limit. The design A has a larger population of samples (i.e., higher probability) around the 'good' nominal target but widely spread samples outside the specification range due to a larger standard deviation. On the other hand, the design B has a shifted and narrower probability distribution curve which gives very few samples meeting the 'good' nominal design target although all of its samples are within the specification range. This example illustrates that 'within specification' and 'on target' do not have the same meaning. Therefore, it is important to consider the probability distribution characteristics when evaluating different designs.

Dr Genichi Taguchi promoted the idea of not just getting all the units within the specification limits but getting all the units on target. He developed the methodology of quality loss function to evaluate the financial impact of the tolerance range. Taguchi defined the quality loss of an off-target product performance, due to deviation from the nominal target, as the life-cycle monetary losses caused by functional variability and harmful side effects related to customer functional tolerance and the cost to society. He defined that the losses represent a summation of rework, repair, warranty cost, customer dissatisfaction, bad reputation, and eventual loss of market share for the manufacturer. He defined that the loss of quality as a cost increases quadratically with the deviation from the target, and the quality reaches the best at the nominal target. He used the quality loss function to quantify the quality of a design and determine the tolerances.

Various models have been proposed to estimate the product quality loss by using the mean value and the standard deviation of the response. Figure 1.14 and Table 1.5 show the quality loss function. Table A.1 also summarizes the related formula used in statistics. The two basic sample statistics to quantify variability are $\bar{Y}$ and σ_{SV}^2, which describe the central location and the width of the probability distribution. A histogram is another way to describe the distribution of the frequency of occurrence. From the average quality loss equation shown in Table 1.5,

$$\bar{L}_q(Y) \approx C_q \cdot [\sigma_{SV}^2 + (\bar{Y} - Y_{target})^2] \qquad 1.1$$

where C_q is a quality loss coefficient, it is observed that minimizing the variance σ_{SV}^2 (i.e., variability) or making the average response $\bar{Y}$ on the target Y_{target} can reduce the quality loss. Such a method developed by Taguchi lays out the theoretical ground of robust design.

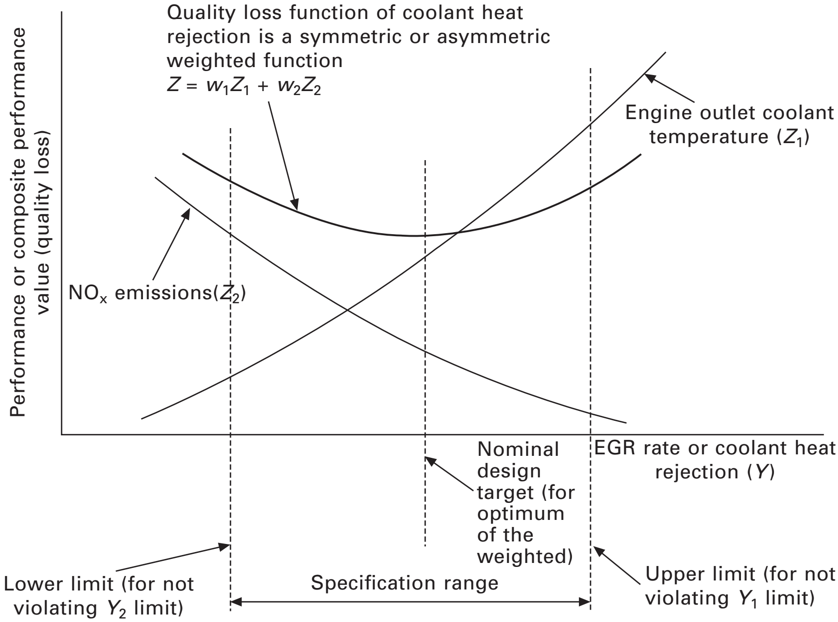
Quality includes not only the conformance of geometric tolerances,

Table 1.5 Formulae of Taguchi's quality loss in robust design

Parameter name	Parameter symbol	Formula
Taguchi's quality loss	$L_q(Y)$	$L_q(Y) = C_q \cdot (Y - Y_{target})^2$
Taguchi's mean square deviation	C_{MSD}	$C_{MSD} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_{target})^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 + (\bar{Y} - Y_{target})^2$
Taguchi's average quality loss	$\bar{L}_q(Y)$	$\bar{L}_q(Y) = C_q \cdot C_{MSD} = C_q \cdot [\sigma^2 + (\bar{Y} - Y_{target})^2]$ $\approx C_q \cdot [\sigma_{SV}^2 + (\bar{Y} - Y_{target})^2]$

but also the conformance of functional requirements. For example, a television's quality can be measured by its appearance and the sharpness of the image. For diesel engine system design, engine product quality refers to the conformance to the requirements of a combination of performance, durability, and packaging. The quality loss function developed by Taguchi is most useful to quantify the deviation of quality in the area of packaging design and manufacturing tolerance. In general, it is difficult to apply the quality loss function directly to diesel engine system design because it is difficult to quantify the monetary cost during the system design for the deviation of a design attribute. An easier way to utilize the concepts of quality and quality loss in system design is to use an attribute parameter in engineering units instead of the cost in monetary units. The quality loss of an engine attribute can be simplified as a continuous function of another dependent parameter. For instance, as shown in Fig. 1.15, if we want to quantify the quality loss of EGR rate or coolant heat rejection, the quality loss function can be constructed as a weighted function of engine outlet coolant temperature and NO_x emissions, which are dependent upon EGR rate and heat rejection. The coolant temperature is an indicator of durability and increases with coolant heat rejection. NO_x emissions are affected by EGR rate, which in turn directly affects the coolant heat rejection. The lower the EGR rate or heat rejection, the higher the NO_x emissions. Another example of quality loss is the effect of air–fuel ratio. A deviation to higher air–fuel ratio than the nominal target causes excessively high peak cylinder pressure, while a deviation to lower air–fuel ratio results in high soot or high exhaust manifold gas temperature. Essentially, the quality loss function in engine system design can be simplified as any continuous composite function that may be used as the objective function in system optimization in order to determine the nominal design target and allowable tolerances.

Another key measure of robustness in Taguchi's robust design for quality improvement is the signal-to-noise ratio (or the S/N ratio). Good quality requires the performance be on target with low variability. Taguchi focused



1.15 Quality function in engine performance.

on reducing variations in design and believed that it might be preferable to have a product that gives consistently good results than one that gives inconsistent results which are sometimes better but worse on average. The details of Taguchi's concepts of quality, quality loss function, signal-to-noise ratio, and robust design were explained in detail by Fowlkes and Creveling (1995).

1.3.6 The concept of reliability

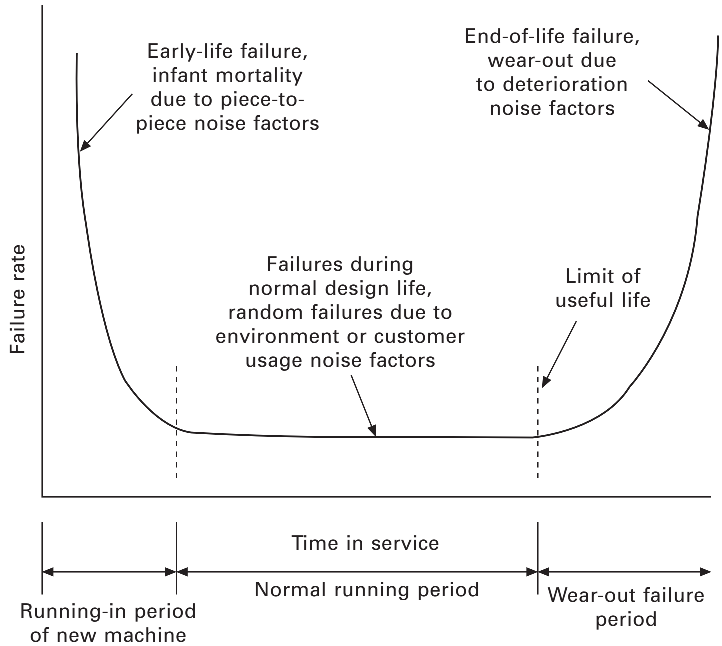
The definition of reliability

Reliability has a great impact on the brand image of a product. Customers expect a product of good quality that performs reliably over time. Issues in durability, quality, and reliability lead to warranty. Manufacturers often suffer high costs from failures under warranty. Good reliability leads to reduced warranty costs and improved brand image. Automotive engines are designed to possess longer warranty periods or mileage targets with lower failure rates for better reliability.

Reliability emerged as a technical consideration after the First World War in airplane engine applications. At that time, reliability was measured as the number of accidents per hour of flight time (Rausand and Hoyland, 2004). Reliability means the dependability of a product related to its failure and time to failure (Fig. 1.13). Reliability is usually concerned with failures in the time domain. In general, reliability is defined as the fulfillment of quality requirements over the life cycle under specific application conditions. According to Chandrupatla (2009), 'reliability is the probability that a system or component can perform its intended function for a specified interval under stated conditions'. According to O'Connor (2002), 'reliability is the probability that an item will perform a required function without failure under stated conditions for a stated period of time'. As summarized by Rausand and Hoyland (2004), until the 1960s reliability was defined as 'the probability that an item will perform a required function under stated conditions for a stated period of time'. They pointed out that a more preferred and more general definition of reliability given by ISO 8402 standard was that 'reliability is the ability of an item to perform a required function, under given environmental and operational conditions and for a stated period of time.' A required function here may be a single function or a combination of functions that is necessary to provide a specified service. A hardware system may pass the initial quality specification test on the factory's manufacturing assembly line, but it may not perform reliably over a specified period of time in customer usage. Rausand and Hoyland (2004) defined that 'according to common usage, quality denotes the conformity of the product to its specification as manufactured, while reliability denotes its ability to continue to comply with its specification over its useful life. Reliability is therefore an extension of quality into the time domain'.

In diesel engine system design, reliability is the ability of how reliable to maintain the quality of performance, durability, and packaging over time after the engine is put into service. Reliability is not a design attribute. Instead, it is a characteristic of the product quality extended into the time domain. By definition, reliability contains two key factors: a probabilistic impact by noises and time. Those factors are illustrated in the 'bathtub' curve of reliability in Fig. 1.16. When a product is produced as marginally acceptable with respect to its specifications, a little additional external noise can cause a failure in service in the early life of the product. The decreasing failure rate in Fig. 1.16 suggests infant mortality, which means defective items fail early and the failure rate decreases over time as they fall out of the population. The constant failure rate characterizes the bottom of the 'bathtub' curve. It is caused by random failure events associated with external or internal environment noise factors or customer usage noise factors. Eventually, the cumulative effects of deterioration noise factors (e.g., wear out) and environment/usage noise factors cause the end-of-life failure, which is characterized by an increasing failure rate as time goes on. The goal of design for reliability is to make the useful product life longer. The goal of robust design is to make the design insensitive to all the noise factors during the product life.

In general, reliability can be classified into three categories: hardware,



1.16 Reliability 'bathtub' curve.

software, and human. The scope of diesel engine system design is primarily concerned with hardware reliability of engine components and system. Reliability can be deduced from a probability distribution function of the time to failure (Rausand and Hoyland, 2004). Reliability is usually expressed by the following means:

1. Failure rate and mean time to failure (Fig. 1.13, for example, 1% of failure in total population by 250,000 miles; or 100,000 miles of passenger car operation with no engine tune-ups; or one million miles of commercial truck operation before the overhaul).
2. Number of failures over a period of time (e.g., R_N/1000 or the number of repairs per thousand units).
3. Probability of the event that the time to failure is longer than the specified useful life (for example, if the specified useful life is three years and the time to failure is four years, the reliability is 100%).
4. $R = 1 - P_{failure}$, where $P_{failure}$ is the probability of failure or the failure rate. The failure rate at a given mileage can be measured from the warranty information of the engines in service.

To further illustrate the noise and time factors in reliability, Fig. 1.13 shows an example of structural reliability in the random probability distribution diagram of the stress–strength interference model. It should be noted that the probability distribution curves of stress and strength shown in the top part of Fig. 1.13 are constructed by considering noise factors, and they may or may not include a time factor (recall that one of the noise factors discussed earlier is changes over time). In the time-in-service domain, the stress or load in usage varies with time in an uncontrolled manner. The strength of the component also varies with time because the component deteriorates in material properties over time due to failure mechanisms like creep and corrosion. This time-dependent characteristic is shown in the bottom part of Fig. 1.13. The reliability expressed by the time-to-failure is then the time until the stress exceeds the strength. It should be noted that, since there is a high level of uncertainty involved, it is very difficult to predict reliability with statistical methods or precise mathematical calculations.

The role of reliability in engine system design

Reliability is affected by all the phases of a product life cycle (engineering design, manufacturing, operation, and maintenance). Reliability engineering is a process to ensure reliability is controlled in the design stage. Reliability engineering efforts control the risks by addressing the causes of failures in order to prevent or minimize their occurrence. Although reliability can be improved by controlling production variability and human variation in manufacturing and quality control or by implementing preventive maintenance strategies, design plays a vital role to determine the reliability. The concept of design for reliability was elaborated by Kececioglu (2003) and Kumar *et al.* (2006). More in-depth discussions on reliability in engine system design are provided in Chapter 2. The fundamentals on reliability can be found in Ireson *et al.* (1996), Bignonnet and Thomas (2001), Kuehnel *et al.* (2005), Dodson and Schwab (2006), Klyatis and Klyatis (2006), Rahman *et al.* (2007), Zhou and Li (2009), and Klyatis (2010).

1.3.7 System design – from 'design for target' to 'design for variability' and 'design for reliability'

Overview of system design methodologies

As discussed earlier, variability and reliability are nondeterministic in nature. Without considering the statistical probability distribution of the attribute the design for a point target would be either an over-design or under-design. A reliable and robust design of a diesel engine system requires the following design methodologies:

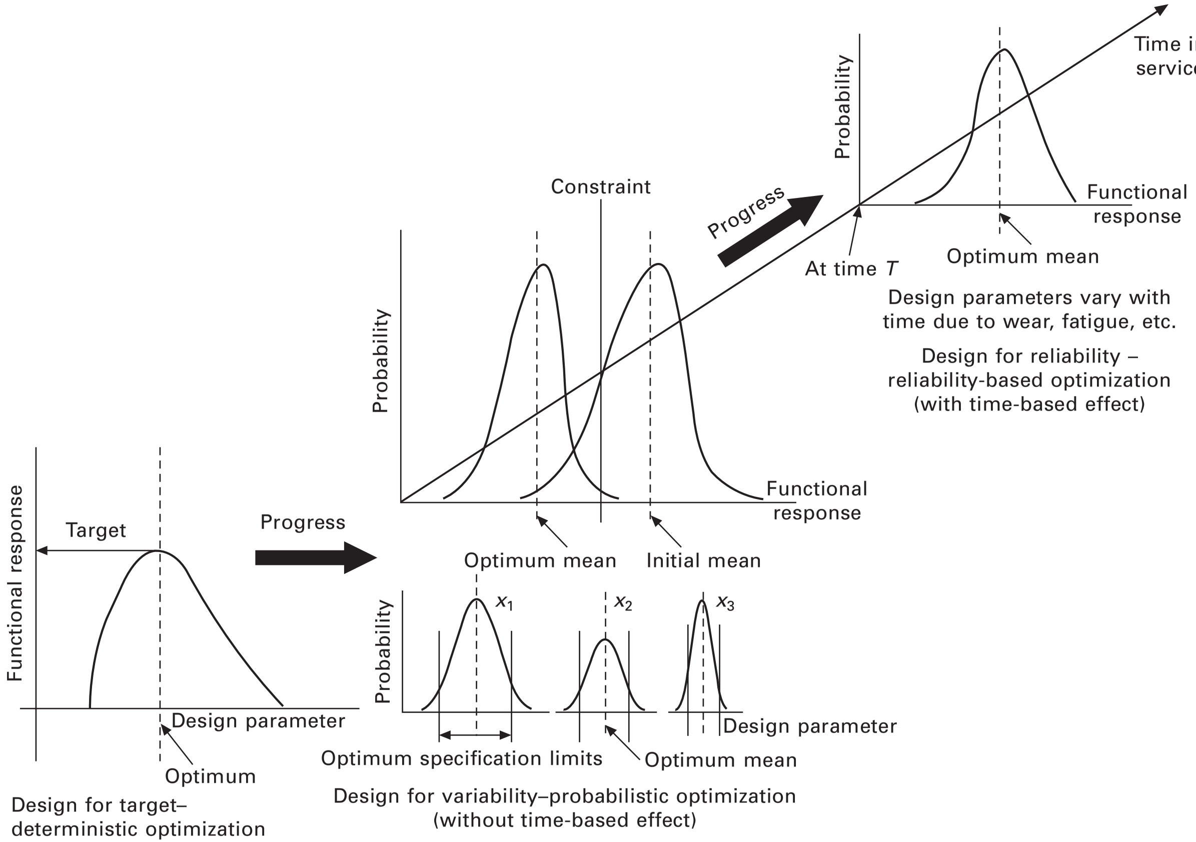
- design for target (i.e., design for a specific point target, either nominal or limit)
- design for variability (i.e., design for both mean value and tolerance

range of the nominal target to achieve a robust and sensitive design)
- design for reliability (i.e., design for time-dependent variability or design-for-time-degradation).

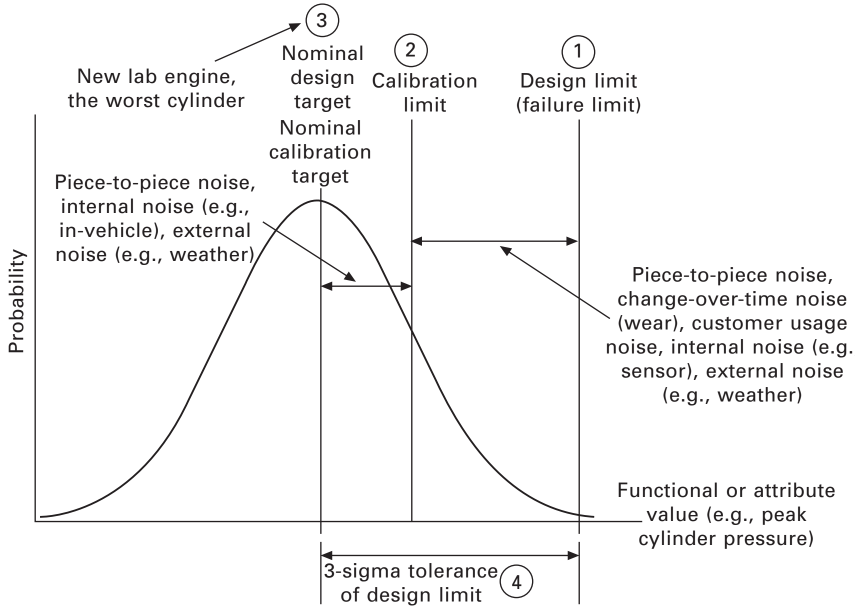
Among the above three methods, design for target is the foundation and the basic technique in system design. It is a deterministic design approach. Many analytical techniques can be applied effectively in this category for a precise design. The rest of the book gives a thorough introduction to this basic technique to strengthen the foundation. Design for variability is a more advanced technique and is nondeterministic. It expands the point design in design for target to a more complex one-dimensional design by adding the probability distribution of the attribute parameter. The probability distribution involves uncertainty and the design loses its precise nature. Inadequate sampling in measurement or test data may make the probability distribution inaccurate. Design for reliability is more complicated because it further expands the one-dimensional design to a two-dimensional design by adding the random factor of time-in-service into the model. Design for reliability is the least precise in nature due to more uncertainties involved. Owing to the difficulty of predicting reliability accurately, the time-dependent probability distribution of the attribute parameter may not be accurate. However, once modeled successfully (with the support of a large amount of lab/field/service data), design for reliability will give the best quality of engine system design. Figure 1.17 illustrates the evolution of these design methodologies. Figure 1.17 shows that the ultimate design goal of diesel engine system design is good reliability in service life.

Design for target

Design for target is the traditional deterministic design approach where a point target is given. It can be further classified into two subsets: design for nominal and design for limit. The target can be either a nominal value (design for nominal) or limit value (design for limit) used in design and calibration. For example, the engine is designed to have the structural strength to sustain a maximum limit of peak cylinder pressure at 200 bar. In order to ensure the pressure does not exceed the limit of 200 bar under any operating condition with all the noise factors, the engine rated power needs to be calibrated at a peak cylinder pressure of only 180 bar as a nominal target under standard lab conditions. The difference of 20 bar serves as a design margin to cover any deviations from the standard nominal (Fig. 1.18). Design for target is a very useful and still prevalent approach in engine system and component designs for both steady-state and transient topics. In the mathematical formulation of design for target, the number of equations matches the number of unknowns so that a deterministic solution can be found. Two examples of design for target are provided below:



1.17 Engine system design optimization – design for target, variability, and reliability.



1.18 Engine system design constraints or limits.

1. At a given engine speed and power, assuming the air–fuel ratio and EGR rate are known requirements, find turbine area and EGR valve opening.
2. For a transient emissions cycle with prescribed engine speed and load varying as a function of time, assuming the transient history of the air–fuel ratio and EGR rate are known, find the transient history of the turbine area and EGR valve opening as a function of time.

Design for target is the fundamental design technique in diesel engine system design and a stepping stone toward the more advanced nondeterministic designs. The shortcoming of design for target is that without the information of the probability distributions of the attribute parameters or reliable previous experience, it is difficult to select an appropriate target value in order to achieve a design that is neither over-designed nor under-designed. Design for variability needs to be used to address this issue.

Design for variability

Design for variability addresses the effects of both control factors and noise factors in order to control both the mean value and the variation range of the attribute parameters by changing the control factors. Design for variability includes two subsets: robust design and sensitive design, one for noise input factors and the other for control input factors. The goal of design for variability is to achieve a reliable design that is insensitive to noise factors but sensitive to control factors. The reliable design means that the probabilistic distribution of the system response has a reasonable mean value and deviation range so that a predetermined percentage of population satisfies the requirements of performance, durability, or packaging without failures. The response can be any parameter of performance, durability, or packaging.

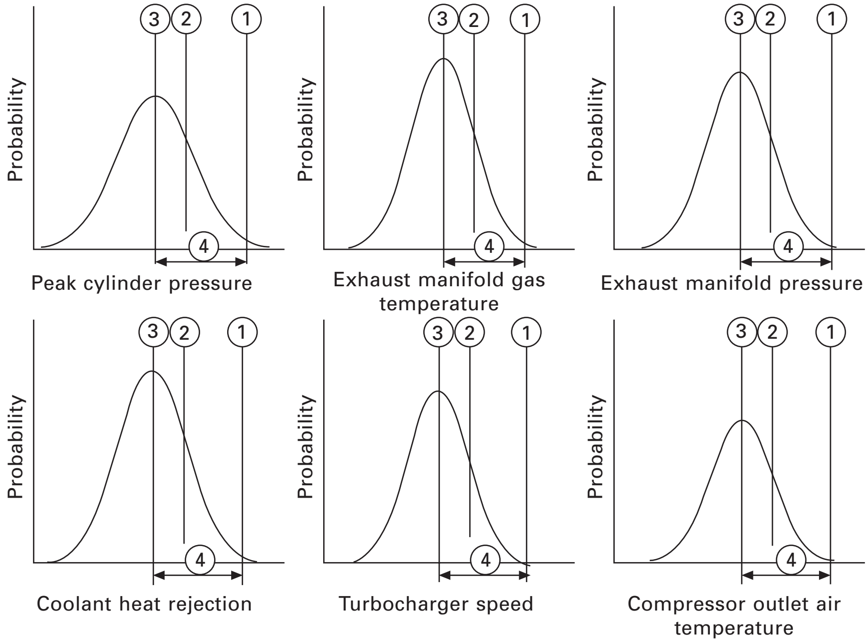
Robustness means that the system/component response is insensitive to, or not adversely affected by, the variation of the input noise factors within a range of circumstances, even though the sources of variability have not been eliminated. Robustness can be measured by the signal-to-noise (S/N) ratio. It is the effect of noise factors that robust design wants to control, via changes in the control factors. Robust design is the process to achieve the defined robustness. Sensitivity analysis, mean value design (for setting the nominal target), and tolerance design (for setting the specification range) are three important tasks in robust design.

Robust design should not be confused with sensitive design. The engine needs to be a sensitive system which responds quickly to the control input factors, for example, to achieve good transient performance. Note that the system needs to be sensitive to the control factors rather than the noise factors. When noise factors present, it is desirable for the engine system to be insensitive to the noises in order to ensure a stable operation under uncertainties.

There are numerous examples of robust design in engine applications. For example, the fuel system of the diesel engine must provide a consistent supply of liquid fuel regardless of external or internal noise factors. Inconsistent fuel delivery may cause engine stumbles when cruising, accelerating and decelerating, rough or rolling engine idle, and other drivability issues. Higher temperatures in the engine may lead to fuel vaporization which may cause insufficient fuel delivery to the injectors. A robust fuel delivery design needs to isolate or minimize the impact of these unfavorable external temperatures.

Statistical probability distributions of random parameters are required in design for variability. Several types of distribution can be used to model the input factors, such as normal, uniform and Beta distributions (Table A.2). The combination of the probability distributions of different input factors generates a probability distribution of the output response. Appropriate mean and standard deviations of the control factors can be searched or optimized simultaneously in order to achieve the desirable distribution of the response. The design solution produced by such a design-for-variability approach is usually a more cost-effective and more robust design than that obtained by using the deterministic design approach.

As shown in Fig. 1.19, when multiple design constraints exist, as is usually the case with engine design, the deterministic approach cannot completely handle design-for-limit if the limit value of only one constraint is used because that single point does not serve as the limiting case for all the design constraints. For example, if a design-for-limit case requires an



1.19 Coordination between different engine system design targets and constraints.

analysis at peak cylinder pressure of 200 bar, this case would give a lower exhaust manifold temperature than a case of 180 bar cylinder pressure due to higher air–fuel ratio. Therefore, this case cannot be used to assess the durability of the exhaust manifold. The probabilistic approach of design for variability does not have this limitation because it produces the statistical distributions of all the design constraints simultaneously so that all the design constraints can be easily identified, as illustrated in Fig. 1.19. A multi-objective optimization can then be conducted to try to seek a solution that satisfies all the constraints. Monte Carlo simulation is an effective tool used in design for variability and this topic is detailed in Chapter 3.

Some research work related to design for variability and robust design was reported in the literature in the areas of emissions, cooling, and vehicle fuel economy. Yan *et al.* (1993) and Dave and Hampson (2003) used the DoE method and Monte Carlo simulation to analyze the robust design of emissions and engine BSFC. Rahman and Sun (2003) analyzed engine cooling system design to control coolant temperatures. Catania *et al.* (2007) presented a robust optimization for vehicle fuel economy analysis.

Design for reliability

Design for reliability is design for variability extended into the time-in-service domain. One example is turbocharger matching to account for the variation in exhaust restriction caused by soot loading changes in the DPF. Exhaust restriction (the pressure drop across the aftertreatment system) significantly affects turbocharger matching and engine system design. When the exhaust restriction becomes higher, the turbine expansion ratio and the engine air flow rate usually reduce. This results in a decrease in peak cylinder pressure but an increase in exhaust manifold gas temperature. The nominal design point of turbocharger matching is dependent on exhaust restriction. Unlike the muffler-only exhaust system that has a deterministic back pressure at the rated power condition, modern diesel engines equipped with a DPF have a randomly fluctuating back pressure over time, increasing during the soot accumulation phase or decreasing after DPF regeneration. Matching the turbocharger for a clean DPF or a fully loaded DPF yields very different results. If the probabilistic distribution of the exhaust restriction during the vehicle lifetime can be found, the turbocharger can be matched better to balance the system efficiency and durability.

Like design for target, design for variability/reliability is often carried out with optimization techniques. In reliability-based design optimization (RBDO), in order to reduce computing time, it is useful to first conduct a deterministic optimization to pre-screen, and then apply the probabilistic variations of the uncertainties to the pre-screened sub-optimal solutions for further optimization. The RBDO usually consists of the following steps:

1. Deterministic design-of-experiments (DoEs) variable pre-screening to identify key factors.
2. Deterministic DoE response surface fit to build emulator models.
3. Non-time-dependent nondeterministic variability analysis with key factors.
4. Time-dependent nondeterministic reliability optimization with key factors.
5. Confirmation runs.

These optimization topics are elaborated in Chapter 3.

1.4 The concept of cost engineering in diesel engine system design

1.4.1 Design for profit and design for value

Cost is one of the four attributes of diesel engine system design. The cost of a part is related to the design in terms of the material required and the processes available to build and maintain it. The central theme of cost engineering is design for profit. The corporate revenue is directly related to product market demand, which is affected by product attributes and market price. Revenue is basically price multiplied by demand quantity. The design of the attributes is accompanied by cost. Profit is the difference between revenue and cost. The precise definitions of these economics terms are provided by Ostwald and McLaren (2004). Design for profit is a process with target costing where the effects of the design are evaluated for demand, revenue, cost, and profit, and the profit is maximized in engineering design and decision making. The essence of design for profit is to ensure that the product meets a target market price to allow product competition. Another important concept in cost engineering is value. Value is defined as a function-to-cost ratio, or cost effectiveness. Maximizing value should be also considered in design.

1.4.2 The need for engine system cost analysis

Many engine development programs failed due to ineffective control of product cost and price. In today's highly competitive market, customer requirements on engine product functionality are increasingly stringent. Better fuel economy, higher power, faster acceleration, lower noise, and longer reliability life are a few examples. On the other hand, the increasingly booming supplier industry is trying to provide engine manufacturers with a wide range of components or modularized subsystems with advanced or luxury features. Two-stage variable-geometry turbocharger, electric boosting, camless valvetrain and hybrid powertrain are several examples. An engine product can be developed by throwing in many advanced features simultaneously to achieve superior functional performance, but it may not be financially viable at all. There is a pressing need for system cost integration and control during engine development in order to design the least expensive product while meeting customer requirements.

Although cost reduction opportunities exist in almost all engine components, there are three main aspects that drive a high product cost, namely emissions compliance, fuel economy improvement, and advanced features.

The components with major cost increase for low-emissions engines include the following (in order from high to low):

- aftertreatment (DPF, LNT or lean NO_x trap, HC dosing, additional sensors about pressure, temperature and NO_x, precious metal loading)
- turbocharger (two-stage turbocharger or VGT)
- EGR system (larger EGR cooler size, more coolers, more EGR valves, intake throttle valve)
- fuel injector (resulting from the requirement of higher injection pressure)
- cylinder head (resulting from increased cylinder pressure)
- piston and piston rings (resulting from increased cylinder pressure and change in piston cooling)
- crankcase (resulting from increased cylinder pressure).

The following technologies may incur high product cost for fuel economy improvement:

- hybrid electric or hydraulic powertrain
- waste heat recovery with Rankine, Brayton or Sterling cycles
- turbocompounding
- variable valve actuation
- variable compression ratio
- variable swirl
- cylinder deactivation
- continuously variable transmission
- flexible cooling.

Advanced engine features associated with high product cost may include the following examples:

- high power output (affecting structural strength or engine size)
- demanding acceleration characteristics and transient response (related to hardware design and software controls)
- powerful engine brake or driveline retarder (for heavy truck applications)
- in-cylinder pressure real-time monitoring and closed-loop combustion controls
- NVH and noise reduction features (e.g., engine encapsulation or material selection for low noise).

Each of the above items has a heavy weight on overall engine product cost. If each of them is not planned and controlled carefully, the entire engine will be at a risk of high cost.

Many engine development programs failed due to the lacking of cost control or cost design. In diesel engine system design there is a need to plan the integrated cost roadmap for various advanced technologies for their entire life cycles and to coordinate/balance the cost structure of the production design between different subsystems or components. During this cost design process, a proper balance among the four attributes (performance, durability, packaging, and cost) needs to be achieved. Normally, an engineer in the areas of performance, durability, or packaging is not sufficiently knowledgeable or qualified to conduct the system cost design. The cost design requires an in-depth and up-to-date knowledge of the competitive benchmarking product costs for all the engine components and advanced technologies, ranging from the overall powertrain cost of OEM products to the cost of global suppliers' component products. The cost design also requires a broad knowledge on the relationship between cost and the other three attributes. Moreover, the trend of modularization and integration of subsystem designs in systems engineering environment requires the cost engineer to examine and propose the most cost-effective approaches on how to conduct the modularization and integration. In addition, the cost engineer also needs to identify the party in charge of the modularization and integration. For example, a supplier may prefer to provide the engine manufacturer with a wholly integrated subsystem of waste heat recovery (consisting of heat exchangers, turbines and electronic controls) and charge a high price for the integrated package. Although it may look appealing to a performance or durability engineer due to the ease or convenience, this may not be a financially viable option for the engine manufacturer. The manufacturer may prefer to modularize the waste heat recovery subsystem in a different and more cost-effective way to different global suppliers and conduct the integration work by itself without the need for paying a high price to others. A system performance or system durability engineer may be capable of proposing a system design plan to meet the functional or reliability requirements, but he or she is usually not eligible or available to optimize the cost structure of the product. All these cost-based activities have to be planned, conducted, and optimized by a highly qualified cost engineer. It is obvious that such a job function of system cost design is an independent full-time engineering function rather than a small supplemental function attached to the design.

1.4.3 The process of design target costing

The cost design activity of a cost engineer in engine development is a part of the overall financial analysis for the new product development. The financial analysis is characterized by four main phases as explained by Menne and Rechs (2002), namely (1) cost target calculation, (2) piece cost calculation, (3) piece cost verification, and (4) monitoring of running costs. Product cost design occurs from program introduction to Job 1 (start of production). The process of design for profit starts at the top with market price and profit goals. At the beginning of the program, from a financial perspective, the target product cost is set as the maximum engine cost which allows the market price of an engine, negotiated with a customer, to achieve a certain level of return on net asset. This financial cost target is the fully loaded cost including purchased materials and parts, product tooling, manufacturing labor, warranty cost, etc. Eventually, the cost target is stripped down to the engine system cost of total product materials and parts. The target is then translated by the engine system design team to a technology roadmap through integration and optimization of subsystem design costs while trying to meet customer functional requirements. There is a negotiation and iteration involved during the cost target setting process in order to close the gaps. Sometimes trade-offs need to be made between functionality and cost. The initial established target cost is then distributed among different phases and work areas, and is cascaded downward to lower subsystem and component levels. The target cost is finalized through iterations and cost reduction activities. Product cost control is conducted by comparing actual reported costs against estimated costs, identifying and focusing attention on those work areas that deviate seriously from the initial estimates.

1.4.4 Objectives of engine system cost analysis

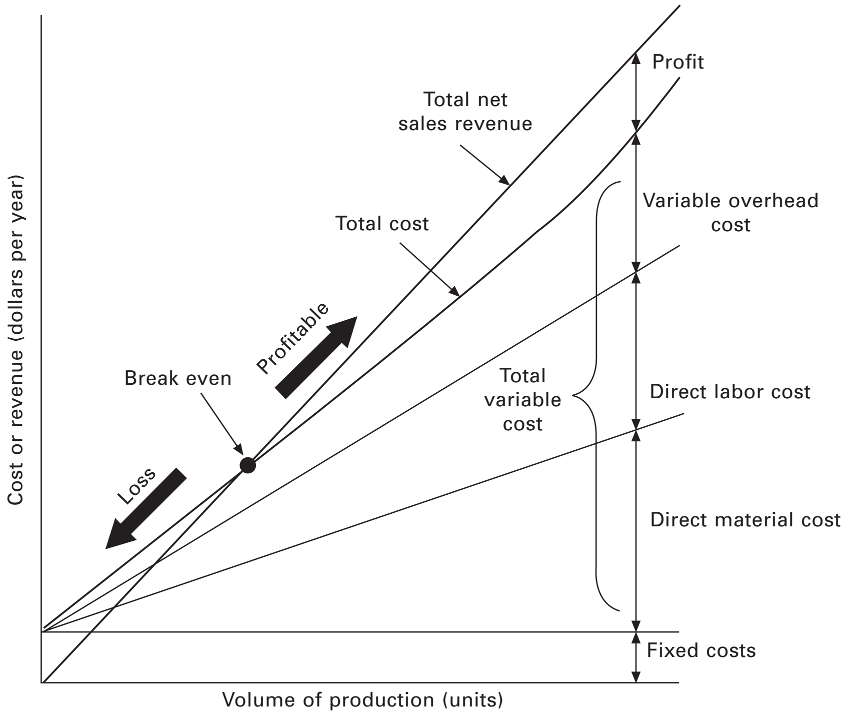
Engineering design drives cost. The cost engineer carries out an engineering design function for the cost–technology structure of the product. The cost engineer is neither a financial controller in program management nor a financial analyst in purchasing or finance. The cost engineer neither plans/controls engine development labor costs, nor calculates pricing and corporate profitability. Instead, the engine system cost engineer focuses on the technical design of the cost structure of the engine, which is mainly related to materials/parts cost and product life. The objectives of engine system cost design/analysis usually consist of the following six types:

1. Corporate cost–benefit analysis for various cutting-edge engine technologies (e.g., aftertreatment, advanced combustion, alternative fuels, fuel economy improvement, or hybrid powertrain technologies).
2. Demand–revenue–cost–profit analysis for engineering design decisions to maximize corporate profitability (i.e., design for profitability).
3. Cost payback analysis to optimize between initial product cost (or standing charge) and running cost (e.g., total cost recovery over time through reduced fuel consumption to offset the higher initial cost).
4. Optimized unit cost structure for selecting engine technologies and balancing subsystem costs with value-based design to meet the functional requirements and the cost target during production engine design.
5. Life cycle cost saving estimate for design improvement.
6. Competitive benchmarking cost analysis.

1.4.5 Classification of cost and factors in cost analysis

Total product cost can be defined as the total of unit manufacturing cost, life cycle cost, and quality loss cost. When using cost to calculate the value associated with a functional requirement (recall that ‘value’ is defined as the ratio between function and cost), it is important to use the total product cost. The unit manufacturing cost includes all the manufacturing costs such as variable cost, fixed cost and tooling cost. The life cycle cost includes all the operational costs and the costs of warranty, repair, and routine maintenance. The quality loss cost takes into account less tangible cost incurred by the manufacturer, the customer, and society.

The product cost can also be broken down into two categories: variable cost and fixed cost. The variable cost includes direct material, direct labor, and



1.20 Illustration of cost, price, volume, and profit.

variable overhead (e.g., direct material or parts cost, supplier transportation cost, warranty cost, direct manufacturing labor cost, non-wage labor cost, manufacturing tooling cost, and overhead costs). The variable cost changes in total dollar amount with volume. The fixed cost is the one which remains constant in the total amount of dollars throughout a specific period of time even if there are changes in volume. Examples of fixed cost include capital expenditure and fixed overheads. The fixed cost is 'diluted' or spread over the volume. The engineering cost can be a part of overheads, or can be handled as a separate item if the expense is exceptionally large. Figure 1.20 illustrates the relationship among costs, price, volume, and profit.

Generally, a cost engineer considers the following factors in order to conduct cost analysis:

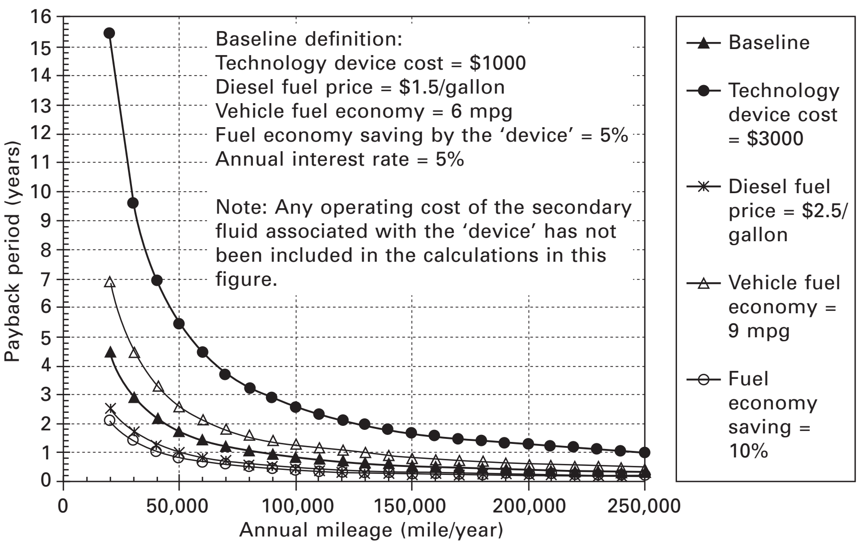
- variable cost
- fixed cost
- production methods
- production volumes
- product life (annual and life cycle mileage, number of hours or years), user usage profiles (e.g., vehicle weight, driving cycle), and operational costs (e.g., fuel cost, oil and maintenance costs, aftertreatment running cost)

- bill of materials (BOM), or a cost bill-of-materials tree
- product specifications and technical drawings
- purchasing and supplier source/cost data
- interest rate, exchange rate, and inflation rate
- depreciation.

Among the above factors, the purchased materials/parts cost and product life are the most important ones for diesel engine system cost engineers. Moreover, many factors contain a range of geographic or temporal variation or uncertainties (e.g., diesel fuel price, vehicle annual accumulated mileage). System cost analysis and estimate is not an exact science. For its statistical/probability analysis the reader is referred to Ostwald and McLaren (2004).

1.4.6 The method of engine system cost analysis

Figure 1.21 illustrates an example of the cost payback analysis with many influential factors included. It is assumed that a device associated with certain engine technology imposes an initial cost for the customer. The device could be any arbitrary one, for example, a waste heat recovery device such as an organic Rankine cycle system, or a diesel aftertreatment device. Engine BSFC or vehicle fuel economy can be improved by using such a device hence the operating fuel cost can be reduced. A straightforward discounted cash flow methodology is used to identify the time period required to recover the cost of the device. The interest rate and net present value are considered to include the time value of the money. The horizontal axis in the figure gives the anticipated annual mileage accrued for a diesel vehicle. The vertical axis



1.21 Cost payback analysis.

shows the time required for the expected fuel savings to offset the device cost. The payback period is the ratio of the amount of money spent on the device to the amount of money saved. It is observed that vehicle mileage has an exponential effect on the payback period. The device cost has a dominant effect on the payback period. Retail diesel fuel price, vehicle fuel economy and fuel economy saving percentage all have very strong effects on the payback period. The general trends that result in a short payback period can be summarized as follows: lower device cost, increased fuel economy, increased mileage, and increased diesel fuel price. For heavy-duty long haul trucks a payback period of less than three years is desirable since many fleets replace trucks after approximately three to four years.

Figure 1.21 is a general purpose illustration to demonstrate the methodology used in cost analysis. It should be noted that in Fig. 1.21 the additional cost associated with any secondary fluid such as the liquid-based urea used in SCR (if the device refers to an SCR system) has not been included. When the operating cost of such a secondary fluid is included, the payback period will be much longer (worse) than that shown in Fig. 1.21.

1.4.7 Review of existing engine cost analysis methods

Fundamental theories of cost analysis and estimating for engineering and management have been provided by Park and Jackson (1984), Ostwald and McLaren (2004), and Humphreys (2005).

In the cost–benefit analysis for automotive technologies, regulatory agencies and research institutes have been largely interested in the cost analysis related to advanced diesel emissions aftertreatment technologies, advanced fuels, well-to-wheel efficiencies of various powertrain technologies, and energy policies. Generally, the conclusions about a particular technology drawn from a global viewpoint of the entire society may not be valid at a corporate level due to the differences in the assumptions of sales volume and other technical or economical factors. Each company has a unique situation, and it is important to develop their own cost–benefit analysis in order to reach a wise business decision.

Walsh (1983, 1984) conducted studies on the costs and benefits of diesel particulate control for the North American market. Browning (1997) provided an overview of the technologies and incremental costs for meeting the US 2004 emissions standards for heavy-duty diesel engines. Burke (2003) used Excel spreadsheet cost models to show the trade-offs between fuel savings and economic attractiveness for different hybrid-electric powertrains (mild and full hybrids) in light-duty vehicles (passenger cars and mid-size SUVs).

State and local agencies use cost-effectiveness as one criterion to decide whether to implement a particular emissions control program. The cost-effectiveness is defined as the ratio of the dollar amount to the unit of effect produced by the cost. In a Texas government program, Prozzi *et al*. (2004) investigated the cost-effectiveness of an emulsified diesel fuel for highway construction equipment fleets by thoroughly quantifying many sources of cost impact (e.g., fuel cost penalty, implementation and conversion cost, fuel economy penalty effects, re-fueling penalty, fuel–emulsion mixing cost, cost of lower equipment productivity due to the loss of torque, increased maintenance cost, fuel storage cost, etc.). They concluded that the fuel tested (Lubrizol's PuriNO$_x$) is a relatively high cost strategy for NO_x reduction, and it will become much less cost-effective in reducing NO_x emissions as the fleet replaces the older engines with new, cleaner electronically-controlled engines. Matthews *et al*. (2005) evaluated the effects of an ultra-low-sulfur diesel fuel on emissions, fuel economy and maintenance cost, and they also calculated the cost-effectiveness in terms of dollars of cost penalty per ton of NO_x removed.

Life cycle assessment is a 'cradle-to-grave' approach for assessing a technology from production and use to final disposal. Ginn *et al*. (2004) presented a comprehensive life cycle economic assessment to compare four alternative technologies to conventional diesel engine idling for heavy-duty vehicles. The long haul trucks are idled for a long period of time (e.g., overnight) to heat or cool the cabin, to keep the engine warm, and to run electrical accessories. The idling time can be 2400 hours per year. It results in a large amount of emissions, fuel consumption, and increased engine wear. Ginn *et al*. (2004) compared four alternatives (auxiliary power unit, direct-fired heater, truck stop electrification, and advanced truck stop electrification) and assessed their emissions benefits, environmental impact, fuel savings, maintenance/wear savings, payback period, net present value, and the emissions savings per dollar cost of technology (i.e., the reciprocal of cost-effectiveness).

Li *et al*. (2004) introduced a demand–cost–profit economic analysis method applied to the design decision making about engine manifold surface finishing. The performance benefits on power and BSFC due to improved manifold surface roughness were simulated with an engine cycle simulation software package. The microeconomic theory and a demand–cost–profit model were introduced along with optimization techniques (to maximize the profit associated with the design) to bridge engine performance simulation, cost analysis and business decision making. This work provided an attempt on design for profit.

Diesel engine system cost analysis is a new and challenging inter-disciplinary field. It will receive increasing attention from the technical community.

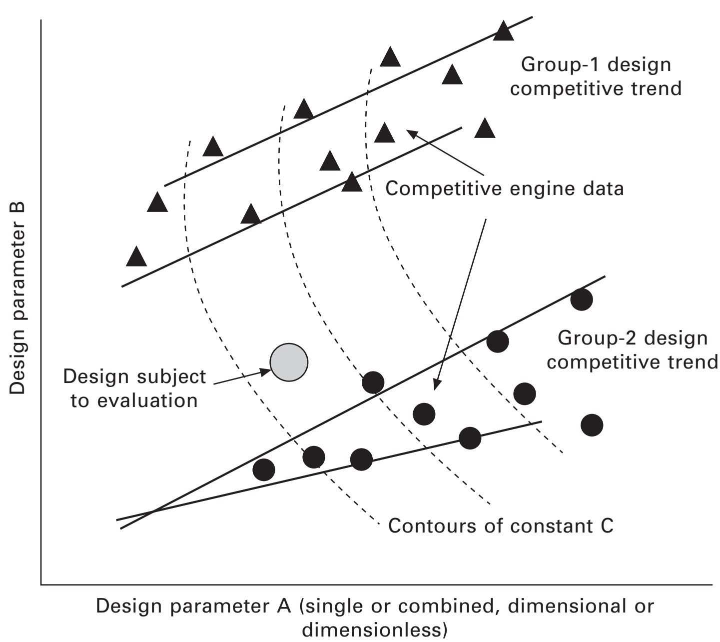
1.5 Competitive benchmarking analysis

1.5.1 The need for competitive benchmarking analysis

Benchmarking is a systematic approach to identifying standards for comparison. It reveals the gaps in design attributes and provides ideas for design improvement. Benchmarking should include best-in-class objective attribute measures and research into how this attribute is achieved.

1.5.2 Methods of competitive benchmarking analysis

Heavy-duty diesel engine design details can be found in numerous books (for example, the overview by Merrion and Weber, 1999; Heisler, 1995; and the vehicular engine design textbook by Hoag, 2006). These works cover empirical design guidelines for engine layout and component details including engine balance, cylinder head, block, water jacket, bearing, gasket, piston, crankshaft, camshaft, etc. Advanced design software and finite element analysis have been widely used in those areas. However, from the point of view of diesel engine system design, another way to enhance the quality of analysis in these traditional design areas is to apply a competitive benchmarking design analysis technique. The technique analyzes one fundamental design parameter against another by using a large amount of different engine design data to form empirical trend curves in order to check whether the subject design falls within or outside the range of the trend. The concept is illustrated in Fig. 1.22. A simple example is to plot stroke-to-bore ratio versus engine displacement for multiple engine power ratings. Such a competitive analysis is very effective for stroke and bore design. In fact, advanced competitive



1.22 Approach of competitive benchmarking analysis.

analysis is much more complex than that, because it requires heuristic modeling and sometimes similarity theory to extract and group apparent design parameters in order to reflect the fundamental physics. Usually, in each complex mechanical system, it is possible to extract one or several fundamental design parameters in a certain combined form that manifests their physical nature of performance or durability in a rather intuitive and simple manner. Through such an analysis, major design problems can be quickly identified for further examination. It is very important to identify those characteristic design parameters for each component and system.

1.5.3 Basic engine system design parameters

Engine system design features are usually measured by certain basic system parameters describing the overall engine performance or characteristics, such as power density, vehicle power-to-weight ratio, high altitude capability without derating, cold start capability, adequate cooling capacity at hot ambient temperature, etc. Basic engine system design parameters are very important for analyzing performance and durability at an overall system level. Some of the commonly used basic engine system parameters are presented as follows:

- The number of cylinders, which affects the compactness of the engine and torque balance.
- Engine displacement, which affects engine power density.
- Cylinder centerline distance, conrod length, and piston compression height.
- Engine cylinder bore diameter, which affects engine weight and wear.
- Engine stroke, which affects stroke-to-bore ratio, mean piston speed, engine speed, and engine friction.
- Stroke-to-bore ratio, which affects bore, stroke, engine speed, combustion chamber clearance, emissions, and heat rejection.
- Engine compression ratio, which affects emissions, thermodynamic cycle efficiency, peak cylinder pressure, engine friction, and cold start.
- Rated power, which usually determines the most severe thermo-mechanical condition for structural design.
- Rated BMEP, which is the rated torque per engine displacement volume and is comparable for different engine sizes.
- Rated speed, which affects engine rated power and engine–transmission matching for vehicles.
- Rated mean piston speed, which affects the gas-flow velocities in the intake/exhaust ports and the cylinder, the thermal load and inertia load in components, engine friction, and wear.
- Peak torque, which determines the torque backup and turbine area.

- Peak torque speed, which affects the speed reserve and engine–transmission matching.
- Peak BMEP, which is the peak torque per engine displacement volume and is comparable for different engine sizes.
- The coefficient of torque backup, which is equal to the ratio of peak torque to the torque at rated power, and affects vehicle drivability and the engine stability to resist a load increase.
- The coefficient of speed reserve, which is equal to the ratio of rated speed to peak-torque speed, and affects vehicle drivability.
- The coefficient of adaptability, which is equal to the coefficient of torque backup multiplied by the coefficient of speed reserve.
- The coefficient of rated intensity, which is equal to rated BMEP multiplied by rated mean piston speed.
- Engine weight.
- Engine specific power (also called power density), which is the rated power per engine displacement volume.
- Engine power per piston area, which basically reflects the thermal load per unit area of the combustion chamber.
- Engine weight per displacement volume (also called weight density), which reflects the effectiveness of structural design and the compactness of the engine.
- Engine specific weight, which is the ratio of the weight density to the power density, i.e., the reciprocal of the power-to-weight ratio.
- Engine volume per displacement volume, which reflects the compactness of engine packaging design. (i.e., volume density).
- Engine specific volume, which is the ratio of the volume density to power density.
- A characteristic ratio of engine power or torque to an engine application parameter, which is a fundamental matching parameter between the engine as a power or torque source and its application. For example, it can be the ratio of engine rated power or peak torque to vehicle weight for on-road trucks.

A brief discussion is provided below to reveal the relationships among some of the above basic engine system design parameters. According to the definitions of BMEP and mean piston speed,

$$\varpi_{BMEP} = \frac{\dot{W}_E n_s}{V_E N_E} \tag{1.2}$$

$$v_{mp} = 2S_E N_E \tag{1.3}$$

engine brake power can be derived as follows, proportional to BMEP, square of bore diameter and mean piston speed (note that $n_s = 2$ for four-stroke engines and $n_s = 1$ for two-stroke engines),

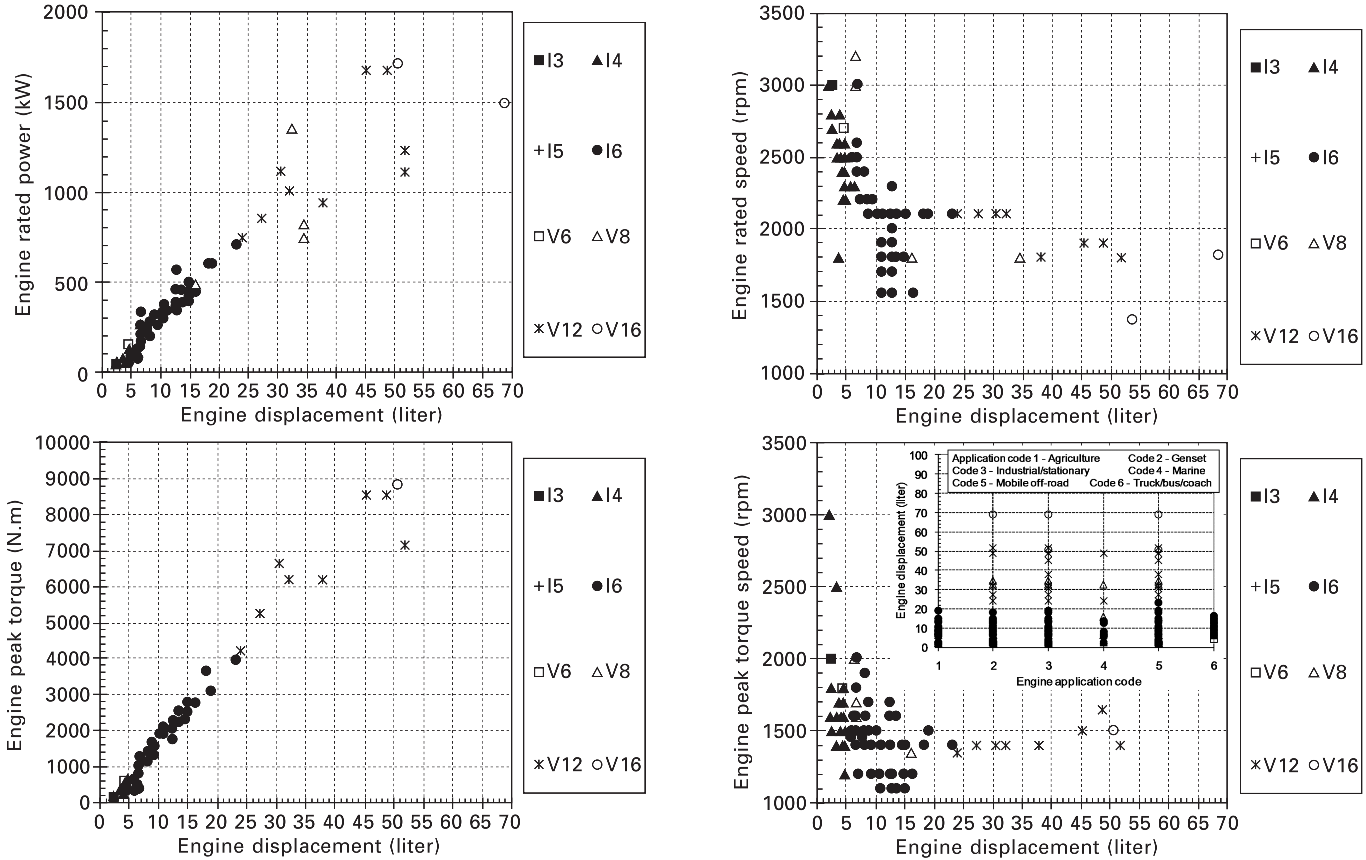
$$\dot{W}_E = \frac{\varpi_{BMEP} V_E N_E}{n_s} = \frac{\varpi_{BMEP}\,\pi B_E^2 n_E S_E N_E}{4 n_s} = \left(\frac{\pi}{8 n_s}\right)\varpi_{BMEP} B_E^2 n_E v_{mp} \qquad 1.4$$

Engine maximum BMEP reflects the structural capability to sustain peak cylinder pressure and thermal load. The maximum allowable BMEP has been constantly increasing over the past several decades in order to meet the demand for higher torque output produced by a smaller engine. From equation 1.4 it is observed that increasing cylinder bore diameter can increase engine power at a given BMEP. Usually, the aerodynamic resistance of intake and exhaust processes and the component stress caused by inertia loading are proportional to v_{mp}^2. Piston-assembly friction force and thermal flux rate are proportional to v_{mp}. The wear rate of piston assembly may increase proportionally or exponentially with v_{mp}. Therefore, an increase in v_{mp} may result in an increase in component stress and thermal loading, a reduction in ϖ_{BMEP} and a decrease in engine life due to an increase in wear. The mean piston speed is a critical parameter affecting engine performance, durability and reliability. It needs to be selected with great care.

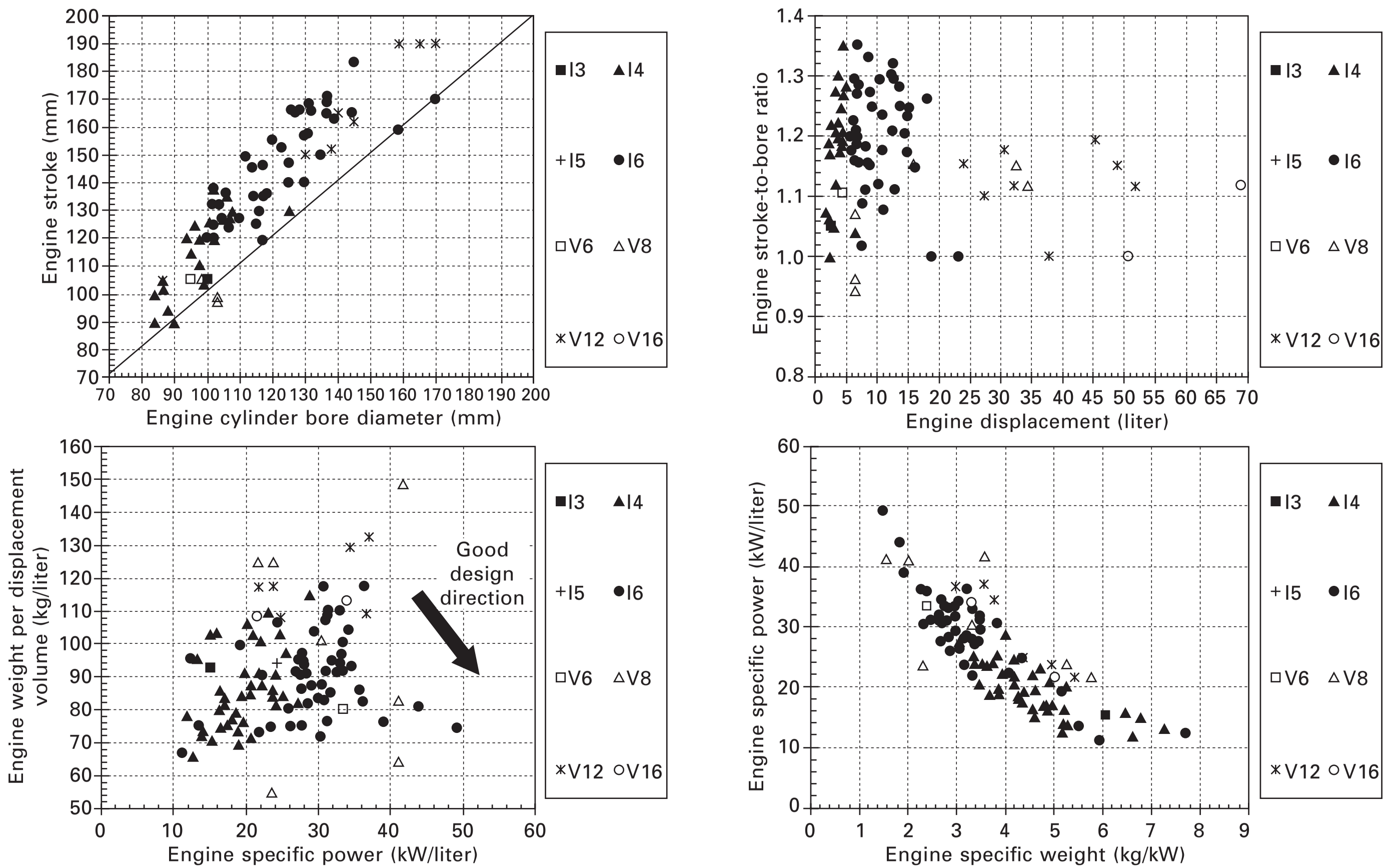
Usually the values of all the above basic system design parameters are available in the literature or engine product catalogs. A competitive design analysis can be readily conducted with this public information.

1.5.4 Competitive benchmarking analysis of engine performance

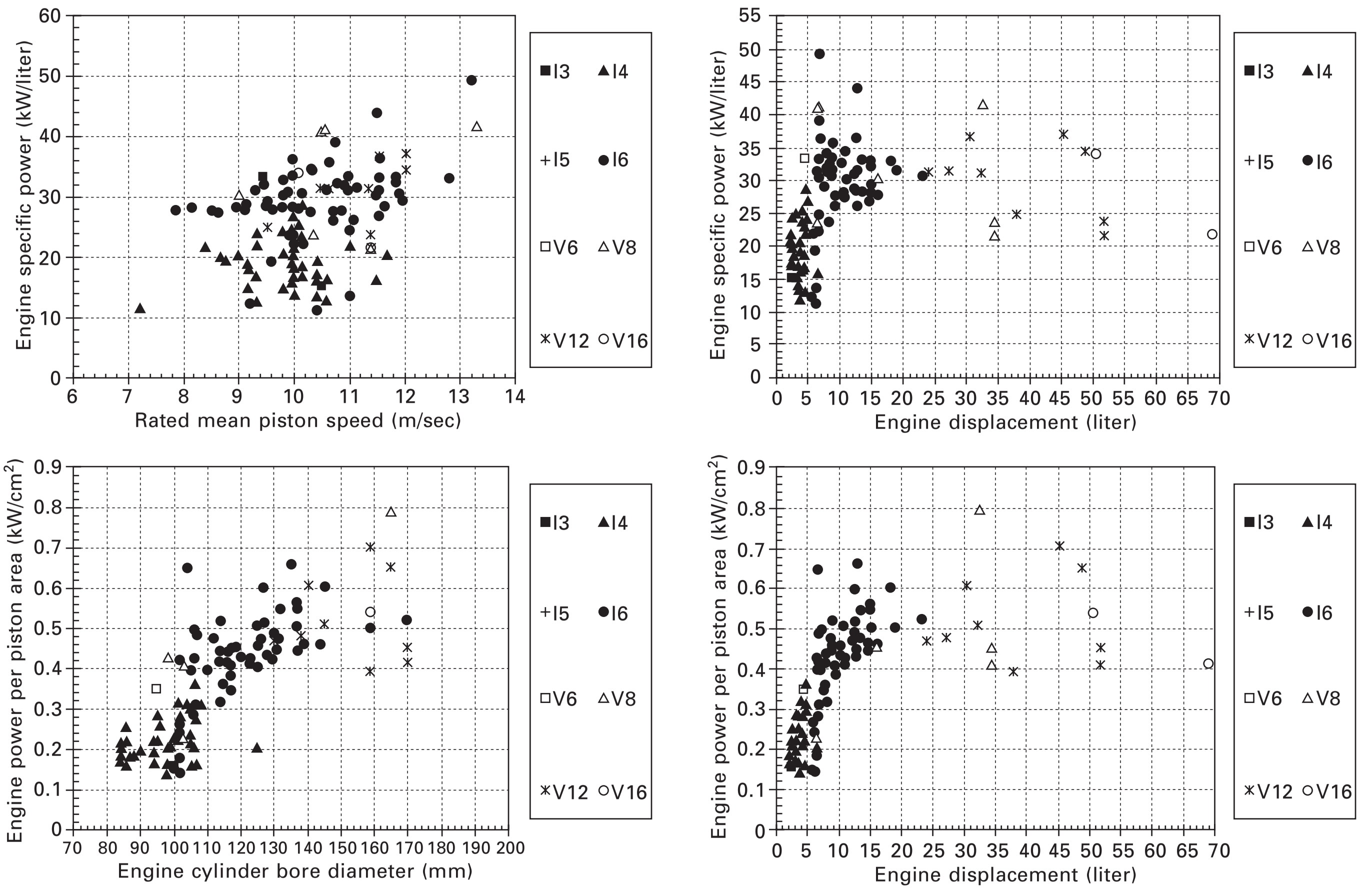
Figure 1.23 illustrates the analysis of the relationships among several basic system design parameters for a large number of heavy-duty diesel engines in North America, including both on-road and off-road applications. It is observed that there is a fairly linear correlation between engine rated power (or peak torque) and engine displacement over a wide range. The majority of the engines are rated around 1500–3000 rpm, with a rated mean piston speed around 8–12 m/s. Most engines with a displacement larger than 7 liters are rated around 1500–2400 rpm. The peak torque speeds are usually around 1100–1700 rpm. Most of the engines have a stroke-to-bore ratio around 1.1–1.3. There is a clear trend of decreasing engine specific power with increasing engine specific weight. There is also a trend of increasing engine power per piston area with increasing cylinder bore diameter. Most of the engines have a coefficient of torque backup around 1–1.45, and a coefficient of speed reserve around 1.3–1.8. There is a clear correlation between engine dry weight and engine displacement. There is a large variation in specific weight from 2 to 8 kg/kW, mainly caused by small engines. Most engines with a displacement larger than 7 liters appear to have a specific weight around 2.8–3.6 kg/kW.



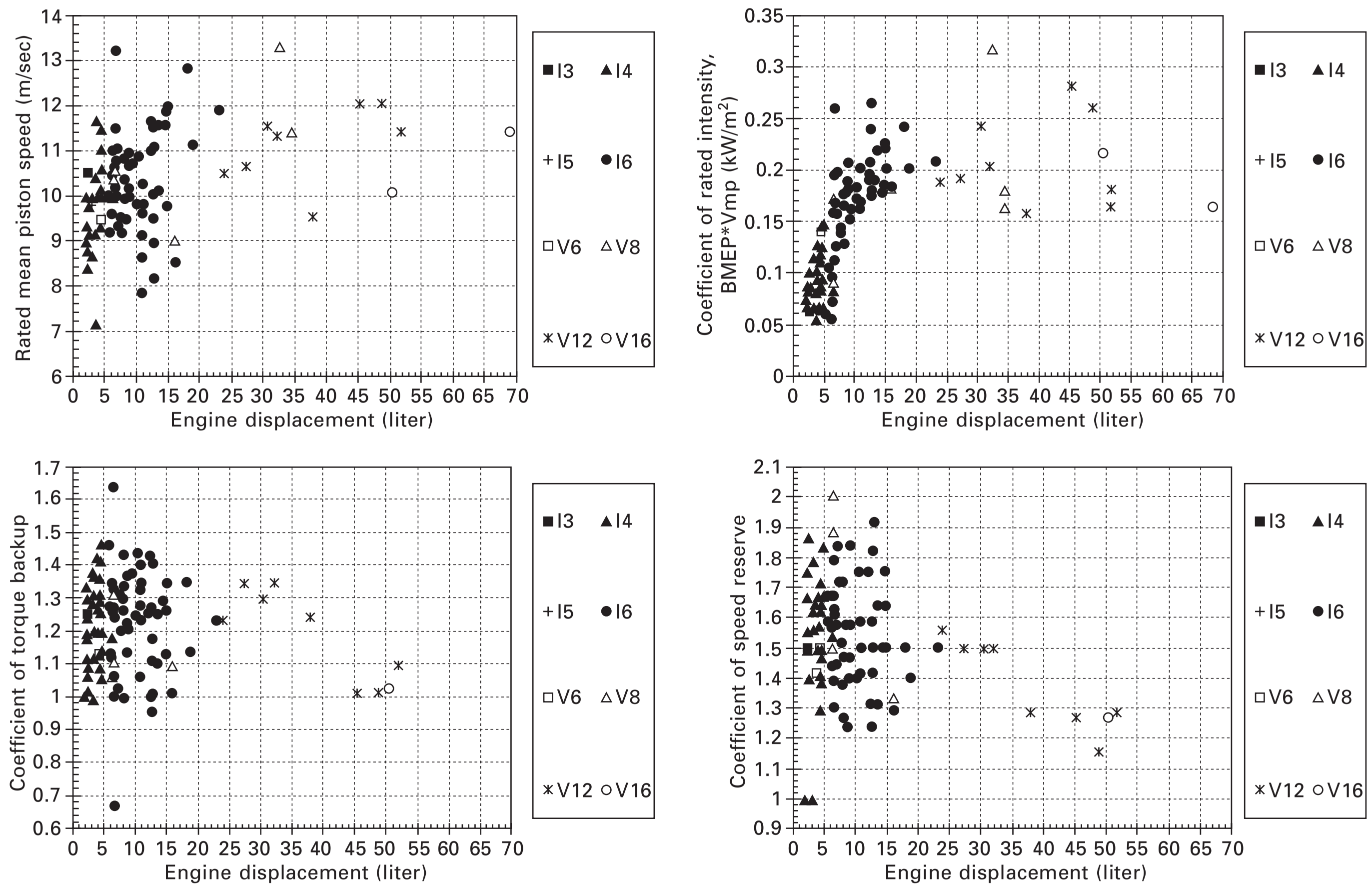
1.23 Competitive benchmarking performance analysis of 2008 North America HD diesel engines.



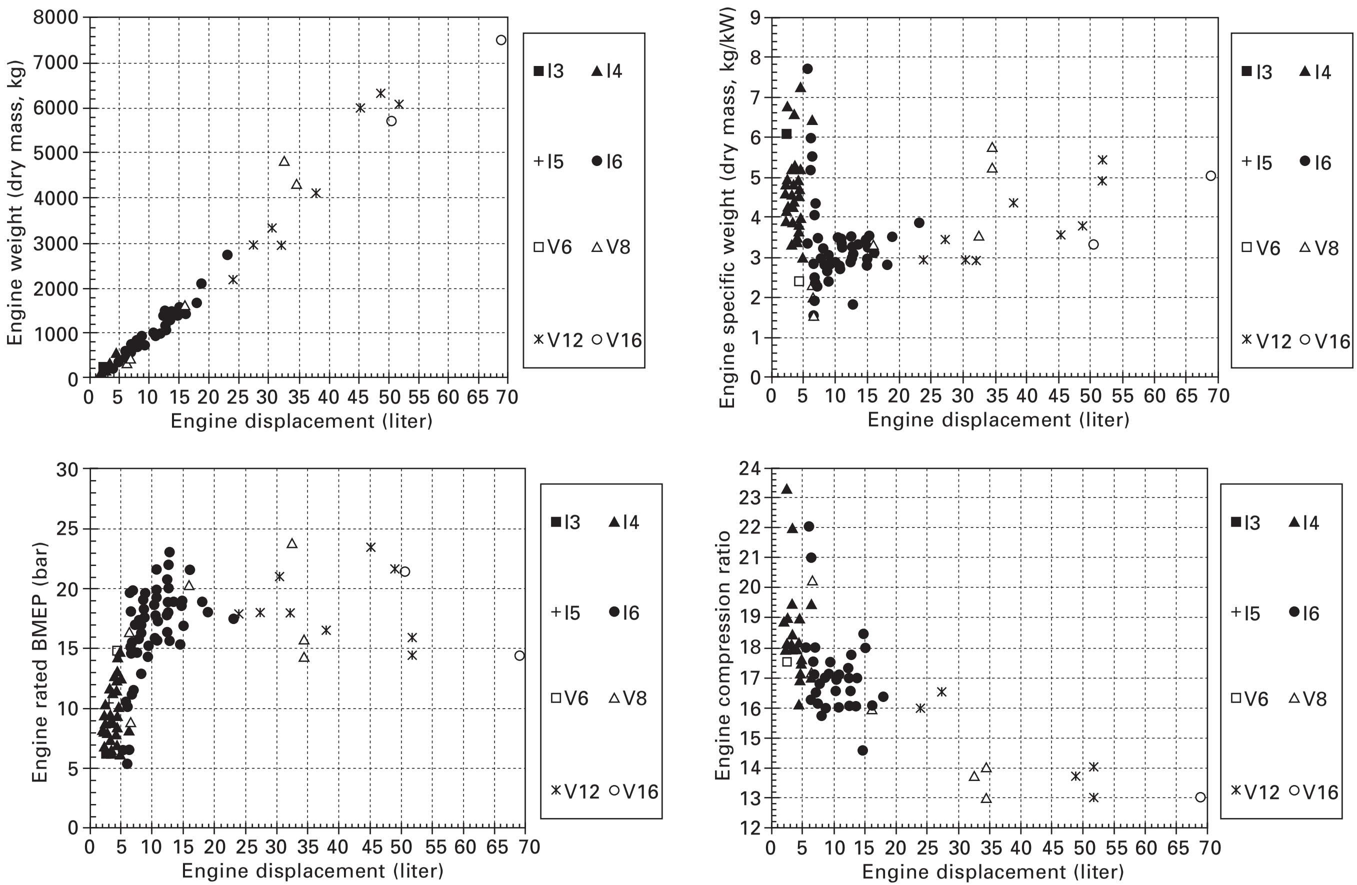
1.23 Continued



1.23 Continued



1.23 Continued



1.23 Continued

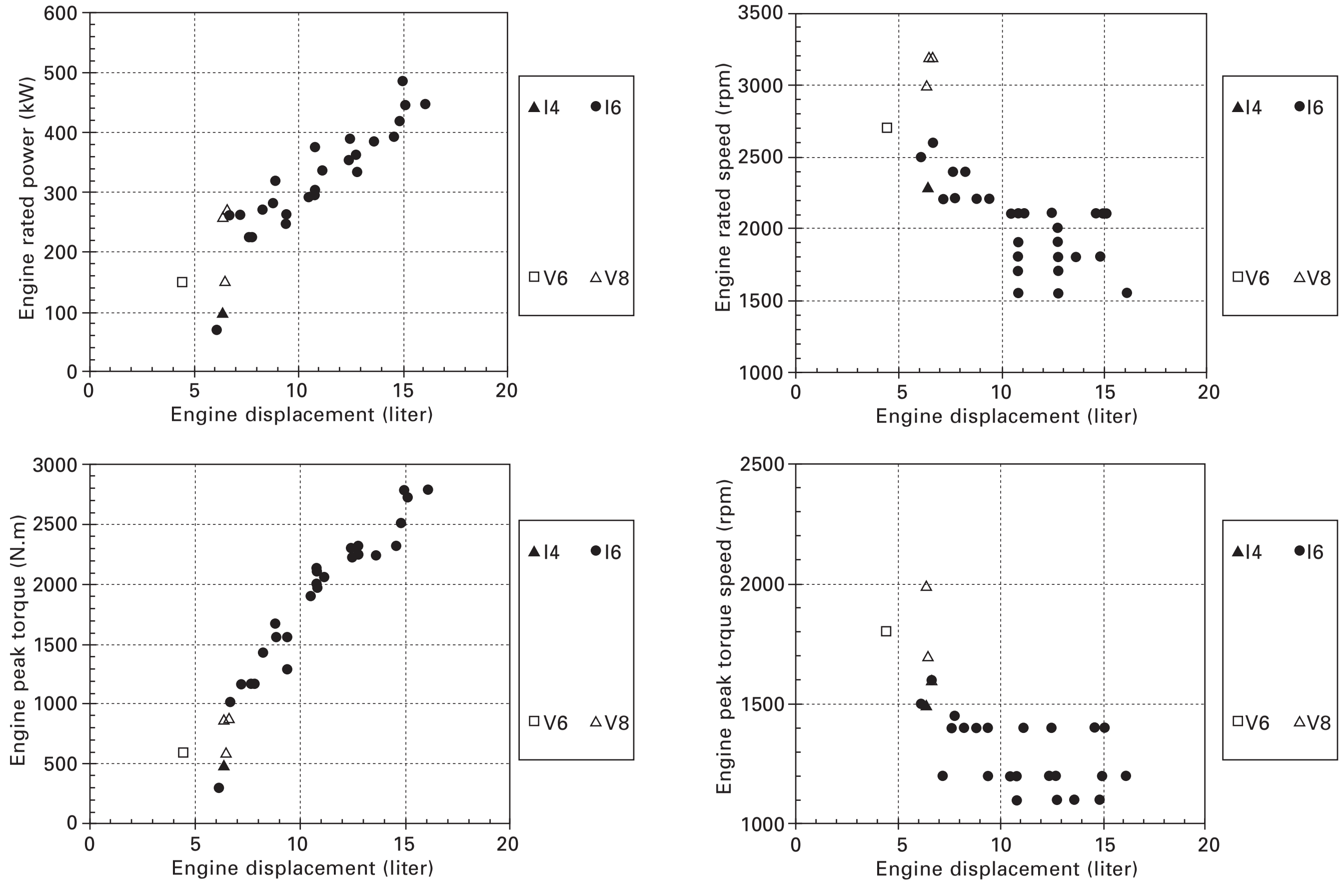
Figure 1.24 illustrates the analysis for the on-highway heavy-duty diesel engines in North America from 5 to 16 liters. There is a clear trend of increasing engine rated power (or torque) with increasing engine displacement. There is also a clear trend of decreasing engine specific power with increasing engine specific weight. Engine dry weight increases almost linearly with bore diameter. Figure 1.25 presents the analysis for the light-duty diesel engines and some heavy-duty diesel engines used in pickup trucks.

1.5.5 Competitive benchmarking analysis in mechanical design

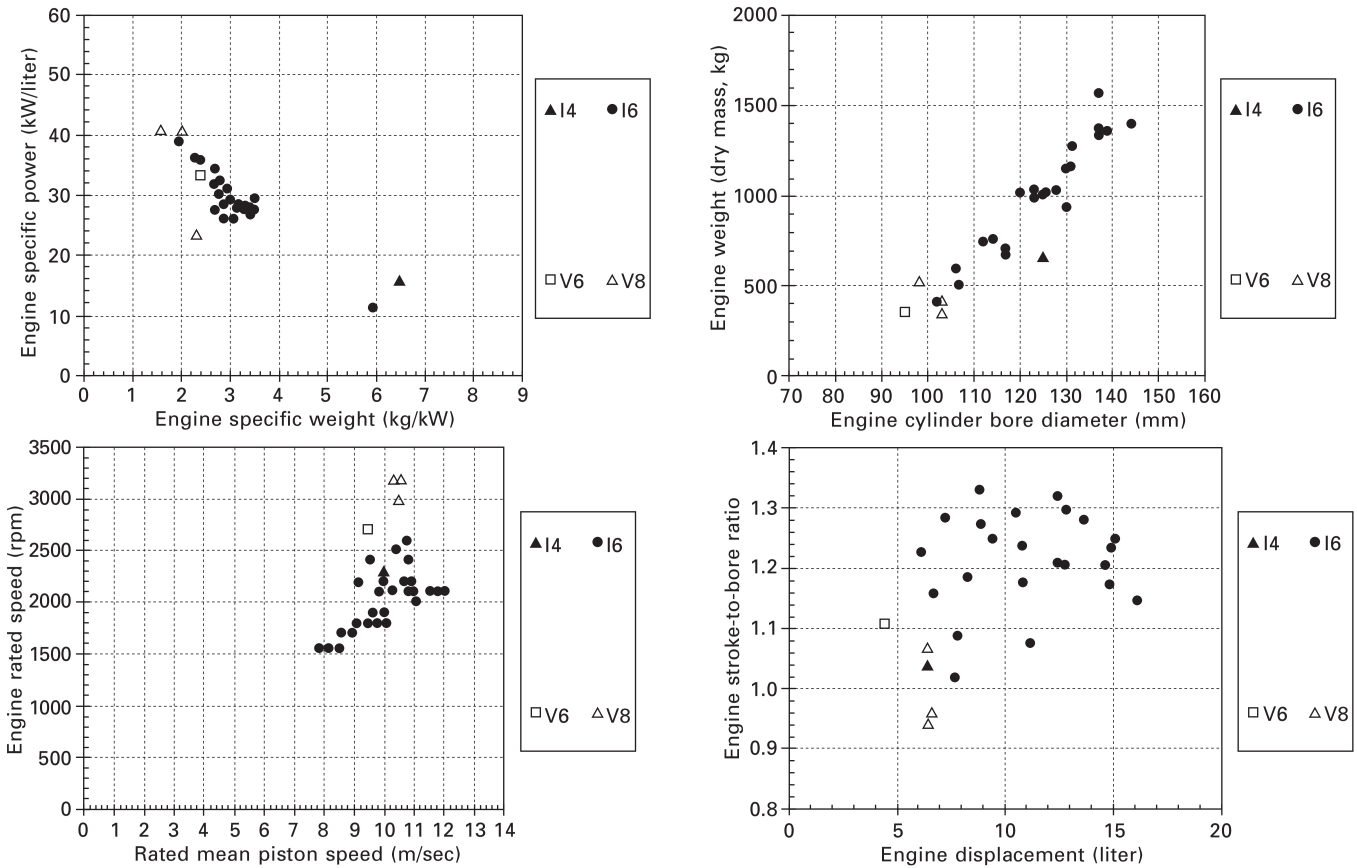
Engine block and crankcase designs

The design issues concerning the number of cylinders, engine balance, Vee bank angle, and uneven firing order were discussed by Hoag (2006). Most diesel engines used for automotive cars and trucks are inline 4-cylinder (I4), inline 6-cylinder (I6), V6, and V8. The reciprocating mass is roughly proportional to the cube of the cylinder diameter. The I4 engine has a shorter block length. Although the secondary reciprocating inertia force of an I4 engine is not balanced, its magnitude is relatively small, being only a fraction of the primary inertia force. For the I4 engines having a small cylinder diameter, the engine balance is still satisfactory. Therefore, I4 engines are popular in passenger cars and light trucks that have stringent requirements on engine size. I6 engines have the best engine balance and strong exhaust pulse energy for turbocharging. I6 is the most popular configuration for commercial truck and bus engines. V8 engines with 90° bank angle are also popular because of good engine balance and low height. V6 engines are the most compact in length and height, and are used in some passenger cars and light trucks. But the V6 engines with 90° bank angle have an unbalanced secondary reciprocating inertia force, and the V6 engines with 120° bank angle have an unbalanced primary inertia force.

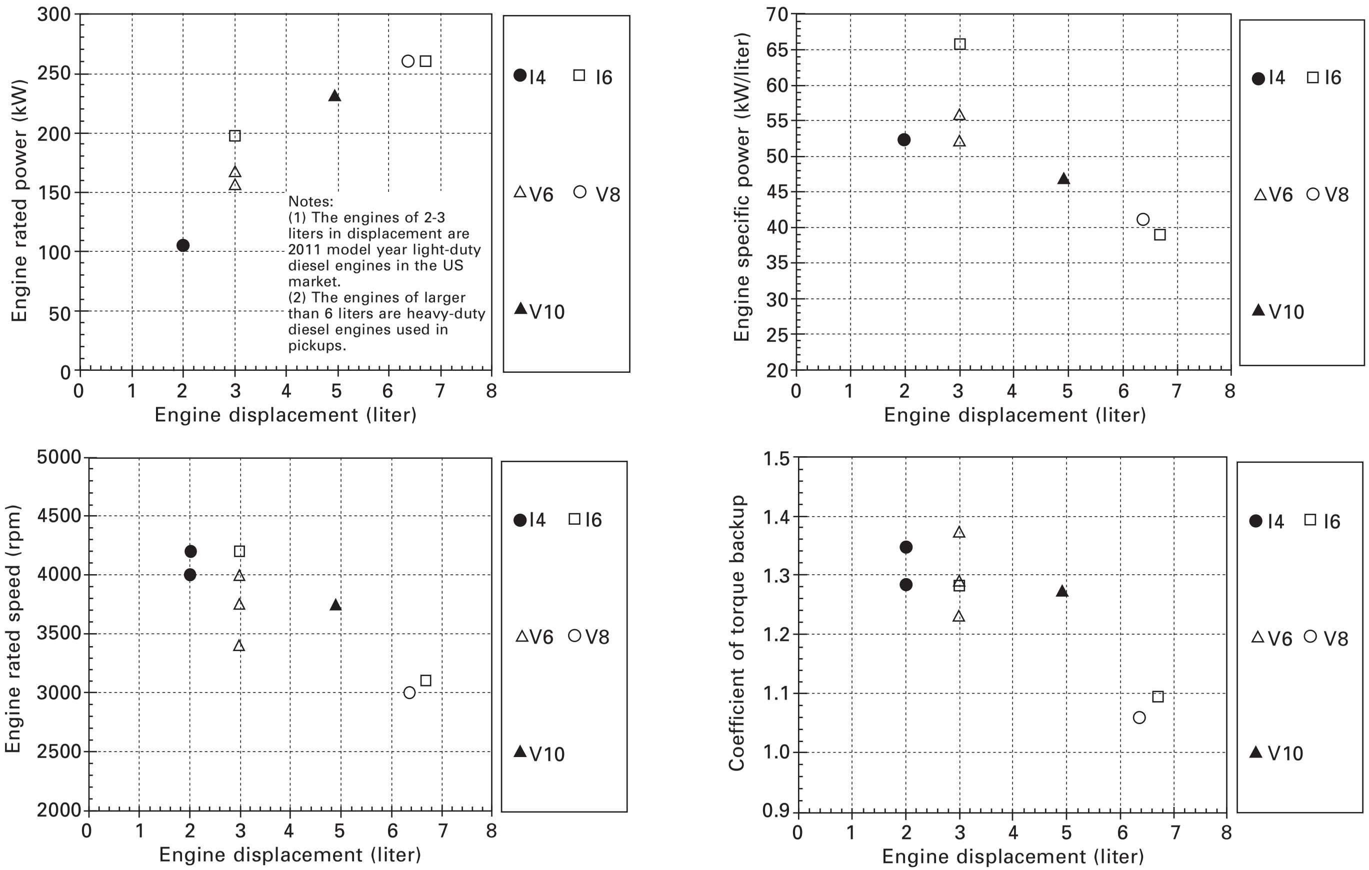
High stiffness and strength, low NVH, and weight are important for engine block design. The potential competitive benchmarking parameters include the following: engine width; the geometric clearances between rotating moving parts and the inner wall of the engine block; engine length; and the 'C_E/B_E' ratio, which is the ratio of the cylinder-to-cylinder centerline distance to the cylinder bore diameter. Note that the C_E/B_E ratio reflects the compactness of the engine length and is affected by crankshaft length, cylinder head design, and liner design. Plotting the C_E/B_E ratio or typical thicknesses of the engine block or head as a function of cylinder bore diameter for various engines is an example of conducting competitive design analysis. Connecting rod length and piston compression height are the two main parameters that affect engine height.



1.24 Competitive benchmarking performance analysis of 2008 North America on-highway HD diesel engines.



1.24 Continued



1.25 Competitive benchmarking performance analysis of North America light vehicle diesel engines.

Cylinder liner design

The following are important design considerations for cylinder liner: low friction and wear occurring on the liner inner surface between the liner and the piston ring or piston skirt; minimum cavitation occurring on the coolant side of the liner caused by piston slap and liner vibration; sufficient cooling at the liner top; proper metal surface temperature across the circumferential and axial directions of the liner; minimum bore distortion; and high stiffness for NVH reduction.

Cylinder head design

The subjects of cylinder head design usually include the following: number of valves; valve diameter and seat angle; port flow discharge coefficient and swirl ratio; intake and exhaust port orientation (to avoid intake port heating by exhaust gas); wall thickness; exhaust port length and heat rejection losses; cooling; fire deck temperature control in the area between the valves in the cylinder head (to prevent thermo-mechanical fatigue); cylinder head gasket design, etc. The basic competitive benchmarking parameters for cylinder head design include:

- the ratio of port diameter to port length;
- the ratio of port length to cylinder bore;
- the characteristic curve of swirl ratio vs. port flow coefficient, which is used for the evaluation of the trade-off between in-cylinder turbulence and volumetric efficiency;
- the number of cylinder head bolts and the head height, which are used for stiffness and sealing evaluation
- the ratio of cylinder head bottom wall thickness to bore diameter which is used for the evaluation of the trade-off between mechanical load and thermal load.

Connecting rod, crankshaft and bearing designs

The design guidelines for power conversion components include the following: sufficient mechanical strength and stiffness in compression, stretch, torsional and bending motions under cylinder gas pressure and inertia loading; light weight and light balancing weight; compact geometry; smooth chamfers and low local stress; proper structural design of the lubricant oil feed hole angle; proper settings of bearing clearances; low bearing wear, friction, impact noise, etc. The basic competitive benchmarking parameters include:

- the connecting rod length, which affects engine height, width, and piston assembly dynamics;

- the ratio of connecting rod length to crank radius, which affects secondary reciprocating force and piston side thrust;
- the crank pin diameter, which affects friction and rotating mass;
- the main bearing diameter, which needs to satisfy strength and lubrication requirements and helps adjust the designs of crank web thickness, crank pin diameter, and cylinder centerline distance;
- the overlap between the main bearing journal and the connecting rod crank pin journal, which affects crankshaft strength; and
- the bearing aspect ratio (i.e., length divided by diameter).

1.6 Subsystem interaction and analytical engine system design process

1.6.1 Engine subsystem interaction

The design of automotive diesel engines has been driven by emissions regulations, fuel economy, NVH, drivability, durability and product cost. Diesel engines are designed specifically for their applications. In on-highway applications, heavy-duty engines are used for heavy weight trucks and buses under highly loaded duty cycles with much time spent at full load. In contrast, automobiles and light trucks usually operate at very light loads and low engine speeds, with high speed and load encountered only at brief transients. The duty cycle difference determines the design features related to durability and fuel economy.

Engine design is one of the most sophisticated industrial design processes due to its complex nature that the functions of many components affect each other. Inside the engine, air is mixed with fuel to combust to produce power and generate emissions. The function and design of the fuel system are closely related to those of the combustion system. Engine system design is largely related to air handling. An engine has several major subsystems related to air delivery and pumping loss: intake manifold, exhaust manifold, cylinder head, valvetrain, turbocharger, EGR circuit, coolers, and exhaust aftertreatment. Their performance is affected by certain key parameters such as air and EGR flow rate, exhaust temperature, peak cylinder pressure, boost and back pressures. The optimization among those subsystems needs to be carried out carefully. Modern electronically controlled engines require an on-target precise design to deliver air–fuel ratio, EGR rate and intake manifold temperature for meeting the nominal design/calibration target at critical operating speed/load modes. Meanwhile, the durability design limits or constraints should not be violated. Neither an over-design nor under-design is acceptable. The following guideline can be used to determine whether a given design parameter should be optimized at a system level or at a subsystem or component level. If the parameter is shared among several subsystems, it is a system level parameter because its change impacts all the related subsystems. For instance, turbocharger matching is a system design issue because it has a direct impact on air–fuel ratio and EGR circuit restriction design. In contrast, an exhaust manifold runner design is more like a local component design issue. As system optimization propagates more widely and deeply throughout the whole engine to address the interactions among different components, the work scope and job boundary between 'system' and 'subsystem/component' could become blurred. The purpose of system engineering is to ensure excellent functional interaction with neighboring subsystems.

Engine subsystem interaction occurs at the interfaces between the individual subsystems, and between the major system and each subsystem. Interactions of engine–vehicle, engine–aftertreatment, engine–turbocharger and combustion–fuel system are just a few examples. The interaction manifests itself at both steady-state and transient operations (i.e., static and dynamic) and at the interface between hardware characteristics and software controls. Therefore, a systematic analysis is necessary in order to understand engine subsystem interactions. The analysis includes engine thermodynamic first law, second law, heat transfer, fluid mechanics, dynamics and electronic controls.

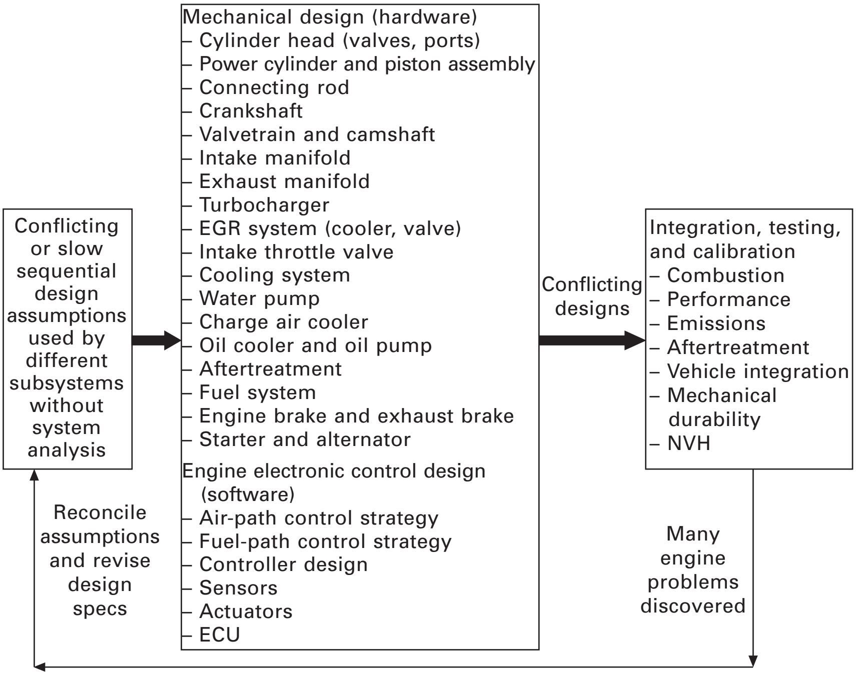
The tasks in engine system integration include the following for each design attribute:

- identify the system design objective parameters that are affected by each subsystem;
- analyze how the system design objective parameters are affected by each subsystem;
- optimize the subsystems and the system;
- optimize among the four design attributes (performance, durability, packaging, and cost).

1.6.2 Empirical engine design process

Figure 1.26 shows an example of the empirical 'trial-and-error' engine design process which lacks systems engineering and concurrent engineering. An empirical design process is characterized by the following:

1. Lack of advanced analysis.
2. Lack of a systems engineering approach to guide the design work. As a result, a precisely defined system design specification is not available at the beginning of the design process.
3. Sequential design. Conflicting requirements or assumptions are used by different subsystem teams.
4. Low efficiency and low quality in design.



1.26 Empirical engine design process.

The diesel engine is a very complex machine and it consists of many subsystems or components (as shown earlier in Fig. 1.5). In the empirical design process the detailed designs of different subsystems start without a formalized requirements analysis and engine system specifications. This leads to duplicated work and waste of resources. As many design problems are discovered during testing and integration at a later development stage, great effort has to be spent to reconcile the conflicting issues at the system level. Some component design work may have to start over again.

1.6.3 Advanced analytical engine system design process

Simultaneous engineering process

The advanced analytical engine system design process is characterized by systems engineering, simultaneous (or concurrent) engineering and advanced product quality planning (APQP). Simultaneous engineering is a design and/or manufacturing process where cross-functional teams strive for a common goal. It reduces development cycle time by replacing the sequential series of phases by a simultaneous engineering process. In the sequential process, the results are relayed from one area to the next area for execution. In the simultaneous process, all the subsystem/component areas carry out their design work concurrently with a unified system design specification.

APQP is a structured process of defining and establishing the steps necessary to assure a product satisfies customer expectations (APQP reference manual, 1995). In the initial phase of the program great effort is spent in engine system design by using advanced simulation tools to analyze the functional requirements. Then, the system design specifications are issued to the subsystem design teams. Because the primary risk factors have been analyzed and resolved, and the critical interface and functional analysis have been completed by the system design team, the subsystem engineers can start their designs with a much lower risk of failure. APQP uses process planning to ensure the design changes or the corresponding development efforts continuously decrease from the beginning of the engine program to the end for a smooth transition to the start of production.

The engineering product development process has evolved from the traditional slow process to today's accelerated process where the research and advanced engineering functions are combined. Organizational processes and project management for production engineering were introduced by Menne and Rechs (2002) in detail, including a summary of key checkpoints for a program. A simultaneous engineering process of detailed subsystem packaging and structural designs was provided by Dubensky (1993). Design and process tools were elaborated by Carey (1992) and Gale *et al.* (1995). Diesel engine design details were introduced by Merrion (1994).

Engineering functions in engine development

For each engine component, the engineering job functions can be classified into three types, namely analysis, design, and testing. The analysis function can be further divided into two main subareas: thermal/fluid performance and mechanical structure. It is usually impossible for one person or one department to carry out all the three job functions for all the components. In large engineering organizations usually the work functions of analysis, design, and testing are carried out in different departments. Cross-functional cooperation is the key to success.

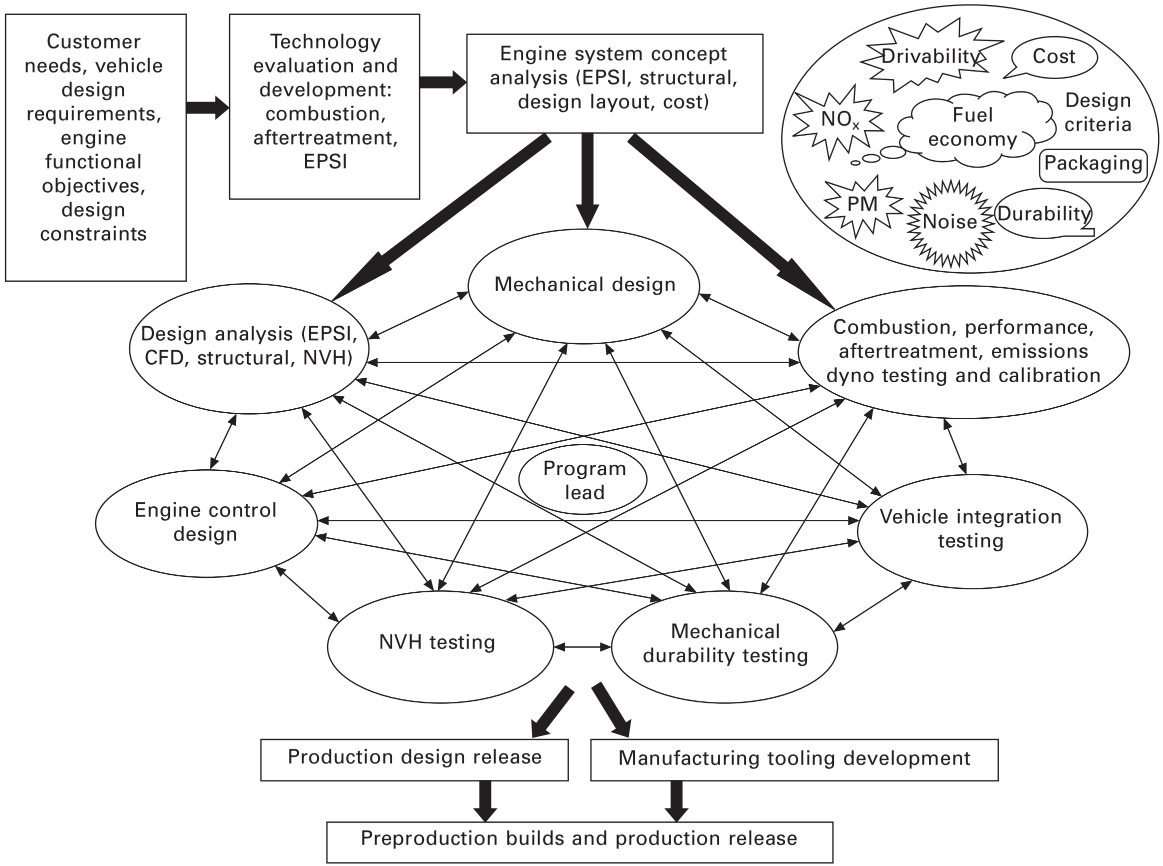
The job function of *analysis* can be defined as an activity to generate design specifications and analyze problems for the four engine design attributes (performance, packaging, durability, and cost) with advanced calculation tools. The traditional job function of *design* can be defined as an activity to realize the design specifications with computer aided design (CAD) tools, drawings, and prototypes. The job function of *testing* can be defined as an activity to validate the design specifications with experimental means.

In a product development process, design stays as a central role receiving support from analysis and testing. Simulation is a part of analysis. It supplies detailed information about many parameters which are difficult or impossible to measure, for example, heat transfer in all components, composition of gas flows, and instantaneous pressure or flow through valves or in manifolds. When setting up an engine test, it is often beneficial to conduct simulation before testing to verify whether the test objectives can be met. Testing can confirm engine system design specifications or identify deficiencies. Many factors can render inaccurate system modeling results: model quality, assumptions, incorrect input data of the component or subsystem, etc. Simulation model tuning and comparison with experimental data are always necessary. Only after a successful model validation can the model be used with confidence to produce system design results. The analysis and testing functions should be partners reinforcing each other. The system design engineer needs to actively participate in the planning of testing to help define the objectives and procedures of the testing in order to discover any problems as early as possible. Test planning and analysis is a primary responsibility for system engineers.

Diesel engine system design is a new integrated function. The systems engineering theory (Armstrong, 2002) believes that the system engineers must release the system from the design departments. Because the majority efforts of diesel engine 'system design' are placed on producing system performance (or functional) design specifications, it is natural to consider the simulation analysis work conducted by a 'system analysis' engineer as the diesel engine 'system design' work instead of 'system analysis' work. In other words, most system engineers are design engineers, and they are granted design authority for the entire system. They use advanced simulation software as their design tools. They are not just the simulation or analysis engineers in a supporting role. As the importance of diesel engine system design emerges in product development, the job function of engine performance analysis assumes a dominant role in the engine design process. The system testing engineers run test cells to validate the whole engine system. Their work complements the work of system design engineers.

Advanced analytical engine system design process

A work flow process of advanced analytical engine design and development from concept to production is illustrated in Fig. 1.27. The input and output of the functional areas are interactive and affect each other, as shown by the double-headed arrows in the figure. The program management located in the middle controls the specified program schedule and budget. The efficiency of the engine development process depends on how clearly technical specifications or design/development changes are organized and cascaded by the system team to the component teams within a given organization. Although the technical communication channels between different functional areas are complex, overall there is an optimum 'top–down' process as shown by the thicker block arrows. In the old era when technical specifications were generated with crude hand calculations, precise design and optimization were



1.27 Simultaneous engineering processes.

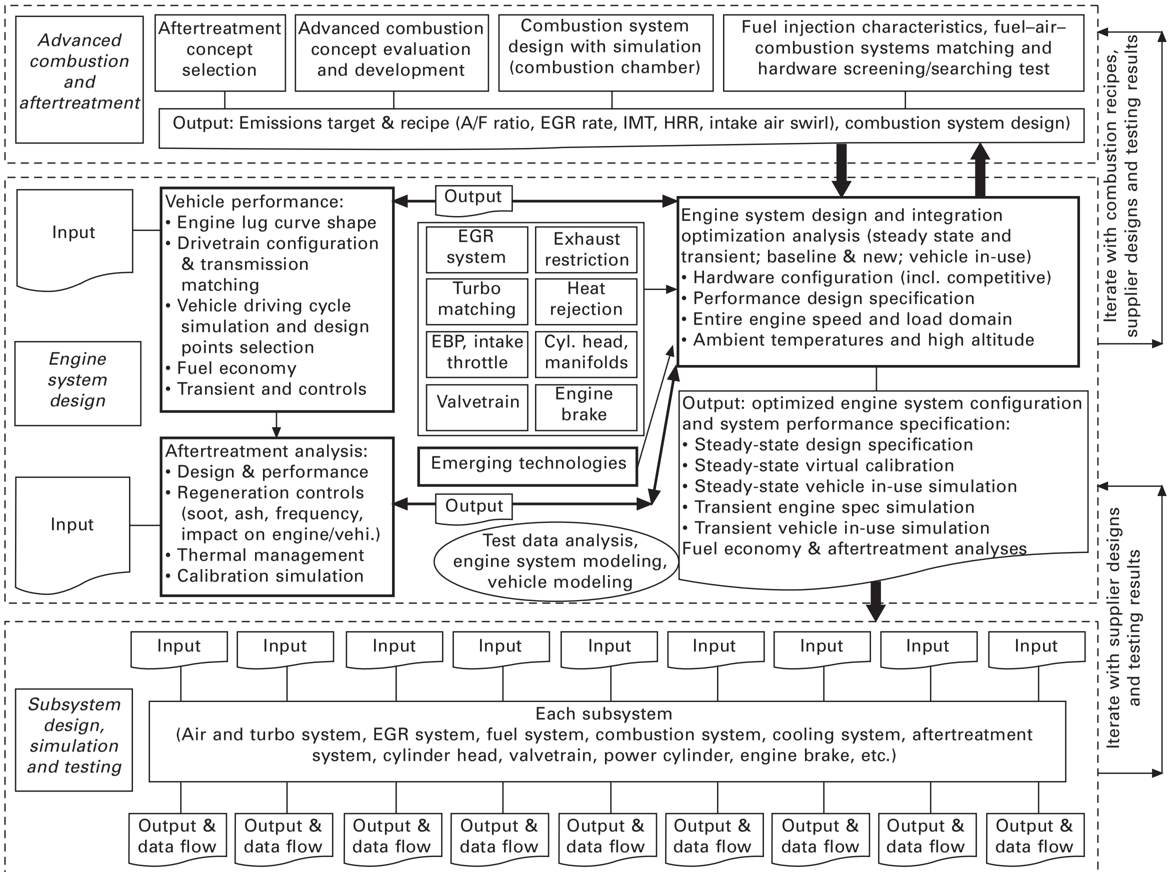
impossible. Today, with the effective use of computer simulations of engine thermodynamic cycle performance, it is feasible to accurately predict air system performance and subsystem interactions. Such a function of analytical system integration needs to be placed at the very top of the development chain. A common objective in engine development is to continue improving the upstream analysis capability and minimizing the costs in the prototype testing stages, as pointed out by Hoag (2006) in his detailed description of the engine development path. Figure 1.28 shows a more detailed process of analytical engine system design. Figure 1.29 illustrates the technical scope and examples of diesel engine system design.

Figure 1.30 illustrates the process of automotive powertrain definition, which links the engine system design to its upper level, the powertrain system design. Planning is the first step in a typical engine program. The vehicle-level requirements are cascaded down to generate the design targets at the engine level. After the engine system design specifications are determined, the design targets are then broken down further to the individual component level. A wide range of criteria must be considered, including performance (acceleration, fuel economy), emissions, cost, weight, packaging, and reliability. Often there are trade-offs between different criteria, for example, between 0–60 mph acceleration and fuel economy at different engine displacement. A larger swept volume (displacement) usually gives better naturally aspirated breathing capability so that the vehicle acceleration is faster. But during the driving cycle the larger engine runs more frequently at a lower BMEP level, hence the fuel consumption becomes worse. In this 'top-down' process, changing program targets or system design specifications will disrupt the development process and require extra modifications outside the agreed scope of modification freedom. Therefore, it is essential to define the program targets carefully by foreseeing future requirements and produce engine system design specifications accurately so that the extra modifications can be avoided as much as possible. Figure 1.31 shows an example of a powertrain design decision tree.

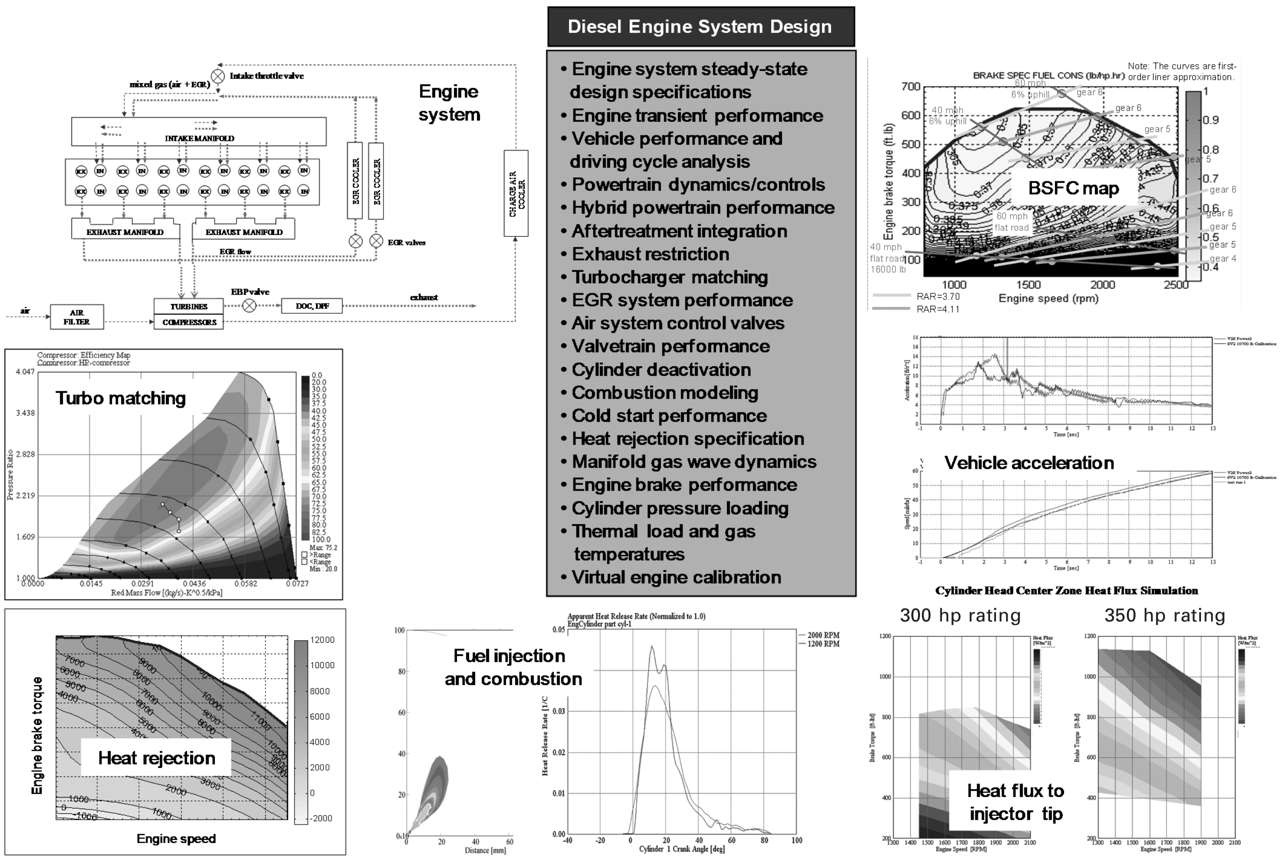
1.7 Engine system design specifications

1.7.1 Overview of engine design specifications

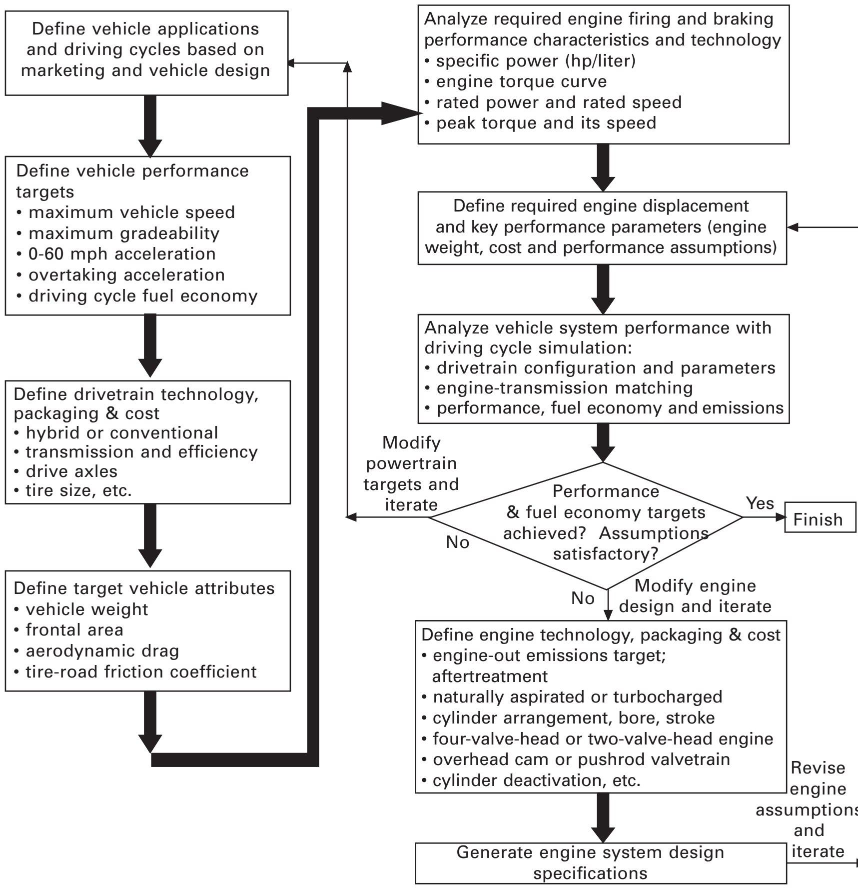
Design documents include schematic diagrams, functional block diagrams, computer simulation results, reports, drawings or graphics, bill of materials, traceability matrices, engineering specifications, material specifications, etc. Engineering design specification refers to a description of the subject – what it does, how it works, and how it is built. Engine design specifications include a complete set of information about how the engine is designed, operated, maintained, and repaired. Technical publications include operations manual, maintenance manual, service manual, diagnostics manual, etc.



1.28 Analytical diesel engine system design process.



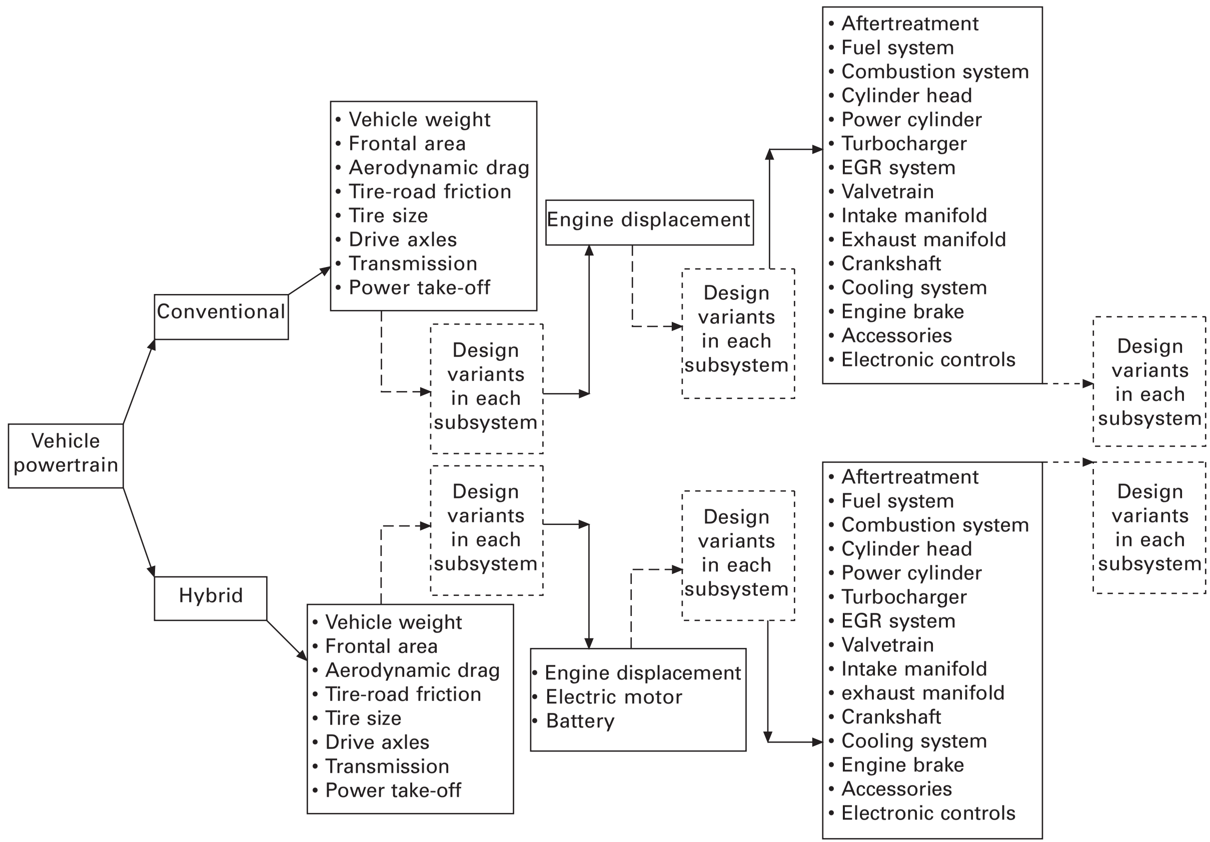
1.29 Illustration of diesel engine system design.



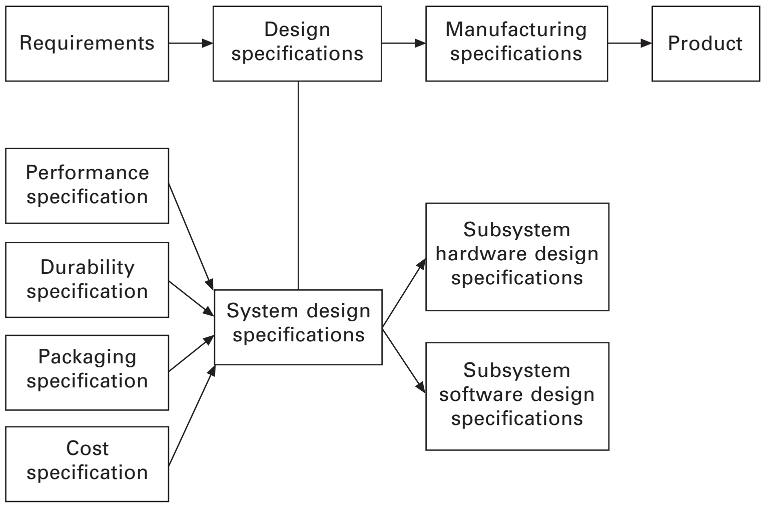
1.30 Powertrain system definition process.

System functional specification refers to a statement describing completely and concisely all the functions of the system to fulfill its operational requirements. The specification, as a definition of the system, should include mission, concept of operation, configuration, system interfaces, functions, system targets, requirements, hardware/software performance characteristics, and a history of document revision and methodology used. Design specifications ensure technical disciplines in the processes and coordination between different functional areas. The system design specifications cascade from the system level down to the subsystem and component levels, and they evolve during the course of engine development. Engine design process is a process to generate, implement and validate the design specifications continuously and iteratively (Fig. 1.32).

The system specifications include identification and description of all functions to be provided, along with the associated quantitative requirements



1.31 Powertrain design decision tree.



1.32 Design specifications.

to be met by each subsystem (Kossiakoff and Sweet, 2003). Another classical definition of general system specifications was given by Armstrong (2002):

> System specifications state the technical and functional requirements for a system, allocates requirements to functional areas, documents design constraints, and defines the interfaces between and among the functional areas and other systems. This specification will also identify necessary performance requirements and test provisions necessary to ensure that all requirements are achieved. The requirements for use of any existing equipment will also be identified.

The fact is that there is no commonality or consistency with regard to deliverables or standard design documents (including system specifications) for engine products within the industry or even within the same organization. This raises the demand for a common process with identified metrics or deliverables.

Diesel engine development programs usually start with a document of functional objectives that describes the system requirements. The functional objectives need to be maintained up-to-date and followed by all the parties throughout the program. The functional objectives include the basic definitions of the engine architecture (e.g., displacement, Vee or inline configuration, valvetrain type), performance targets (e.g., rated power, fuel economy, vehicle acceleration, emissions, noise), overall packaging geometry and weight, durability target (e.g., B10 life, peak cylinder pressure) and overall cost target of the engine.

The necessity for the system design specifications can be understood by the fact that different groups working in various phases of the program need to be coordinated by a detailed and clearly defined design document as the design target. In diesel engine system design, the functional objectives are translated into more detailed measurable specifications for design implementation and testing validation. For example, the functional requirements of engine power, fuel economy and emissions can be translated to a set of parameters of engine gas flow rate, pressure, temperature and heat rejection at different engine speed and load. Those parameters will be used by each subsystem for hardware sizing. The requirements of transient emissions and vehicle acceleration can be translated to the required functions of engine controllers and control software. The requirements of hybrid powertrain fuel economy and emissions can be translated to functional block diagrams of supervisory control strategies. The 'translation' is conducted by simulation analysis or testing.

A design specification usually consists of a nominal target and an allowable tolerance range (i.e., an upper limit and a lower limit). They ensure control factors are designed properly with respect to their means and standard deviations at the presence of noise factors. One example is the EGR rate at a given engine speed and load mode. A nominal target of EGR rate needs to be achieved in order to control NO_x emissions. The EGR rate can vary due to the noises in EGR valve opening and the pressure differential across the EGR circuit. But the variation range needs to be controlled within a certain tolerance range. Diesel engine system design specifications can be classified into four areas: performance, packaging, durability, and cost. They are introduced in the following sections.

1.7.2 System performance specifications

Introduction of system specifications for hardware and software

The design specification for diesel engine system performance is expressed in terms of both performance parameters and hardware (or calibration) parameters. The examples of performance parameters include the following: engine speed, power, fuel flow rate, air flow rate, BSFC, air–fuel ratio, EGR rate, intake manifold pressure and temperature, exhaust manifold pressure and temperature, engine delta P, volumetric efficiency, intake and exhaust restrictions, and heat rejections. The hardware or calibration parameters are used to achieve the performance. Examples of key hardware parameters at the engine system level include engine displacement, compression ratio, engine valve size and cam timing (affecting volumetric efficiency), turbine area (affecting air–fuel ratio and EGR rate), EGR cooler flow restriction and effectiveness (affecting intake manifold gas temperature), aftertreatment pressure drop (affecting exhaust restriction), etc. The calibration parameters refer to the adjustable ones with electronic controls, and examples include

VGT vane opening, EGR valve opening, fuel injection timing, etc. Examples of engine concept layout analysis have been provided by Mikulec *et al.* (1998) and Delprete *et al.* (2009).

Output of engine system design

The engine system design analysis generates three types of output. They are:

1. performance (or attribute) sensitivity data and optimization results
2. system performance (or attribute) design specifications
3. root cause analysis of particular problems.

Specifically, the system specification refers to a predicted list of all critical steady-state and transient engine performance and emissions parameters for a given concept configuration in the entire engine speed and load domain, and it is especially required at critical modes such as rated power, peak torque and driving part load at various ambient temperatures and altitudes. The specification also defines the data for turbocharging, EGR circuit design, engine heat rejection and electronic controls to be used in each subsystem design by suppliers and customers. The specification needs to cover both target (for on-target nominal design) and limits/range (for variability/reliability design).

The performance sensitivity and optimization refer to any steady-state and transient simulation data of parameter sweeping or design-of-experiments (DoE) optimization for comparing configuration concepts or justifying design specifications of hardware sizing. For example, the sensitivity study may be conducted for the maximum achievable rated power, or the optimum turbine nozzle areas in two-stage turbocharging, or transient vehicle acceleration simulation as a function of time with different vehicle weights. The specification should develop from optimization.

The engine system specification needs to cover the following five aspects:

1. Steady-state performance design specification.
2. Steady-state virtual calibration.
3. Steady-state vehicle in-use simulation (e.g., variations in charge air cooling, exhaust restriction, radiator performance and underhood thermal conditions compared to engine test cell conditions).
4. Transient specification simulation.
5. Transient vehicle in-use simulation.

The above-mentioned 'root cause' analysis refers to a simulation on any particular issue or failure. For example, insufficient EGR flow at engine peak torque is caused by inadequate engine delta P due to an excessively large turbine nozzle area. Another example is an excessively high exhaust manifold gas temperature caused by cooler failure or low air–fuel ratio. Generally, the need of root cause analysis should be minimized as much as possible through successful up-front specification design with system optimization. The system performance specification is usually the most important of the four system specifications in engine design (performance, durability, packaging, and cost).

During the process of generating engine system design specifications, often off-the-shelf solutions or design need to be considered in order to simplify the product design. The off-the-shelf design usually has been fully tested on a specific engine. Its cost and manufacturing method are also known. Engine system design needs to verify the selected off-the-shelf solution prior to testing validation.

1.7.3 System durability

The durability design specification is expressed in terms of both 'stress' and 'strength'. The stress refers to any general loading, usually coming from the performance specification (e.g., peak cylinder pressure, cylinder heat flux, exhaust manifold temperature, compressor outlet temperature). The strength refers to the desirable structural design parameters representing the capability of the system or component. The strength is also used by the performance area as a design constraint for iteratively refining the performance specification. Examples of preliminary engine structural calculations were provided by Makartchouk (2002) and Delprete *et al*. (2009).

1.7.4 System packaging

System packaging addresses the issues with weight, size, shape, component location, and clearance between the components. Good engine packaging requires low weight, compactness, suitable shape fit in the engine compartment, reasonable relative locations and clearances between the components without functional or geometrical interferences. Engine packaging design is handled mainly by three-dimensional solid modeling.

Engine weight can be estimated by using the volumes of the components. The representative dimensions of major components can be classified into two categories: (1) the dimensions determined by the basic design parameters of the engine such as cylinder bore diameter and engine stroke; and (2) the dimensions determined by the durability requirements that are related to engine operating speed and cylinder pressure loading. Weight, size, and cost are especially important for NVH design when seeking an acoustically-effective solution to package. Schuchardt *et al*. (1993) introduced the concepts of packaging quantity and quality and the relationship between packaging and performance in their study on engine intake noise control.

Engine and component size and the overall shape are strongly affected by the engine configuration (i.e., number of cylinders, inline, Vee, or opposed arrangement), bore, stroke, connecting rod length, valvetrain type (pushrod or overhead cam), turbocharger type (single-stage, two-stage, or twin-parallel), EGR cooler size, and FEAD. Component size is also related to performance. This is especially true for diesel aftertreatment devices. The device size is usually limited by the available spacing of the under-floor for a given vehicle. A smaller aftertreatment component has higher space velocity and may have lower operating efficiency. For example, there is a trade-off between LNT size and fuel consumption penalty. A smaller size gives higher fuel consumption penalty.

The design of relative locations and clearances between the components deserves great attention. The performance and durability characteristics of the engine components are often subject to the harsh thermo-mechanical boundary conditions. They are affected by the heat transfer and vibration of the neighboring components. A primary focus of system packaging work is to optimize the relative positions of all the subsystems or components in order to minimize the negative impact of all thermo-mechanical effects as a whole.

Lastly, it should be noted that in modern diesel engine system design, design for manufacturability and design for serviceability are two important trends in packaging design. They represent the processes to optimize the relationships among design function, manufacturability, ease of assembly, and ease of maintenance.

1.7.5 System cost

Engine system cost can be estimated by using bill of materials and operating cost. The attributes of performance, durability and packaging all directly affect cost. The engine system cost specification plans the capital cost and operating cost of the engine technologies adopted. It also coordinates between the subsystems on the maximum allowable cost for each subsystem in order to control the total engine cost.

1.8 Work processes and organization of diesel engine system design

1.8.1 Characteristics and principles of diesel engine system design

Academic background and characteristics of diesel engine system design

A diesel engine is a mechanical system. According to the definition of ‘mechanical system’ given by the National Science Foundation (Panel Steering

Committee, 1984), a mechanical system can be defined as an interconnection of mechanical and/or electromechanical components, coordinated and controlled by computational and informational networks (and often humans), which accomplishes dynamic tasks involving mechanical forces and motions and energy flows. The field of mechanical systems can be further classified into the following four major academic disciplines: design methodology and interactive graphics; dynamic systems and control; machine dynamics; and tribology. The discipline of design methodology deals with computer-aided engineering, optimization, and other generic techniques. The discipline of dynamics systems and control deals with the dynamic behavior and control of the interacting subsystems or components, including their energy flows, motions, and forces. Machine dynamics involves kinematics, dynamics, solid mechanics, acoustics, and finite element methods, and are generally related to reliability/durability. Tribology is the science and technology of lubrication, friction, and wear.

Diesel engine system design is a multi-disciplinary application field which possesses the features of the above academic disciplines. First of all, it requires an optimization approach supported by systems engineering and robust engineering. Secondly, it requires an approach of system dynamics to handle both transient (dynamic) performance/control and steady-state (static) performance. Thirdly, it requires dynamic motion analysis for the critical systems such as vehicle longitudinal dynamics, piston-assembly dynamics and valvetrain dynamics. Lastly, it is related to friction analysis of the engine components since friction directly affects engine efficiency.

A diesel engine system is also a combination of hardware and software control. It requires proper hardware sizing and matching to achieve target performance, and it also requires electronic software control to realize the performance during fast-changing and highly-nonlinear transient events. A correct system design can be achieved only when high-quality analytical tools or simulation models are available.

Moreover, a diesel engine is a thermodynamic energy system which produces power by a working fluid medium. The system performance design specifications address mainly the gas pressures, temperatures, and flow rates inside the engine. Advanced energy system analysis and design require using both the thermodynamic first law and second law. The first law provides the energy distribution inside the system, while the second law reveals the irreversibility, availability and the maximum potential of each component and process of the system.

Diesel engine system design and EPSI

Diesel engine system design includes four major branches: performance, durability, packaging, and cost. Performance is the leading one and carries the majority of the work in the entire system design. The complete name of the performance branch in diesel engine system design is engine performance and system integration (EPSI). The function of EPSI was illustrated earlier in Fig. 1.27. EPSI analyzes and integrates the performance of various engine subsystems or technologies, and properly matches them with optimization approaches. It conducts large-scale sophisticated computation in the areas related to thermal conditions, fluids, dynamics, and controls to derive precise system design solutions by using analytical tools such as engine cycle simulation software or analytical models and advanced data processing techniques.

System design is very important for the integration of simultaneous engineering processes for diesel engine product design, ranging from high-level product strategy planning to detailed production design. The missions of diesel engine system design include:

- providing engine system design and analysis;
- developing system analysis methodologies and simulation techniques; and
- developing core engine technologies with the viewpoint of system integration.

Engine system design covers a wide range of technical specialties including vehicle–engine–aftertreatment integration, thermodynamic cycle performance, air system design and turbo matching, powertrain dynamics and electronic controls.

1.8.2 Theoretical foundation and tools for diesel engine system design

There are several tools used in diesel engine system design. The first tool is engine cycle simulation. Complete cycle simulation for engine performance emerged in the early 1960s when digital computers became available. The computer codes were pioneered by Benson and Woods (1960) and Borman (1964). The initial use of engine cycle simulation was simple and restricted mainly to research groups. As a result, engine design decisions at that time were primarily driven by testing rather than computation. Since the 1980s, engine cycle simulation has gradually moved from the research groups into the production design and development process to support design (Morel and LaPointe, 1994). Today's engine development process requires optimized design with much shorter development time than ever for much more sophisticated engines. The analytical design process inevitably demands engine cycle simulation to become a part of the standard design tools used for production engine system design.

The foundation of engine system design analysis is built on thermodynamic cycle simulation. Engine cycle performance models are either zero-dimensional (i.e., spatially homogeneous, based on ordinary differential equations) or one-dimensional (based on partial differential equations of manifold wave dynamics). The input of the model includes engine geometry, subsystem characteristics and engine calibration parameters such as fuel injection timing and EGR valve opening. There are two types of output: the crank-angle based instantaneous values (e.g., gas pressure, temperature, and flow rate); and the cycle-average macro system performance parameters (e.g., engine torque, air–fuel ratio, and coolant heat rejection). Both steady-state and transient performance can be computed at various ambient conditions. There are several commercial software packages of cycle simulation, such as Gamma Technologies' GT-POWER (Morel *et al.*, 1999), Ricardo's WAVE and AVL's BOOST. Figure 1.33 shows an example of the engine cycle simulation model. The key issues in the modeling for engine system design are as follows:

- intake restriction (characterized by pressure drop vs. intake flow rate)
- exhaust restriction (characterized by pressure drop vs. exhaust flow rate)
- engine intake and exhaust valves (characterized by instantaneous effective valve flow area)
- intake and exhaust manifolds (characterized by volume, heat transfer, and pressure drop frictional losses)
- cylinder (characterized by volumetric efficiency, mechanical friction, and base engine heat rejection)
- coolers (characterized by flow restriction and thermal effectiveness)
- EGR valve and intake throttle valve (characterized by flow restriction through an orifice)
- turbine (characterized by effective area and efficiency)
- compressor (characterized by flow range and efficiency).

More discussions on engine cycle simulation are provided in Chapter 4.

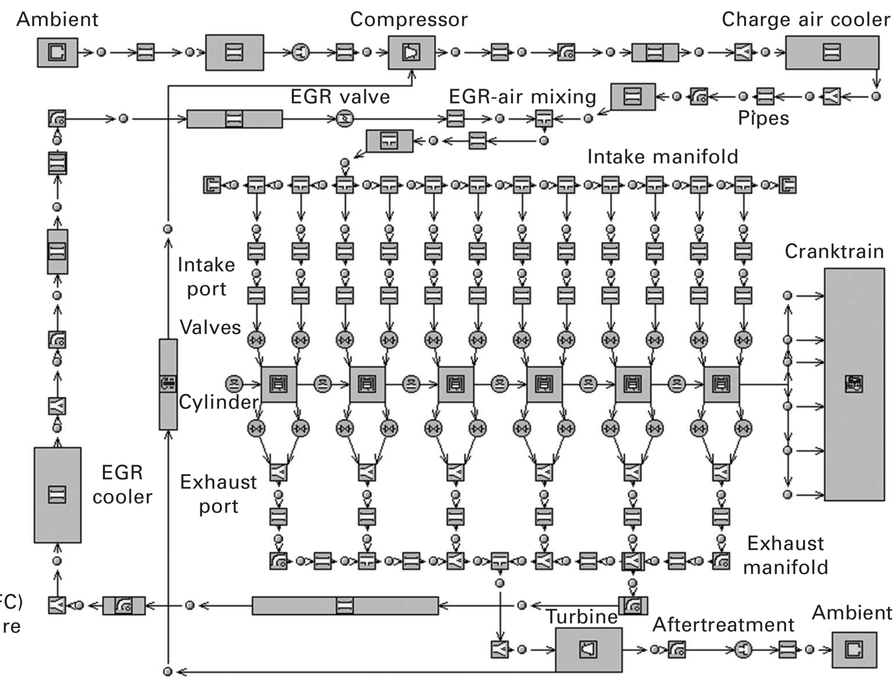
The second tool in engine system design is vehicle and powertrain modeling. The model can be used to calculate the forces and torques in an engine–drivetrain system and the transient motion of the vehicle. The main analysis objectives include engine–transmission matching, hybrid powertrain supervisory control, vehicle transient acceleration performance and driving cycle fuel economy. The model accounts for road grade, rolling resistance, aerodynamic drag, brake, powertrain inertia, clutch, torque converter and transmission characteristics, drive axles, driver behavior, powertrain controls, and engine performance characteristics (e.g., map-based mean-value model or high-fidelity crank-angle-resolution model). The model usually can be run in two different methods: forward (dynamic) or backward (kinematic). In the forward method, the equations of motion for the drivetrain components and the vehicle are numerically integrated in time to obtain the transient speeds

Cycle simulation model

Theory of the model
- Engine cycle thermodynamics
- Mass and energy conservation
- In-cylinder gas property changes
- Cylinder and pipe heat transfer
- Turbocharger principle
- Combustion and heat release rate
- 1-dimensional gas wave dynamics
- Acoustics of intake and exhaust
- Piston-assembly dynamics
- Controls linked to Matlab/Simulink

Input of the model
- Engine geometry
- Valve lift profile and port flow C_f
- Turbocharger maps and efficiency
- Base engine heat rejection percentage
- Engine mechanical friction
- EGR cooler and CAC characteristics
- Fuel injection or combustion timing

Output of the model
- Cylinder average parameters (e.g., BSFC)
- Instantaneous gas pressure, temperature and mass flow rate inside the engine



1.33 Engine cycle simulation model (GT-POWER).

and torques in the system. One example is to predict 0–60 mph acceleration with gear shifting events. In the backward method, the vehicle speed driving profile is the input. The required speeds and torques of the powertrain components to drive the vehicle are calculated based on the vehicle force balance equation. One example is to calculate engine operating points for a given driving cycle and the corresponding fuel consumption and emissions. Typical vehicle simulation software packages include Gamma Technologies' GT-DRIVE (Morel *et al.*, 1999) and AVL's Cruise. Other powertrain dynamics models based on Simulink programming were introduced by Moskwa *et al.* (1999) and Assanis *et al.* (2000). More discussions on engine–vehicle matching are provided in Chapter 5.

The third tool is diesel aftertreatment modeling. The primary objectives of aftertreatment analysis in diesel engine system design include the following:

- simulating tailpipe emissions;
- configuration architecture selection;
- component sizing;
- precious metal loading;
- controlling DPF regeneration at all vehicle operating conditions for safe and fuel efficient operation; and
- optimizing the interaction with other engine subsystems.

Over the past several years, considerable progress has been made to model individual aftertreatment components such as DOC, DPF, LNT, and SCR. For system design a more useful tool needs to be a system-level aftertreatment model that is integrated with the engine model. Such integrated diesel aftertreatment modeling was developed in Gamma Technologies' GT-POWER by Tang *et al.* (2007) and Wahiduzzaman *et al.* (2007). Other integrated models based on Simulink were introduced by Rutland *et al.* (2007) and He (2007). The DOC model simulates the exhaust gas temperature raise due to hydrocarbon oxidation caused by fuel dosing for DPF regeneration. The DOC model includes chemical reaction kinetics for the oxidation of CO, HC, and NO to CO_2, H_2O, and NO_2. The DPF model can be either a zero-dimensional (also called lumped-parameter) model or a one-dimensional model. The one-dimensional model considers the axial changes of exhaust gas dynamics and particulate matter oxidation. The input to the model is exhaust gas flow from the turbine with heat losses through the tailpipe. The DPF model simulates the deep-bed filtration, soot-cake filtration, particulate matter deposition inside the filter and the resulting pressure drop. The DPF regeneration process can be simulated by the model for thermal and catalytic oxidations of the trapped soot. More discussions on engine–aftertreatment integration are provided in Chapter 8.

The fourth tool is optimization and probabilistic analyses. The DoE method has been widely used to generate a matrix containing partial-factorial combinations of input factors for efficient data gathering. Each case (or run) in the DoE matrix can be simulated with engine cycle simulation software. The output data are called responses. Mathematical models (polynomial emulators) are built with response surface methodology (RSM (Myers and Montgomery, 2002)) to link each response parameter with the factors. Then, optimization is conducted by using the emulators to search the global optima under certain constraints by minimizing or maximizing a functional target (e.g., minimizing BSFC). Probabilistic analysis can handle the system design task of variability- or reliability-based optimization (Baker and Brunson, 2000). The probability distributions of random input factors are generated with the Monte Carlo simulation by producing thousands of cases. Each case is run using a simulation model to produce response data. Then the thousands of response data are processed with statistical software to obtain the probability distribution of the response. Variability- or reliability-based design constraints can be applied to conduct optimization in order to determine the optimum mean and tolerance range of the control factors. The optimization and probabilistic analysis methods will be introduced in detail in Chapter 3.

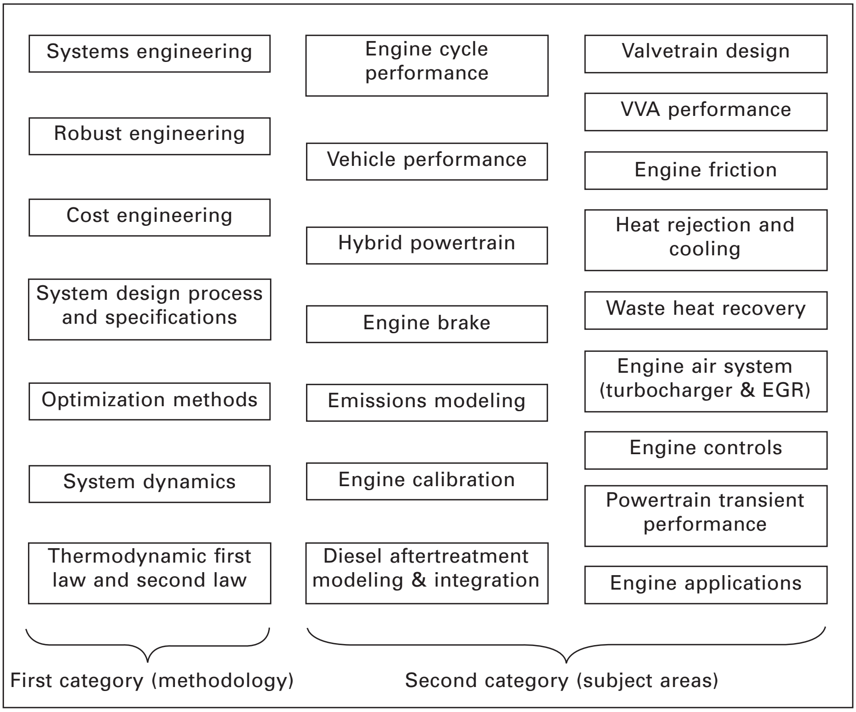
The trend in the vehicle industry is toward integrated engine–powertrain–vehicle design and electronic controls. It is desirable to have an integrated simulation environment containing all the engine system design tools for seamless connections between the systems and easy data sharing. Gamma Technologies has made impressive progress to integrate such tools within the GT-SUITE environment (Ciesla *et al.*, 2000; Silvestri *et al.*, 2000; Morel *et al.*, 2003).

1.8.3 Technical areas of engine performance and system integration

For the areas in diesel engine system design related to performance, two categories of research are identified (Fig. 1.34). The first category links the system design to generic design methodologies, for example, how to apply the principle of reliability engineering in diesel engine system design, or how to formulate system dynamics models effectively to enhance the quality of system design. The second category of research addresses the analysis and modeling of each key subject area, and their interactions. Table 1.6 shows the required skill sets in engine system design.

The functional areas in engine performance and system integration include the following:

1. *Technical field 1: Engine system thermodynamic cycle performance*
   - Basic engine operation (power rating, torque curve, etc.)
   - Basic engine design configuration (cylinder bore, stroke, etc.)



1.34 Technical areas in diesel engine system design.

- Valvetrain performance (valve timing, cam, valvetrain dynamics)
- Variable valve timing
- Cylinder deactivation
- Exhaust restriction
- Combustion, fuel injection, and emissions modeling
- Cold start performance
- Turbocharging
- EGR system performance
- Engine heat rejection
- Cooling circuit performance
- Intake and exhaust manifold gas wave dynamics
- Engine friction
- Waste heat recovery
- Fuel economy improvement roadmap
- Engine brake performance
- Transient engine performance (load response, turbo lag, warm-up, emissions cycles, etc.)

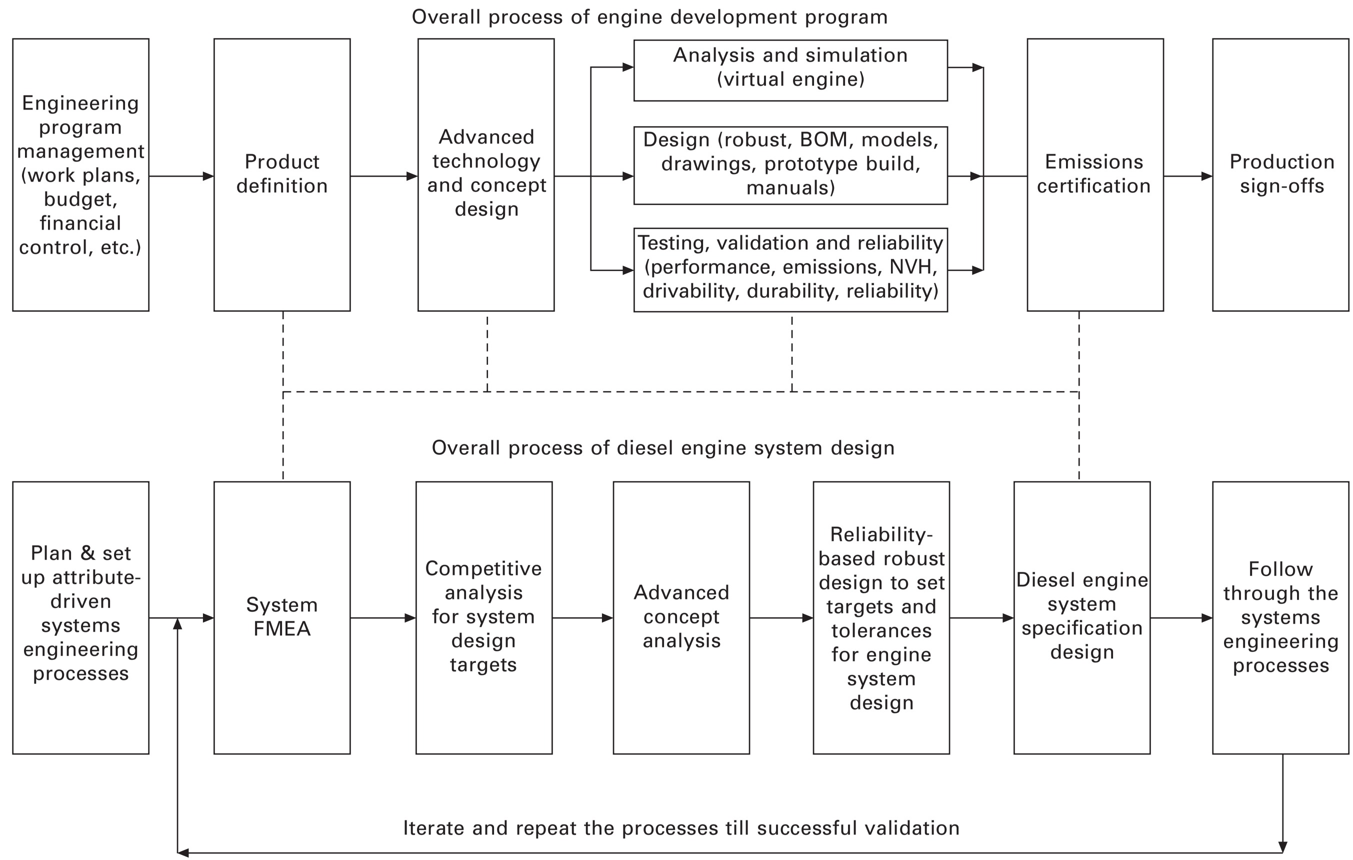
2. *Technical field 2: Vehicle and powertrain performance*
    - Vehicle acceleration, launch, gradeability, and drivability

Table 1.6 Skill sets for each attribute in diesel engine system design

Index	Performance	Durability	Packaging	Cost
1	Thermodynamics			
2	Combustion			
3	Fluid mechanics			
4	Heat transfer	Heat transfer	Heat transfer	
5	Tribology	Tribology		
6	Vibration	Vibration	Vibration	
7	System dynamics	System dynamics		
8	Controls			
9	Acoustics	Acoustics	Acoustics	
10	Probability and statistics	Probability and statistics	Probability and statistics	Probability and statistics
11		Materials	Materials	
12		Solid mechanics		
13				Cost engineering
14		Finite element analysis		
15		3-D solid modeling	3-D solid modeling	
16	Optimization methods	Optimization methods	Optimization methods	Optimization methods
17	Engine system design theory	Engine system design theory	Engine system design theory	Engine system design theory
18	Numerical simulation analysis	Numerical simulation analysis		Numerical simulation analysis
19	Engine design	Engine design	Engine design	Engine design
20	Engine testing	Engine testing		
21	Engine applications	Engine applications	Engine applications	Engine applications

- Drivetrain configuration and drive axle ratio
- Engine–transmission matching
- Vehicle driving cycle fuel economy
- Transient powertrain dynamics
- Hybrid powertrain performance and supervisory controls

3. *Technical field 3: Aftertreatment system integration*
   - Aftertreatment performance calibration and matching
   - Aftertreatment emissions modeling
   - Aftertreatment controls
4. *Technical field 4: Model-based engine calibration and optimization*
5. *Technical field 5: Engine controls modeling*
   - Thermal/fluids performance modeling for model-based controls

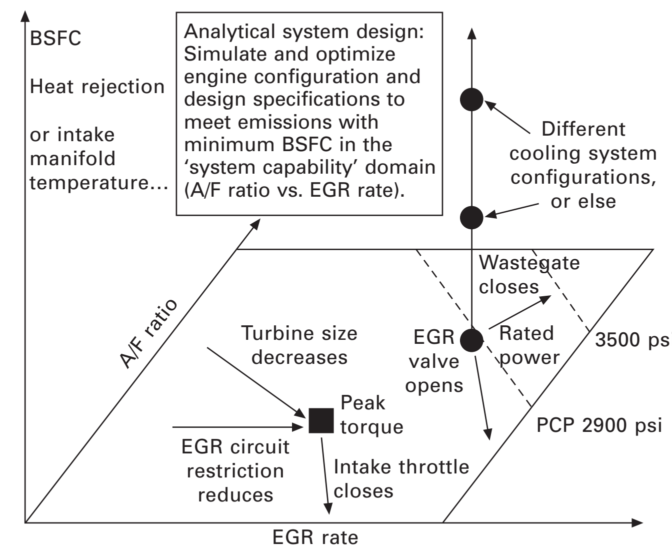


1.35 Overall process of diesel engine system design.

- Performance models for virtual sensors
- Evaluation of engine control strategies and algorithms.

1.8.4 Work processes of diesel engine system design

Engine system performance analysis is characterized by large-scale complex data handling across a wide range of technical specialties. The following are five key factors for successfully implementing system design analysis in the engine product development process:



Step 1: Understand the current engine's performance
- List all design constraints
- Analyze performance test data and subsystem characteristics
- Analyze emissions test data
- Calculate heat release rate
- Build well-tuned simulation model
- Identify transient problems by analyzing test data

Step 2: Check sensitivity and define new engine system specification
- Define engine performance functional targets
- Explore parametric performance sensitivity by simulation
- Generate system design specification and obtain consensus
- Issue specification to subsystem teams and suppliers to start design
- Simulate to check proposed design hardware and controls
- Analyze test data of engine hardware and calibration
- Revise functional targets and iterate system specification analysis

1.36 Tasks in diesel engine system design.

1. Effectively communicate technical data across different functional areas to ensure timely information sharing.
2. Establish analysis task categories and corresponding standard analysis procedures to ensure quality and consistency.
3. Develop advanced data processing templates to maximize efficiency.
4. Provide effective training for system design engineers.
5. Conduct organized research in all system design areas to develop advanced analysis methodologies.

The essence of diesel engine system design is to reveal the fundamental parametric dependency among various design and attribute parameters in order to derive an optimized system solution. Figures 1.35–1.36 summarize the processes and tasks involved in executing system specification design.

1.9 References and bibliography

Advanced Product Quality Planning (APQP) and Control Plan (1995), reference manual, Chrysler Corporation, Ford Motor Company, and General Motors Corporation.

Arcidiacono G and Campatelli G (2004), 'Reliability improvement of a diesel engine using the FMETA approach', *Quality and Reliability Engineering International*, 20, 143–154.

Armstrong L (1996), 'Systems engineering as a structured design process', SAE paper 962178.

Armstrong L R (2002), *Electronic System Integration*, 47th L. Ray Buckendale Lecture, SP-1727, SAE International, Warrendale, PA.

Asmus T (1999), 'A manufacturer's perspective on IC engine technology at century end', ASME paper No. 99-ICE-202, 1999 Fall Technical Conference, ICE-Vol. 33-1, pp. 1–9.

Assanis D, Filipi Z, Gravante S, Grohnke D, Gui X, Louca L, Rideout G, Stein J and Wang Y (2000), 'Validation and use of SIMULINK integrated, high fidelity, engine-in-vehicle simulation of the International class VI truck', SAE paper 2000-01-0288.

Austin T E (2007), 'Why have a systems engineering (SE) capability for automotive product development? – questions and answers', SAE paper 2007-01-0782.

Baker J and Brunson D (2000), 'Statistical modelling of engine systems', in Edwards S P, Grove D M and Wynn H P (eds), *Statistics for Engine Optimization*, Professional Engineering Publishing, Suffolk, pp. 61–82.

Benson R S and Woods W A (1960), 'Wave action in the exhaust system of a supercharged two-stroke engine model', *International Journal of Mechanical Science*, 1, 253–281.

Bignonnet A and Thomas J J (2001), 'Fatigue assessment and reliability in automotive design', SAE paper 2001-01-4061.

Boesel J, Romeo S and Brotherton T (2003), 'Advanced fuels and technology forecasts for heavy-duty vehicles', SAE paper 2003-01-3404.

Boggs D, Belaire R C, Bartunek B, Dürnholz M and Ecker H J (1997), 'A small displacement DI diesel engine concept for high fuel economy vehicles', SAE paper 972680.

Borman G L (1964), *Mathematical Formulation of Internal Combustion Processes and Performance Including Comparison with Experiment*, Ph.D. dissertation, University of Wisconsin, Madison, WI.

Brady R N (1996), *Modern Diesel Technology*, Prentice-Hall, Englewood Cliffs, NJ.
Browning L H (1997), 'Technologies and costs for on-road heavy-duty engines meeting 2004 emissions standards', SAE paper 973256.
Burke A (2003), 'Saving petroleum with cost-effective hybrids', SAE paper 2003-01-3279.
Buuck B and Hampton K (1997), 'Engine trends and valvetrain systems for improved performance, fuel economy, and emissions', *Proceedings of the International Symposium on Valvetrain System Design and Materials*, April 14–15, 1997, Dearborn, MI.
Carey W R (1992), *Tools for Today's Engineer: Strategy for Achieving Engineering Excellence*, the 38th L. Ray Buckendale Lecture, SAE SP-913, SAE paper 920040, SAE International, Warrendale, Pennsylvania.
Catania A E, d'Ambrosio S, Guenna G M, Petrolo D, Pilo L and Strazzullo L (2007), 'New modeling for reliable evaluation of parameter variability effects on vehicle fuel consumption', SAE paper 2007-01-0328.
Chandrupatla T (2009), *Quality and Reliability in Engineering*, Cambridge University Press, New York.
Ciesla C, Keribar R and Morel T (2000), 'Engine/powertrain/vehicle modeling tool applicable to all stages of the design process', SAE paper 2000-01-0934.
Cummins Jr. C L (1993), *Diesel's Engine: Volume One, From Conception to 1918*, Carnot Press, Wilsonville, OR.
Daubercies F, Geslin C and Morin G (2009), 'Cylinder head reliability assessment: from numerical method development to validation based on high mileage field data', SAE paper 2009-01-0203.
Dave A and Hampson G J (2003), 'Robust engine design using engine simulations', SAE paper 2003-01-0371.
Delprete C, Pregno F and Rosso C (2009), 'Internal combustion engine design: a practical computational methodology', SAE paper 2009-01-0477.
Dodson B and Schwab H (2006), *Accelerated Testing: A Practitioner's Guide to Accelerated and Reliability Testing*, SAE International, Warrendale, PA.
Dopson C, Taitt D and Sandford M H (1995), 'Powertrain systems definition process', SAE paper 950811.
Dubensky R G (1993), 'Simultaneous engineering of powertrain subsystems – design aspects', SAE paper 932894.
Eberhardt J J (1999), 'Overview of the DOE heavy vehicle technologies R&D program', SAE paper 1999-01-2235.
Ecker H J, Schwaderlapp M and Gill D K (2000), 'Downsizing of diesel engines: 3-cylinder / 4-cylinder', SAE paper 2000-01-0990.
Escobar L A and Meeker W Q (2006), 'A review of accelerated test models', *Statistical Science*, 21, 552–577.
Evans J L, Vogt N, Thompson J R and Coit D (2005), 'Correlating field requirements to accelerated life testing for vehicle electronics', SAE paper 2005-01-1492.
Feng M and Chen Y (2004), 'Fixed emitting point method – an accelerated test program to determine fatigue limit of automotive components', SAE paper 2004-01-1646.
Fowlkes W Y and Creveling C M (1995), *Engineering Methods for Robust Product Design: Using Taguchi Methods in Technology and Product Development*, Addison-Wesley Publishing Company, Reading, MA.
Gale N F, Widener S K, Sui P C and Zhang H (1995), 'Analytical engine design methods: a review', SAE paper 950806.
Ginn J, Toback A, Hearne J, Marchese A J, Hesketh R P and Amundsen C (2004), 'Life

cycle and economic analysis of heavy duty diesel vehicle idling alternatives', SAE paper 2004-01-0637.

Grosser M (1978), *Diesel, the Man & the Engine*, Newton Abbot, UK.

Hazelrigg G A (1998), 'A framework for decision based engineering design', *Journal of Mechanical Design*, 120, 653–658.

He Y (2007), 'Development of an integrated diesel exhaust aftertreatment simulation tool with applications in aftertreatment system architecture design', SAE paper 2007-01-1138.

Heisler H (1995), *Advanced Engine Technology*, SAE International, Warrendale, PA.

Hikosaka N (1997), 'A view of the future of automotive diesel engines', SAE paper 972682.

Hoag K L (2006), *Vehicular Engine Design*, SAE International and Springer-Verlag, Vienna.

Humphreys K K (ed.) (2005), *Project and Cost Engineers' Handbook*, 4th edn, Marcel Dekker, New York.

Ireson W G, Coombs C F and Moss R Y (1996), *Handbook of Reliability Engineering and Management*, 2nd edn, McGraw-Hill, New York.

Jackson M R, Gunderson M L and Maanen K D V (1991), 'A systems engineering approach to subsystem design: automotive transmissions', SAE paper 911936.

Kanefsky P, Nelson V A and Ranger M (1999), *A Systems Engineering Approach to Engine Cooling Design*, the 44th L. Ray Buckendale Lecture, SP-1541, SAE paper 1999-01-3780, SAE International, Warrendale, PA.

Kececioglu D B (2003), 'Keynote address: robust engineering design-by-reliability', SAE paper 2003-01-0141.

Kestly M, Popielas F, Grafl D and Weiss A (2000), 'Accelerated testing of multi-layer steel cylinder head gaskets', SAE paper 2000-01-1188.

Killmann I G and Tholen P (1999), 'Aircooled engines', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 424–448.

Klyatis L M (2010), *Accelerated Reliability and Durability Testing Technology*, Wiley-Blackwell, Oxford.

Klyatis L M and Klyatis E (2006), *Accelerated Quality and Reliability Solutions*, Elsevier Science, Oxford.

Kokkolaras M, Mourelatos Z, Louca L, Filipi Z, Delagrammatikas G, Stefanopoulou A, Papalambros P and Assanis D (2005), 'Design under uncertainty and assessment of performance reliability of a dual-use medium truck with hydraulic-hybrid powertrain and fuel cell auxiliary power unit', SAE paper 2005-01-1396.

Kossiakoff A and Sweet W N (2003), *Systems Engineering: Principles and Practice*, John Wiley and Sons, Hoboken, NJ.

Kuehnel W, Weise S, Krueger U, Sekler H and Stephenson P (2005), 'The validation process – one key element of reliability management', SAE paper 2005-01-1778.

Kumar U D, Crocker J, Chitra T and Saranga H (2006), *Reliability and Six Sigma*, Springer Science + Business Media, New York.

Li Z, Georgiopoulos P, Papalambros P Y, Filipi Z, Wu G and Yang X (2004), 'Model based analysis of performance-cost tradeoffs for engine manifold surface finishing', SAE paper 2004-01-1561.

Majewski W A and Khair M K (2006), *Diesel Emissions and Their Control*, SAE International, Warrendale, PA.

Makartchouk A (2002), *Diesel Engine Engineering: Thermodynamics, Dynamics, Design, and Control*, Marcel Dekker, New York.

Matthews R, Hall M, Prozzi J, Machemehl R, Anthony J, Ullman T, Baker R and Lewis D (2005), 'The Texas diesel fuels project, part 4: fuel consumption, emissions, and cost-effectiveness of an ultra-low-sulfur diesel fuel compared to conventional diesel fuels', SAE paper 2005-01-1724.

Menne R J and Rechs M N (2002), *The System Integration Process for Accelerated Development*, SAE International, Warrendale, PA.

Merrion D F (1994), *Diesel Engine Design for the 1990s*, the 40th L. Ray Buckendale Lecture, SAE SP-1011, SAE paper 940130, SAE International, Warrendale, PA.

Merrion D and Weber K E (1999), 'Trucks and buses', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 553–585.

Mikulec A, Reams L, Chottiner J, Page R W and Lee S (1998), 'Cranktrain component conceptual design and weight optimization', SAE paper 980566.

Morel T and LaPointe L A (1994), 'Concurrent simulation and testing concept in engine development', SAE paper 940207.

Morel T, Keribar R, Silvestri J and Wahiduzzaman S (1999), 'Integrated engine/vehicle simulation and control', SAE paper 1999-01-0907.

Morel T, Keribar R and Leonard A (2003), 'Virtual engine/powertrain/vehicle simulation tool solves complex interacting system issues', SAE paper 2003-01-0372.

Morin G, Nicouleau-Bourles E, Simon F and Prince O (2005), 'Reliable diesel engine design based on a new numerical method', SAE paper 2005-01-1762.

Moskwa J J, Wang W, Chen S X and Munns S A (1999), 'Using dynamic modular diesel engine models to understand system interactions and performance', SAE paper 1999-01-0976.

Myers R H and Montgomery D C (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd edn, John Wiley & Sons, New York.

Nelson W B (2004), *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*, 2nd edn, John Wiley & Sons, Hoboken, NJ.

Niewczas A and Koszalka G (2002), 'Method of accelerated evaluation of automotive engine durability', SAE paper 2002-01-0337.

O'Connor P (2002), *Practical Reliability Engineering*, John Wiley & Sons, Chichester.

O'Day J and Bunch H M (1981), 'Safety and economic benefits of heavy truck retarder use', SAE paper 811260.

Ostwald P F and McLaren T S (2004), *Cost Analysis and Estimating for Engineering and Management*, Pearson Education, Upper Saddle River, NJ.

Panel Steering Committee for the Mechanical Engineering and Applied Mechanics Division of the National Science Foundation (1984), 'Research needs in mechanical systems – report of the select panel on research goals and priorities in mechanical systems', *Transactions of the ASME, Journal of Tribology*, 106, 2–13.

Park W R and Jackson D E (1984), *Cost Engineering Analysis: A Guide to Economic Evaluation of Engineering Projects*, 2nd edn, John Wiley & Sons, New York.

Pendola M, Morin G, Mohamed A and Ragot P (2003), 'Reliability based analysis of exhaust manifold lifetime simulation: an essential tool for risk management in industrial product development', SIA Congress, Poissy, France.

Prince O, Morin G and Jouzeau C (2005), 'Validation test optimization based on a statistical approach for diesel engine cylinder heads reliability', SAE paper 2005-01-1780.

Prozzi J, Machemehl R, Matthews R, Baker R, DeFries T H and Lewis D (2004), 'The

Texas diesel fuels project, part 3: cost-effectiveness analyses for an emulsified diesel fuel for highway construction equipment fleets', SAE paper 2004-01-0086.
Rahman S and Sun R (2003), 'Robust engineering of engine cooling system', SAE paper 2003-01-0149.
Rahman S, Kayupov M, Li J and Mourelatos Z P (2007), 'Balance between reliability and robustness in engine cooling system optimal design', SAE paper 2007-01-0594.
Rakopoulos C D and Giakoumis E G (2009), *Diesel Engine Transient Operation: Principles of Operation and Simulation Analysis*, Springer-Verlag London Limited, London.
Rausand M and Hoyland A (2004), *System Reliability Theory: Models, Statistical Methods, and Applications*, John Wiley & Sons, Hoboken, NJ.
Reinhart T E, Sampath A, Bagga K S and Leistensnider G W (2003), 'NVH variations in diesel engine populations', SAE paper 2003-01-1723.
Rutland C J, England S B, Foster D E and He Y (2007), 'Integrated engine, emissions, and exhaust aftertreatment system level models to simulate DPF regeneration', SAE paper 2007-01-3970.
SAE PT-87 (2003), *Truck Systems Design Handbook*, Volume II, SAE International, Warrendale, PA.
Schuchardt M E, Dear T A and Ingard K U (1993), 'An air induction system for a four-cylinder engine: design, noise control and engine efficiency', SAE paper 931317.
Seger J P, Hwang L K, Dendrick B, McIntire K, Geng H and Bartholomew B (2010), '2010 medium and heavy duty vehicle integration', SAE paper 2010-01-1956.
Silvestri J, Morel T, Lang O, Schernus C and Rauscher M (2000), 'Advanced engine/drivetrain/vehicle modeling techniques', *9th Aachener Kolloquium Fahrzeug–und Motorentechnik*.
Stamatis D H (2003), *Failure Mode and Effect Analysis: FMEA from Theory to Execution*, ASQ Quality Press, Milwaukee, WI.
Steinberg P and Goblau D (2004), 'Fuel consumption', in Basshuysen R and Schafer F (eds), *Internal Combustion Engine Handbook*, SAE International, Warrendale, PA, pp. 737–751.
Stroia B J, Currier N W, Li J, England R D, Bush J W and Hess H (2008), 'Critical performance and durability parameters of an integrated aftertreatment system used to meet 2007 Tier II emission standards', SAE paper 2008-01-0769.
Tang W, Wahiduzzaman S and Leonard A (2007), 'A lumped/1-D combined approach for modeling wall-flow diesel particulate filters – applicable to integrated engine/aftertreatment simulations', SAE paper 2007-01-3971.
Wahiduzzaman S, Tang W, Wenzel S and Leonard A (2007), 'Modeling of integrated aftertreatment systems', SAE paper 2007-01-4127.
Walsh M P (1983), 'The benefits and costs of light duty diesel particulate control', SAE paper 830179.
Walsh M P (1984), 'The costs and benefits of diesel particulate control II', SAE paper 840177.
Xin Q (2010), 'Heavy-duty diesel engine system design', in Zhao H (ed.), *Advanced Direct Injection Combustion Engine Technologies and Development, Volume 2: Diesel Engines*, Woodhead Publishing Limited, Cambridge, UK.
Yan J, Rogalla R and Kramer T (1993), 'Diesel combustion and transient emissions optimization using Taguchi methods', SAE paper 930600.
Zhan R, Eakle S, Spreen K, Li C G and Mao F F (2007), 'Validation method for diesel particulate filter durability', SAE paper 2007-01-4086.
Zhou J and Li D (2009), 'Reliability verification: plan, execution, and analysis', SAE paper 2009-01-0561.

2
Durability and reliability in diesel engine system design

Abstract: Consideration of durability and reliability is necessary at the earliest stage of system design. This chapter presents the theory and analysis methods of durability and reliability in diesel engine system design. It begins by describing engine durability issues, followed by an elaboration on the relationship between performance and durability through the discussions on system-level loading and durability design constraints. It then provides a systematic introduction on the fundamentals of thermo-mechanical failures and the applications on diesel engine cylinder head, exhaust manifold, valvetrain, piston, turbocharger and aftertreatment devices, followed by discussions on cylinder liner cavitation, engine wear, and EGR cooler durability. An integrated analysis approach on system durability–reliability is finally summarized.

Key words: durability, reliability, thermo-mechanical failures, damage, fatigue, wear, safety factor, probabilistic design, system durability–reliability optimization model.

2.1 Engine durability issues

Engine durability is very important since customers demand longer service intervals and engine lifetime. Increasing demands on higher engine power density, lower emissions, and longer lifetime impose great challenges on durability for modern diesel engines due to increased mechanical and thermal loading. Engine structural durability issues are important constraints in system design for performance. The durability issues can be mainly classified into the following categories based on failure modes:

- fracture (acute rupture and chronic cracking)
- buckling
- thermo-mechanical failures, mainly including ablation, excessive thermal deformation, fracture and fatigue at high temperatures, creep, oxidation, and corrosion
- cavitation in cylinder liners, fuel system, cooling system, etc.
- wear and lubricant oil degradation at all tribological contacts (piston assemblies, bearings, valvetrain, fuel injector, etc.)
- EGR cooler fouling, boiling, and corrosion
- fouling and deposits
- hydrogen embrittlement.

Note that the durability issues related to engine electronics such as overheating are not covered here. Among the above mechanisms, fracture, thermo-mechanical failures, wear, and cooler fouling are closely related to the design constraints or limits used in diesel engine system design, hence will be explained in detail in later sections. The rest are briefly introduced here for completeness. Buckling is a failure of elastic components in terms of instability subject to high compressive stresses at the two ends of a long and slender object such as a pushrod or a long connecting rod. The buckling stress is less than the ultimate compressive strength. Cavitation is a damaging phenomenon caused by the high energy explosion from the ruptured vapor bubbles which are originally formed by the undesirable dynamic pressures in the fluid system. Fouling and deposits may occur in situations such as injector coking, carbon soot deposits built up in the combustion chamber and at the valve seat, and deposits in the piston top land resulting in associated cylinder bore polishing. Hydrogen embrittlement refers to a situation where the metal material becomes more brittle when the hydrogen generated from water decomposition enters the metal, for example during the welding process. The material becomes less capable of resisting shock loading, deformation, and low cycle fatigue.

There are numerous failure modes for all the engine components. Most of the failures in modern diesel engines usually occur in the following components: fuel injection system, turbocharger, EGR system, aftertreatment, cylinder head, piston assembly, exhaust manifold, and valvetrain. Sometimes failures may also occur in electronic sensors, lubrication and cooling systems, etc. Examples of durability failures include: the cracks in the piston, turbocharger, cylinder head, exhaust manifold, diesel particulate filter, rocker arm fulcrum plate, and crankcase; valve head and spring breakage; piston ablation and scuffing; bore polishing on the cylinder liner; VGT vane sticking; thermal distortion of the EGR valve; turbine blade–casing interference due to high temperature creep; pushrod buckling; injector spool valve and coil failures; fuel injector coking (soot deposit formation); valve sticking and leaking caused by carbon deposits on the valve seat and valve stem; leakages at the gasket, water jacket, turbocharger, fuel line, and oil cooler; liner cavitation erosion; unusual wear in the piston ring and bearings; fuel dilution in engine oil; EGR cooler fouling due to carbon soot and unburnt hydrocarbons; EGR cooler film boiling; and corrosion in the intake manifold.

Engine durability issues can be either an acute catastrophic failure or a chronic mild failure. The engine components subject to normal wear need to be replaced after a scheduled service period. Engine durability life is usually defined by the engine overhaul point, the life-to-overhaul. When an engine has excessive wear, oil consumption or blow-by, the engine needs an overhaul. Scuffing, scratches and unusual wear patterns and discoloration are signs of chronic durability problems.

2.2 System design of engine performance, loading, and durability

2.2.1 Engine system-level loading and durability design constraints

In general, all durability problems are related to four elements: loading, component structural design, material, and manufacturing. Failures occur when either the loading is too high or the structural design and material are not sufficiently strong. While the structural design and material strength are mostly component-level topics, the loading is usually a system-level parameter. The system design engineers determine the load that is then cascaded to the component design teams for the selection of a structural design with the suitable material in order to sustain the load. The load here refers to any mechanical, thermal or flow-related parameters in a general sense. For instance, cylinder pressure and valve seating impact force are mechanical loads; cylinder head heat flux is a thermal load; and engine-out soot is a flow-related load to the DPF.

The following engine performance parameters in system design are related to the load and subject to durability design constraints or limits:

- engine brake torque
- mean piston speed and piston-assembly inertia load
- peak cylinder pressure, temperature, and heat flux
- exhaust manifold pressure and temperature
- compressor outlet air temperature and compressor pressure ratio
- turbocharger speed
- coolant heat rejection and engine outlet coolant temperature
- charge air cooler and EGR cooler outlet gas temperatures
- valvetrain load
- piston slap kinetic energy
- engine-out soot load for lubricant oil and DPF regeneration.

The major design constraints are listed in Table 2.1. Other miscellaneous components or subsystems that are not directly related to system design loads are not shown, for example electronics and sensors, other components in the fuel, cooling and lubrication systems, pumps, seals, breathers, etc. The engine-out soot flow rate affects the lubricant oil degradation and component wear. It also affects the regeneration frequency and durability of the DPF. The exhaust manifold pressure limit refers to both the cycle-average and the maximum instantaneous pressure pulsation.

The load is a system parameter for three reasons. First of all, most of the loads acting on different components are essentially produced by the engine gas-side performance requirements (e.g., gas pressure, temperature, and flow rate). Therefore, they are determined at the design stage of system sizing.

Table 2.1 System performance parameters as durability design constraints

	Transmission and drivetrain	Cylinder head	Piston assembly	Cylinder liner	Valvetrain (e.g., valves, cam)	Connecting rod and crankshaft	Bearings	Crankcase	Intake manifold	Exhaust manifold	Turbocharger	EGR system (valve and cooler)	Charge air cooler	Fuel injector	Aftertreatment (e.g., DPF)
Engine brake torque	x														
Mean piston speed			x		x	x	x								
Peak cylinder pressure		x	x			x	x	x							
Peak cylinder temperature and cylinder heat flux		x	x		x									x	
Exhaust manifold temperature										x	x	x			
Exhaust manifold pressure										x	x	x			
Compressor outlet air temperature													x		
Compressor pressure ratio											x				
Turbocharger speed											x				
Heat rejection and engine outlet coolant temperature		x	x												
Cooler outlet gas temperature			x									x	x		
Piston slap kinetic energy				x											
Valvetrain loading					x		x								
Engine-out soot loading									x						x
Oil soot percentage			x		x	x	x								

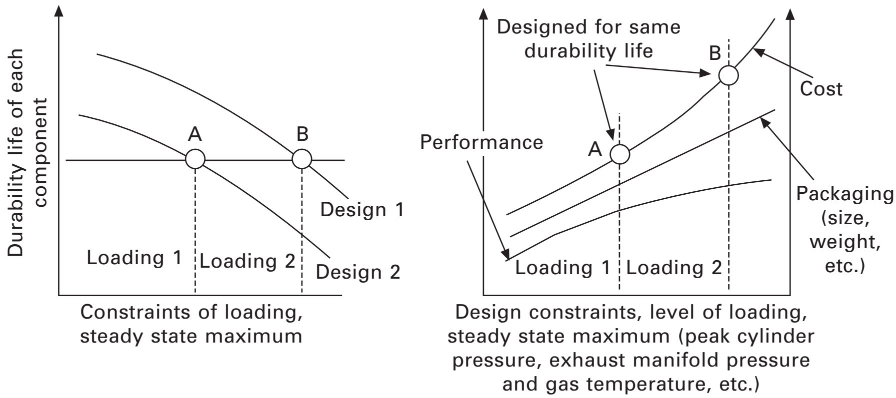
A typical example is the peak cylinder pressure that is determined by the required power density, emissions level and optimum engine compression ratio. Another example is the need to increase the design limit for the maximum allowable exhaust manifold pressure for achieving high air–fuel ratio in order to reduce soot at rated power. A third example is that the EGR cooler outlet gas temperature cannot be designed too low due to concerns about corrosion and hydrocarbon fouling. The system engineer needs to understand the reliability consequences of the proposed limits.

Secondly, many loads are related to each other inherently via engine thermodynamic processes and therefore require system-level coordination. For example, when the air–fuel ratio is increased, the peak cylinder pressure increases but the exhaust manifold temperature may decrease. Another example on a system-level balance between different durability constraints dated back to the 1970s when Zinner (1971) compared the design solutions of increasing BMEP and mean piston speed in order to increase the diesel engine power for a given engine displacement. For automotive, industrial and marine applications, the increase in continuous power output is always desirable. The engine components are subject to increased mechanical and thermal loads at higher power. Based on a simplified analysis, Zinner (1971) concluded that increasing BMEP is a simpler, cheaper and more reliable approach than increasing the mean piston speed to meet the higher power requirement.

Thirdly, the overall design and sizing of many subsystems need to be conducted and coordinated at a system level in order to ensure the entire system is optimized and the loads for each subsystem or component are well controlled. One example is the cam profile design and the valve spring load selection. Cam profile affects engine performance. The corresponding valvetrain load matched for the cam needs to be determined by the system engineer. Another example is to select the size of the inter-stage cooler for a two-stage turbocharger in order to control the air temperature at the high-pressure-stage compressor outlet. The cooler sizing is the result of the overall coordination of comparing and balancing different design solutions related to the cooler, the turbocharger and EGR in order to control the air temperature.

2.2.2 Iterative design of system performance and durability

There are usually trade-offs among performance, durability, packaging, and cost, as shown in Fig. 2.1. For example, the engine structure can always be designed strong enough to sustain a very high cylinder pressure, but the penalty is heavy weight and high cost. If the maximum allowable system loads can be specified accurately at the early stage of the design, it will be easier to



2.1 Durability analysis for engine system design.

close the loop of re-iteration from the system level to the component level to ensure the loading parameters match well with the component strength at affordable cost and packaging.

System performance and durability engineers need to ensure the loads designed at the system level are the reasonable and sustainable design targets to be cascaded to each subsystem to be realized through detailed component-level design. A thorough understanding of how engine durability issues are generated, analyzed, and resolved is required for a system engineer in order to determine or propose appropriate design limits. Engine durability validation usually requires prolonged experimental testing at the late stage of the development program. A detailed component-level finite-element structural analysis usually requires long computational time to simulate the failures and estimate the lifetime of the components. In order to consider structural robustness and reliability at the early engine system design stage, analytical models need to be developed to determine the appropriate limits of durability design constraints. A computationally efficient system-level durability and reliability model is highly desirable in order to estimate the impact of system loads on engine lifetime. Due to the extreme complexity of durability issues, the values of durability design constraints (e.g., maximum allowable cylinder pressure and exhaust temperature) to be used in engine system design as design limits often have to rely on empirical experience.

2.2.3 The role of system durability engineers

In the iterative design process to achieve performance and durability, a key issue between the system and component engineers is to determine the reasonable loading parameters or system design constraints/limits. As shown in Fig. 2.1, the system engineer selects the point on the curves as a system design specification to cascade to the component engineer. If the loads are not reasonable from a durability standpoint, they need to be revised iteratively with the component engineers during the course of design.

In order to conduct system optimization, the performance engineer needs design maps of durability as a function of loads or design constraints (Fig. 2.1). The map should include parametric sensitivity variations of key design and material parameters, preferably for each failure mode and each major component. Similar demands exist for the attributes of packaging and cost since the system performance engineer needs the maps from all the attributes in order to integrate them together. One major requirement for a system engineer is to generate the durability constraint maps of various components based on experimental data and numerical simulations, and then assemble a durability map for the whole engine system through appropriate integration.

Currently, the engine system design in the performance area has reached a level of highly precise on-target design, under the assumption of pre-selected durability design constraints. The validity and accuracy of the design constraints naturally become an issue for further examination. If the durability design constraints corresponding to satisfactory reliability are wrongly determined, the precise design capability in the performance area cannot ensure good design outcome in the durability/reliability area. Such a mismatch may become the bottleneck for the overall engine system design from the standpoint of design for reliability.

In the early stage of the design it is unlikely that detailed finite element analysis or extensive experimental work will be conducted to search for appropriate maximum allowable design constraints. Fast and effective heuristic modeling and benchmarking analysis are required to fulfill the above demand. Although very challenging, better predicting engine durability and reliability is the future direction for improving the quality of diesel engine system design.

2.3 The relationship between durability and reliability

As discussed in Chapter 1, the relationship between durability and reliability can be summarized as follows. Reliability is a probability reflecting the possibility of failure of any quality issue such as durability, performance, packaging, or manufacturing after the product is put in service. The probability of failure can be caused by the variations in production population or environmental changes. Durability is a product attribute reflecting structural endurance. If a device never functions properly, it is a quality issue. If the device works well once but fails at the second time, it is a reliability issue. The issue can be related to performance (e.g., device not functioning due to excessive friction or heat without structural failure) or durability (e.g., device failing due to rupture or fatigue). Reliability can be assessed by using customer service data and/or engine development data in the design/validation stage. Durability development needs to screen and solve most, if not all, of the problems that would otherwise be encountered in service. A durability issue sometimes may be addressed with only one prototype in the engine test cell or may be studied with many samples and environmental noise factors in terms of probability of failure. When a performance or durability issue in the design stage is studied by using the statistical probabilistic approach, it essentially reflects the concept of 'design for reliability', which is the ultimate design goal.

Durability study focuses on how to solve the structural problems in the design stage, while reliability focuses on counting how many problems still show up in service after the durability development. For example, if the work is to revise a piston design to prevent premature rupture, it is called durability work. On the other hand, if the focus is piston failure rate in production population with different customer usage after one million miles, it is a reliability topic. This difference is reflected by the fact that in the structural durability area the deterministic approach using safety factors has been convenient and acceptable (although simplified), while in the reliability area the probabilistic approach has always been required.

2.4 Engine durability testing

Engine durability testing is the most important development work to validate the design after the prototype is available. Heavy-duty vehicle endurance testing was summarized by Murphy (1982). Diesel engine durability testing was introduced by Goshorn and Krodel (1978) and Alcraft (1984). The fundamentals of material properties of the automotive engine components were introduced by Yamagata (2005).

Each engine manufacturer has developed proprietary durability test cycles based on their experience to validate structural components and the engine system. A durability test is conducted to let the engine or the component undergo sufficiently high mechanical and thermal loads (stresses) and a sufficient number of fatigue cycles (e.g., hundreds of hours). Durability tests are custom made and depend on engine specific requirements. Typical engine durability tests include full-load test in the lab, over-fueling test, load cycle tests, field test in vehicles, etc. Specific cycles include low-temperature environment, thermal cycling, deep thermal shock, resonance cycle, overload cycle, high load factor in combination with superimposed thermal cycle, engine brake cycle, etc. For example, thermal fatigue cycles are run from no load to full load repeatedly for several hundred hours to validate the sealing capability of the cylinder head gasket joint and the durability of the exhaust manifold. The test is conducted by exposing the engine components under deep thermal gradient conditions. Thermal shock tests are conducted by quickly changing the rate of temperature change in thermal cycles. Note that many durability tests for alternative fuels have been run according to the cycles suggested by the EMA (Engine Manufacturers Association) to determine the difference in wear and carbon deposits.

Moreover, modern diesel engines are equipped with EGR systems and many control valves. Those new devices need to be validated with respect to long-term durability. Some of the tests for those devices need to be conducted at low speeds and low loads (e.g., soot accumulation test, DPF regeneration test) instead of the traditional high-speed and high-load tests.

Depending on failure mechanisms, different engine components may encounter their worst durability conditions at different engine speeds/loads. The peak stress may not necessarily occur under the steady-state full load or maximum power condition. For instance, the thermal load may be higher at peak torque than rated power. Maximum cam stress might occur at the cranking speed rather than rated speed. The oil control problem may become worst at low idle and part load. The worst crack due to low cycle fatigue of the cylinder head may occur after a thermal cycle instead of steady-state full load. In other words, different driving cycles may produce different durability issues because the failure mechanism is dependent on engine speed, load, and cycle history.

It is important to investigate the correlation between the engine dynamometer test cycles and the real world usage profiles in order to avoid over-design or under-design of the engine. The load cycles used in dynamometer durability testing need to be representative of real world usage in service. In order to save testing time, sometimes accelerated durability tests are conducted in the lab to accelerate the wear and structural failures by intentionally increasing the engine speed, load or loading cycle frequency.

In-vehicle field durability tests are conducted under various climate conditions such as normal ambient temperature, hot humid, hot dry, cold humid, cold dry, sea level, high altitude, etc. The tests need to represent real world driving conditions including the worst case scenarios such as highway, full load and low load. In vehicle field tests, some important engine parameters need to be recorded, including vehicle speed, fuel economy, oil consumption, used oil data (for wear analysis), engine speed and torque, accelerator pedal position, fuel injection pressure and quantity, exhaust manifold pressure, turbine outlet pressure, EGR valve duty cycle, turbocharger control duty cycle, and fluid temperatures (fuel, oil, coolant, and air).

Component bench testing, metallurgical analysis and engine oil analysis are also important in determining design acceptance related to material selection, wear, and reliability. Regular oil sampling and analysis can prevent major repairs and catastrophic failures. At the end of the engine durability test, acceptance is determined based on experience and design margins are validated.

It is always important to correlate the following three sets of data in order to predict reliability: the predicted durability life, the predicted reliability, and the collected reliability service data. Durability assessment and prediction rely heavily on experimental endurance tests, either in engine laboratory with specially designed accelerated testing or on vehicle in the field. A durability target of B10 or B50 life can be derived by using a Weibull diagram which depicts the failure rate (usually on the vertical axis of the Weibull chart) as a function of product life, number of running cycles or vehicle distance traveled (on the horizontal axis of the Weibull chart). The failure characteristics of the components (e.g., early failure, random failure) can be evaluated based on the beta slope of the Weibull curve.

2.5 Accelerated durability and reliability testing

Accelerated durability or reliability testing is a fast growing area because it enables the designer to determine the durability or reliability of a product more quickly. The goal of controlled accelerated laboratory testing is to achieve the results that are representative of real world usage at a lower testing cost and reduced testing time. By exposing a design to a combined set of amplified stresses, multiple failure modes and their sequence and distribution may be obtained in a short time. The same failures which typically show up in service over time at much lower stress levels show up quickly in the short term overstress condition in the lab. However, it should be noted that generally the greater the acceleration, the less realistic the test result.

Engine durability accelerated testing is very complex due to multiple types of loading and multiple failure modes involved. It is difficult to decompose them to simpler cases or cycles. Appropriate acceleration factors need to be determined to use in the test. Engine durability development often employs accelerated testing by using specially designed durability cycles and load factors. For the need of engine system design, it is desirable to find out the correlation between the accelerated testing and the real world usage scenarios for critical system design constraints. For example, if an accelerated durability test concludes that 200 bar peak cylinder pressure needs to be used as the design constraint, the question to confirm is the level of confidence that the 200 bar is representative in the non-accelerated real world. It is desirable to possess a certain level of understanding on how accelerated tests are designed and how the results are interpreted.

For the fundamental theories on accelerated testing, the reader is referred to Nelson (2004), Dodson and Schwab (2006), Escobar and Meeker (2006), Klyatis and Klyatis (2006), and Klyatis (2010). More detailed information on accelerated testing in engine or automotive applications was provided by Indig and Williams (1984), Goshorn and Krodel (1978), Krivoy *et al.* (1986), Dystrup *et al.* (1993), Vertin *et al.* (1993), Davis and Christ (1996), Kestly *et al*. (2000), Niewczas and Koszalka (2002), Feng and Chen (2004), Evans *et al*. (2005), and Franke *et al*. (2007).

2.6 Engine component structural design and analysis

The objective of engine component structural design and analysis is to achieve light and durable design by the optimization of shape and the selection of material with low manufacturing cost. General design guidelines to ensure satisfactory thermo-mechanical structural durability include the following: increase structural stiffness, avoid high gradients in stiffness and temperature, avoid high strains, and reduce temperatures by cooling. General guidelines for good wear durability include reducing the load and promoting hydrodynamic lubrication.

Engine structural analysis usually includes the following areas:

- Static analysis of deflection, plasticity, stress, and strain
- Thermo-mechanical fatigue
- Wear
- Cavitation
- Fatigue life analysis
- Multi-body dynamic vibration and modal analysis
- Transient structural analysis
- Fluid-structure interaction.

Finite element analysis (FEA) (Haddock, 1984) coupled with advanced constitutive material models is the routine analysis tool used for structural design. It can handle complex component geometry and identify local stress concentration. Safety factors have been widely used in structural design. Safety factors are calculated at every node in the FEA model based on the stress calculated so that a spatial distribution of the safety factor can be obtained in order to view the most critical location at a given loading condition.

The capability of today's structural analysis in component design has reached an advanced level so that it can predict reasonably well the critical failure area and the fatigue life. The trend of analysis is to integrate individual components and individual failure modes to achieve a full system validation, and also to include a probabilistic approach for estimating the structural reliability.

2.7 System durability analysis in engine system design

With the system loads used as input, durability analysis in system design prior to the detailed component-level design serves three purposes: to provide preliminary structural sizing and strength evaluation on stress and strain for major engine components; to estimate durability/reliability life; and to help the performance area determine the durability design constraints/limits. A detailed component-level design analysis is certainly too time-consuming hence not realistic at the early stage of the system-level design. The system durability calculation normally does not require a full sophisticated finite element analysis. Therefore, the detailed information of stress concentration due to complex geometry is generally ignored. The analysis has two types of approaches: the deterministic approach with the concept of safety factor, and the probabilistic approach with the concept of reliability or failure rate. The hand-calculation formulae used in concept-level structural evaluation for engine components with the deterministic approach were summarized by Makartchouk (2002). The structural analysis with the more advanced probabilistic approach was introduced by Kececioglu (2003).

For the preliminary structural calculation of the connecting rod, crankshaft, bearing and valvetrain, the structural loads can be calculated by using engine cycle simulation and multi-body dynamics of piston assembly, crankshaft, or valvetrain to obtain the cylinder gas pressure load, the component inertia load, and the vibration load. In the thermo-mechanical fatigue problems for the piston, the cylinder head, and the exhaust manifold, the mechanical and thermal loads are calculated by using engine cycle simulation. The boundary conditions of gas temperature and heat transfer coefficient can also be calculated. A simple FEA can be conducted with the cycle simulation software (e.g., GT-POWER) to obtain the metal temperature distribution in the components. The fatigue damage models are detailed in the later sections. For the cylinder liner cavitation problem, piston-assembly dynamics can be used to calculate the piston slap kinetic energy (detailed in Chapter 11). The wear at the piston ring, the cam and the valve seat can also be modeled (also detailed in later sections). System durability modeling requires the development of fast and simplified surrogate durability models to replace the full FEA models.

Structural analysis usually starts with a deterministic model to analyze a particular pair of stress–strength relationship and the safety factor. However, it is not sufficient to only consider a nominal case (i.e., a combination of mean material properties, average usage parameters, and pre-selected safety factors). The engineer must also consider the nondeterministic probability distribution of all the noise factors such as load changes, potential variations in material properties, different customer usages and manufacturing tolerances.

From a probabilistic standpoint structural failures can be illustrated by the commonly used stress–strength probability distribution curve or the stress–strength interference model (Fig. 1.13 in Chapter 1). Here the 'stress' refers to any type of load in general, and 'strength' refers to the maximum structural capability of the component in general. Failure occurs in the over-stressed area where the stress is higher than the strength. Mechanical failures (e.g., fracture or fatigue) are caused by overload or strength degradation. Overload can be caused by misuse or insufficient design limit adopted. The strength degradation is caused by material property degradation, high-temperature creep, fatigue, corrosion, or chemical attack. The probabilistic analysis will be detailed in the later sections related to reliability.

2.8 Fundamentals of thermo-mechanical failures

2.8.1 Overview of thermo-mechanical structural concepts

System durability analysis is an important part of diesel engine system design. The failure mechanisms and the concepts in the area of structural durability analysis are very complex. Most system engineers coming from the engine performance area are not familiar with the structural analysis concepts. Compared to other mechanical failure mechanisms such as wear and cavitation, the theory of thermo-mechanical failures has a paramount importance for engine system durability because two critical performance parameters are directly related to them: peak cylinder gas pressure and exhaust manifold gas temperature.

This section lays a foundation for system durability analysis by summarizing the key concepts and the available analysis approaches related to thermo-mechanical failures. Detailed explanation of the fundamentals can be found in numerous textbooks (e.g., Draper, 2008) and will not be covered here. The three most basic concepts are stress, strain and strength. They are important because stress and strain are used differently in different failure mechanisms as the control criterion. Strength is related to the structural design limit selected. The next important concept is damage. In the damage model the failures due to different mechanisms or types of loading cycles can be combined in order to assess the reliability life. Thermo-mechanical failure modes are introduced in this section, mainly including fracture, fatigue and creep. Fatigue is elaborated, focusing on thermal fatigue, high cycle fatigue (HCF), and low cycle fatigue (LCF). In-depth discussions on structural durability analysis are presented in Rie and Portella (1998), Nicholas (2006), and Yu (2002). For details of thermo-mechanical structural analysis, the reader is referred to a series of research work conducted by LMS International (Nagode and Zingsheim, 2004; Nagode and Hack, 2004; Nagode and Fajdiga, 2006, 2007; Nagode *et al.*, 2008, 2009, 2009a, Rosa *et al.*, 2007; and Seruga *et al.*, 2009). SAE procedures also provide useful information on durability and reliability (SAE J450, J965, J1099, J2816, JA1000-1, and JA1000).

2.8.2 Fundamental concepts of mechanical failures

Stress

When the stress is higher than the yield limit the material works in the plastic domain. The ratio of the ultimate strength or yield strength to the actual stress is defined as the factor of safety. In a one-dimensional model where the effects of the other two dimensions are neglected, the stress tensor is reduced to only one component and indistinguishable from a scalar stress. The stress in this case is called a uni-axial stress. In engineering applications of elastic materials it is usually assumed that the cross-sectional area of the structural element remains constant during deformation when the element is elongated or compressed. In fact, the cross-sectional area changes by a small amount depending on the Poisson's ratio of the material. The stresses determined under the assumption of constant cross-sectional area are called engineering stress or nominal stress. The stress calculated by considering the cross-sectional area change is called true stress. The true stress in uni-axial compression is less than the nominal stress. The true stress in uni-axial tension is greater than the nominal stress.

The loads can create multi-axial stresses in the engine component due to external loads, residue stresses and the geometry of the structure (Bignonnet and Thomas, 2001). Although in many cases a calculation of uni-axial stresses is sufficient for preliminary analysis, FEA is required to calculate the stresses under multi-axial loads together with the complex geometry of the component.

Examples of stresses include the residual stresses as a result of material casting process and heat treatment, residual stresses due to inelastic deformation, pre-stresses due to press-fit and tightening of the bolts (bolt loads), dynamic mechanical stresses due to gas and inertia loading, compressive stresses caused by constrained thermal expansion due to different thermal expansion coefficients of the parts, and temperature gradients in the components, etc. The residual stresses are those that remain after the original cause of the stresses has been removed. Uncontrolled residual stresses are undesirable.

Stress relaxation refers to the behavior that the stress in the material decreases when time increases and the strain is unchanged. Stress relaxation may affect the loading behavior of fasteners and springs.

Stress or strain concentration is a common problem in component design resulting in failures even under low loads. Both loads and component geometry design details determine the location and the level of stress concentration. In the system-level stress estimation during the system design stage where the component design details are not available, it is reasonable to assume the component design will solve any stress concentration problems successfully so that they will not become the bottleneck for system performance and the associated required loads.

Strain

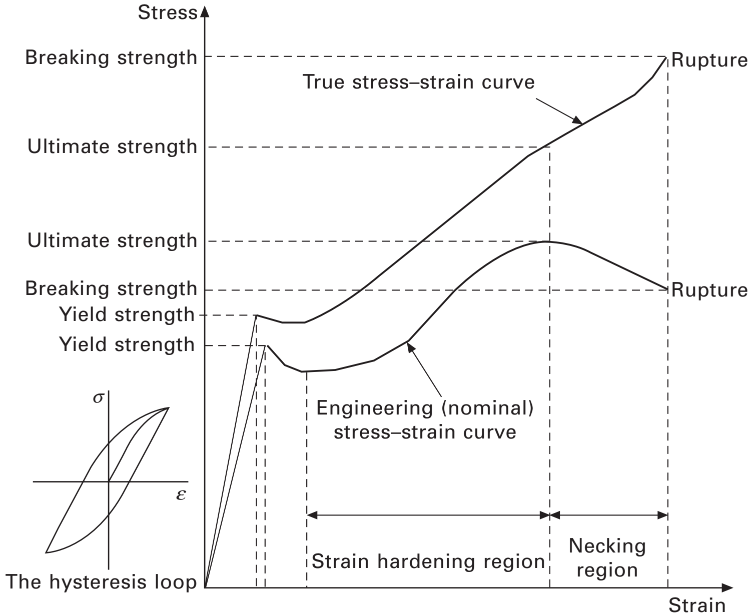
Mechanical strain is a geometric measure of deformation representing the relative displacement between particles in a material body. Strain is caused by external constraints or loads. There are two types of strain: elastic and plastic. Stress–strain curves are usually used to characterize the material structural behavior (e.g., the hysteresis loop, Fig. 2.2) in a load cycle. The stress–strain curves together with the material properties obtained by thermal fatigue testing can be used to ensure the accuracy of finite element structural analysis.

There are two types of models of stress–strain relationship. In the stress–strain model based on the deformation theory of plasticity, the strain is expressed as a function of stress, temperature, and time. In the unified stress–strain model based on the incremental theory of plasticity, the plastic and creep strain increments are treated as one inelastic strain increment. Typical models of the above two types are the Robinson model and the Sehitoglu model, respectively, quoted by Ogarevic *et al.* (2001). The life time of the component can be predicted based on the plastic strain amplitude or the stress amplitude, or both.

Strength

The yield strength is the stress at which the material strain changes from elastic deformation to permanent plastic deformation. Below the yield strength all the deformation is recoverable because it is within the elastic range. For applications where plastic deformation is not acceptable, the yield strength is used as the design limit. The yield point of some ductile metals is defined by using the rule of '0.2% offset strain'.

The ultimate strength is the maximum stress a material can withstand at compression, tension or shearing loads before fracture occurs (i.e., the peak on the engineering stress–strain curve in Fig. 2.2). The ultimate strengths for compression and tension may be similar or very different in value, depending on the material. For instance, the ultimate compressive and tensile strengths of steel (a ductile material) are similar, while the compressive strength of cast iron (a brittle material) is much greater than its ultimate tensile strength. Ultimate tensile strength indicates when necking (i.e., reduction in cross-sectional area due to plastic flow) will occur under a tensile load. Note that necking is not observed for materials subject to compressive loading. After the period of necking in tension loading, the material will rupture and the stored elastic energy will be released as noise and heat. The stress at the time of material rupture is known as the breaking strength. A typical example of an engine component subject to both compressive and tensile stresses periodically is the connecting rod.



2.2 Stress–strain property of steel.

Both yield strength and ultimate strength are functions of temperature of the material. Beyond a certain temperature (e.g., above 375°C for iron), the ultimate strength starts to decrease. Generally, tensile stress is detrimental to the component while compressive stress may relieve the fatigue of the material (Kim *et al.*, 2005).

2.8.3 Damage and damage models

Damage is defined as the ratio of the actual number of cycles (or time) to the required number of cycles (or time) to reach a pre-defined macroscopic failure (e.g., crack length due to LCF or HCF; wear amount). The concept of damage in structural analysis has usually referred to the failures of thermo-mechanical fatigue, but it can easily be extended to cover the damage caused by other mechanical failures such as wear and cavitation. Fatigue damage mainly includes three mechanisms: fatigue, creep, and oxidation. The damage can be caused by a single type of cyclic loading or multiple types of cyclic loading. For complex stress time histories of driving/usage profiles (e.g., obtained from thermo-mechanical simulation or test data), the commonly used approach to decompose them into a set of simple fatigue cycles is the 'rainflow decomposition' method (Masuishi and Endo, 1968).

The simplest way to estimate the fatigue damage under a combination of multiple types of loading cycle is to assume a linear accumulation of the fatigue damage due to the contributions from each type of cyclic loading as

$$D = \sum_{i=1}^{m} \frac{n_{li}}{N_{fi}} = 1 \tag{2.1}$$

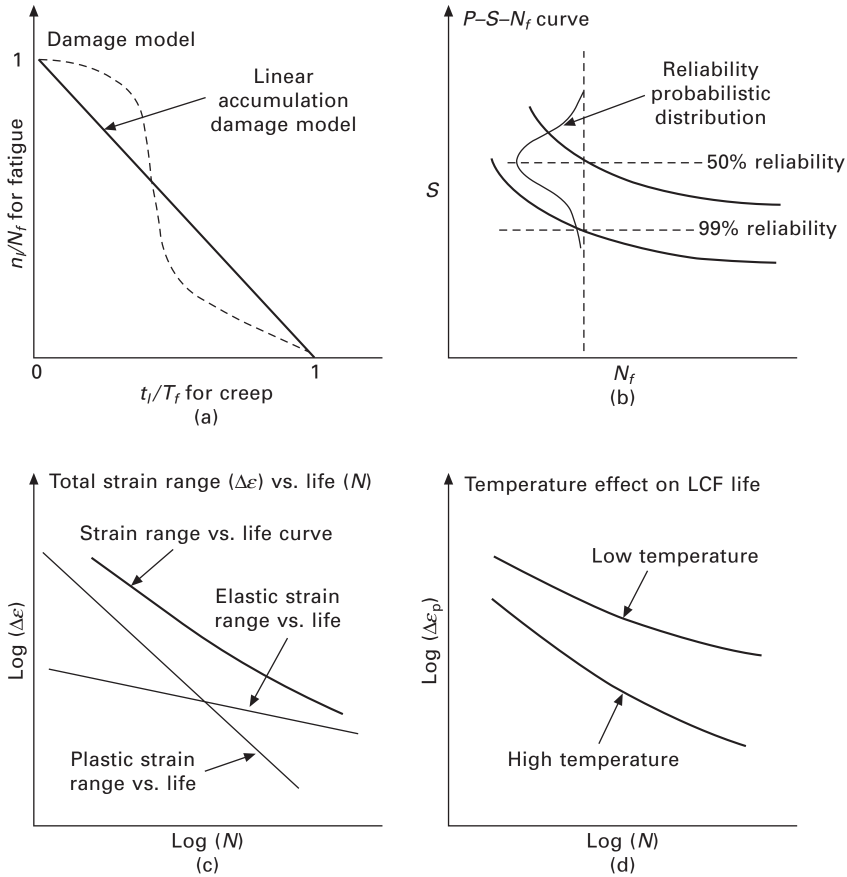
where D is the total accumulated damage over the entire stress–strain history, N_{fi} is the number of cycles required to reach the pre-defined failure criterion under the loading of the ith type of decomposed simple fatigue cycles, n_{li} is the actual number of loading cycles under the ith type of loading, and m is the number of types of the loading cycle. When the damage $D = 1$, it is assumed that fatigue failure occurs. In fact, D can be greater than or less than 1 in reality.

The above damage model can be extended to include the independent superposition of various damage mechanisms in order to account for the combined effect of fatigue, creep, oxidation, etc. The following model has been widely used:

$$\begin{aligned} D &= D_1 + D_2 + D_3 + \ldots + D_1 D_2 + \ldots \\ &= \sum_{i=1}^{m_1} \frac{n_{li,1}}{N_{fi,1}} + \sum_{i=1}^{m_2} \frac{t_{li,2}}{T_{fi,2}} + \sum_{i=1}^{m_3} \frac{t_{li,3}}{T_{fi,3}} \\ &\quad + \ldots + C_{1,2}\left(\sum_{i=1}^{m_1} \frac{n_{li,1}}{N_{fi,1}} \cdot \sum_{i=1}^{m_2} \frac{t_{li,2}}{T_{fi,2}} \right) + \ldots \end{aligned} \tag{2.2}$$

where D is the total damage, D_1 is the damage due to fatigue, D_2 is the damage due to creep, D_3 is the damage due to oxidation or degradation, etc. D_1D_2 is the interaction term of damage between the two mechanisms. $C_{1,2}$ is a constant characterizing the importance of the interaction effect (e.g., $C_{1,2} = -0.5 \sim 0.5$). T_{fi} is the required time to reach creep failure for the ith type of temperature or mechanical loading for creep, t_i is the actual time of creep under the ith type of loading for creep. The value of D can be calculated by using equation 2.2 based on experimental data or numerical simulation models. When $C_{1,2} = 0$ and $D = 1$, it means the accumulation of the damages due to different failure mechanisms follow the simplest model of linear accumulation of damage. Experimental evidence showed that the total damage due to the combined effects of fatigue, creep, and oxidation does not always give $D = 1$. Instead, the actual measured data of D fluctuate around the line of $D = 1$ (Fig. 2.3a).

The stress–strength model has traditionally been used in the probabilistic reliability evaluation for structural durability. Since many thermo-mechanical fatigue problems require using plastic strain (e.g., for low cycle fatigue problems) instead of stress as an evaluation criterion, the damage parameter



2.3 Damage model and durability life analysis.

has been commonly used to replace the stress in the stress–strength model in this scenario. For each material, component design and a usage loading profile at a given vehicle mileage or engine hours, one damage indicator can be calculated. Such a damage calculation can be repeated with the Monte Carlo simulation for the entire production population including all the variabilities in order to construct the stress–strength curve.

The damage models for life prediction (i.e., to determine the total required number of cycles to failure) are either stress-based (e.g., the Chaboche method) or strain-based (e.g., the Sehitoglu method, and the total strain–train range partitioning method or also known as the TS–SRP method). Ogarevic *et al*. (2001) concluded in their review article of thermo-mechanical fatigue that the three most widely accepted life prediction methods are the TS–SRP method developed by Manson and Halford, the Sehitoglu method, and the Chaboche method.

The TS–SRP method is one of the earliest methods capable of predicting thermo-mechanical fatigue life. The method is based on the observation that the strain range of any stress–strain hysteresis loop associated with thermo-mechanical fatigue can be partitioned into four basic strain-range elements in order to account for their different impact on the fatigue life. The partitioning depends on whether creep occurs and whether the strain is in compression or tension. The four strain range components are:

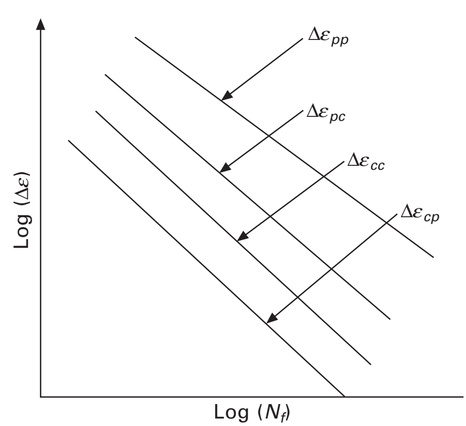
1. plastic strain in tension and plastic in compression, $\Delta\varepsilon_{pp}$
2. plastic in tension and creep in compression, $\Delta\varepsilon_{pc}$
3. creep in tension and plastic in compression, $\Delta\varepsilon_{cp}$
4. creep in tension and creep in compression, $\Delta\varepsilon_{cc}$.

The relationship between the strain range $\Delta\varepsilon$ and fatigue life N_f is introduced in Section 2.8.5. The fatigue lifetimes of the four components are ranked from low to high as: $N_{f,cp}$, $N_{f,cc}$, $N_{f,pc}$, and $N_{f,pp}$ (Fig. 2.4). The total life $N_{f,fc}$ due to fatigue and creep is given by the single-cycle damage model as:

$$\frac{1}{N_{f,fc}} = \frac{1}{N_{f,pp}} + \frac{1}{N_{f,pc}} + \frac{1}{N_{f,cp}} + \frac{1}{N_{f,cc}} \qquad 2.3$$

Equation 2.3 can be further modified by including weighting factors or proportional coefficients in order to make the life prediction more accurate. The total damage is the superposition of these four types of mechanisms. The history and details of the TS–SRP method were reviewed by Ogarevic *et al.* (2001).

In the Sehitoglu method, the total damage of the thermo-mechanical fatigue cycle is the sum of the damage due to fatigue, oxidation, and creep. The



2.4 Total strain–strain range partitioning.

fatigue life is calculated by a strain-based approach. The Chaboche method is a stress-based approach. In fact, it is generally believed that a strain-based method is more suitable than a stress-based method for LCF problems because LCF is controlled mainly by the plastic strain rather than the elastic stress as encountered in the case of HCF. As an example, in the Chaboche's model, the number of cycles to failure for each decomposed simple fatigue cycle due to pure fatigue, N_{fi}, is calculated based on stresses as below (for example, used by Morin *et al.* (2005) for a diesel cylinder head):

$$N_{fi} = \left(\frac{s_u - s_{i,max}}{s_{i,max} - s_{fl}}\right)\left(\frac{\Delta s_i}{C_1}\right)^{-C_2} \tag{2.4}$$

where i indicates the ith decomposed simple fatigue cycle, s_u is the ultimate stress, $s_{i,max}$ is the maximum stress, s_{fl} is the fatigue limit stress, Δs_i is the stress amplitude, and C_1 and C_2 are material constants.

These three prevailing models of life prediction, along with their corresponding material testing procedures, were elaborated by Ogarevic *et al.* (2001). It should be noted that the three methods are not interchangeable. Therefore, once a method is chosen, it is not easy to change to another one because the empirical database is usually built with one method.

2.8.4 Thermo-mechanical failure modes

Mechanical failures in engines include mainly fracture (rupture and crack), buckling, fatigue, creep, corrosion, oxidation, cavitation, wear, fouling, boiling, and deposits. Fracture can be caused by mechanical overload (e.g., an instant catastrophic failure due to rupture or breakage) or fatigue (i.e., macroscopic large broken cracks resulting from the fatigue-initiated microscopic small cracks which chronically propagate and grow). Fatigue includes high cycle fatigue and low cycle fatigue. Buckling is a special failure mode for long and slender components (e.g., the pushrod). Creep is a chronic plastic deformation of the material under stress. The causes of fracture, fatigue, and creep are stresses.

Thermal failures refer to any failures caused or aggravated by temperature effects such as thermal stress, thermal expansion, and material degradation at high temperatures. Thermo-mechanical failures refer to the combined effects of mechanical and thermal failures. Only the components exposed to the normal ambient temperature in the engine may experience pure mechanical failure with negligible thermal effects (e.g., the compressor inlet pipe, the engine mount). Most engine components experience both mechanical and thermal effects and may be subject to thermo-mechanical failure. Most commonly encountered failures are in the cylinder head, the exhaust manifold, or the piston bowl. For example, the gas-side of the cylinder head may have high temperature LCF failures, while the coolant-side of the cylinder head may have high temperature HCF failures. Which failure occurs first depends on particular engine applications. There are interactions between the mechanical and thermal failure modes, also between the various more detailed failure modes such as fatigue, creep, and oxidation. Material strength decreases due to the accelerated aging of the mechanical properties under elevated temperatures. The interactions affect the lifetime of the component. For a detailed introduction to engine thermal loading, the reader is referred to Heywood (1988) and French (1999).

Thermal failures

Thermal failures are caused by the thermal load which is essentially the metal temperatures, temperature gradients and thermal stresses in the component. Thermal load in engine components is related to gas temperature, heat flux, component design, and material properties. The components exposed to excessively high temperatures may have failure due to ablation, distortion, corrosion, creep, relaxation, and thermal fatigue. Material property degradation at high temperatures reduces the material strength to resist the failures. Excessive thermal deformation may induce the problems of scuffing and clearance interference, and aggravate the wear of tribological components. Alternating thermal stresses and strains cause thermal fatigue, especially under the effects of creep and relaxation. Thermal stresses may also aggravate mechanical fracture failures when the total of mechanical and thermal stresses exceeds the ultimate tensile strength of the material. As engine power density increases, thermal failures have become more difficult to control than mechanical failures. Thermal failures have become the primary limiting factor for engine reliability.

In-cylinder gas temperature and exhaust manifold gas temperature are two most important indicators of the thermal load for engine components. In-cylinder gas temperature affects the metal temperatures of the cylinder head, the injector tip, and the piston. Exhaust gas temperature affects the metal temperatures of the exhaust valve, the valve seat, and the exhaust manifold. It should be noted that the peak in-cylinder gas temperature and the exhaust manifold gas temperature are not always coherent. At a fixed engine speed, a higher peak in-cylinder temperature resulting from higher power or lower air–fuel ratio corresponds to a higher exhaust temperature. However, a lower peak in-cylinder temperature resulting from retarded fuel injection timing causes an increase in the exhaust temperature. The heat transferred to the component at different engine speeds is dependent on the timescale rather than the crank angle scale. Moreover, using the maximum temperature in a thermo-mechanical cycle to predict isothermal fatigue life may not be the safe or conservative method (Ogarevic *et al.*, 2001).

Ablation and excessive thermal deformation

Examples of ablation include unusual burn-out due to over-fueling for power cylinder components, backfire, or uncontrolled burning of soot in the DPF. Excessive thermal deformation usually occurs in the piston, the cylinder head and the exhaust manifold, and is caused by high in-cylinder gas temperatures or exhaust manifold gas temperatures. For instance, the piston temperature at the top ring position should not become excessively high, otherwise the lubrication of the ring may fail, resulting in ring scuffing or sticking in the ring groove.

Fracture and rupture

Fracture refers to the separation of a material into two or more pieces under the loading of stress. Fracture can be an instant or acute rupture of a component caused by mechanical overload in a single event, or as a final result of the chronic crack initiated and propagated due to long-term fatigue. Mechanical overload occurs when the applied load is greater than the ultimate strength (tensile, compressive, or shear) of the component. It can result in a tension failure, compression failure, shear failure, or bending failure (with both tensile and compressive forces). Ductile rupture is the ultimate failure of ductile materials as a consequence of extensive plastic deformation due to tension. In brittle rupture, no plastic deformation occurs before rupture. Rupture under tension occurs after the following steps: plastic deformation, necking, void nucleation, crack formation, crack propagation, and surface separation. The mechanism of cracking due to fatigue is different. At high temperatures the ultimate strength of the metal material decreases so that fracture occurs more easily. Fracture is a common structural durability problem in engine components.

Fatigue

According to the source of stress, fatigue can be classified into mechanical fatigue and thermal fatigue. For engine components the effects are usually combined as thermo-mechanical fatigue. According to the number of cycles to failure, fatigue includes high cycle fatigue and low cycle fatigue. HCF produces failures after a high number of cycles (e.g., greater than 10^4 cycles) with low stresses and elastic deformation. LCF produces failures after only a low number of cycles (e.g., smaller than 10^4 cycles) with high stresses and plastic deformation. Note that both mechanical and thermal fatigue can be either HCF or LCF. The timescale of one cycle can vary greatly from one engine cycle (e.g., for the loading cycle in HCF due to cylinder pressures) to several hours (e.g., for the loading cycle in LCF such as slow thermal cycles). Although the damage per cycle of LCF is greater than that of HCF, if the occurrence frequency of LCF is much lower than that of HCF, the component may reach a HCF failure earlier than a LCF failure. Different failure modes (e.g., different crack locations) may have different fatigue mechanisms (i.e., LCF or HCF, thermal stress induced or mechanical stress induced). The discussions on the mechanism and modeling of fatigue are detailed later.

Creep

Creep is a slow time-dependent irreversible process of plastic deformation for a metal material under the influence of stresses which are lower than the yield strength of the material. Creep results from long-term stresses and generates chronic strain accumulation or stress relaxation. Creep is generally damaging and related to inter-granular cracking and void growth over time.

The strain increase in creep is usually nonlinear with time. Both temperature and mechanical stress can generate creep, with temperature as the primary factor. Creep is especially significant at in-phase-loading conditions and less significant in the situation of out-of-phase loading. The creep under alternating mechanical or thermal stress is larger than the static creep. Creep increases with temperature. The rate of creep in terms of strain rate increases exponentially with metal temperature at high temperatures. The effect of creep becomes noticeable at approximately 30% of the melting temperature of metals. Large creep strain may cause cracks and fracture. When the temperature is sufficiently high, even if the stress is designed much lower than the yield strength, creep may occur as a plastic deformation to cause failure. Creep and plastic deformation occur at elevated temperatures in thermal cycles. This produces tensile stresses after the thermal load is removed.

Creep needs to be considered in the thermal failures encountered in high temperature operations of the engine. The plastic strain due to creep can be modeled as a function of time, temperature, and stress. The rate-dependent visco-plasticity theory needs to be included in the stress analysis for creep.

Corrosion

Chemical corrosion in the engine occurs due to the corrosive combustion products (e.g., sulfur in the fuel) in the exhaust gas in the combustion chamber, the exhaust and intake systems, the EGR system, and in the lubricant oil. Corrosion aggravates fatigue and wear of the component.

Oxidation

Oxidation damage is caused by repeated formation of an oxidation layer at the crack tip and its rupture. The rate of growth of oxidation layer thickness is proportional to the square root of time in the absence of cyclic loading; and the rate of growth is much higher in cyclic loading conditions where the oxidation layer repeatedly breaks and the fresh surface is exposed to the environment (Ogarevic *et al.*, 2001). The details of the modeling of oxidation damage and creep damage were provided by Su *et al.* (2002) in their investigation of cylinder head failures.

2.8.5 Fatigue

Material fatigue

Fatigue is a slow cycle-number-dependent irreversible process of plastic deformation when a material is under cyclic loading. It is a progressive and localized type of structural damage. Fatigue is caused by stress levels less than the ultimate tensile strength or even below the yield strength. Fatigue is embodied as macroscopic crack initiation and propagation induced by microscopic trans-granular fracture in the material. The causes and processes of fatigue can be explained by fracture mechanics as several stages: crack nucleation and initiation, crack growth, and ultimate ductile failure. Material behavior such as cyclic hardening, creep and plasticity has an important impact on fatigue. Fatigue is different from creep. Creep is embodied by the deformation and growing of the cavities inside the material primarily at high temperatures caused by inter-granular fracture. Creep is dependent on time rather than the number of loading cycles.

Fatigue life

Fatigue life is defined as the number of loading (stress) cycles of a specified character that a specimen sustains before failure of a specified nature occurs. The number of cycles is related to engine speed. It can be converted to equivalent durability hours. Fatigue life is affected by cyclic stresses, residual stresses, material properties, internal defects, grain size, temperature, design geometry, surface quality, oxidation, corrosion, etc. The fatigue life of a component under the following different fatigue mechanisms can be ranked from low to high as: thermal shock, high temperature LCF, low temperature LCF, and HCF. In the assessment of the risk of fatigue failure, it may be assumed that the component is safe for an infinite number of cycles if it does not show failures after more than ten million cycles.

The total fatigue life is equal to the life of crack formation and crack propagation. Fatigue life is dependent on the cycle history of the loading magnitude since crack initiation requires a larger stress than crack propagation.

The fatigue life of the component can be determined by the strain, stress, or energy approach. Fatigue is a very complex process affected by many factors. It is usually more effective to use a macro phenomenological method to model the effects of fatigue mechanisms on fatigue life rather than using a microscopic approach.

Fatigue strength and fatigue limit

Fatigue strength is defined as the stress value at which fatigue failure occurs after a given fatigue life. Fatigue limit is defined as the stress value below which fatigue failure occurs when the fatigue life is sufficiently high (e.g., 10–500 million cycles). Ferrous alloys and titanium alloys have a fatigue limit below which the material can have infinite life without failure. However, other materials (e.g., aluminum and copper) do not have such a fatigue limit for infinite life and will eventually fail even with small stresses. For these materials, a number of loading cycles is chosen as a design target of fatigue life.

Thermal fatigue

Thermal fatigue is a fatigue failure with macroscopic cracks resulting from cyclic thermal stresses and strains due to temperature changes, spatial temperature gradients, and high temperatures under constrained thermal deformation. Thermal fatigue may occur without mechanical loads. The constraints include external ones (e.g., bolting load) and internal ones (e.g., temperature gradient, different thermal expansion due to different materials connected). Compressive stresses are produced by the bolting load at high temperatures, or generated in the material having high coefficient of thermal expansion. Tensile stresses are produced when the component cools, or generated in the material having low coefficient of thermal expansion. Thermal stresses are produced by cyclic material expansion and contraction when temperature changes under geometric constraints. A crack may develop after many cycles of heating and cooling. The failure indicator or criterion of thermal fatigue is usually strain rather than stress. Thermal fatigue life is determined mainly by material ductility rather than material strength.

Thermal fatigue can be HCF or LCF, depending on the magnitude of thermal stress compared to the yield strength of the material. Thermal fatigue life can be predicted by using either stress (for HCF) or plastic strain (for LCF) as a criterion. The thermal fatigue in engine applications usually refers to the thermo-mechanical fatigue problems where thermal fatigue plays a dominant role.

Anisothermal fatigue can sometimes be more damaging than isothermal fatigue. Isothermal fatigue occurs when tension or compression cycles are imposed at a constant temperature. Anisothermal fatigue occurs when the component temperature and strain vary simultaneously. An engine may operate at isothermal conditions in steady state for a long period of time (e.g., in stationary power application or durability testing). Automotive engines often encounter anisothermal fatigue during largely varying thermal cycles. Anisothermal fatigue is more complex to model than isothermal fatigue because of the varying temperatures within the cycle.

The mechanical properties of the material deteriorate with time when the material is exposed above certain level of temperature. The ultimate strength of the material decreases due to the aging of mechanical properties at high temperatures. This aggravates the occurrence of plastic deformation in thermal fatigue. Experimental work has confirmed that the maximum component temperature in a thermal cycle (e.g., cycling from high engine speed-load modes to low speed-load modes) has a much greater influence on thermal fatigue life than the minimum or cycle-average component temperatures. The maximum temperature is also more important than the temperature range of the cycle for the reason that the fatigue-resistance property of the material deteriorates quickly at high temperatures. This means in engine system design the maximum gas temperature and heat flux should be used as design constraints in most cases.

Thermal fatigue life can be improved by reducing the temperature and temperature gradient or alleviate the geometric constraints. For example, reducing metal wall thickness can reduce the gas-side surface temperature and thermal expansion hence increase fatigue life. Using slots or grooves in the component may eliminate the constraints for thermal expansion.

Thermo-mechanical fatigue

The thermal fatigue in engines is usually accompanied by both thermal and mechanical stresses. The compressive and tensile stresses often exceed the yield strength of the material in much thermo-mechanical fatigue. Three typical engine components subjected to thermo-mechanical fatigue failures are the cylinder head, the piston, and the exhaust manifold.

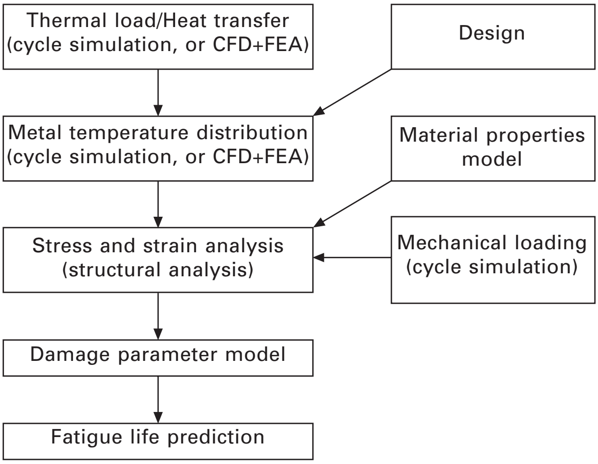
Thermo-mechanical fatigue (either HCF or LCF) failures consist of the accumulated damage due to three major mechanisms: mechanical or thermal fatigue; oxidation and degradation; and creep. Oxidation is caused by environmental changes. Degradation refers to chemical decomposition and the deterioration of material strength due to temperature change or mechanical fatigue. Material aging also affects damage but as a secondary effect.

Engine system design, component design and durability testing are three closely related areas to achieve successful design and prediction of thermo-mechanical fatigue life of the engine. Complete thermo-mechanical fatigue analysis includes both stress–strain and life predictions. The key elements in the analysis include the following (Fig. 2.5):

- dynamic thermal loading
- dynamic mechanical loading
- transient component temperature distribution
- material constitutive law (behavior) under both low and high temperatures
- stress and strain
- fatigue criteria and damage indicator
- component lifetime prediction
- statistical probabilistic prediction to account for the variations in the population.

The temperature field calculation is dependent on the thermal history and thermal inertia of the engine as well as the gas temperature and mass flow rate during the transient cycles. The effects of three-dimensional stresses and anisothermal cycle may be important in damage indicator calculation for detailed component-level design analysis.

A comprehensive review of engine thermo-mechanical fatigue was provided by Ogarevic *et al.* (2001). The simulation methodologies of thermo-mechanical fatigue life prediction were presented by Swanger *et al.* (1986), Lowe and Morel (1992), Zhuang and Swansson (1998), and Ahdad and Soare (2002). A discussion of diesel engine cumulative fatigue damage was provided by Junior *et al.* (2005). Deformation and stress analysis was reviewed by Fessler (1984).



2.5 Thermo-mechanical fatigue analysis process.

HCF

High cycle fatigue is a type of fatigue caused by small elastic strains under a high number of load cycles before failure occurs. The stress comes from a combination of mean and alternating stresses. The mean stress is caused by the residual stress, the assembly load, or the strongly non-uniform temperature distribution. The alternating stress is a mechanical or thermal stress at any frequency. A typical loading parameter in engine HCF is the cyclic cylinder pressure load or component inertia load.

HCF requires a high number of loading cycles to reach fatigue failure mainly due to elastic deformation. It has lower stresses than LCF, and the stresses are also lower than the yield strength of the material. HCF usually does not have macroscopic plastic deformation as large as that in LCF. The dominant strain in HCF is mainly elastic. In contrast, the dominant strain in LCF is plastic. Typical HCF examples are the cracks at the tangential intake ports connected to the flame deck on the water side of the cylinder head, the valve seat area, the piston, the crankshaft, and the connecting rod.

Because HCF is governed by elastic deformation, stress is usually a more convenient parameter than strain to be used as the failure criterion. The HCF life of the component is usually characterized by a stress–life curve (i.e., the s–N_f curve, Fig. 2.3b), where the magnitude of a cyclic stress is plotted versus the logarithmic scale of the number of cycles to failure. For constant cyclic loading, an empirical formula of the HCF s–N_f curve is usually expressed as

$$s^{C_1} N_f = C_2 \tag{2.5}$$

where C_1 and C_2 are constants, s is the maximum stress in the load cycle, N_f is the HCF life in terms of the number of cycles to failure. Higher stress results in a shorter life. Higher material strength can improve the HCF life.

LCF

LCF is a type of fatigue caused by large plastic strains under a low number of load cycles before failure occurs. High stresses greater than the material yield strength are developed in LCF due to mechanical or thermal loading. For instance, high tensile stresses are developed after the hot compressed area has cooled down during one loading cycle. The stresses may exceed the yield strength and cause large plastic deformation. Like other failures the crack due to LCF is usually initiated in the small areas having stress–strain concentration. The failure criterion of LCF can be a macroscopic crack with certain length or depth or a complete fracture of the component. Typical examples of LCF failure are the cracks at the inter-valve bridge area in the cylinder head and in the exhaust manifold after thermal cycles.

The lifetime of the component in LCF can be predicted by either plastic strain amplitude or stress amplitude, with the former being more appropriate and commonly used. In fact, both maximum tensile stress and the strain amplitude are useful in LCF life prediction. In general, larger plastic strains cause a shorter life. Better material ductility can improve the LCF life. Higher material strength may actually reduce the component life if it is subject to LCF failure since higher material strength usually reduces ductility.

Under a single type of cyclic loading (i.e., the plastic strain range remains constant in every cycle of the loading) the plastic strain deformation is usually predicted by the Manson–Coffin relation developed in the 1950s. The formula characterizes the relationship between the plastic strain range and LCF life as follows:

$$\Delta\varepsilon_p N_f^{-C_3} = C_4 \qquad \text{i.e.,} \qquad \Delta\varepsilon_p = C_4 N_f^{C_3} \tag{2.6}$$

where $\Delta\varepsilon_p$ is the plastic strain range, N_f is the number of load cycles to reach fatigue failure. Note that $2N_f$ is the number of reversals to failure in the stress–strain hysteresis loop. C_3 and C_4 are empirical material constants. C_3 is known as the fatigue ductility exponent, usually ranging from –0.5 to –0.7. Higher temperature gives a more negative value of C_3. C_4 is an empirical constant known as the fatigue ductility coefficient, which is closely related to the fracture ductility of the material. It is observed from equation 2.6 that a larger plastic strain range leads to a lower number of cycles of the fatigue life.

A more general relationship was later proposed by Mason by using the total strain, including both elastic and plastic strains, as the indicator of LCF failure:

$$\Delta\varepsilon_{total} = \Delta\varepsilon_p + \Delta\varepsilon_e = C_4 N_f^{C_3} + C_6 N_f^{C_5} \tag{2.7}$$

where $\Delta\varepsilon_e$ is the elastic strain range which is equal to the elastic stress range divided by the Young's modulus, $\Delta\varepsilon_p$ is the plastic strain range, C_6 is a coefficient related to the fatigue strength, C_3 and C_5 are material constants. Equation 2.7 can be used to construct the strain–life diagram for the component on a logarithmic scale (Fig. 2.3c).

The LCF in diesel engines is usually caused by large thermal stresses at high component temperatures which are higher than the creep temperature. The creep temperature is usually equal to 30–50% of the melting temperature of the metal in Kelvin. Many factors such as creep, relaxation, oxidation, and material degradation start to play important roles at high temperatures. At low temperatures the fatigue mechanism is predominant, while at high temperatures creep may become more important. The life of high temperature LCF is usually significantly lower than the life at lower temperatures (Fig.

2.3d). Ductility is a major factor to determine the LCF life. The material ductility is affected by temperature. Moreover, if the frequency of the loading cycle becomes slower, high temperature LCF life will reduce since creep and oxidation play more prominent roles at lower cycle frequencies. When the creep effect is considered, equation 2.6 is modified as below, and such a modification has been widely used to estimate the LCF life:

$$\Delta\varepsilon_{pc} = C_4 (N_f f_{ql}^{C_7 - 1})^{C_3} \qquad 2.8$$

where $\Delta\varepsilon_{pc}$ is the inelastic strain range including the plastic strain range and the creep strain range, C_7 is a material constant, f_{ql} is the cyclic loading frequency which accounts for the time-related effects of creep (e.g., the creep hold time) and relaxation, and so on. The elastic strain term in equation 2.7 can be modified similarly.

Barlas *et al.* (2006) provided an analysis to calculate the number of cycles to failure due to pure creep as an integral over a time period of t_1–t_2 for the cylinder head:

$$\frac{1}{N_f} = \frac{1}{C_8 + 1} \int_{t_1}^{t_2} \left(\frac{s}{C_9}\right)^{C_{10}} \mathrm{d}t \qquad 2.9$$

where C_8, C_9 and C_{10} are material coefficients for pure creep, and s is an equivalent stress as a linear combination of the Von Mises stress, the principal maximum stress and the trace of the stress tensor.

Strictly speaking, the classical LCF laws (e.g., the Manson–Coffin relation) based on the isothermal and uni-axial conditions are not valid for the anisothermal and multi-axial problems in reality. Therefore, the available prevailing analysis method is more or less a simplified approximation for the real world durability problems encountered in engines. Predicting the LCF life in the anisothermal and multi-axial conditions is very challenging, especially for the concept-level structural/life analysis in engine system design. Lederer *et al.* (2000) provided an in-depth discussion on the limitations of the classical CLF theories for fatigue life prediction. Lederer *et al.* (2000) also showed that, using a reasonably simple anisothermal LCF criterion, their simulation produced good agreement between the predicted and the tested critical zones of failure.

Thermal shock

Thermal shock refers to the process that the component experiences suddenly changed thermal stresses and strains of large magnitude when the heat flux and component temperature gradient change abruptly. Thermal shock produces cracks as a result of rapid component temperature change. The stresses generated in thermal shock are much greater than those in normal loading cycles, and even greater than the ultimate strength of the material. Thermal shock can be regarded as a severe type of LCF although it has its unique characteristics. The criterion used to analyze thermal shock failures can be strain or stress, with strain being more appropriate.

Thermal shock can make the material lose ductility and shorten the normal LCF life and the thermal fatigue life of the component accordingly. It may also cause brittle fracture which has a much shorter life than the normal LCF life. The materials having low thermal conductivity and high thermal expansion coefficient are vulnerable to thermal shock. Thermal shock can be prevented by reducing the thermal gradient through changing the temperature more slowly, or by improving the robustness of a material against thermal shock through increasing a 'thermal shock parameter'. The parameter is an indicator of the capability to resist thermal shock, and is proportional to the thermal conductivity and the maximum tension that the material can resist, and inversely proportional to the thermal expansion coefficient and the Young's modulus.

2.9 Diesel engine thermo-mechanical failures

2.9.1 Cylinder head durability

Failure mechanisms of cylinder head

Cylinder head fatigue and cracking are among the most important issues in engine durability development. Cylinder head durability is limited by thermo-mechanical fatigue. Controlling the maximum metal temperature and temperature gradient in the cylinder head is the key to solving thermal fatigue problems. The cylinder head has two major failure modes: the HCF crack due to cylinder pressure loading (high stress) at the coolant side, and the LCF crack caused by severe thermal loading (high plastic strain) at the gas side. Kim *et al*. (2005) pointed out that HSDI diesel cylinder head tends to have more cracks in the water jacket area due to HCF than the cracks on the gas-side fire deck due to LCF. Maassen (2001) summarized the failure modes for the diesel cylinder head.

There are four types of load in the cylinder head as follows:

- Preload stresses coming from the manufacturing process, such as the residual stresses due to molding, heat treatment (e.g., quenching) and machining processes.
- Assembly loads such as bolting load and press-fits. The residual stresses and the assembly loads affect the value of the mean stress hence the fatigue life and cylinder head durability.
- Mechanical operating load due to cylinder gas pressure.
- Thermal load caused by high in-cylinder gas temperatures and large variation of temperatures during operating duty cycles.

HCF in cylinder head

Because the diesel engine operates under high peak cylinder pressures, HCF cracks occur frequently, especially in the water jacket area in the cylinder head. High tensile residual stresses generated from the heat treatment processes were found to cause cracks at the foot of the long intake port with the lowest safety factor (Kim *et al.*, 2005). They also found that the bolting load of the cylinder head induces bending deformation and tensile stresses at the crack location. The firing gas pressure load and thermal load increase the tensile stresses.

There could be a large variation in stress levels from cylinder to cylinder due to inhomogeneous component temperature distributions. Design solutions for the cracking problem include increasing the bending stiffness at the cracking location in order to reduce the amount of bending, and reducing the tensile residual stresses with better heat treatment process during manufacturing. Design solutions for cylinder head HCF problems were elaborated by Hamm *et al.* (2008) with the use of minimum safety factors as design evaluation criteria to compare different designs.

LCF in cylinder head

The LCF cracks in the cylinder head occur mainly on the gas-side fire deck at the inter-valve bridge area between the valves, the area between the injector hole and the exhaust valve seat, and the area between the glow plug bore and the exhaust valve seat, especially in the inter-valve bridge area between the intake and exhaust valves. The cracks are caused mainly by large thermal loading, temperature gradients, and thermal stresses in the relatively slow varying thermal cycles and the decreased material strength at elevated temperatures. Compressive stresses are generated in the cylinder head under high temperatures due to constrained thermal expansion. When the cylinder head temperature reduces at the low-load conditions, the material contracts. This results in high tensile stresses. Large temperature variations combined with the effects of creep and material thermal aging can make the local tensile stresses exceed the yield strength of the material under the hot condition to produce plastic strain amplitude. After a number of such hot–cold thermal cycles, cracking is initiated and propagates as the damage cumulates. A typical example of such a thermal fatigue cycle is the engine start–stop.

Although the metal temperature in the bridge area between the intake and exhaust valves is colder than that in the bridge area between the two exhaust valves, the intake–exhaust bridge is usually weaker in LCF due to a larger temperature gradient (Zieher *et al.*, 2005). The LCF thermal fatigue in that area imposes a constraint for engine system design in terms of thermal loading encountered by the engine during thermal cycles. One way to alleviate the thermal LCF fatigue is to prevent the metal temperature from increasing to a point where the elastic limit (yield strength) of the material starts to decrease (Gale, 1990).

Thermo-mechanical fatigue life prediction of cylinder head

Diesel engine cylinder head design was reviewed by Gale (1990). Design solutions to the thermo-mechanical fatigue problems of diesel cylinder heads were provided by Kim *et al.* (2005) and Hamm *et al.* (2008). Maassen (2001) provided a detailed discussion on residual stresses, HCF and LCF for cylinder heads. The structural modeling of thermo-mechanical fatigue failures of diesel cylinder heads was investigated by Koch *et al.* (1999), Lee *et al.* (1999), Maassen (2001), Su *et al.* (2002), Zieher *et al.* (2005), and Barlas *et al.* (2006). Su *et al.* (2002) analyzed the cast aluminum cylinder head, while Zieher *et al.* (2005) analyzed the cast iron cylinder head (gray cast iron, compacted graphite iron, ductile iron). Koch *et al.* (1999) developed a simulation tool to calculate the strains and stresses as a function of time for a defined engine operating cycle. Cyclic thermal load of temperature variation is applied to the FEA models of the cylinder head.

The thermo-mechanical damage of the diesel cylinder heads that are made of cast aluminum usually includes the accumulation of fatigue, oxidation, and creep (Su *et al.*, 2002). The damage of the cylinder heads that are made of gray cast iron includes one more mechanism – embrittlement. Zieher *et al.* (2005) described a constitutive model, a damage model and a life prediction model for the cast iron cylinder heads by considering three mechanisms: fatigue, creep, and embrittlement. They pointed out that for gray cast iron, brittle damage plays an important role in thermal cycles with large thermal strains.

Su *et al.* (2002) introduced a unified visco-plastic constitutive model for the material behavior, and extended the uni-axial thermo-mechanical fatigue life analysis to a three-dimensional analysis of stress and fatigue damage. Barlas *et al.* (2006) accounted for the effect of material aging in their constitutive model, which is important for diesel engine cylinder head durability analysis.

2.9.2 Exhaust manifold durability

Exhaust manifold loading and impact on engine performance

The exhaust manifold consists of the inlet flange, exhaust pipes (also called runners) and the outlet flange with gaskets and bolts. There are four types of loads acting on the exhaust manifold:

1. The thermal load (stress) caused by exhaust manifold gas temperatures and the associated cyclic varying metal temperature field (both temperature level and gradient).
2. Stresses generated by the external constraints such as the local restraining bolt forces and different thermal expansion. The first and second types of load are usually the primary loads for the exhaust manifold.
3. The dynamic excitations from the external attached components such as a turbocharger. The dynamic excitation affects NVH. The mechanical stresses sometimes cause HCF. In fact, HCF failure problems are rarely encountered with the exhaust manifold and are mainly due to improper design of the bracket or mounting.
4. Exhaust manifold gas pressure. This type of load has negligible effect on exhaust manifold durability.

Thermal cycles are encountered when the engine warms up and cools down as the engine is started and stopped or running at different speeds and loads. A thermal load may result in sufficiently high thermal stresses to exceed the yield strength of the exhaust manifold material to cause LCF.

Exhaust manifold gas temperature is a critical system design parameter that affects turbocharger performance. The exhaust temperature is influenced by fueling rate, intake manifold temperature, air–fuel ratio, EGR rate, fuel injection timing, exhaust port, and manifold heat transfer. In many cases a low peak cylinder gas temperature correlates to a low exhaust manifold gas temperature (for example, when the EGR rate is increased). Consequently, the exhaust manifold gas temperature is often used as a convenient indicator of the thermal load acting on the in-cylinder components. For example, engine derating may occur based on exhaust temperature due to the concern of durability of the cylinder head rather than the exhaust manifold. However, it should be noted that such a correlation does not always hold true since the peak in-cylinder gas temperature is affected by the engine thermodynamic cycle process during the compression and expansion strokes, while the exhaust manifold gas temperature is affected primarily by the in-cylinder gas temperature during the exhaust stroke.

Impact of exhaust manifold pressure on durability

Exhaust manifold pressure is one of the most important engine system design parameters. A thorough understanding of the impact of this parameter on durability is required in order to specify an appropriate design limit for the exhaust manifold pressure. Exhaust manifold pressure is governed mainly by engine exhaust flow rate, turbine area, and exhaust manifold volume. It affects turbine pressure ratio and hence intake manifold boost pressure. It also largely affects the engine delta P (i.e., exhaust manifold pressure minus intake manifold pressure). As the engine power density increases and the allowable engine-out soot level decreases, higher intake manifold pressure is required at the rated power condition. This results in the challenging demand on the engine structural strength to sustain the higher peak cylinder pressure, higher compressor outlet pressure and temperature, and higher exhaust manifold pressure. (Note that the compressor out temperature affects the LCF life of the compressor wheel and housing, the rubber seals, and the charge air cooler.)

There are two different limits that should not be exceeded in engine design and performance/emissions calibration due to durability concerns: the engine delta P and the exhaust manifold pressure. The engine delta P is an indicator of pumping loss. It also affects exhaust valve floating off the valve seat during the intake stroke. During the fast transient acceleration, operation in cold climate or exhaust braking, the EGR valve may stay closed so that a large amount of exhaust gas flows through the turbine to create a very high engine delta P. A sufficiently high exhaust valve spring preload needs to be used in order to prevent the exhaust valve from floating off the valve seat. If the spring preload is too low, excessive exhaust valve bouncing may occur after the valve floating. If the spring preload is too high, valvetrain friction and cam stress may be too high.

Exhaust manifold pressure usually has large pulsating amplitude due to gas wave dynamic effects. A very high exhaust manifold pressure itself does not cause structural failure of the manifold because the manifold wall is usually sufficiently thick. A gasket is used between the exhaust manifold and the cylinder head to prevent high temperature exhaust leaks or air seepages from the outside. The gasket can tolerate a much higher pressure than the normal operating pressure level of the exhaust manifold. The gasket leakage problem is usually not caused by the manifold pressure. Instead, the leakage is caused by the deformation due to thermal fatigue. A high exhaust manifold pressure itself also does not adversely affect the exhaust valvetrain when the engine delta P is under control. The maximum allowable exhaust manifold pressure, often occurring at full load or exhaust braking condition, is usually not constrained or determined by the following: EGR valve, EGR cooler pickup joint at the exhaust manifold, turbine shaft bearing, thrust bearing and seals, and turbocharger oil leakage. Instead, the maximum allowable design limit of the exhaust manifold pressure is usually determined by the following factors:

- fatigue of the thin EGR cooler tubes
- valve stem seals and valve guides
- the thin flexible bellows in the exhaust or EGR pipes at the turbine inlet
- turbine wheel HCF life (related to strong shockwave pressure pulses)
- sensors and their mounting in the exhaust manifold.

In fact, all the above five items can be improved in design in order to sustain a higher exhaust manifold pressure so that they do not become a bottleneck for raising the allowable manifold pressure limit. It should be noted that the maximum allowable exhaust manifold pressure, being one of the most important engine system design constraints, should not be casually imposed by a supplier (e.g., EGR system or turbocharger supplier) without reasonable justifications. Otherwise, the engine manufacturer would be unnecessarily over-constrained in engine design and performance/emissions calibration.

Thermo-mechanical LCF failure of exhaust manifold

Thermo-mechanical LCF is the primary failure mechanism in the exhaust manifold. The failures are reflected by manifold cracking and gasket leakage, caused mainly by high exhaust gas temperatures and high temperature gradients in the metal. Cracks most likely occur in the manifold region where the deformation is restrained or the temperature gradient is large (e.g., at the fillets). Exhaust manifold leakage is usually caused by plastic deformation due to the thermal load at the interface between the manifold and adjacent components. The high compressive stresses in the exhaust manifold are caused by the high material temperature and restricted thermal expansion by the screwed connections. When the stress exceeds the yield strength of the material, plastic deformation or strain occurs. After the engine is cooled down, the compressive stress changes to local tensile stress which may exceed the tensile yield strength. The repeated large plastic strain amplitude induced by the cyclic temperature loading results in thermo-mechanical LCF failure of the manifold. The high metal temperature affects material strength. Three mechanisms are involved in the following order of importance: fatigue due to plastic deformation, creep, and oxidation. At the full-load high-temperature conditions, the exhaust manifold is subject mainly to compressive loads (i.e., out-of-phase load) where creep has a secondary influence.

Design solutions for exhaust manifold durability

The fundamental design issue in the exhaust manifold is to minimize thermal stresses and thermal deformation in order to avoid large plastic strain amplitudes and fatigue cracks. Exhaust manifold durability is generally a component-level issue, especially for new engines, since proper material selection and local design details can always solve almost all the problems related to the gas temperature. However, the material selection is a trade-off between cost and thermal capability. The exhaust manifold gas temperature of diesel engines is generally lower than that of gasoline engines.

Based on the maximum exhaust manifold gas temperature requirement, the manifold type and material can be properly selected. Better material can improve thermal fatigue life and reduce creep effects. The oxidation problem can often be solved by using a material with higher resistance to oxidation but often at a trade-off against high temperature strength (Gocmez and Deuster, 2009). Exhaust manifold material selection and the properties of high temperature oxidation and thermal fatigue were discussed by Park *et al*. (2005).

Exhaust manifold LCF life is strongly affected by local design details (e.g., shape, assembly constraints) and gas temperature. If all the design solutions are exhausted and the LCF plastic strains still cannot be controlled, the exhaust manifold gas temperature has to be reduced in engine operation. This certainly affects engine performance. Gocmez and Deuster (2009) provided a review of exhaust manifold failure modes, lifetime prediction, and design optimization.

Thermo-mechanical fatigue life prediction of exhaust manifold

Thermal fatigue life testing of the exhaust manifold takes a long time to complete. Predicting thermal fatigue life of the exhaust manifold using simulation is important at both the concept design stage and the detailed design stage. Simulation can help determine the acceptable gas temperature level, identify critical failure locations, and guide component design to remove local structural weakness.

Full structural FEA of exhaust manifold includes the evaluation of bolt force, metal temperature field, stress–strain response, fatigue life, and gasket pressure distribution. The metal temperature distribution is usually the most important boundary condition for LCF analysis. The concern with hardware sizing in engine system design is mostly at the full-load condition where the exhaust manifold gas temperature reaches the highest. A simplified durability analysis may use the highest temperature that is encountered during thermal cycles to evaluate the exhaust manifold fatigue life. It should be noted that the highest exhaust gas temperature does not necessarily cause the highest stress or strain in the exhaust manifold. Sometimes the lowest temperature during a thermal cycle can cause the highest tensile stress. The real question for a system engineer is how high the exhaust temperature can be designed at steady-state full-load conditions so that the corresponding thermal strains produced during both full-load durability validation test and slow varying transient thermal cycles are acceptable. Moreover, advanced exhaust manifold durability and fatigue life modeling should consider all three major mechanisms involved in high temperature LCF: fatigue (due to cyclic plasticity), creep, and oxidation.

Simplified early-stage exhaust manifold thermo-mechanical modeling can be conducted to screen the concept designs by minimizing a thermal stress index parameter (Park *et al*., 2006), which is the ratio of the elastic effective stress to the yield strength. This is a quicker calculation than the full FEA approach.

Exhaust manifold thermo-mechanical fatigue analysis has been investigated by several researchers during the past decade. Watanabe *et al*. (1998) developed an analytical method to use the plastic strain as the evaluation criterion to judge the thermal fatigue life of gasoline engine exhaust manifolds. Their simulation correlated well with the engine durability test results. Lederer *et al*. (2000) provided a detailed description of the analysis methodology for the thermo-mechanical fatigue of diesel engine exhaust manifolds, including thermal boundary calculations, material characteristics, and fatigue life estimation. Mamiya *et al*. (2002) discussed a detailed process of exhaust manifold temperature calculation with heat transfer models for both the inner and outer walls. They used a simplified approach to estimate the LCF life by assuming the dominant factor contributing to the cracks was the large thermal plastic strain range (derived from FEA). The strain concentrated regions were the critical locations subject to failure. They used the classical Manson–Coffin relation to model the fatigue life. Their simplified methodology with acceptable accuracy was validated by comparison with the fatigue life test data. Ahdad and Soare (2002) proposed using an energy-based thermo-mechanical fatigue criterion to analyze exhaust manifold life. Delprete and Rosso (2005) detailed a transient structural simulation (as a function of time) with FEA for exhaust manifold stress–strain analysis. They found that the maximum equivalent stress (the Von Mises stress) in the exhaust manifold occurred at the end of the thermal cycle where the metal temperature falls to the minimum level within the cycle. They concluded that a simplified FEA model was able to simulate the thermo-structural behavior of the exhaust manifold with short computing time and thus was attractive to engine system design.

2.9.3 Engine valve durability

Valvetrain durability issues

Engine valvetrain durability issues include the following: thermo-mechanical fatigue of the valves; mechanical fatigue of the valve spring; fatigue and wear of the cam; valve seat wear; valve stem scuffing and wear; and wear at other interface (e.g., rocker shaft, rocker pad). Many valvetrain durability issues are related to system-level loading or design parameters, for example, exhaust temperature, valve spring force, and valve seating velocity. Higher power density of the engine imposes a higher cylinder pressure load and thermal load on the valves and valve seats. The requirements for lower emissions and fuel consumption demand more aggressive cam lift profiles and the associated better valvetrain motion control by using higher valve spring load. The requirement for extended durability demands the valvetrain design to be more durable. Profitability requires the valvetrain design to be cost effective in material selection and design complexity. All these demands require the system engineer to consider valvetrain durability in the early stage of the diesel engine design.

Valve material and design

Common valve failure modes include the following five phenomena: valve head cracking, valve stem fracture, valve stem sticking and scuffing, valve dishing, and face guttering (Johnson and Galen, 1966). In the past half a century, exhaust valve materials and design have evolved to a great extent, while intake valve materials and design have evolved at a much lesser pace (Schaefer *et al.*, 1997). The changes in modern diesel engine valve design are driven by performance requirements. For example, better engine breathing demands the use of four-valve-head design. Using four valves per cylinder results in smaller size for each valve. Better valvetrain dynamics demands a lighter valvetrain. Valve mass has been minimized by reducing valve stem diameter.

The most important parameters for valve durability are the operating temperature, the imposed stress level, and the corrosive combustion products to which the valve is exposed. The magnitude of the cyclic tensile stress in the valve head is a function of the peak cylinder pressure. Exhaust valve metal temperature is closely dependent upon the exhaust gas temperature and cooling. Modern diesel engines have two performance characteristics related to the exhaust temperature. First, as the EGR rate is increased and NO_x emissions become lower, the exhaust temperature becomes lower (if the power density does not increase). Second, there is very little or no gas scavenging from the intake port directly to the exhaust port due to the need to drive EGR flow with the exhaust manifold pressure higher than the intake manifold pressure. Therefore, compared with the non-EGR engines, the air cooling effect from scavenging fresh air for the exhaust valve is much less in EGR engines. Moreover, note that the intake valve temperature is much cooler than the exhaust valve temperature. This temperature difference results in different valve thermal expansion and valvetrain lash for the intake and exhaust valves.

The temperature distribution inside the valve, and the temperature effect on valve steel properties, thermal/mechanical stresses and valve seat design have been thoroughly researched in the past. The modern engine requirements on valve design have been focused on the high-temperature fatigue strength and the wear resistance to achieve longer life of the valve and enable higher power density of the engine. Valve material selection is a component-level design detail. Sufficient strength of the material is the key to resist high-temperature fatigue. The material selection is a trade-off between performance/durability and cost.

During the process of selecting valve alloy and heat treatment method, the corrosion resistance of the material needs to be considered since corrosion significantly reduces the fatigue strength of the material. Valve corrosion is affected by fuel property. The use of low-sulfur diesel fuel and low-sulfated-ash oil causes less sulfidation corrosion at high temperatures for the intake and the exhaust valves. EGR gas may cause acid-based corrosion for the intake valve.

The durability of internal combustion engine valves has been extensively researched in the past several decades (Newton, 1952; Newton *et al.*, 1953; Tauschek, 1956; Johnson and Galen, 1966; Giles, 1966; Wang, 2007). An important comprehensive review on the heavy-duty diesel engine valve materials and designs in the last fifty years was provided by Schaefer *et al.* (1997). The review covered all major design guidelines for valve durability related to material and design.

Valve seating velocity and valve stem fracture

The durability risk of valve stem fracture due to fatigue is governed by valve seating impulse (i.e., valve velocity multiplied by mass), which is largely controlled by cam profile design. The cam design also affects engine performance and valvetrain dynamics. A cyclic tensile stress in the valve head and the valve stem is generated by the valve seating event. At high-load conditions, due to large thermal distortion of the cylinder head the valve does not seat uniformly along the circumferential direction on the valve seat. This creates additional seating stresses on the valve and the seat, and also creates bending stresses at the stem–retainer interface.

The valve seating velocity is affected primarily by cam acceleration profile and secondarily affected by cam closing ramp design. Valve seating velocity also affects engine NVH. The intake valve seating velocity may become too high if the cam acceleration is too aggressive (for achieving high volumetric efficiency) or poorly designed. The exhaust valve seating velocity may become too high for the following reasons: very aggressive cam acceleration for achieving low pumping loss; very aggressive exhaust cam acceleration at the cam closing side for reducing the high recompression pressure that is built up by a small turbine area used in a high-EGR diesel engine; poorly designed exhaust cam acceleration shape; and poorly designed cam closing ramp.

2.9.4 Cam fatigue and stress

Introduction of cam stress

Most of the valvetrain cam and follower distress can be classified into two main categories: fatigue and wear. Cam contact fatigue is often referred to as spalling (or pitting, flaking). Spalling is a metal failure due to fatigue under high contact stresses with internal cracks spreading up to the surface with the consequent material removal. Spalling of the contact surface can be avoided by strengthening the material or reducing the maximum Hertzian elastic stress. The Hertzian stress is a good indicator of this type of HCF fatigue failure.

Scuffing is the result of micro-welding of the contact surfaces caused by oil film breakdown due to high contact pressures. Scuffing occurs when the contact becomes partially plastic. Severe adhesive wear is commonly known as scuffing in valvetrain. Scuffing and wear are associated with the lubrication condition and the temperature in the contact area.

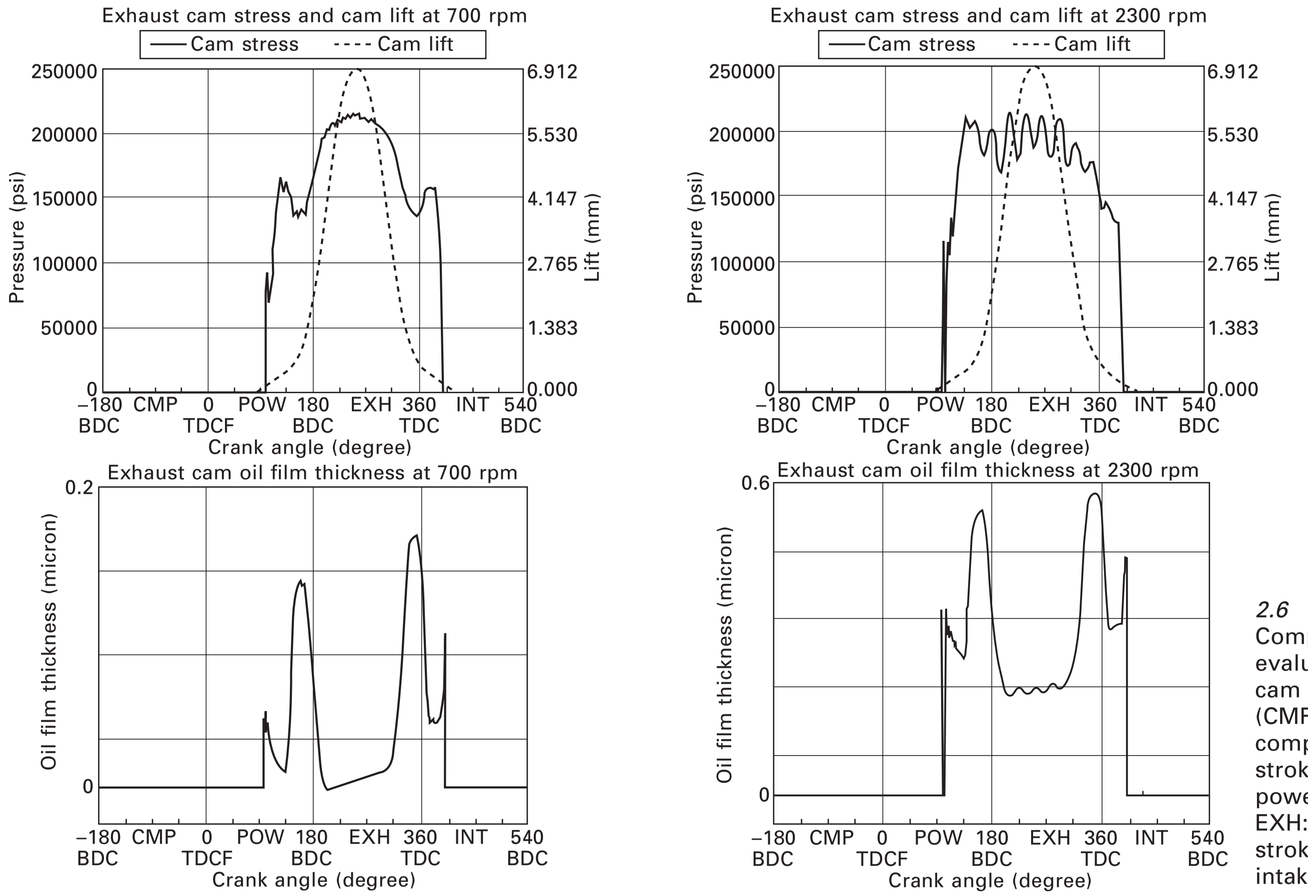
It should be noted that cam stress is not the only parameter to be used to judge the cam–follower durability. The other two important parameters are oil film thickness (to be detailed in Chapter 10) and flash temperature (for flat-faced followers, discussed in Section 2.11). Figure 2.6 shows the calculated cam stress and oil film thickness for a pushrod valvetrain in a heavy-duty diesel engine. It is observed that the oil film thickness is lower at the lower speed (700 rpm). The oil film thickness is low at the cam nose where the cam stress is the highest. The oil film thickness reaches almost zero on the cam opening flank at approximately 40° cam angle before the cam nose where the lubricant entraining velocity is zero. At that location the cam stress remains high.

The state of deformation and the stress existing between two elastic bodies in contact under load were established by Hertz. The Hertzian stress (a compressive stress) is the only principal stress used for cam stress evaluation. Cam stress is an important engine system design parameter which affects the design decisions of rocker arm ratio, cam profile shape, cam acceleration, spring force, cam base circle diameter, roller diameter, etc. Cam stress is determined mainly by cam force and cam radius of curvature (Fig. 2.7). Higher spring force, vibration force, and gas load make the cam stress higher. The Hertzian stress is proportional to the square root of the cam force. At low engine speeds, the cam stress as a function of cam angle has a smooth distribution shape with the peak stress occurring at the cam nose (Fig. 2.8). The cam nose stress may reach the highest value at the engine cranking speed where the dynamic inertia effect is negligible. At high engine speeds or under high gas loading on the valvetrain, the cam stress exhibits vibratory characteristics and may have peak values in the cam flank region.

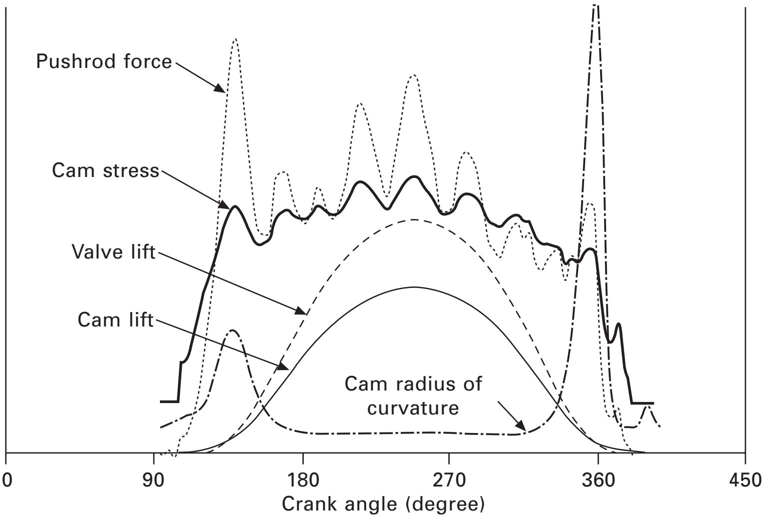
Cam Hertzian stress calculation was discussed in great detail by Turkish (1946), Korte *et al.* (1997, 2000), and Krepulat *et al.* (2002), regarding different materials' cam stress limits and design guidelines.

Cam stress with roller follower

Roller follower has been increasingly used to avoid the problem of excessively high flash temperatures encountered in the sliding contact with flat followers.



2.6 Comprehensive evaluation of cam durability. (CMP: compression stroke; POW: power stroke; EXH: exhaust stroke; INT: intake stroke)



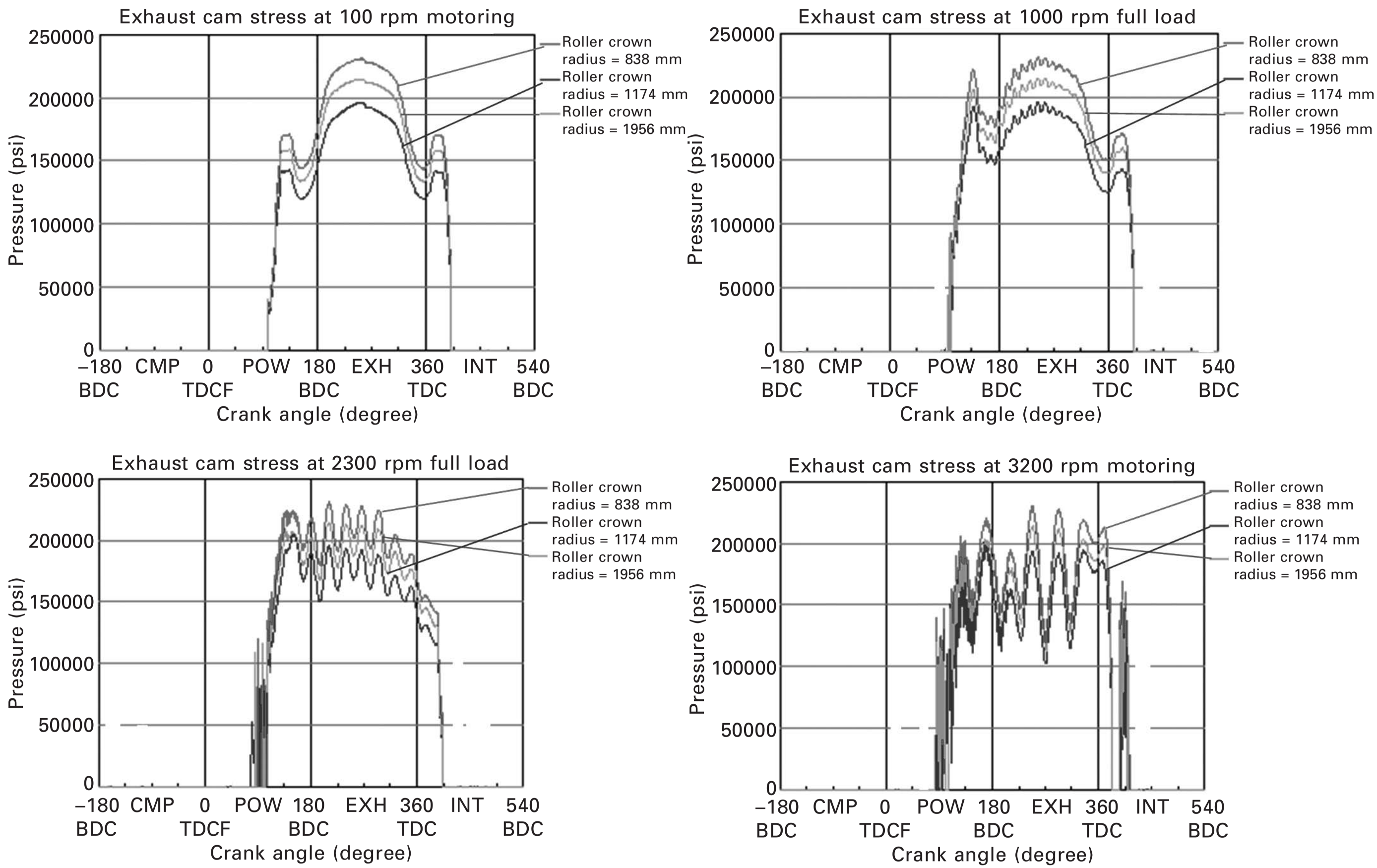
2.7 Illustration of pushrod force and cam stress.

Roller follower has also been used to solve the problem of increased stresses due to higher cam load. A rolling contact design is required when the cam load exceeds the limits of stress and flash temperature for a sliding contact. Roller follower can also greatly reduce valvetrain frictional losses. Moreover, it enables the use of a negative radius of curvature to achieve a more aggressive cam acceleration profile for better engine breathing.

Rolling contact causes an increase in contact stress between the cam and the follower. However, the maximum allowable stress limit of a rolling contact is much greater than the limit of a sliding contact. If a crowned roller is used in order to avoid the edge loading caused by the misalignment between the roller and the cam lobe, the maximum allowable cam stress limit can be further increased. The stress limit is different for various materials of different costs. The limit is also related to the lubrication condition on the contact surface. Different engine companies have established different design criteria of the maximum allowable cam stress limit for their products. The textbook values (e.g., Turkish, 1946) can only be used as a reference guideline.

Smaller cam radius of curvature and smaller roller crown radius (i.e., larger crown height) make the cam stress higher (Fig. 2.8). A larger roller crown radius enlarges the longer axis of the contact ellipse between the cam and the roller. The calculated contact ellipse area should be kept within the width of the cam lobe and should not exceed the roller width by more than 10–20%. Roller follower design and durability details were provided by Korte *et al*. (2000).

An in-depth discussion about the comparison of the cam stress calculations between the analytical Hertzian equation and the more advanced FEA approach



2.8 Cam stress sensitivity to the effect of roller crown radius (CMP: compression stroke; POW: power stroke; EXH: exhaust stroke; INT: intake stroke).

was provided by Krepulat *et al.* (2002). The FEA model can handle more complex calculations that cannot be conducted using the Hertzian equation due to restricted assumptions in the Hertzian equation. Such restrictions may include the edge loading effect, nonlinear and plastic material behavior, complex geometric form of contact bodies such as the logarithmic roller profile (Fujiwara and Kawase, 2007), or measured irregular roller crown profiles. Krepulat *et al.* (2002) also pointed out the pitfalls of incorrectly applying the Hertzian equation of a cylindrical roller to the case of a crowned roller or crowned cam lobe. They showed that the miscalculated cam stress as such could be 30% lower than the correct value.

The skew (tilting) motion of the roller affects the cam elastic deformation and stress. The dynamic skew motion as a function of cam angle was investigated by Ito (2006). He found that the skew motion was affected by the cam profile, the roller profile, the follower–guide misalignment, and the moment of inertia of the follower. Moreover, roller slip may cause wear and durability problems. Roller skew motion and roller slip are two major research topics related to roller follower simulation. Roller slip friction is detailed in Chapter 10.

2.9.5 Engine piston and crankshaft durability

The piston is probably the engine component whose structural durability has been the most extensively investigated. The thermo-mechanical durability of the piston is affected mainly by peak cylinder pressure, thermal loading and piston cooling. The heat transfer rate and piston temperature increase with engine speed and torque. The maximum piston temperature usually occurs at rated power. The effects of oil jet cooling and engine operating condition on piston temperature distribution were investigated by Roehrle (1978) and Thiel *et al.* (2007). The temperature distribution of piston and piston rings without forced cooling for the under-crown surface of the piston was experimentally investigated by Furuhama and Suzuki (1979). The calculation of piston temperature distribution was conducted theoretically by Woschni (1979) and numerically by Wu and Chiu (1986). For more information on diesel engine piston structural calculation, design, fatigue life, material, and durability analysis, the reader is referred to the following: Makartchouk (2002), Munro (1999), Spengler and Young (1986), Myers (1990), Keribar *et al.* (1990), Afonso *et al.* (1991), Myers and Chi (1991), Castleman (1993), Vertin *et al.* (1993), Barnes and Lades (2002), Reichstein *et al.* (2007) and Cha *et al.* (2009).

Crankshaft durability investigation is presented in Law (1984), Henry *et al.* (1992), Park *et al.* (2001), Zoroufi and Fatemi (2005), Williams and Fatemi (2007), Montazersadgh and Fatemi (2007, 2008), Choi and Pan (2009), and Çevik *et al.* (2009).

2.9.6 Turbocharger durability

Turbocharger design is handled primarily by specialized suppliers. Turbocharger durability issues in diesel engine system design are related mainly to LCF in the compressor wheel and HCF in the turbine wheel. Compressor durability is related to the wheel and housing temperatures and rotational shaft speed. The wheel and housing temperatures are dominated by the compressor outlet air temperature. The LCF of the compressor occurs due to alternating stresses caused by the variation in wheel speed when engine speed/load changes. The HCF in the turbine is related to the blade vibration and sometimes resonance at the natural frequencies of the blades. In conditions of very high turbine inlet pressures, strong shockwave pressure pulses are generated at the turbine nozzles or the throat. The gas pulses impinge on the downstream turbine wheel to create excitations that lead to the HCF failure of the wheel. Exhaust manifold gas temperature is usually not a major concern for turbine durability because diesel engines operate with high air–fuel ratio and hence the turbine is exposed to cooler exhaust gas temperatures compared to gasoline engines.

Compared to the wastegated turbine, VGT offers the following benefits: reduced engine pumping loss and fuel consumption, flexibility of regulating EGR rate and air–fuel ratio, faster transient response, and enhanced engine braking. VGT has been gaining in popularity for all different sizes of diesel engines (e.g., from passenger cars to Class-8 trucks) and for different applications (e.g., from on-road to off-road such as agricultural and construction equipment and marine engines). However, VGT generally tends to have more durability issues than fixed geometry or wastegated turbines due to more moving elements used to control the turbine area. The durability issues of VGT usually include the following: thermal distortion, interference or seizure, thermo-mechanical fatigue, creep, wear, and corrosion. Lighter, less costly, and more thermally durable VGT is the direction for future design improvement.

Turbocharger durability discussions were extensively covered by Watson and Janota (1982), Japikse and Baines (1997), Baines (2005), Japikse (1996) and Moustapha *et al.* (2003). The compressor LCF life and the methodology of incorporating lifetime calculation in turbocharger matching were discussed by Engels (2002), Ryder *et al.* (2002) and Christmann *et al.* (2010). The analysis methodology of thermo-mechanical fatigue life prediction for turbine housing was provided by Ahdad and Soare (2002), Längler *et al.* (2010), and Bist *et al.* (2010). Thermo-mechanical analysis of the turbine wheel was presented by Heuer *et al.* (2006). The turbine HCF due to wheel blade vibration was analyzed by Kitson *et al.* (2006), Chen (2006) and Kulkarni *et al.* (2010). The reliability data of the turbochargers used in marine engines were reviewed by Banisoleiman and Rattenbury (2006). VGT reliability was discussed by Furukawa *et al.* (1993).

2.9.7 Durability of diesel oxidation catalyst and particulate filter

Modern diesel engines use aftertreatment devices (e.g., DOC, DPF) to reduce tailpipe emissions. Aftertreatment durability issues include thermal, mechanical, and chemical failures. Thermal degradation includes sintering, alloying, and oxidation. Mechanical failure includes physical breakage, thermal shock, blockage of pore structure, and fouling. Chemical degradation includes catalyst poisoning, inhibition, and life cycle stability. The durability of diesel oxidation catalyst was investigated by Takahashi *et al.* (1995) and Shinozaki *et al.* (1998).

Many durability issues of the DPF are related directly to its regeneration performance. Regeneration increases the thermo-mechanical load and the durability risk on both the base engine and the DPF itself. DPF durability issues include mainly the following:

1. Substrate and washcoat failures caused by excessive thermal stress during DPF runaway regeneration or uncontrolled regeneration.
2. Catalyzed DPF performance degradation caused by catalyst poisoning (e.g., by the sulfur in the diesel fuel or lube oil).
3. Substrate or matting material failure due to excessive vibration.
4. DPF thermal aging.
5. Structural failures of the DPF assembly.

Ambient temperature and in-use duty cycle are important noise factors for DPF durability, for example, the severe use duty cycles in cold climate.

Zhan *et al.* (2006) presented methodologies to control DPF regeneration to preserve durability life. Zhan *et al.* (2007) provided a comprehensive summary of the durability failure modes and durability validation testing methods for DPF. They introduced various durability test procedures for DPF substrate and the entire DPF assembly, including on-engine accelerated DPF thermal aging test cycle, accelerated ash accumulation test, hot and cold vibration tests, stepped vibration test, water quench test, thermal profile test, and on-vehicle DPF durability test. Stroia *et al.* (2008) discussed durability test procedures for a diesel aftertreatment system consisting of DOC, LNT, and CDPF. DPF durability analysis on stress and fatigue life was conducted by Gulati *et al.* (1992), Umehara and Nakasuji (1993) and Kuki *et al.* (2004). Other DPF durability investigations were carried out by Locker *et al.* (2002), Fayard *et al.* (2005) and Sappok *et al.* (2009).

2.10 Heavy-duty diesel engine cylinder liner cavitation

2.10.1 Cylinder liner cavitation failure

The fluid cavitation erosion in wet cylinder liners is another chronic failure mode for heavy-duty diesel engines. It causes pitting and perforation on the liner at the coolant side after several hundred hours of operation via the formation and violent collapse of coolant vapor bubbles. The resulting cavity holes on the liner cause the engine coolant to mix with the lubricant oil hence the engine is severely damaged. The loss of material due to liner cavitation may occur faster than the normal component wear so that the cavitation can be a limiting factor of engine life and reliability.

Liner transverse vibration induces coolant pressure fluctuations in the form of expansion and compression waves. When the local dynamic coolant pressure during the expansion phase of the wave becomes very low, the coolant vapor pressure level may be reached and then vapor bubbles start to form and grow. Note that this condition may happen even if the local liner vibration velocity is zero. In the following wave compression phase the pressure wave causes the bubbles to become unstable. Then the bubbles collapse or implode and release a great amount of energy via micro jets. If the bubble collapses near the liner surface, the intense micro jet can impinge the liner wall at a high speed with a high localized shock pressure wave. The sudden collapse of the bubbles produces intense and impulsive load on the liner in a cyclic nature. When the pressure of the micro jet exceeds the material strength, damage accumulates quickly and finally results in cavitation failure.

2.10.2 Impact factors and design solutions for liner cavitation

Liner cavitation was reviewed by Hercamp (1993) and Demarchi and Windlin (1995). Zhou and Hammitt (1990) summarized the possible mechanisms of liner cavitation. They pointed out that cavitation erosion is different from corrosion since the metal is removed as much larger particles in cavitation than in the case of corrosion. Liner cavitation mechanisms in diesel engines were also explained by Hosny *et al*. (1996) with a six-step process: (1) engine loads; (2) piston dynamics; (3) liner vibration; (4) pressure fluctuation in the coolant jacket; (5) bubble formation and collapse; and (6) liner material surface damage.

The primary influential factors for liner cavitation are summarized as follows. The cavitation is caused by liner vibration, which is excited mainly by piston slap. As a dynamic response to piston slap, the transverse vibration of the liner is excited and this results in dynamic cycles of compression and expansion waves in the coolant. Yonezawa and Kanda (1985) studied piston slap and liner vibration. Demarchi and Windlin (1995) obtained good correlations on the causal relationship between piston slap and liner cavitation by testing large piston clearances (oversize bore) to confirm the liner cavitation in an accelerated manner. Bederaux-Cayne (1996) confirmed that piston pin offset has a large impact on coolant pressure fluctuation and hence liner cavitation. Minimizing piston slap kinetic energy through better designs of piston mass, pin offset, clearance, and skirt stiffness may alleviate liner cavitation. Piston slap is detailed in Chapter 11.

Liner cavitation becomes more severe at higher engine speed and load or higher power density. Hosny *et al*. (1996) concluded that cavitation was severe enough to cause damage only at high-speed and high-load conditions. The reason was partly due to the increased peak cylinder pressure and the associated increased piston slap intensity at those conditions.

Liner vibration is affected by its natural frequency and stiffness. Increasing liner wall thickness and liner stiffness (e.g., via better designs of liner support, sealing rings and dampers) can reduce the transient vibrational surface velocities and the cavitation of the liner. However, excessively thick liner wall affects its heat transfer and thermo-mechanical durability adversely. Increasing water jacket thickness (width) can reduce coolant pressure wave fluctuations by changing the wave dynamics so that the cavitation can be suppressed. Liner surface material property at the coolant side is also important in terms of hardness, tensile strength, cavitation resistance coating, thermal conductivity, and corrosion resistance. The liner surface needs to be smooth in order to avoid the rough spot that can nucleate the bubble formation.

Coolant pressure has a direct impact on cavitation. A higher mean pressure makes the fluctuating dynamic pressure more difficult to reach down to the coolant vapor pressure level to form bubbles. The vapor pressure of the coolant is strongly affected by coolant temperature and properties such as coolant composition, the acoustic velocity of the fluid, coolant additives, and so on. Coolant composition affects the surface tension and coolant viscosity, which in turn affect the bubble growth size and the collapse process. For example, water can aggravate the cavitation compared to water–glycol mixture due to the larger bubble size and more violent bubble collapse of water. It was found that coolant flow rate and velocity did not have a significant impact on cavitation because the dynamic coolant pressure drop created solely by coolant flow velocity is negligible compared to the pressure fluctuations induced by the liner vibration (Hosny *et al*., 1996).

2.10.3 Experimental investigation on liner cavitation

Liner cavitation can be quantified and detected by finding out mass loss in laboratory vibration bench test, tracing liner materials loss in the coolant with radioactive method, or measuring liner outer surface vibration and coolant fluctuating pressure. In addition to the above macro processes related to the vibration and pressure signals, new detection methods were developed with micro processes that are related to bubble dynamics and collapse in order to detect the liner cavitation intensity. Cavitation intensity or severity was characterized by measuring the high frequency acoustic implosion signals of coolant vapor bubbles on the liner surface (Hosny, 1996; Hosny *et al.*, 1996). The fluid cavitation intensity measured by such a method was found to increase as a function of the liner vibration level. It should be noted that the measurement of cavitation noise can be problematic due to the difficulty of isolating the fluid cavitation noise from other noises such as the combustion noise in the same frequency range.

2.10.4 Analytical prediction of liner cavitation

Like other chronic durability problems such as fatigue, wear, and creep, cylinder liner cavitation in heavy-duty diesel engines is an important issue to address in the early stage of engine design. To reduce the time and cost of the endurance test to reveal the cavitation problem, it is desirable to develop advanced simulation tools to quantify the parametric design effects and predict cavitation severity and damage. Liner cavitation is governed mainly by piston slap. The simulation methodology of liner cavitation may largely employ the analysis techniques that are used in the area of piston–liner NVH, specifically in the part of piston-assembly dynamics and liner vibration response (detailed in Chapter 11 on piston slap noise). The unique characteristics of liner cavitation analysis are the coolant dynamic pressure fluctuations and bubble dynamics.

In diesel engine system design, the piston dynamics data including piston secondary motions can be easily generated. Such data, based on certain piston clearance assumptions, reflect the system loading acting on the piston and the liner. These data include the effects of engine speed, load, and cylinder pressure. They can be shared in the simulation models for both NVH and liner cavitation in order to address these performance and durability problems together in the early stage of the design.

The simulation methodology consists of the following five steps for liner cavitation modeling:

1. Piston-assembly dynamics modeling to predict the source of excitation and the impact kinetic energy, and generate a piston impact forcing function vs. time.
2. FEA liner modal analysis or the analysis with a simplified liner dynamics model to predict the vibration velocity of the liner outer surface in a transient and spatial distribution.
3. Coolant flow modeling to predict the dynamic coolant pressure wave

propagation and fluctuation in the coolant jacket, and predict the onset location of cavitation.

4. Bubble dynamics modeling to predict the bubble growth, collapse and the micro jet behavior of liner wall impingement.
5. Liner lifetime prediction for pitting and perforation failures.

The simulation involves engine operating conditions, piston and liner design parameters, coolant jacket design parameters, and coolant properties. Step 4 mentioned above is usually difficult to achieve due to the complexity of bubble dynamics. Steps 2 and 3 can be coupled to account for the coolant–liner interaction effect (i.e., the liner damping effect due to the coolant) in order to make the model more accurate in liner transient response. Step 3 can sometimes be omitted for simplicity and a pre-determined velocity threshold of the liner wall can be used instead. Above the critical threshold, cavitation can be assumed to occur. The threshold is dependent on coolant properties. Obtaining high critical threshold of the liner surface velocity and reducing the actual transverse vibration velocities of the liner are design goals in order to control the liner cavitation problem. A cavitation safety factor has been commonly used as an output in Step 2. The safety factor is defined as the ratio of the critical threshold velocity to the calculated cycle maximum vibration velocity of the liner at each spatial node on the liner.

Katragadda and Bata (1994) investigated the bubble formation and collapse for liner cavitation. Their study covered the theoretical fluid mechanics analysis of the cavitation phenomena in fluid flow and bubble dynamics modeling. Lowe (1990) proposed an analytical simulation of liner cavitation to predict the transient response of the liner. The critical cavitation velocity was reported. Hosny and Young (1993) developed an analytical predictive model of liner cavitation that can be used in the early design stage. The input of the model was the impact force from piston dynamics, and the output of the model was liner vibration and the prediction on pitting. The cavitation intensity map in the engine speed–load domain as proposed by Hosny *et al*. (1996) is a good way to convey the boundary limit curve for the region of cavitation in the engine operating domain. Green and Engelstad (1993) simulated coolant pressure wave propagations and attempted to predict the bubble formation location and the cavitation location. They emphasized the importance of including the coolant flow and pressure wave propagations in the liner FEA vibration model in order to account for the coolant–liner interaction.

2.11 Diesel engine wear

2.11.1 Fundamentals of wear

Wear is a progressive loss of material over time. Although engine wear often appears to be a local component-level issue, it needs to be considered in the early stage of engine system design because the wear life of the component is closely related to system loading. For example, peak cylinder pressure affects the wear life of the valve seat, the piston assembly and the engine bearings. A system engineer needs to consider all of the related parameters that affect engine wear and be proficient with the modeling techniques for predicting the wear life.

The wear of engine components usually undergoes three stages: (1) a running-in period where the rate of change is high; (2) a stabilized normal running period where a steady rate of wear is maintained; and (3) a wear-out failure period where a high rate of wear leads to rapid failures. Generally, there are four common types of wear in the engine: adhesive, abrasive, corrosive, and impact wear. Adhesive wear is the most common form of wear and is sometimes called sliding wear. It occurs at the interfaces with high contact stresses and significant sliding. Oil film thickness is a good indicator of adhesive wear. Adhesive wear occurs in mixed and boundary lubrication regimes where metal surface asperity contact occurs. (Lubrication and friction are elaborated in Chapter 10.) Abrasive wear occurs when a hard rough surface or hard solid particles slide on a softer surface. It may occur in hydrodynamic lubrication due to the particles in the engine oil. The number of particles is dependent mainly on the engine soot emission level and oil aging. In this case, the abrasive wear is three-body wear where the soot particles are free to roll and slide down a surface to scratch and score. Corrosive wear occurs when the component is exposed to the corrosive combustion product such as sulfuric acid (e.g., in valve face guttering). Impact wear occurs at the impact surfaces such as the valve seat. Detailed mechanisms of wear were explained by Kato (2002). A study on engine wear rate was provided by Macian *et al.* (2003) based on oil analysis.

According to the Archard equation (1953), the adhesive wear volume is equal to the product of the contact load at the sliding surfaces, the sliding distance, an adhesive wear coefficient, and the reciprocal of combined hardness of both surfaces, i.e.,

$$V_{w,sl} = \frac{f_w F_n l_{w,sl}}{h_w} \qquad 2.10$$

where $V_{w,sl}$ is the wear volume, f_w is the wear coefficient, F_n is the contact force, $l_{w,sl}$ is the sliding distance, and h_w is the penetration hardness of the softer surface of the two contact surfaces (N/m^2). The Archard equation is commonly used to calculate wear volume in sliding contact.

One key element in wear calculation is to determine the wear coefficient, which lumps many complex effects (e.g., material, metallurgical, lubrication, anti-wear additive) into one coefficient. The adhesive wear coefficients of typical materials were provided by Booser (1997). The value of wear coefficient is dependent on the materials of the two surfaces, their lubrication regime and other environmental factors. In hydrodynamic lubrication there is no or very little wear, therefore, $f_w = 0$. The value of the wear coefficient in mixed lubrication stays between the values of hydrodynamic and boundary lubrication regimes. For simplicity it can be assumed that the wear coefficient approximately follows a linear relationship as a function of the oil film thickness-to-roughness ratio (i.e., the Lambda ratio).

Equation 2.10 was originally developed for adhesive wear. It has also been used successfully to model both abrasive wear (Suh and Sridharan, 1975) and fretting wear (Stower and Rabinowicz, 1973), as pointed out by Lewis and Dwyer-Joyce (2002). Pint and Schock (2000) proposed a generalized total wear coefficient in a similar model form, which includes the contributions of all three types of wear for piston rings (i.e., adhesive, abrasive, and corrosive wear).

Impact wear can be modeled as follows (Fricke and Allen, 1993; also used by Lewis and Dwyer-Joyce, 2002), for an example of valve seating impact wear:

$$V_{w,im} = f_{w,im} n_l E_{k,im}^C = f_{w,im} n_l \left(\frac{1}{2} m v_{im}^2 \right)^C \qquad 2.11$$

where $V_{w,im}$ is the impact wear volume, n_l is the number of loading cycles, E_k is the impact energy per cycle, $f_{w,im}$ and C are empirically determined wear constants. When equation 2.11 is used for valve seating impact, m can be taken as the equivalent valvetrain mass and v_{im} is the valve seating velocity.

2.11.2 Engine bearing wear

Diesel engine bearing failure analysis was reported by McGeehan and Ryason (1999). Traditional assessment of bearing durability focused on the prediction of minimum oil film thickness and maximum oil film pressure. Xu *et al*. (1999) and Ushijima *et al*. (1999) developed a wear model and a fatigue parameter to characterize engine bearing failure mechanisms. In their model, the wear amount was assumed proportional to the intensity of the metal asperity contact and the frequency of sliding motion. Their fatigue parameter was proposed to be proportional to a wear parameter (which corresponds to the density of surface-originating cracks), peak oil film pressure, peak frictional force, and bearing surface area. The fatigue parameter is inversely proportional to an oil film thickness limit, a mean sliding velocity, the time of one loading cycle, material strength, the rotational speed of the journal, and the mean frictional torque.

2.11.3 Piston ring and piston pin wear

A piston ring wear model was briefly introduced by Taylor (1998), which was analogous to the model developed for cams and followers by Colgan and Bell (1989). Pint and Schock (2000) simulated piston ring wear by using a generalized total wear coefficient to account for all the wear mechanisms involved, including adhesive, abrasive, and corrosive wear. The wear mechanisms of piston rings and cylinder liner with the use of high-sulfur diesel fuel were also studied by Takakura *et al*. (2005). They believed that the wear mechanisms with the use of high-sulfur fuel were dominated by a combination of corrosive wear and abrasive wear caused by the formation of aqueous sulfuric acid via acid condensation at the EGR cooler outlet. The wear at the piston pin and the connecting rod small-end bush was studied by Merritt *et al*. (2008).

2.11.4 Cam wear

Cam wear mechanisms

The primary wear mechanism in valvetrain cam is adhesive wear. The central issue in cam wear analysis is to predict the wear location on the cam surface in the design phase. Previous experimental work has confirmed that a cam having higher Hertzian stress and thicker oil film thickness may exhibit less wear than a cam having lower Hertzian stress but thinner oil film thickness. This indicates that cam fatigue (mainly related to Hertzian stress) and wear (mainly related to oil film thickness) have different mechanisms. This also indicates the paramount importance of good lubrication design in cam wear control.

Wear usually occurs around the cam nose and at certain locations on the cam flanks (about 40° from the cam nose). Cam lobes are prone to wear at certain locations of the opening and closing flanks where the elastohydrodynamic lubricating oil film thickness and lubricant oil entraining velocity reach a minimum theoretically (near zero). Note that the entraining velocity characterizes the amount of oil 'pumped' along the surfaces towards the contact. Experimental evidence showed that wear on the cam lobe usually occurs at the locations where the oil film thickness reaches the lowest.

Narasimhan and Larson (1985) presented the fundamentals of wear and a comprehensive review of the valvetrain wear and material performance. Purmer and van den Berg (1985) measured cam wear from cam lobe profiles and suggested that the maximum cam wear occurred at the locations where the elastohydrodynamic lubrication condition became critical. They pointed out that relatively small changes in valvetain geometry and kinematics may significantly change the oil entraining velocity and hence the wear pattern and the wear location on the cam.

Cam wear mechanism was experimentally researched by Ito *et al*. (1998) by using friction and wear measurement. They found that the cam wear was confined across the cam nose within the region between the two locations having the theoretical minimum (zero) oil film thickness which corresponded to theoretical zero entraining velocity. They found that cam wear started from the locations corresponding to the minimum oil film thickness position (about ± 40° cam angle) and propagated toward the cam nose direction. At those two locations, the measured friction coefficient reached the peak value within one cam cycle, indicating severe boundary lubrication. They found that mixed and boundary lubrication regimes predominated around the cam nose, while elastohydrodynamic lubrication occurred on the cam flanks. Their test results revealed that oil deterioration and high temperature of oil supply were the two factors that strongly correlated to cam wear for the flat-faced follower.

It should be noted that in highly loaded engines the exhaust cam force and stress at the timing of exhaust valve opening are very high due to the high gas load acting on the exhaust cam. The gas load affects the Hertzian stress and may also affect the oil film thickness. As a result, the exhaust cam lobe is prone to wear also at the location corresponding to the exhaust valve opening timing.

Ceramic materials have excellent properties of wear and seizure resistance compared to the steel parts under high contact pressure and poor lubrication conditions. Ceramic materials are able to prevent micro-welding in adhesive wear and scuffing. The amount of wear at the cam–follower interface can be significantly reduced by using a ceramic roller to replace the conventional steel roller (Kitamura *et al*., 1997). The wear characteristics of a roller follower in a variable valve timing system was researched by Leonard *et al*. (1995).

Cam wear modeling

The minimum oil film thickness on the cam can be calculated based on the elastohydrodynamic lubrication theory (detailed in Chapter 10). The oil entraining velocity as a function of cam angle can be calculated based on valvetrain kinematics. The entraining velocity largely determines the effectiveness of the elastohydrodynamic lubrication.

Yang *et al*. (1996) developed a lubrication and surface temperature model based on the flash temperature concept for a cam with a flat-faced follower. They indicated that only calculating the oil film thickness and the maximum Hertizain stress is not sufficient to explain or predict the cam wear pattern observed from their experiments. Therefore, cam–follower surface temperature must be calculated. They believed there was a strong relationship between the surface temperature and cam wear because their engine durability test results showed that the wear started approximately from the locations corresponding to the maximum surface temperatures. The calculated cam surface temperature reached multiple maxima (peaks) at the two positions of zero oil film thickness on the cam flanks and at the cam nose. They concluded that cam wear initiated from the locations having the maximum surface temperature and then propagated toward the cam nose direction where the surface temperature remained relatively high. This theory was in line with the observations in their cam durability test results.

Colgan and Bell (1989) proposed an important valvetrain wear model to calculate the wear on the cam and the follower. A simpler valvetrain adhesive wear model was proposed by Coy (1997) based on the Archard wear law in which the wear coefficient was assumed as a linear function of the oil film thickness throughout the mixed lubrication regime.

2.11.5 Cam flash temperature

The method of calculating the flash temperature at cam–follower contact was introduced by Dyson and Naylor (1960). The maximum cam surface temperature strongly correlates to the location of cam wear (Yang *et al*. (1996). Cam flash (or surface) temperature calculation needs to be included in cam design and durability analysis for flat-faced followers. The cam flash temperature is one of the three most important design criteria in cam design: the Hertzian stress, the flash (or surface) temperature, and the oil film thickness. The cam flash temperature usually reaches a maximum at three locations on the cam lobe: the cam nose, and the two maximum eccentricity locations that have zero cam acceleration. Cam flash temperature calculations were also discussed by Ito *et al*. (2001) and Yang *et al*. (1996).

2.11.6 Valve stem wear

The interface between the valve stem and the valve guide and the interface between the valve seat and the seat insert have become more challenging in modern engine design. Lack of lubrication at the valve seat due to tighter lube oil consumption control and using valve stem seals may increase valve seat wear and stem–guide wear. The widely used solution for the valve stem scuffing problem is to use chromium plating on the stem. In fact, many of today's intake and exhaust valve stems are chromium plated to resist stem scuffing. The use of non-guided valve bridge for cost savings in the four-valve head design may increase the side load and bending load acting on the valve stem and hence induce or increase valve stem scuffing (Schaefer *et al*., 1997).

2.11.7 Valve seat wear

Valve seat recession and wear mechanism

Valve seat wear is a combined effect of impact wear and sliding (adhesive) wear that leads to the issues known as seat recession or seat pounding. Valve seat recession is a major problem in modern diesel engines. As pointed out by Schaefer *et al.* (1997), higher contact stress and increased seat sliding motion caused by increased peak cylinder pressures in combination with reduced lubrication at the valve seating surface due to tighter lube oil consumption control will increase the propensity of seat recession. Larger distortion of the cylinder head and valve seat and the corresponding larger misalignment also increase the sliding wear on the valve seat. Both impact and sliding can be important for valve seat wear. The most influential factors on seat wear include valve seating velocity, valve mass, valve seat face angle, valve head stiffness, hardness of seat material, peak cylinder pressure, and the impact wear constant. The wear constant is related to the resistance to impact wear of the seat material. The wear of the exhaust and the intake valve seat and insert was investigated by Dissel *et al.* (1989), Wang *et al.* (1995, 1996), and Lewis and Dwyer-Joyce (2002).

Lewis and Dwyer-Joyce (2002) conducted a comprehensive investigation on the relationship between the engine design parameters and the valve seat wear. They found that diesel engine intake valve recession increased with combustion loading, valve closing velocity, and misalignment of the valve. The seat recession originated from the impact of the valve during seating. The recession also originated from the sliding of the valve against the seat as the valve head deflected and wedged into the seat under the combustion pressure. The valve seating impact created a series of ridges and valleys circumferentially around the axis of the valve seating face. The valve seating impact also created cracking on the seat insert. The ridges and valleys were caused by the deformation or gouging process during the valve impact seating against the seat inserts. They found that the valve recession amount increased as the valve seating velocity and the number of loading cycles increased. Misalignment of the valve relative to the seat reduced the initial contact area between the valve and the seat when valve seating occurred. Therefore, the misalignment increased the deformation of the seat at seating impact. They found valve recession (or wear amount) was approximately proportional to the square of seating velocity for the cast insert material.

Valve seat angle

Valve seat angle is related to both engine breathing performance and valve seat wear durability. Larger valve seat angle can improve valve flow but may cause more lateral sliding movement and seat adhesive wear than smaller (flatter) seat angle under cylinder pressure loading. The intake valve seat wear is more severe than the exhaust valve seat wear partly due to less lubrication available at the interface between the intake valve and seat insert than the exhaust valve. Therefore, the intake seat angle (30° as typical) is usually smaller than the exhaust seat angle (30–45° as typical) in automotive diesel engines. Although using better seat facings or seat inserts may reduce valve seat wear, flattening the seat angle is probably the most cost-effective way to minimize valve seat wear, as pointed out by Schaefer *et al.* (1997).

Valve seat wear modeling

Very few studies exist on valve seat wear modeling in the literature. A valve seat durability model allows evaluation of the effect of cylinder pressure and valve seating velocity on valve seat wear and recession in the early design stage. Such modeling can be useful to reveal the relative or qualitative comparison between design concepts.

Lewis and Dwyer-Joyce (2002) proposed a semi-empirical model to predict intake valve recession and wear. In their model, total valve seat wear is the sum of the impact wear and the sliding wear. Impact wear was related to the valve seating impulse. Sliding wear was related to peak cylinder pressure. Therefore, the total valve seat wear was a function of valve seating velocity, cylinder pressure, and the number of loading cycles. They found that the contribution of impact wear to the total wear was greater than that of sliding wear. However, the relative contributions varied with valve seat materials. For instance, the contribution of sliding wear to the total varied from 30% for a cast seat insert to 2% for a sintered seat insert.

2.11.8 EGR engine component wear

Impact of EGR on engine oil

During the expansion stroke carbon soot is exposed to the lubricant oil in the cylinder. The effects of engine speed, load, air–fuel ratio, EGR rate and blow-by all affect oil soot loading. Oil aging refers to an increase in the soot level, an increase in oil viscosity and a decrease in the total base number (TBN) of the oil. The viscosity increase may be due to the insolubles and the soot (Sasaki *et al.*, 1997; George *et al.*, 2007). The increase in viscosity due to the soot may negate the fuel economy gains from the light-weight oil. The percentage of soot in the oil increases almost linearly with engine operating hours (or vehicle mileage) and is affected by the soot flow rate and the driving cycle. At very high mileage the engine oil may have a 'gel-like' appearance when the soot percentage in the oil is high. Previous theory in the literature indicated that at low load conditions or mixed cycles, EGR has no effect on oil aging and engine wear; but EGR may increase component wear at high-load conditions (Dennis *et al*., 1999). Engine soot flow rate usually increases with engine speed and load. At high speeds and high loads, higher EGR rate generally tends to increase the soot flow rate.

Based on the NO_x–soot trade-off in diesel engines, one may think that the retarded fuel injection timing and increased EGR rate used for NO_x control result in an increase in soot emission, and the soot is transported into the lubricant oil to accelerate oil aging. However, such a trade-off may not hold true when more advanced high-pressure fuel systems and air handling systems are applied. In other words, modern high-EGR diesel engines may achieve low engine-out soot level with high EGR rate via higher-pressure fuel systems and advanced air systems without much degradation in oil soot loading and wear. Engine-out soot control is an important system design topic for engine wear durability. If it is difficult to reduce the engine-out soot level at increased EGR rate, larger oil sump volume, reduced oil change service interval and better oil film thickness in tribological design can be considered in order to accommodate the increased soot level in the oil. Some chemically treated oil filters can keep soot from agglomerating in the oil for soot dispersion so that a long oil drain interval can be maintained. Enhanced oil filters may also help TBN retention.

If proper design measures are not taken to reduce, control, or accommodate soot emission, increasing EGR usually increases the rate of oil aging and hence increase engine wear on the tribological components to a certain extent. In this case, a combination of both high engine load and high EGR rate may increase engine wear. However, at low load or low EGR rate at full load, the usage of EGR does not usually raise concerns about engine wear.

Impact of carbon soot on abrasive wear

Three commonly identified mechanisms of abrasive wear are plowing, cutting, and fragmentation. It was found that the lubricant contaminated with carbon soot resulted in increased wear (Mainwaring, 1997). Experimental work with fundamental specimens was conducted by Green *et al*. (2006) on the effect of soot-contaminated oil on valvetrain component wear. They confirmed that the wear volume increased with the level of carbon black content in the oil due to metal sliding wear, soot particle abrasion, and starvation of lubricant from the contact. Ishiki *et al*. (2000) conducted engine tests to find that the piston frictional force increased with higher EGR rate in the latter half of the compression stroke due to the soot deposited in the top ring groove, while the piston frictional force decreases in the middle of the stroke with higher EGR rate. They found that the piston ring wear essentially was a function of the soot loading in the oil and the soot loading in the combustion chamber rather than simply the EGR rate level. Dennis *et al*. (1999) reviewed the hypothetical wear mechanisms of EGR, including antiwear additive absorption, abrasive wear with carbon soot particles, corrosion/acid induced wear (i.e., corrosion and abrasive wear mechanisms). It was believed that the corrosion-induced wear with EGR was not significant.

Impact of sulfur on corrosive wear

The total base number (TBN) of the oil decreases with oil aging due to acids associated with EGR gas. A high TBN can neutralize the acid effects. Reducing the sulfur content in the fuel can defer the decrease of the TBN. High-sulfur fuel may increase engine wear due to the acid effect. Furuhama *et al*. (1991) found that the low-temperature corrosive wear was not primarily caused by the decrease in the oil TBN. Instead, it was caused by strong acidity of condensed water at the top land space of the piston. Dennis *et al*. (1999) experimentally found that fuel sulfur level had a large effect on power cylinder wear at the piston ring and the liner, and the effect was aggravated when the oil became aged. The researchers at the Musashi Institute of Technology conducted a series of experiments to try to understand the causes of the abnormal wear in diesel engines with EGR. They found that the acid components in the EGR gas may result in ring wear for the given fuel tested (Furuhama *et al*., 1991; Urabe *et al*., 1998; Ishiki *et al*., 2000).

2.12 Exhaust gas recirculation (EGR) cooler durability

EGR is used for NO_x control and sometimes for exhaust temperature control. EGR cooler durability issues usually include fouling, plugging, condensation, boiling, corrosion, and coolant leak.

2.12.1 EGR cooler fouling and plugging

EGR cooler fouling refers to the performance degradation over time in heat transfer capability and gas-side flow restriction due to deposits. EGR cooler plugging refers to a large increase in the gas-side pressure drop caused by severe fouling. EGR cooler deposits are a combination of thermophoretic carbon soot deposition, condensed hydrocarbons (HC), and acids occurring on the cooled surface inside the cooler tubes.

EGR cooler fouling may cause significant deterioration in heat transfer performance, sometimes in the order of 20–30% (Hoard *et al*., 2008). This causes an increase in the intake manifold gas temperature and affects the NO_x emissions. Gas-side pressure drop may increase slightly (e.g., a few percent) or significantly (plugging). When EGR cooler plugging occurs, if the engine control system cannot compensate for the high flow restriction in the EGR cooler, lower EGR flow than desirable results, and the NO_x emissions target cannot be met.

EGR cooler fouling characteristics change over time. The deposits initially build up quickly with a fast deposition rate and then stabilize after 50–200 hours (Hoard *et al.*, 2008). As a result, the rate of change in cooler performance degradation due to fouling is initially quite large. However, the cooler effectiveness may asymptotically approach a stabilized constant value after some time (Zhang *et al.*, 2004). The stabilization is believed due to certain mechanisms related to deposit removal and deposition rate change.

The mechanism of EGR cooler fouling is complex. It depends largely on both design and engine operating conditions such as the amount and composition of soot/HC/acid, condensate, exhaust gas temperature, gas flow velocity, and flow pressure pulsation. Severe EGR cooler fouling and plugging may occur when excessive carbon soot is generated during combustion, or an excessively large amount of HC is produced by low-load operation, misfire or late post-injection used to assist DPF regeneration. A larger amount of condensate makes the cooler deposits worse in fouling since heavy wet soot/HC deposits are worse than dry fluffy soot deposits. The problem of soot and HC fouling may be aggravated by the presence of liquid film in the cooler such as water condensate, leaking coolant, unburnt and condensed HC.

EGR cooler configuration (fin type or tube-in-shell type) affects its fouling characteristics. High gas flow velocity in the cooler tube may reduce the deposition but it generally causes a larger pressure drop. Lower gas flow velocity increases the amount of deposits. The most dangerous condition for soot fouling is at the low speeds/loads. Deposit accumulation is worse at the lower speed or load conditions where the EGR flow rate and gas temperature are lower. At these conditions, it is also more difficult to blow off the deposits with the low gas flow velocity. The higher exhaust gas velocity under some driving conditions may have an abrasive self-cleaning effect to remove the soot deposits. This event can help avoid tube clogging and maintain sufficient performance of heat transfer and pressure drop.

The EGR fouling and plugging problems are complex issues whose solution requires coordination of the efforts from the teams of cooler design, combustion development, emissions calibration, aftertreatment and electronic controls. The selection of EGR cooler fin density is critical for alleviating the fouling problem. The cooler size in terms of cooling capacity and flow restriction needs to be determined based on the stabilized fouled (seasoned) condition rather than the clean condition with a sufficient design margin reserved. Hoard *et al.* (2008) suggested that in order to consider the worst-case service operation condition the EGR coolers need to be over-sized in the order of 30% to obtain the required cooling capacity when fouled. In fact, the specific percentage margin reserved for fouling should be determined based on specific cooler design details and engine operating conditions. A simulation model of cooler soot/HC fouling coupled with engine–vehicle driving cycle analysis is desirable in order to provide guidance on cooler sizing in system design. The simulation model may predict the deterioration of cooler performance. Using DOC and DPF in the EGR system upstream of the EGR cooler may remove a large amount of HC and soot to reduce the cooler deposits. Moreover, the use of EGR can be minimized to reduce the EGR cooler usage in order to alleviate the cooler fouling problem. For example, at cold intake manifold temperatures or in the scenarios that are not frequently encountered during driving (e.g., outside the FTP, SET, or NTE zones) the usage of EGR may be reduced in order to achieve better durability of the EGR cooler.

EGR cooler fouling was reviewed by Hoard *et al.* (2008). The mechanisms of soot deposition and EGR cooler fouling control were experimentally investigated by Zhang *et al.* (2004), Bravo *et al.* (2005, 2007), Zhan *et al.* (2008), Mulenga *et al.* (2009), and Chang *et al.* (2010). EGR cooler fouling models were developed by Abarham *et al.* (2009a, 2009b), Teng and Regner (2009), and Teng (2010). The effects of liquid film, gas and coolant temperatures and Reynolds number on the deposits in coolers were discussed by Lepperhoff and Houben (1993).

2.12.2 EGR cooler boiling

EGR cooler boiling is a failure mode where film boiling occurs in the cooler so that the cooler tube metal temperature becomes unacceptably high. The boiling is the result of a mismatch between a high heat rejection and a low coolant flow rate.

When the cooler tube is hot enough or the coolant boiling temperature is low enough, the liquid coolant that is in contact with the tube instantly vaporizes. This creates a film of vapor. The vapor film barrier between the tube metal surface and the coolant results in poor heat transfer. As a result, the tube metal temperature quickly approaches the gas temperature and causes damage (e.g., tube annealing). At the same time, EGR cooler outlet gas temperature increases rapidly as well. Film boiling occurs in the following situations: (1) the EGR cooler coolant flow rate is too low; (2) the coolant pressure is too low; (3) the coolant temperature is too high; and (4) the heat rejection is too high.

Film boiling is different from nucleate boiling. The latter improves heat transfer and does not result in excessively high metal temperature. The maximum allowable cooler tube metal temperature is restricted by nucleate boiling, beyond which failure occurs. Between nucleate and film boiling there is an unsteady transition boiling condition, where the coolant vapor film forms and collapses in an unsteady manner. The transition boiling can also cause high metal temperature failures.

The theoretical boiling temperature of the coolant can be calculated based on coolant pressure. An increased coolant pressure may increase the boiling temperature threshold. A less restrictive coolant circuit in the EGR cooler may help preserve desirable high boiling threshold (e.g., using less restrictive baffle design in the EGR cooler). Coolant composition also affects the boiling temperature. For example, water as the coolant is more prone to boiling than the 50/50 mixture of water and ethylene glycol. A sufficient margin is required in the EGR cooler design between the theoretical boiling temperature threshold of the coolant and the actual coolant temperature in order to cover product variability and extreme ambient operations in the entire engine speed–load domain.

Film boiling is usually prone to occur at engine full load and low or medium speed conditions, for example, near peak torque, where the heat rejection is relatively high but the coolant flow rate is relatively low. Boiling may occur at both steady-state and transient conditions. In the transient thermal cycles when the engine load is reduced from high load to low load or idle, the thermal inertia in the cooler tubes may cause boiling. Proper protective measures are needed in order to prevent coolant boiling during such transients.

The maximum allowable heat load for a given EGR cooler is the heat rejection at which coolant boiling occurs. The worst case of cooler boiling usually occurs at the conditions of low coolant pressure and excessively high exhaust temperature, for example at high altitude or with high intake or exhaust restriction. The EGR cooler needs to be designed to accommodate the required heat rejection without film boiling. This should be done by supplying a sufficiently high coolant pressure and flow rate under all operating conditions. On the other hand, the coolant flow rate should not be over-designed too high with an excessively large margin with respect to boiling.

2.12.3 EGR cooler corrosion and condensation

Acidic corrosion and condensate

Modern diesel engines use high EGR rate and low EGR gas temperature via EGR cooling to control NO_x. The EGR flow contains corrosive gases having sulfur and nitrogen compounds produced from combustion. Corrosion occurs with the onset of condensation in the EGR cooler.

Fuel sulfur level affects the corrosion in EGR cooler and in engine power cylinder components. Most of the sulfur in the fuel is converted into gaseous SO_2 or absorbed onto the particulate matters. The gaseous SO_2 reacts with oxygen in the exhaust to form SO_3. With cooled EGR these corrosive gases are returned to the intake manifold. If the EGR cooler outlet gas temperature is very low, condensate of highly corrosive compounds may form, especially sulfuric acid. The SO_3 reacts with the water vapor to form sulfuric acid (H_2SO_4). The presence of acid increases the corrosion at lower component temperatures close to the acid dew points. The sulfuric acid in the exhaust gas will condense at near 150°C; and the dew point margin is believed to be the most important variable that affects corrosion (Kass *et al.*, 2005).

The corrosion rate is dependent on the condensation and formation of sulfuric acid, which is a function of fuel sulfur level, EGR rate, dew point temperature margin and ambient humidity. The diesel fuels with high sulfur level (e.g., 350 ppm sulfur) combined with condensation at the EGR cooler outlet produce high corrosion rates in the engine components, while the ultra-low-sulfur diesel fuel (with less than 15 ppm sulfur) does not produce significant corrosion (Kass *et al.*, 2005). Sulfuric corrosion can be minimized by using the ultra-low-sulfur fuel and by limiting the condensation of water to avoid acid condensation in the EGR circuit. Sulfuric acid condensation in EGR cooler was discussed by McKinley (1997), Kreso *et al.* (1998) and Mosburger *et al.* (2008). Acidic condensation for diesel and biodiesel fuels was investigated by Moroz *et al.* (2009).

Water condensate

There are two types of condensate in the engine: acid and water. Condensate may cause three durability problems: in-cylinder lubricant oil dilution by water, acidic corrosion, and aggravation of cooler fouling. Water condensation may occur at any cooler outlet (e.g., charge air cooler, EGR cooler, inter-stage cooler) and eventually in the intake manifold. EGR cooler condensation occurs when the EGR cooler outlet gas temperature is lower than the local condensation temperature. The condensation temperature is affected by the gas pressure and humidity. At higher ambient humidity (higher dew point temperature of water), higher EGR gas pressure or colder EGR cooler outlet gas temperature, it is easier for the water in the EGR gas flow to condense. The condensate rate is equal to the difference between the mass flow rate of the H_2O in EGR and the mass flow rate of saturated water vapor. The H_2O flow rate in EGR can be calculated based on the EGR rate and the H_2O flow in the exhaust gas. The H_2O in the exhaust comes from the water contained in the intake ambient air and the water formed during the combustion process. The saturated vapor flow rate is calculated based on the saturation pressure of the water vapor, which in turn varies with the gas temperature.

For heavy-duty engines, high water condensate rate at the EGR cooler outlet usually occurs in regions of low speed and high load in the engine speed–torque domain (e.g., around peak torque) due to the combination of a relatively high EGR gas flow rate, high EGR gas pressure, and cold EGR cooler outlet gas temperature. At rated power or low load, the water condensate rate is usually lower. The acceptable level of condensate is determined by engine durability testing.

The sulfuric acid vapor condenses at a higher temperature than the water vapor does. EGR gas contains acidic combustion products. Therefore, the EGR water condensate is acidic and can cause corrosion problems. The gas temperature at the EGR cooler outlet and the intake manifold needs to be set sufficiently high for a given fuel sulfur level in order to minimize the durability risk due to water and sulfuric acid condensate. The EGR cooler capacity is finally a design balance between the durability requirement of water/acidic condensation and the emissions requirement of NO_x control.

In summary, appropriate design targets should be selected and matched properly in EGR cooler design, including carbon soot and HC flow rate, gas–coolant temperature gradient, cooler outlet gas temperature, and condensate rate.

2.13 Diesel engine system reliability

2.13.1 Objectives of engine system reliability analysis

Reliability has a great impact on brand image. Good reliability can reduce product warranty costs. The objectives of engine system reliability analysis include the following:

1. use the concept of design for reliability to ensure durability constraints are properly selected in the early stage of diesel engine system design;
2. examine the reliability and lifetime of each subsystem or component, and conduct reliability allocation to optimize the system lifetime and cost.

For the first objective, the concept of design for reliability requires a system reliability target to be specified for a given period of mission time for the population of the engine product to cover the variability. The probabilistic nature of the reliability problems requires a nondeterministic approach to be used in the analysis. System durability models are required in order to estimate the lifetime of the engine or estimate the reliability for the statistical population of the engine at a specified lifetime.

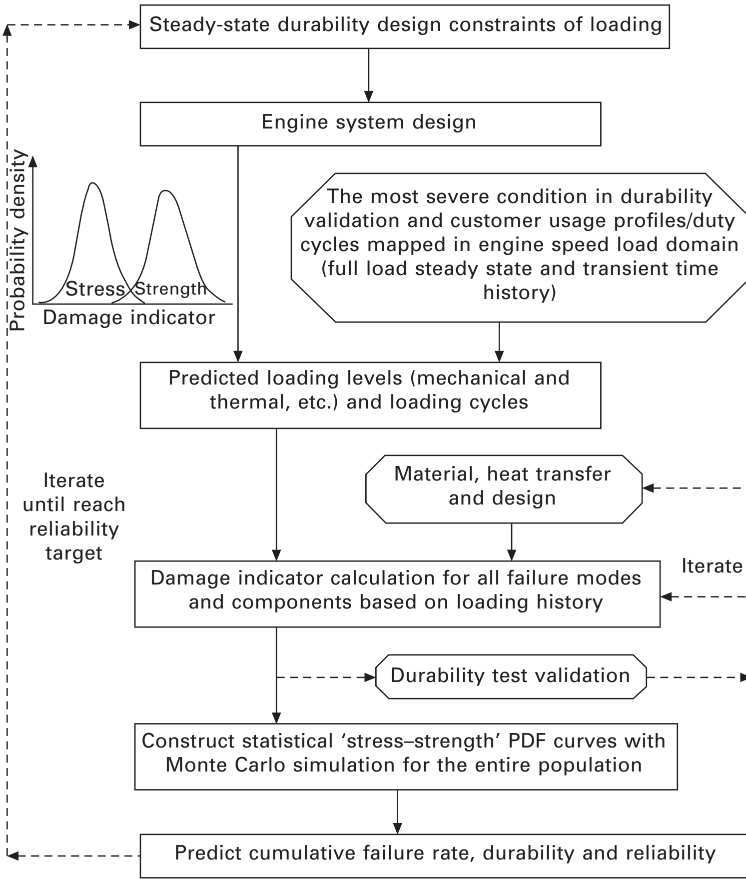
Depending on the maturity of the durability model for each component, the second objective can be realized in either the early stage of system design or a later stage of detailed design via system integration when the component prototypes are available. A system reliability model is needed as a function of component reliability. Reliability allocation needs to be conducted between the components in order to meet the system reliability target while satisfying the minimum requirement on the reliability target for each component. Reliability importance and the impact of reliability improvement on cost need to be evaluated. Nonlinear constrained optimization can be conducted to minimize system cost. If some components are over-designed for lifetime or reliability due to a great mismatch compared with other on-target design components, their lifetime or reliability can be reduced in order to save cost and reach a balanced lifetime between the components. The following sections discuss the key concepts involved in these processes.

2.13.2 Reliability engineering in diesel engine system design

Reliability engineering is a well-developed discipline closely related to statistics and probability theory. There are many areas in reliability engineering, for example: reliability data analysis with the time-domain probabilistic models of reliability, failure rate, and hazard rate by using time as the random variable to address the probability of failure as a function of mission time (e.g., analysis with the Weibull distribution); the stress–strength probabilistic interference model by using design parameters as the random variable; reliability networks of series, parallel, series–parallel and standby systems; FMEA and fault tree analysis; and reliability-based optimization and design. Note that structural durability failure mechanisms are usually not a major subject in reliability engineering. When the principles of reliability engineering are applied to diesel engine system design, the focus is to link durability with reliability and develop probabilistic models of durability life estimation for major engine components. Figure 2.9 illustrates the methodology of design for reliability in engine system design.

The role of reliability in the design process was introduced by Dhillon (1999, 2005), Kumar *et al*. (2006), Dietrich (2008), and Chandrupatla (2009). The probability analysis in reliability engineering was elaborated by Kapur (1992). Kuehnel *et al*. (2005) described a durability validation process in reliability management for cooler development. Zhou and Li (2009) summarized the key concepts in reliability validation processes in engineering design.

Durability and reliability are two different concepts as discussed in Section 2.3. However, the concepts of durability and reliability become consistent when the probability of failure is used. For example, a durability specification of B10 life at one million miles (or equivalent number of engine hours) represents that 10% of the engine population will fail within one million miles. The equivalent reliability specification can be stated as the reliability is 90% or the probability of failure is 10% at one million miles. An engine system reliability model can be built using the probability distribution equation with time as the random variable. The durability or reliability life can be calculated by solving this equation for time, provided the failure



2.9 Methodology of design for reliability in engine system design.

rate or reliability is a known input. It is desirable to include a reliability requirement in the engine system design specification so that reliability is designed-in at the beginning of product development.

2.13.3 Testing methods for reliability assessment

The reliability of a component, subsystem and system can be validated by the methods of 'pass/fail testing' and 'test to failure' (Baker and Brunson, 2000). The pass/fail testing method is also known as 'test to bogey'. Derived from the binomial distribution model, the probability of no failure within a specified test time duration is given by

$$P_{pass} = R^n = 1 - C_{cl} \qquad 2.12$$

where R is reliability, n is the sample size, and C_{cl} is a confidence level which is equal to the probability of one or more failures (Ireson *et al.*, 1996). The sample size required to prove the reliability level is then given by

$$n = \frac{\ln(1 - C_{cl})}{\ln R} \qquad 2.13$$

With the pass/fail testing method the failure distribution and the system life at the end of test time are unknown. Moreover, the sample size required is usually rather large. In contrast, although it takes longer to test, the method of 'test to failure' can use a much smaller sample size (typically reduced by in excess of 50% according to Baker and Brunson, 2000) and offer the information of failure rate as a function of test time so that the probabilistic distributions of the reliability of different systems can be analyzed and compared. For example, the failure rate data can be fitted with a Weibull distribution model.

The 'test to failure' method is destructive in order to obtain the failure rate information. The test is expensive and often still requires a large number of samples in order to gain statistical confidence. The smaller the sample size, the higher the uncertainty. Probabilistic simulation may assist the evaluation.

2.13.4 Probabilistic design for reliability

Time-based and stress- or damage-based probabilistic models

The time-based reliability model is shown below where the Weibull statistical distribution is used as an example (see Table A.2(f) in the Appendix):

$$R(t) = e^{-(t/\beta)^{\gamma}} \qquad 2.14$$

where time t is the random variable, β is the scale parameter and γ is the shape parameter. Note that this model does not contain the detailed design information of why the failure occurs. Such a model can be obtained by using a statistical distribution to fit the reliability warranty/service data, lab test data, or simulation data. For high mileage engines, warranty/service data collection can be very time consuming. As conducted by Daubercies *et al.* (2009), repurchasing customer vehicles to analyze the reliability by part inspection is one way to obtain real world usage data. Appropriate durability testing in engine test cells or field test vehicles, preferably with accelerated testing, may provide a good estimate of the failure rate to substitute service reliability data. The current trend is to use numerical simulation to predict failure rate in the early stage of the design. The simulation model needs to be tuned by using durability test data or reliability service data.

A more detailed reliability model at a specified vehicle mileage or operation time is the classical 'stress–strength interference' probabilistic (nondeterministic) model (Fig. 1.13 in Chapter 1). The stress–strength model was initially applied to fatigue reliability analysis by Freudenthal *et al.* (1966). This model has been widely used to calculate the reliability of a mechanical item when the associated stress and strength probability density functions are known. The stress here is referred to the structural durability load encountered in usage, and the strength is referred to the limit of the ability of the part/material to resist failure. Unlike the time-based model in equation 2.14 where the random variable is time-to-failure, the stress–strength model uses stress and strength as the random variables. The statistical variability caused by noise factors such as product population, hardware tolerance, usage style and environmental condition is represented. The stress–strength model is usually in the following form:

$$\begin{aligned} R &= \int_{-\infty}^{+\infty} \left[\int_{S}^{+\infty} f_{PDF,\,strength}(S)\,\mathrm{d}S \right] f_{PDF,\,stress}(s)\,\mathrm{d}s \\ &= \int_{-\infty}^{+\infty} f_{PDF,\,stress}(x)\, f_{PDF,\,strength}(x)\,\mathrm{d}x \end{aligned} \qquad 2.15$$

where $f_{PDF,stress}(s)$ is the probability density function of the stress, and $f_{PDF,strength}(S)$ is the probability density function of the strength. Note that for the Beta distribution (see Table A.2(c) in the Appendix), the upper limit of the integrals in equation 2.15 may be a finite value (Kececioglu, 2003). The reliability is defined as the probability of stress less than strength. The failure rate or risk of failure is defined as the probability of strength less than stress, essentially reflected by the overlapping area under the probability density function curves of stress and strength. The reliability and the failure rate are related by the following relationship:

$$R = 1 - P_{failure} = 1 - P(S < s) \qquad 2.16$$

where $P(S < s)$ indicates the probability of strength less than stress. The reliability model can be applied to either a component or an entire system.

The degree of stress–strength interference is dependent on time because the durability parameters such as accumulated damage and material strength are related to time. A stress–strength model can be used to calculate reliability values for different specified time as input. Then, the reliability data obtained as such may form a time-based reliability model.

The stress–strength model contains the detailed design information. Therefore, it should be the focus of reliability modeling in engine system design. The probabilistic sensitivity analysis using such a stress–strength model can identify the most important parameters that affect the structural reliability.

The stress–strength interference model can be used at either steady-state full load conditions or in transient driving cycles to evaluate the appropriate design constraints that need to be used in system design such as peak cylinder pressure or exhaust manifold gas temperature. In the transient cycle analysis, after the time history of component temperature is obtained, the time history of stress and strain can be calculated. When the failure can be evaluated based on the load or stress (such as in the case of steady-state HCF problems), stress may be used as the random variable in the stress–strength model. However, in many other cases, the stress or a cycle average load may not be the convenient or appropriate parameter to use. For example, in LCF the dominating failure criterion is usually strain rather than stress. Moreover, the load can come from either normal use or accidental misuse. The misuse may only occur a dozen times at most during the lifetime of the engine so that its load may not be representative enough. Furthermore, in transient usage cycles a cycle-average stress or load also may not be representative or conveniently available. In these cases, a more general damage–strength interference model should be used as below:

$$R = \int_{-\infty}^{+\infty} \left[\int_{S_D}^{+\infty} f_{PDF,strength}(S_D)\,\mathrm{d}S_D \right] f_{PDF,damage}(D)\,\mathrm{d}D \qquad 2.17$$

where D is a damage indicator, and S_D is the strength in terms of critical damage threshold for failure. The concept of damage was introduced in Section 2.8.3, and the damage can refer to any failure. The value of S_D is usually around 1 for a given failure mode. The random variable of damage may represent the cumulative damage due to all failure modes. The failure rate due to all failure modes of the component or the system at a given target mileage or operating hours is equal to the probability that the cumulated damage exceeds the critical damage threshold. The strength distribution for the critical damage threshold can be obtained by durability testing or numerical simulation. Reliability target can be defined for each component and each failure mode in order to analyze them separately. Reliability target can also be defined for the entire engine system and a combination of all the failure modes in order to characterize the overall system behavior.

The reliability analysis can be conducted by using a sampling method such as the Monte Carlo simulation (to be detailed in Chapter 3). The probability of failure can then be estimated by the ratio of the number of failures to the total number of simulation samples. The probability density functions of the stress, damage, and strength can be expressed by the normal, lognormal, Weibull, or beta distributions. The method of using the stress–strength model for reliability prediction based on stress profiles from real world usage was discussed by Mettas (2005).

The probabilistic distribution of strength can be obtained experimentally, for example by observing the failure mode. It may also be estimated with probabilistic numerical simulation. In the case of partial failures in the test (e.g., the crack propagation in the inter-valve bridge of the cylinder head at a initial 5 mm crack length instead of the 10 mm full failure limit), the strength needs to be derived by extrapolating the test result of partial failures to the time for which the component would have failed completely. An example technique of such an extrapolation was demonstrated by Prince *et al.* (2005) and Daubercies *et al.* (2009) by using their statistical model of crack propagation rate.

The analysis of probabilistic distribution of stress or load in engine system design needs to consider the worst case of damage to be encountered in engine durability validation test and customer real world usage. The analysis needs to consider both steady-state full load (e.g., several hundred hours running at rated power) and transient cycles (e.g., thermal cycles). In system design, most durability constraints are specified at full load, such as peak cylinder pressure, exhaust manifold gas temperature and turbocharger speed. The full-load operation affects the maximum pressure and temperature and their occurrence frequency within a transient cycle. However, continuous full-load operation does not necessarily mean more damage than that produced by a transient driving cycle. Depending on the failure mechanism (e.g., LCF or HCF), the contribution of full-load operation to component failure may be different. For example, cylinder head durability is largely determined by the inter-valve bridge temperature history due to LCF mechanism. In this case, the thermal load cycling from low idle to rated power is more contributing to component failure than continuous full load. Two other considerations in selecting the pressure or temperature load profile in reliability analysis are the effects of ambient condition change and the regeneration of diesel particulate filter. For instance, the cylinder gas and exhaust temperature can become hotter at part load when the engine air flow is throttled or post fuel injection is applied to assist DPF regeneration. Morin *et al.* (2005) provided an example of the adverse impact of DPF regeneration on the time history of the cylinder head temperature and the associated thermo-mechanical damage. The mechanical and thermal loads can be calculated using cycle simulation, multi-body dynamics, and sometimes FEA. When transient metal temperature history is required in the reliability analysis, thermal inertia of the component should also be considered. The temperature time history needs to be converted to mechanical stress history of the component in order to compute the damage.

The integral in the stress–strength interference models can be calculated by numerical integration (e.g., the Simpson's rule) or the Monte Carlo simulation (Kececioglu, 2003). If the probability density functions are in normal distribution, the reliability may also be calculated analytically, as elaborated by Kececioglu (2003) and Zhou and Yu (2006).

The reliability value is sensitive to the errors in the probability density function curves and the relative scattering between the stress and strength curves. Moreover, the reliability value is dependent on the selected confidence level and the number of tested samples. The estimate of the confidence interval for the predicted reliability or failure rate is important when the strength distribution is obtained from a small sample of durability test data. The confidence interval can be established by a Monte Carlo simulation of the failure rate. A detailed procedure of determining the confidence level was provided by Kececioglu (2003) and Daubercies *et al.* (2009).

The above interference models are one-dimensional models. Zhou and Yu (2006) developed a two-dimensional stress–strength model by adding another dimension to cover the variability of the stress asymmetric ratio. They applied their model to diesel engine crankshaft reliability analysis. The stress asymmetric ratio was defined as the ratio of the minimum stress to the maximum stress during a loading cycle. They illustrated that the two-dimensional model could calculate reliability values more accurately than the one-dimensional model.

Safety factors and safety margins

The design strength is usually established *a priori* based on past experience. It must be high enough to cover the uncertainties and variability in operating stress or load, material, and manufacturing processes. In the traditional deterministic (i.e., single-valued) design approach, a safety margin is reserved in order to achieve a reliable design. By definition, the safety factor and safety margin are arbitrary multipliers used to ensure the reliability of mechanical items during the design phase (Dhillon, 2005).

The mean (central) safety factor is usually defined as the ratio of mean strength to mean stress. The mean safety factor is a good measure of safety if the probability data spread of the stress and strength is not large (e.g., both having the normal distribution). The safety margin is defined as the safety factor minus one. Another commonly used safety factor is the ultimate safety factor, defined as the ratio of the ultimate strength of the material to the working stress. Moreover, the extreme safety factor is the ratio of the minimum strength to the maximum stress. In the deterministic approach, the maximum stress is obtained by multiplying a factor by the nominal stress, and the minimum strength is obtained by multiplying another factor by the nominal strength. Then the extreme safety factor can be expressed in terms of probabilistic distribution parameters as follows (Kececioglu, 2003):

$$f_{sf,extreme} = \frac{\overline{S} - \mathrm{C}\sigma_{Sh}}{\overline{s} + C\sigma_{ss}} \qquad 2.18$$

where $\overline{S}$ and $\overline{s}$ are the mean values of strength and stress distributions, respectively; σ_{Sh} and σ_{ss} are the standard deviations of the strength and stress distributions, respectively; and C is a value selected by the designer (usually between three and six).

Kececioglu (2003) illustrated that the concept of safety factor can be fallacious because it may not properly reflect the reliability. For instance, when the means of the stress and strength distributions are kept constant but the standard deviations are varied, the overlapping area (i.e., failure rate) of the two distribution curves will change, and this will indicate different reliabilities. However, because the means of the distributions are unchanged, the mean (central) safety factor remains the same and does not reflect the reliability change properly. Moreover, when both means and standard deviations are changed, reliability obviously changes; however, the safety factors may not respond to the change accordingly. This indicates the serious deficiency of the deterministic design approach by using safety factors in reliability design.

A high value of the safety factor does not always ensure the system is safe or reliable. On the other hand, it may cause over-design and excessive cost. As illustrated by Kececioglu (2003) in his keynote summary on the probabilistic design approach of design for reliability in robust engineering, traditional deterministic design is not sufficient. A probabilistic design method needs to be used. A more advanced and economical design approach based on probability or reliability is more realistic for engine performance and durability predictions. The nondeterministic probabilistic design methodology enables the designer to design the component or system to a desired reliability goal at a desired confidence level.

2.13.5 Probabilistic reliability analysis in engine design

The tasks of reliability-based design optimization in diesel engine system design generally include the following:

- correlate engine performance and durability test results with in-service reliability data in order to develop accelerated testing and predict reliability more effectively;
- conduct probabilistic simulation to check design sensitivity on reliability;
- translate a reliability design target to performance and durability attributes by using probability analysis;
- generate durability constraints for engine system design;
- optimize engine system reliability.

If the reliability does not meet the requirement, an optimization on material, design, and load needs to be conducted to either increase the strength or reduce the stress or damage. If the component engineer has to request a change in the system loading parameters (e.g., pressure, temperature, heat rejection), this request is passed to the system engineer to revise the system design specifications. Sometimes the stress can be reduced by the design of another component. For example, a modification in engine coolant passage may increase the heat rejection and reduce the cylinder head metal temperature and the stress. Very often design modifications affect both stress and strength, for instance, modifying the width of the inter-valve bridge area in the cylinder head. When the stress and strength are changed, the probability density function curves are shifted or stretched, and the overlapping area is changed. Reliability analysis is also helpful for determining a wise target used in durability validation testing in order to prevent under-design or over-design.

Reliability simulation for diesel engine design became more active during the last decade. Gale *et al*. (1995) introduced a probabilistic reliability analysis methodology for engine durability. Their analysis had a failure model, which covered crack, fatigue, wear, or other structural issues. They believed that in dynamics and fatigue areas analytical prediction of durability has been well established; however, for thermal and wear-related mechanisms the models still relied on empirical correlations. Their analysis output a cumulative distribution function which represents the probability of failure of the component for a certain lifetime. The analysis predicted the B10 or B50 durability life of the component. The effects of design change and noise factor control (e.g., geometry, surface roughness) on the B10 or B50 life of the component were demonstrated in their analysis. They claimed that the Monte Carlo simulation could obtain the same result but with 500 times longer computing time than their Fast Probability Integration software developed at the Southwest Research Institute.

Bignonnet and Thomas (2001) used the statistical stress–strength model to predict the reliability of automotive components related to fatigue. Niewczas and Koszalka (2002) presented a method for reliability prediction related to engine cylinder liner wear and power cylinder durability. They stated that predicting engine wear reliability with ten-times shortened on-road accelerated test data was possible. Their method also provided a conversion factor between the engine test bed time and the vehicle mileage based on the comparison of wear intensity (e.g., one hour of engine test bed time was reported to be equivalent to 140 km of vehicle mileage).

Aldridge (2003) presented a methodology to predict engine component/system reliability and durability by using the Weibull chart. The method was applied to an electronic diesel fuel injector. The correlations between the component and system reliability requirements were studied. The method was able to predict the reliability and durability affected by engine system performance calibration and manufacturing variations.

Probably the most important engine reliability and durability simulation work to date, at least from the viewpoint of engine system design, is the series of work published by Prince *et al*. (2005), Morin *et al*. (2005) and Daubercies *et al*. (2009). They developed a methodology to estimate and improve the reliability of a diesel engine cylinder head during the design stage. Their work focused on the thermo-mechanical damage of the cylinder head. Their method started with a time history of engine transient speeds and loads during driving cycles obtained from durability test and in-service data. A thermal analysis of component temperature history and a thermo-mechanical stress calculation were conducted. Failure rate was estimated in their damage–strength reliability model. Design iterations were conducted to achieve the reliability target. They used lognormal distributions to represent the stress and strength distributions. They developed simplified surrogate computing models that were derived from the more complex FEA model. The simplified model could simulate the inter-valve bridge stress history of a 10,000–30,000 km vehicle journey in less than one hour of computing time. Their model was validated with the reliability service data of high-mileage vehicles. The methodology provided a major advance from empirical design for reliability to advanced analytical design.

Other engine reliability studies can be found in Pendola *et al.* (2003), Rahman *et al.* (2007), and Kokkolaras *et al.* (2005).

2.13.6 Reliability allocation and system optimization

System reliability

Reliability can be evaluated for each failure mode or the combination of all the failure modes. Reliability can also be evaluated for a subsystem/component or a system. The reliability of each subsystem/component determines the overall reliability of the entire system. If the subsystem/component malfunction leads to the entire system malfunction, the system is classified as a series system in terms of its connectivity. The reliability of a series system, R_S, is equal to the product of the reliabilities of each independent component R_i as follows:

$$R_S(t) = \prod_{i=1}^{n} R_i(t) \qquad 2.19$$

where t is the time-to-failure. When a series system comprises a large number of components, the system reliability may be rather low even if the individual components have high reliabilities. This fact highlights an important guideline for system design – design for simplicity (i.e., using fewer components).

On the other hand, if the entire system can function properly with at least one component functioning properly, the system is classified as a parallel system in terms of its connectivity. In other words, a parallel system loses its function only when all the components fail. The reliability of a parallel system is given as follows:

$$R_S(t) = 1 - \prod_{i=1}^{n} (1 - R_i(t)) \qquad 2.20$$

Engine components (or subsystems) can be classified into two types based on their different effects on diesel engine system performance:

1. The malfunction of the component results in unrecoverable malfunction of the entire engine (e.g., piston crack, skirt scuffing).
2. The malfunction of the component can be compensated by adjusting its own or other components' characteristics. The entire engine system performance is maintained as a result.

The diesel engine is usually considered as a series system for the major components (or subsystems) when subjected to system design. System reliability fundamentals were provided by Tillman *et al.* (1980), Kececioglu (1991), Leemis (1995), O'Connor (2002) and Rausand and Hoyland (2004).

Reliability allocation and impact on cost

Reliability allocation refers to the optimization process on the reliabilities of all or some of the components of a given system in order to meet the target of overall system reliability with minimum cost (Mettas and Savva, 2001). Reliability allocation needs to occur when the estimated or designed system reliability is not sufficient, or when the reliabilities of the components are severely imbalanced, causing a large over-design and waste of lifetime for some components. In diesel engine system design, reliability allocation and optimization is essentially allocating the risks of different design constraints. For example, at rated power there is a trade-off between peak cylinder pressure and exhaust manifold gas temperature. When fuel injection timing is advanced, the peak cylinder pressure may increase and the exhaust temperature may decrease. The pressure and temperature affect the durability of the cylinder head and exhaust manifold, respectively. Properly assessing the component reliability and allocating the reliability can achieve an optimum balance at the system level. Reliability allocation helps to achieve 'design for reliability' from the beginning of an engine program in the system design stage.

Reliability importance ranks the impact of a component's reliability change on the overall system's reliability. It is defined as follows (Leemis, 1995):

$$I_{R,i} = \frac{\partial R_S(t)}{\partial R_i(t)} \qquad 2.21$$

Cost can be modeled as a function of a component's reliability. The overall system cost is the summation of each component's cost. In general, cost increases exponentially with reliability. The trade-off between cost and reliability is managed in reliability allocation and system optimization. There are two types of guidelines for reliability allocation:

1. The components with high reliability importance can be assigned a high

reliability since a high importance indicates the component has a large impact on the overall system reliability.
2. The most costly component can be assigned the lowest increase in reliability.

The topic of reliability allocation for diesel engines can be formulated as a constrained nonlinear optimization problem as follows:

$$\begin{cases} \text{Minimize cost } C_{c,S} = \sum_{i=1}^{n} C_{c,i}(R_i) \\ \text{subject to } \underset{i=1,2,\ldots,n}{R_S} \geq R_{S,target},\ R_{i,min} \leq R_i \leq R_{i,max} \\ \text{given } R_S(t) = \prod_{i=1}^{n} R_i(t) \quad R_i = R_i(t) \end{cases} \qquad 2.22$$

This optimization formulation is to achieve a minimum total system cost $C_{c,S}$ under system reliability constraints. When the cost is minimized, the reliability of each component tends to be reduced to their respective lowest value in order to avoid over-design for lifetime or a large imbalance of lifetime between the components. To solve the above optimization problem, one needs to obtain the time-based probability equation of component reliability as a function of time-to-failure, and the equation for cost as a function of component reliability. Note again that the time-based reliability model (e.g., a Weibull distribution function) usually does not contain engine design parameters. Once the reliability target is allocated, the method to increase the reliability of the engine system is usually to re-design the components.

More detailed discussions on reliability allocation and cost–reliability models are provided by Mettas (2000) and Mettas and Savva (2001). They present the concept of reliability allocation and optimization, as well as a detailed cost model as a function of four parameters: (1) feasibility of increasing a component's reliability; (2) current reliability; (3) minimum required reliability; and (4) maximum achievable reliability of the component.

2.13.7 Engine system durability–reliability optimization model

There are generally two types of calculations for durability and reliability in engine system design:

1. With a given damage–strength interference model, calculate reliability.
2. With a given reliability target value (e.g., R = 99%), determine the corresponding required damage–strength distribution curves. The

system engineer is responsible for optimizing the required system load, strength, and operation time associated with the reliability. The required load directly affects both durability and engine performance. The final design specification may require several iterations in the design process by using the damage–strength model.

In engine system design, the second type of calculation is more frequently encountered. The reliability allocation model in equation 2.22 focuses on the reliability variation with respect to time and the reliability allocation from the system level cascaded down to the component level. Equation 2.22 does not address how to realize the durability design to achieve the component reliability target because the design parameters such as stress, strength, and damage are not explicitly involved. Note that the reliability target is related to two key elements: a failure rate and at a specified lifetime. The reliability value is defined as 'one minus the overlapping area under the damage–strength distribution curves' (equation 2.16).

The earlier damage–strength model shows that stress, strength, or operation time (related to damage accumulation) can all be altered in order to reach a given reliability target. A damage-based durability–reliability optimization model is developed as follows. The model consists of two types of sub-models: a deterministic durability damage model and a nondeterministic reliability damage–strength model. In the model, the system reliability target first needs to be allocated to each component to define the component reliability targets. The damage of a given failure mode for a given component is calculated with a deterministic model by using an initial estimated operation time based on a given strength and stress (load). Such a calculation is repeated with the Monte Carlo simulation to cover all the statistical variations in order to generate the probability density function curve of the damage. Then, the component reliability can be calculated with the damage–strength interference model. If the calculated reliability value does not match the goal value, the estimated operation time needs to be iterated until they match. It is possible that the operation time obtained from this damage–strength model is different from the lifetime calculated by using the time-based reliability model in the form of equation 2.14 due to different assumptions used by these models. In order to make the operation time match with the goal of a desirable reliability/durability lifetime of the engine, optimization is required to change the stress or strength for a given component. The durability–reliability optimization model structure is:

$$\begin{cases} D_{i,HCF} = D_{i,HCF}^{fatigue} + D_{i,HCF}^{creep} + D_{i,HCF}^{oxidation} + \ldots = f(t_{i,HCF}) \\ R_{i,HCF} = \int_{-\infty}^{+\infty} \left[\int_{S_D}^{+\infty} f_{PDF,strength}(S_D, t_{i,HCF}) \mathrm{d}S_D \right] f_{PDF,damage}(D, t_{i,HCF}) \mathrm{d}D \\ D_{i,LCF} = D_{i,LCF}^{fatigue} + D_{i,LCF}^{creep} + D_{i,LCF}^{oxidation} + \ldots = f(t_{i,LCF}) \\ R_{i,LCF} = \int_{-\infty}^{+\infty} \left[\int_{S_D}^{+\infty} f_{PDF,strength}(S_D, t_{i,LCF}) \mathrm{d}S_D \right] f_{PDF,damage}(D, t_{i,LCF}) \mathrm{d}D \\ D_{i,wear} = D_{i,adhesive} + D_{i,abrasive} + D_{i,corrosive} + \ldots = f(t_{i,wear}) \\ R_{i,wear} = \int_{-\infty}^{+\infty} \left[\int_{S_D}^{+\infty} f_{PDF,strength}(S_D, t_{i,wear}) \mathrm{d}S_D \right] f_{PDF,damage}(D, t_{i,wear}) \mathrm{d}D \\ \ldots \\ D_{i,other} = D_{i,1} + D_{i,2} + D_{i,3} + \ldots = f(t_{i,other}) \\ R_{i,other} = \int_{-\infty}^{+\infty} \left[\int_{S_D}^{+\infty} f_{PDF,strength}(S_D, t_{i,other}) \mathrm{d}S_D \right] f_{PDF,damage}(D, t_{i,other}) \mathrm{d}D \\ i = 1,2,3,\ldots, n \end{cases} \quad 2.23$$

where t is the operation time, and i represents a component or subsystem such as cylinder head, piston, exhaust manifold, valvetrain, piston ring, cylinder liner, crankshaft, bearing, connecting rod, compressor, turbine, DPF, etc. The unknowns in equation 2.23 are $t_{i,HCF}$, $t_{i,LCF}$, $t_{i,wear}$, …, $t_{i,other}$. Each component may be subject to multiple failure modes, for example, HCF crack at one location and LCF crack at another location. For each failure mode, there may be multiple failure mechanisms to accumulate the damage in the durability model in 2.23, for example, fatigue, creep, and oxidation. The probability density functions used in the reliability model in 2.23 can be evaluated by the Monte Carlo simulation to account for statistical probabilistic variations. Equation 2.23 can be formulated for each component to cover all the failure modes and failure mechanisms. The operation time of each component for the reliability goal is determined as the minimum of all failure modes:

$$t_i = \min(t_{i,HCF}, t_{i,LCF}, t_{i,wear}, \ldots, t_{i,other}) \quad 2.24$$

The engine system's operation time for the reliability goal is given by:

$$t_S = \min(t_1, t_2, t_3, \ldots, t_n) \quad 2.25$$

The operation time may not meet the desirable reliability lifetime or durability life, $t_{S,goal}$. The durability–reliability design optimization is formulated as:

$$\begin{cases} \text{Minimize component life difference } \sum_{i=1}^{n} (t_i - t_{S,goal})^2 \\ \text{subject to } t_S = t_{S,goal}, \; t_i \geq t_{S,goal} \\ i = 1,2,3, \ldots, n \\ \text{by changing stress or strength in design} \end{cases} \quad 2.26$$

By changing system loading or component strength, the system operation time can be designed on the target requirement. Meanwhile, the lifetime of the over-designed components is reduced by changing the stress or strength (if possible) so that all the components may ideally have the same or a similar lifetime at their own component reliability target with minimum system cost.

In summary, the durability–reliability optimization model covers the following key elements for engine system design: balance between the system and components; balance among operation time, stress, and strength; failure modes; failure mechanisms; damage accumulation; and probabilistic distribution.

2.14 References and bibliography

Abarham M, Hoard J, Assanis D, Styles D, Curtis E W, Ramesh N, Sluder C S and Storey J M E (2009a), 'Numerical modeling and experimental investigations of EGR cooler fouling in a diesel engine', SAE paper 2009-01-1506.

Abarham M, Hoard J, Assanis D, Styles D, Curtis E W, Ramesh N, Sluder C S and Storey J M E (2009b), 'Modeling of thermophoretic soot deposition and hydrocarbon condensation in EGR coolers', SAE paper 2009-01-1939.

Afonso A, Ferran G and Chi F (1991), 'Development of fiber reinforced aluminum alloy for diesel piston applications', SAE paper 910632.

Ahdad F and Soare M A (2002), 'Prediction of duration of life of automotive components under thermomechanical fatigue', IMechE paper C602/020/2002, *IMechE Conference Transactions*, *Seventh International Conference on Turbochargers and Turbocharging*, PEP, Bury St Edmunds and London, pp. 311–320.

Alcraft D A (1984), 'Ensuring the reliability of diesel engine components', in Haddad S D and Watson N (eds), *Design and Applications in Diesel Engineering*, Ellis Horwood Limited, Chichester.

Aldridge D S (2003), 'Component and system life distribution prediction using Weibull and Monte Carlo analysis with reliability demonstration implications for an electronic diesel fuel injector', SAE paper 2003-01-1363.

Archard J F (1953), 'Contact and rubbing of flat surfaces', *Journal of Applied Physics*, 24(8), 981–988.

Baines N C (2005), *Fundamentals of Turbocharging*, Concepts NREC, White River Junction, VT.

Baker J and Brunson D (2000), 'Statistical modelling of engine systems', in Edwards S P, Grove D M and Wynn H P (eds), *Statistics for Engine Optimization*, Professional Engineering Publishing, Suffolk, pp. 61–82.

Banisoleiman K and Rattenbury N (2006), 'Reliability trends, operating issues and acceptance criteria related to exhaust gas turbochargers used in the marine industry – a classification society view', *8th International Conference on Turbochargers and Turbocharging*, Woodhead Publishing Limited, Cambridge, pp. 289–303.

Barlas B, Massinon D, Meyer P, Cailletaud G, Guillot I and Morin G (2006), 'A phenomenological model for fatigue life prediction of highly loaded cylinder heads', SAE paper 2006-01-0542.

Barnes S J and Lades K (2002), 'The evolution of aluminum based piston alloys for direct injection diesel engines', SAE paper 2002-01-0493.

Bederaux-Cayne W S (1996), 'Quantifying parameters relevant to liner cavitation on a diesel engine', SAE paper 960881.

Bignonnet A and Thomas J J (2001), 'Fatigue assessment and reliability in automotive design', SAE paper 2001-01-4061.

Bist S, Kannusamy R, Tayal P and Liang E (2010), 'Thermomechanical fatigue crack growth and failure prediction for turbine housings', *Proceedings of 9th International Conference on Turbochargers and Turbocharging*, May 1–19.

Booser E R (ed.) (1997), *Tribology Data Handbook*, CRC Press, Boca Raton, FL.

Bravo Y, Lázaro J L and García-Bernad J L (2005), 'Study of fouling phenomena on EGR coolers due to soot deposits: development of a representative test method', SAE paper 2005-01-1143.

Bravo Y, Moreno F and Longo O (2007), 'Improved characterization of fouling in cooled EGR systems', SAE paper 2007-01-1257.

Castleman J L (1993), 'Power cylinder design variables and their effects on piston combustion bowl edge stresses', SAE paper 932491.

Çevik G, Tuncal Z and Duran E T (2009), 'A study on the diesel engine crankshaft fatigue performance optimization', SAE paper 2009-01-0261.

Cha S-W, Ha E-J, Lee K-W and Chang H (2009), 'Development of fatigue durability analysis techniques for engine piston using CAE', SAE paper 2009-01-0820.

Chandrupatla T (2009), *Quality and Reliability in Engineering*, Cambridge University Press, New York.

Chang D K, Sobh A, Tjong J S, Styles D and Szente J J (2010), 'Diesel EGR cooler fouling with Ni-Fe-Cr-Al DPF at freeway cruise', SAE paper 2010-01-1955.

Chen H (2006), 'Turbine wheel design for Garrett advanced variable geometry turbines for commercial vehicle applications', *8th International Conference on Turbochargers and Turbocharging*, Woodhead Publishing Limited, Cambridge, pp. 317–327.

Choi K S and Pan J (2009), 'Simulation of stress distributions in crankshaft sections under fillet rolling and bending fatigue tests', *International Journal of Fatigue*, 31, 544–557.

Christmann R, Längler F, Habermehl M, Fonts P-M, Fontvieille L and Moulin P (2010), 'Low-cycle fatigue of turbocharger compressor wheels – online prediction and lifetime extension', *Proceedings of 9th International Conference on Turbochargers and Turbocharging*, May 1–19.

Colgan T and Bell J C (1989), 'A predictive model for wear in automotive valve train systems', SAE paper 892145.

Coy R C (1997), 'Practical applications of lubrication models in engines', *IMechE Proceedings of the First World Tribology Congress on New Directions in Tribology*, September, pp. 197–209.

Daubercies F, Geslin C and Morin G (2009), 'Cylinder head reliability assessment: from numerical method development to validation based on high mileage field data', SAE paper 2009-01-0203.

Davis G D and Christ R J (1996), 'A comparison of engine coolant in an accelerated heavy duty engine cavitation test', SAE paper 960883.

Delprete C and Rosso C (2005), 'Exhaust manifold thermo-structural simulation methodology', SAE paper 2005-01-1076.

Demarchi V and Windlin F L (1995), 'Cavitation erosion on cylinder liners', SAE paper 952250.

Dennis A J, Gardner C P and Taylor D H C (1999), 'The effect of EGR on diesel engine wear', SAE paper 1999-01-0839.

Dhillon B S (1999), *Design Reliability: Fundamentals and Applications*, CRC Press, Boca Raton, FL.

Dhillon B S (2005), 'Reliability in the mechanical design process', in Kutz M (ed.) *Mechanical Engineers' Handbook – Volume 1*, 3rd edn, Wiley, Hoboken, NJ, pp. 1000–1023.

Dietrich D L (2008), 'Reliability from design inception to product retirement', *2008 Proceedings Annual Reliability and Maintainability Symposium*, Las Vegas, NV, January 28–31.

Dissel R V, Barber G C, Larson J M and Narasimhan S L (1989), 'Engine valve seat and insert wear', SAE paper 892146.

Dodson B and Schwab H (2006), *Accelerated Testing: A Practitioner's Guide to Accelerated and Reliability Testing*, SAE International, Warrendale, PA.

Doehrle M D (1978), 'Thermal effects on diesel engine pistons', SAE paper 780781.

Draper J (2008), *Modern Metal Fatigue Analysis*, EMAS Publications, Warrington.

Dyson A and Naylor H (1960), 'Application of the flash temperature concept to cam and tappet wear problems', *Proceedings Institution of Mechanical Engineers (A.D.)*, No. 8, pp. 255–280.

Dystrup J A, Lane W H, Timmons J P and Smith A L (1993), 'The development of a production qualified catalytic converter', SAE paper 930133.

Engles B (2002), 'Lifetime prediction for turbocharger compressor wheels – why use titanium?', IMechE paper C602/037/2002.

Escobar L A and Meeker W Q (2006), 'A review of accelerated test models', *Statistical Science*, 21(4), 552–577.

Evans J L, Vogt N, Thompson J R and Coit D (2005), 'Correlating field requirements to accelerated life testing for vehicle electronics', SAE paper 2005-01-1492.

Fayard J-C, Joubert E and Seguelong T (2005), 'A new active DPF system for 'stop & go' duty cycle vehicles: durability and improvements', SAE paper 2005-01-1754.

Feng M and Chen Y (2004), 'Fixed emitting point method – an accelerated test program to determine fatigue limit of automotive components', SAE paper 2004-01-1646.

Fessler H (1984), 'Deformation and stress analysis of engine components using models', in Haddad S D and Watson N (eds), *Design and Applications in Diesel Engineering*, Ellis Horwood Limited, Chichester, pp. 110–124.

Franke M, Kusters A, Rinkens T, Maassen F and Bruggemann H (2007), 'Mechanical testing – still necessary!', SAE paper 2007-01-1768.

French C C J (1999), 'Thermal loading', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 106–121.

Freudenthal A M, Carrelts M and Shinozuka M (1966), 'The analysis of structural safety', *ASCE*, *Journal of Structure Division*, 92, 267–325.

Fricke R W and Allen C (1993), 'Repetitive impact-wear of steels', *Wear*, 193, 837–847.

Fujiwara H and Kawase T (2007), 'Logarithmic profiles of rollers in roller bearings and optimization of the profiles', *NTN Technical Review*, 75, 140–148.

Furuhama S and Suzuki H (1979), 'Temperature distribution of piston rings and piston in high speed diesel engine', *Bulletin of the JSME*, 22(174), 1788–1795.

Furuhama S, Suzuki M, Ishikawa H and Watanabe T (1991), 'The primary cause for abnormal wear in diesel engines with EGR systems', *Tribology Transactions*, 34(1), 86–92.

Furukawa H, Yamaguchi H, Takagi K and Okita A (1993), 'Reliability on variable geometry turbine turbocharger', SAE paper 930194.

Gale N F (1990), 'Diesel engine cylinder head design: the compromises and the techniques', SAE paper 900133.

Gale N F, Widener S K, Sui P C and Zhang H (1995), 'Analytical engine design methods: a review', SAE paper 950806.

George S, Balla S, Gautam V and Gautam M (2007), 'Effect of diesel soot on lubricant oil viscosity', *Tribology International*, 40, 809–818.

Giles W S (1966), 'Fundamentals of valve design and material selection', SAE paper 660471.

Gocmez T and Deuster U (2009), 'An integral engineering solution for design of exhaust manifolds', SAE paper 2009-01-1229.

Goshorn K D and Krodel A L (1978), 'Accelerated durability test program for diesel truck engines', ASME paper 78-DGP-23.

Green D A, Lewis R and Dwyer-Joyce R S (2006), 'Wear of valve train components due to soot contaminated lubricant', SAE paper 2006-01-1098.

Green G W and Engelstad R L (1993), 'A technique for the analysis of cylinder liner vibrations and cavitation', SAE paper 930582.

Gulati S T, Lambert D W, Hoffman M B and Tuteja A D (1992), 'Thermal durability of a ceramic wall-flow diesel filter for light duty vehicles', SAE paper 920143.

Haddock A K (1984), 'Practical applications of finite elements in the stressing of diesel engine components', in Haddad S D and Watson N (eds), *Design and Applications in Diesel Engineering*, Ellis Horwood Limited, Chichester.

Hamm T, Rebbert M, Ecker H-J and Grafen M (2008), 'Cylinder head design for high peak firing pressures', SAE paper 2008-01-1196.

Henry J-P, Toplosky J and Abramczuk M (1992), 'Crankshaft durability prediction – a new 3-D approach', SAE paper 920087.

Hercamp R D (1993), 'Overview of cavitation corrosion of diesel cylinder liners', *ASTM Special Technical Publication* (1192), pp. 107–127.

Heuer T, Engels B, Heger H and Klein A (2006), 'Thermomechanical analysis of a turbo charger turbine wheel based on CHT-calculations and measurements', *8th International Conference on Turbochargers and Turbocharging*, Woodhead Publishing Limited, Cambridge pp. 235–250.

Heywood J B (1988), *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.

Hoard J, Abarham M, Styles D, Giuliano J M, Sluder C S and Storey J M E (2008), 'Diesel EGR cooler fouling', SAE paper 2008-01-2475.

Hosny D M (1996), 'Real time cavitation detection method', SAE paper 960878.

Hosny D M and Young R W (1993), 'A system approach for the assessment of cavitation corrosion damage of cylinder liners in internal combustion engines', SAE paper 930581.

Hosny D M, Tibbetts D and Luenz R (1996), 'Cavitation intensity measurements for internal combustion engines', SAE paper 960884.

Indig H and Williams T R (1984), 'Exhaust system accelerated durability testing', SAE paper 840503.

Ireson W G, Coombs C F and Moss R Y (1996), *Handbook of Reliability Engineering and Management*, 2nd edn, McGraw-Hill, New York.

Ishiki K, Oshida S, Takiguchi M and Urabe M (2000), 'A study of abnormal wear in power cylinder of diesel engine with EGR – wear mechanism of soot contaminated in lubricating oil', SAE paper 2000-01-0925.

Ito A (2006), 'A study on the measurement and estimation method of skew motion of roller-tappets in an OHV type DI diesel engine', SAE paper 2006-01-3348.

Ito A, Yang L and Negishi H (1998), 'A study on cam wear mechanism with a newly developed friction measurement apparatus', SAE paper 982663.

Ito A, Yang L and Negishi H (2001), 'A study of the measurement of surface temperature at a cam/tappet contact', *Tribotest Journal*, 7(3), 203–218.

Japikse D (1996), *Centrifugal Compressor Design and Performance*, Concepts ETI, Inc., Wilder, VT.

Japikse D and Baines N C (1997), *Introduction to Turbomachinery*, Concepts ETI, Inc. and Oxford University Press, White River Junction, VT.

Johnson V A and Galen C W (1966), 'Diesel exhaust valves', SAE paper 660034.

Junior W B, Raminelli L F and Antonelli J (2005), 'Cumulative fatigue damage in the diesel engines application', SAE paper 2005-01-4110.

Kapur K C (1992), 'Reliability engineering', in Evans D H, *Probability and Its Applications for Engineers*, Marcel Dekker, New York.

Kass M D, Thomas J F, Wilson D, Lewis S A and Sarles A (2005), 'Assessment of corrosivity associated with exhaust gas recirculation in a heavy-duty diesel engine', SAE paper 2005-01-0657.

Kato, K (2002), 'Classification of wear mechanisms/models', *Proc. IMechE*, *Part J: Journal of Engineering Tribology*, 216(6), 349–356.

Katragadda S and Bata R (1994), 'Cavitation problem in heavy duty diesel engines: a literature review', *International Journal of Vehicle Design*, 1(3), 324–346.

Kececioglu D B (1991), *Reliability Engineering Handbook*, *Volume 2*, Prentice-Hall, Englewood Cliffs, NJ.

Kececioglu D B (2003), 'Keynote address: robust engineering design-by-reliability', SAE paper 2003-01-0141.

Keribar R, Morel T and Toaz M W (1990), 'An investigation of structural effects of fiber matrix reinforcement in aluminum diesel pistons', SAE paper 900536.

Kestly M, Popielas F, Grafl D and Weiss A (2000), 'Accelerated testing of multi-layer steel cylinder head gaskets', SAE paper 2000-01-1188.

Kim B, Chang H, Lee K and Kim C (2005), 'Improvement of durability in HSDI diesel cylinder head', SAE paper 2005-01-0655.

Kitamura K, Takebayashi H, Ikeda M and Percoulis H M (1997), 'Development of ceramic cam roller follower for engine application', SAE paper 972774.

Kitson S T, Clay D C, Brown D H, Evans R O, Eastwood D M and Tootill P K (2006), 'Improving analysis capability in order to reduce turbine HCF', *8th International Conference on Turbochargers and Turbocharging*, Woodhead Publishing Limited, Cambridge, pp. 261–271.

Klyatis L M (2010), *Accelerated Reliability and Durability Testing Technology*, Wiley-Blackwell, Oxford.

Klyatis L M and Klyatis E (2006), *Accelerated Quality and Reliability Solutions*, Elsevier Science, Oxford.

Koch F, Maassen F, Deuster U, Loeprecht M and Marckwardt H (1999), 'Low cycle fatigue of aluminum cylinder heads – calculation and measurement of strain under fired operation', SAE paper 1999-01-0645.

Kokkolaras M, Mourelatos Z, Louca L, Filipi Z, Delagrammatikas G, Stefanopoulou A, Papalambros P and Assanis D (2005), 'Design under uncertainty and assessment of performance reliability of a dual-use medium truck with hydraulic-hybrid powertrain and fuel cell auxiliary power unit', SAE paper 2005-01-1396.

Korte V, Barth R, Kirschner R and Schulze J (1997), 'Camshaft/follower-design for different stress behavior in heavy duty diesel engines', SAE paper 972776.

Korte V, Glas T, Lettmann M, Krepulat W and Steinmetz C (2000), 'Cam roller follower design for heavy duty diesel engines', SAE paper 2000-01-0525.
Krepulat W, Dusik M and Korte V (2002), 'Advanced calculation method of the contact stress in roller follower valve train systems', SAE paper 2002-01-0852.
Kreso A M, Johnson J H, Gratz L D, Bagley S T and Leddy D G (1998), 'A study of the vapor- and particle-phase sulfur species in the heavy-duty diesel engine EGR cooler', SAE paper 981423.
Krivoy R, Colvill J K and Parsons D H (1986), 'An investigation of piston ring groove wear in heavy duty diesel engines using statistically designed accelerated testing', *Journal of Materials for Energy Systems*, 8(2), 168–175.
Kuehnel W, Weise S, Krueger U, Sekler H and Stephenson P (2005), 'The validation process – one key element of reliability management', SAE paper 2005-01-1778.
Kuki T, Miyairi Y, Kasai Y, Miyazaki M and Miwa S (2004), 'Study on reliability of wall-flow type diesel particulate filter', SAE paper 2004-01-0959.
Kulkarni A, Tisserant D, Hosny D, Liang E and Krishnaraja N (2010), 'Turbine wheel high cycle fatigue reliability prediction', *Proceedings of 9th International Conference on Turbochargers and Turbocharging*, May 1–19.
Kumar U D, Crocker J, Chitra T and Saranga H (2006), *Reliability and Six Sigma*, Springer Science + Business Media, New York.
Längler F, Scholz A, Aleksanoglu H and Mao T (2010), 'Validation of a phenomenological lifetime estimation approach for application on turbine housings of turbochargers', *Proceedings of 9th International Conference on Turbochargers and Turbocharging*, May 1–19.
Law B (1984), 'Crankshaft loading and bearing performance analysis', in Haddad S D and Watson N (eds), *Design and Applications in Diesel Engineering*, Ellis Horwood Limited, Chichester.
Lederer G, Charkaluk E, Verger L and Constantinescu A (2000), 'Numerical lifetime assessment of engine parts submitted to thermomechanical fatigue, application to exhaust manifold design', SAE paper 2000-01-0789.
Lee K S, Assanis D N, Lee J and Chun K M (1999), 'Measurements and predictions of steady-state and transient stress distributions in diesel engine cylinder head', SAE paper 1999-01-0973.
Leemis L M (1995), *Reliability*, *Probabilistic Models and Statistical Methods*, Prentice-Hall, Englewood Cliffs, NJ.
Leonard H J, Stone C R and Donaldson C P (1995), 'Wear characteristics of a roller follower variable valve timing system', SAE paper 952100.
Lepperhoff G and Houben M (1993), 'Mechanisms of deposit formation in internal combustion engines and heat exchangers', SAE paper 931032.
Lewis R and Dwyer-Joyce R S (2002), 'Wear of diesel engine inlet valves and seat inserts', *Proc. IMechE*, *Part D: Journal of Automobile Engineering*, 216, 205–216.
Locker R J, Gunasekaran N and Sawyer C (2002), 'Diesel particulate filter test methods', SAE paper 2002-01-1009.
Lowe A S (1990), 'An analytical technique for assessing cylinder liner cavitation erosion', SAE paper 900134.
Lowe A S H and Morel T (1992), 'A new generation of tools for accurate thermo-mechanical finite element analyses of engine components', SAE paper 920681.
Maassen F J (2001), 'Simulation of endurance and thermo cycle testing for highly loaded HSDI diesel cylinder heads', SAE paper 2001-01-3226.
Macian V, Tormos B, Olmeda P and Montoro L (2003), 'Analytical approach to wear

rate determination for internal combustion engine condition monitoring based on oil analysis', *Tribology International*, 36, 771–776.
Mainwaring R (1997), 'Soot and wear in heavy duty diesel engines', SAE paper 971631.
Makartchouk A (2002), *Diesel Engine Engineering: Thermodynamics, Dynamics, Design, and Control*, Marcel Dekker, New York.
Mamiya N, Masuda T and Noda Y (2002), 'Thermal fatigue life of exhaust manifolds predicted by simulation', SAE paper 2002-01-0854.
Masuishi M and Endo T (1968), 'Fatigue of metals subjected to varying stress', Japanese Society of Mechanical Engineers.
McGeehan J A and Ryason P R (1999), 'Million mile bearings: lessons from diesel engine bearing failure analysis', SAE paper 1999-01-3576.
McKinley T L (1997), 'Modeling sulfuric acid condensation in diesel engine EGR coolers', SAE paper 970636.
Merritt D, Mian O, Praca M, Wang D and Zhu G (2008), 'Elasto-hydrodynamic lubrication analysis and wear prediction for a connecting rod small-end bush and piston pin interface', SAE paper 2008-36-0068.
Mettas A (2000), 'Reliability allocation and optimization for complex systems', *2000 Proceedings Annual Reliability and Maintainability Symposium*, Los Angeles, CA, January 24–27, 2000.
Mettas A (2005), 'Reliability prediction based on customer usage stress profiles', *2005 Proceedings Annual Reliability and Maintainability Symposium*, Alexandria, VA, January 24–27, 2005.
Mettas A and Savva M (2001), 'System reliability analysis: the advantages of using analytical methods to analyze non-repairable systems', *2001 Proceedings Annual Reliability and Maintainability Symposium*, Philadelphia, PA, January 22–25, 2001.
Montazersadgh F H and Fatemi A (2007), 'Dynamic load and stress analysis of a crankshaft', SAE paper 2007-01-0258.
Montazersadgh F H and Fatemi A (2008), 'Optimization of a forged steel crankshaft subject to dynamic loading', SAE paper 2008-01-0432.
Morin G, Nicouleau-Bourles E, Simon F and Prince O (2005), 'Reliable diesel engine design based on a new numerical method', SAE paper 2005-01-1762.
Moroz S, Bourgoin G, Lujan J M and Pla B (2009), 'Acidic condensation in low pressure EGR systems using diesel and biodiesel fuels', SAE paper 2009-01-2805.
Mosburger M, Fuschetto J, Assanis D, Filipi Z and McKee H (2008), 'Impact of high sulfur military JP-8 fuel on heavy duty diesel engine EGR cooler condensate', SAE paper 2008-01-1081.
Moustapha H, Zelesky M F, Baines N C and Japikse D (2003), *Axial and Radial Turbines*, Concepts NREC, White River Junction, VT.
Mulenga M C, Chang D K, Tjong J S and Styles D (2009), 'Diesel EGR cooler fouling at freeway cruise', SAE paper 2009-01-1840.
Munro R (1999), 'Pistons, rings and liners', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 348–369.
Murphy R W (1982), 'Endurance testing of heavy duty vehicles', SAE paper 820001.
Myers M R (1990), 'Damage accumulation during high temperature testing of a squeeze formed aluminum piston alloy', SAE paper 900132.
Myers M R and Chi F (1991), 'Factors affecting the fatigue performance of metal matrix composites for diesel pistons', SAE paper 910833.
Nagode M and Fajdiga M (2006), 'Temperature-stress-strain trajectory modelling

during thermo-mechanical fatigue', *Fatigue & Fracture of Engineering Materials & Structures*, 29(3), 175–182.

Nagode M and Fajdiga M (2007), 'Coupled elastoplasticity and viscoplasticity under thermomechanical loading', *Fatigue & Fracture of Engineering Materials & Structures*, 30(6), 510–519.

Nagode M and Hack M (2004), 'An online algorithm for temperature influenced fatigue-life estimation: stress-life approach', *International Journal of Fatigue*, 26(2), 163–171.

Nagode M and Zingsheim F (2004), 'An online algorithm for temperature influenced fatigue-life estimation: strain-life approach', *International Journal of Fatigue*, 26(2), 151–161.

Nagode M, Hack M, Dedene L and Fajdiga M (2008), 'Development of reliable fatigue life prediction processes for lightweight exhaust systems', *Proceedings of Sixth International Conference on Low Cycle Fatigue*, DVM Berlin, September 8–12, pp. 663–668.

Nagode M, Hack M and Fajdiga M (2009), 'High cycle thermo-mechanical fatigue: Damage operator approach', *Fatigue & Fracture of Engineering Materials & Structures*, 32(6), 505–514.

Nagode M, Hack M and Fajdiga M (2009a), 'Low cycle thermo-mechanical fatigue: Damage operator approach', *Fatigue & Fracture of Engineering Materials & Structures*, 33(3), 149–160.

Narasimhan S L and Larson J M (1985), 'Valve gear wear and materials', SAE paper 851497.

Nelson W B (2004), *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*, 2nd edn, John Wiley & Sons, New York.

Newton J A (1952), 'Engineering of the poppet valve', SAE Technical Meeting, West Coast Section, November.

Newton J A, Palmer J L and Reddy V C (1953), 'Factors affecting diesel exhaust valve life', SAE paper 530182.

Nicholas T (2006), *High Cycle Fatigue*: *A Mechanics of Materials Perspective*, Elsevier, Oxford.

Niewczas A and Koszalka G (2002), 'Method of accelerated evaluation of automotive engine durability', SAE paper 2002-01-0337.

O'Connor P (2002), *Practical Reliability Engineering*, John Wiley & Sons, Chichester.

Ogarevic V, Whittle B, Lin X and Anderson R (2001), 'Thermal fatigue of automotive components', SAE paper 2001-01-0829.

Park H, Ko Y S and Jung S C (2001), 'Fatigue life analysis of crankshaft at various surface treatments', SAE paper 2001-01-3374.

Park K H, Choi B L, Lee K W, Kim K S and Earmme Y Y (2006), 'Modelling and design of an exhaust manifold under thermomechanical loading', *Proc. IMechE*, *Part D: Journal of Automobile Engineering*, 220, 1755–1764.

Park S H, Kim J M, Kim H J, Ko S J, Park H S and Lim J D (2005), 'Development of a heat resistant cast iron alloy for engine exhaust manifolds', SAE paper 2005-01-1688.

Pendola M, Morin G, Mohamed A and Ragot P (2003), 'Reliability based analysis of exhaust manifold lifetime simulation: an essential tool for risk management in industrial product development', *SIA Congress*, Poissy, France.

Pint S and Schock H J (2000), 'Design and development of a software module for analysis of three dimensional piston ring wear', SAE paper 2000-01-0920.

Prince O, Morin G and Jouzeau C (2005), 'Validation test optimization based on a statistical approach for diesel engine cylinder heads reliability', SAE paper 2005-01-1780.

Purmer P D and van den Berg W (1985), 'Measurement of camshaft wear – wear and kinematics of overhead camshafts', SAE paper 850442.

Rahman S and Sun R (2003), 'Robust engineering of engine cooling system', SAE paper 2003-01-0149.

Rahman S, Kayupov M, Li J and Mourelatos Z P (2007), 'Balance between reliability and robustness in engine cooling system optimal design', SAE paper 2007-01-0594.

Rausand M and Hoyland A (2004), *System Reliability Theory: Models, Statistical Methods, and Applications*, John Wiley & Sons, Hoboken, NJ.

Reichstein S, Weiss R, Kenningley S, Lades K, Konrad P and Doernenburg F (2007), 'High-performance cast aluminum pistons for highly efficient diesel engines', SAE paper 2007-01-1438.

Rie K-T and Portella P D (eds) (1998), *Low Cycle Fatigue and Elasto-Plastic Behavior of Materials*, Elsevier, Oxford.

Roehrle M D (1978) 'Thermal effects on diesel engine pistons', SAE paper 780781.

Rosa U, Nagode M and Fajdiga M (2007), 'Strain-life approach in thermo-mechanical fatigue evaluation of complex structures', *Fatigue & Fracture of Engineering Materials & Structures*, 30(9), 808–822.

Ryder O, McKenzie D J A and Johnson S (2002), 'Turbo matching techniques for highly cyclic bus applications', IMechE paper C602/006/2002.

SAE Surface Vehicle Recommended Practice J450 (2002), 'Use of terms yield strength and yield point'.

SAE Surface Vehicle Information Report J965 (1966), 'Abrasive wear'.

SAE Surface Vehicle Information Report J1099 (2002), 'Technical report on low cycle fatigue properties – ferrous and non-ferrous materials'.

SAE Surface Vehicle Information Report J2816 (2009), 'Guide for reliability analysis using the physics-of-failure process'.

SAE Surface Vehicle Aerospace Recommended Practice JA1000-1 (1999), 'Reliability program standard implementation guide'.

SAE Surface Vehicle Aerospace Standard JA1000 (1998), 'Reliability program standard'.

Sappok A, Santiago M, Vianna T and Wong V W (2009), 'Characteristics and effects of ash accumulation on diesel particulate filter performance: rapidly aged and field aged results', SAE paper 2009-01-1086.

Sasaki M, Kishi Y, Hyuga T, Okazaki K, Tanaka M and Kurihara I (1997), 'The effect of EGR on diesel engine oil, and its countermeasures', SAE paper 971695.

Schaefer S K, Larson J M, Jenkins L F and Wang Y (1997), 'Evolution of heavy duty engine valves – materials and design', in Bolton H A and Larson J M (eds), *Proceedings of the International Symposium on Valvetrain System Design and Materials*, April 14–15, 1997, Dearborn, MI, pp. 129–139.

Seruga D, Nagode M, Rosa U and Fajdiga M (2009), 'Creep damage assessment for thermo mechanical fatigue', *Proceedings of Second International Conference on Material and Component Performance under Variable Amplitude Loading*, March 23–26, pp. 487–493.

Shinozaki R, Kishi Y and Tohno H (1998), 'The durability of the evaluation methods for diesel oxidation catalysts', SAE paper 982802.

Spengler W G and Young W B (1986), 'Techniques to upgrade heavy duty aluminum pistons', SAE paper 860162.

Stower I F and Rabinowicz E (1973), 'The mechanism of fretting wear', *Transactions of the ASME*, *Journal of Lubrication Technology*, 95, 65–70.

Stroia B J, Currier N W, Li J, England R D, Bush J W and Hess H (2008), 'Critical performance and durability parameters of an integrated aftertreatment system used to meet 2007 Tier II emission standards', SAE paper 2008-01-0769.

Swanger L, Harris D, Johnston P and Derbalian G (1986), 'Advanced methods for diesel component life prediction', SAE paper 860885.

Su X, Zubeck M, Lasecki J, Engler-Pinto Jr. C C, Tang C, Sehitoglu H and Allison J (2002), 'Thermal fatigue analysis of cast aluminum cylinder heads', SAE paper 2002-01-0657.

Suh N P and Sridharan P (1975), 'Relationship between the coefficient of friction and the wear rate of materials', *Wear*, 34, 291–299.

Takahashi Y, Kishi Y, Fukano I, Sasaki K and Shinozaki R (1995), 'A study of the durability of diesel oxidation catalysts', SAE paper 952650.

Takakura T, Ishikawa Y and Ito K (2005), 'The wear mechanism of piston rings and cylinder liners under cooled-EGR condition and the development of surface treatment technology for effective wear reduction', SAE paper 2005-01-1655.

Tauschek M J (1956), 'Basic factors affecting engine valves', SAE paper 560323.

Taylor C M (1998), 'Automobile engine tribology – design considerations for efficiency and durability', *Wear*, 221, 1–8.

Teng H (2010), 'A semi-empirical model for predicting pressure drops of fouled EGR coolers', SAE paper 2010-01-1948.

Teng H and Regner G (2009), 'Particulate fouling in EGR coolers', SAE paper 2009-01-2877.

Thiel N, Weimar H-J, Kamp H and Windisch H (2007), 'Advanced piston cooling efficiency: a comparison of different new gallery cooling concepts', SAE paper 2007-01-1441.

Tillman F A, Hwang C L and Kuo W (1980), *Optimization of Systems Reliability*, Marcel Dekker, New York.

Turkish M C (1946), *Valve Gear Design: A Handbook for Designers and Engineers to Aid in the Design of Cams, Tappets, and Springs for the Valve Gear of Internal Combustion Engines*, Eaton Manufacturing Company, Detroit, MI.

Umehara K and Nakasuji Y (1993), 'Lifetime prediction of wall-flow type diesel particulate filters using fatigue characteristics', SAE paper 930128.

Urabe M, Tomomatsu T, Ishiki K, Takiguchi M and Someya T (1998), 'Variation of piston friction force and ring lubrication condition in a diesel engine with EGR', SAE paper 982660.

Ushijima K, Aoyama S, Kitahara K, Okamoto Y, Jones G J and Xu H (1999), 'A study on engine bearing wear and fatigue using EHL analysis and experimental analysis', SAE paper 1999-01-1514.

Vertin K D, Keribar R and Ganapathy V (1993), 'Acceleration of piston durability testing in natural gas engines', SAE paper 930275.

Wang Y (2007), *Introduction to Engine Valvetrains*, SAE International, Warrendale, PA.

Wang Y S, Schaefer S K, Bennett C and Barber G C (1995), 'Wear mechanisms of valve seat and insert in heavy duty diesel engine', SAE paper 952476.

Wang Y S, Narasimhan S, Larson J M, Larson J E and Barber G C (1996), 'The effect of operating conditions on heavy duty engine valve seat wear', *Wear*, 201, pp. 15–25.

Watanabe Y, Shiratani K, Iwanaga S and Nishino K (1998), 'Thermal fatigue life prediction for stainless steel exhaust manifold', SAE paper 980841.

Watson N and Janota M S (1982), *Turbocharging the Internal Combustion Engine*, Macmillan, London.

Williams J and Fatemi A (2007), 'Fatigue performance of forged steel and ductile cast iron crankshafts', SAE paper 2007-01-1001.

Woschni G (1979), 'Prediction of thermal loading of supercharged diesel engines', SAE paper 790821.

Wu H and Chiu C (1986), 'A study of temperature distribution in a diesel piston – comparison of analytical and experimental results', SAE paper 861278.

Xu H, Jones G J, Aoyama S, Ushijima K, Okamoto Y and Kitahara K (1999), 'Simulation of bearing wear and its influence upon bearing performance based on elastohydrodynamic analysis', SAE paper 1999-01-1522.

Yamagata H (2005), *The Science and Technology of Materials in Automotive Engines*, Woodhead Publishing, Cambridge.

Yang L, Ito A and Negishi H (1996), 'A valve train friction and lubrication analysis model and its application in a cam/tappet wear study', SAE paper 962030.

Yonezawa T and Kanda H (1985), 'Analysis of cavitation erosion on cylinder liner and cylinder block', SAE paper 850401.

Yu M-H (2002), 'Advances in strength theories for materials under complex stress state in the 20th century', *Applied Mechanics Review*, 55 (3), 169–218.

Zhan R, Huang Y and Khair M (2006), 'Methodologies to control DPF regenerations', SAE paper 2006-01-1090.

Zhan R, Eakle S, Spreen K, Li C G and Mao F F (2007), 'Validation method for diesel particulate filter durability', SAE paper 2007-01-4086.

Zhan R, Eakle S T, Miller J W and Anthony J W (2008), 'EGR system fouling control', SAE paper 2008-01-0066.

Zhang R, Charles F, Ewing D, Chang J-S and Cotton J S (2004), 'Effect of diesel soot deposition on the performance of exhaust gas recirculation cooling devices', SAE paper 2004-01-0122.

Zhou J and Li D (2009), 'Reliability verification: plan, execution, and analysis', SAE paper 2009-01-0561.

Zhou X and Yu X-L (2006), 'Reliability analysis of diesel engine crankshaft based on 2D stress strength interference model', *Journal of Zhejiang University Science A*, 7(3), 391–397.

Zhou Y-K and Hammitt F G (1990), 'Cavitation damage of diesel engine wet-cylinder liners', SAE paper 900437.

Zhuang W Z and Swansson N S (1998), 'Thermo-mechanical fatigue life prediction: a critical review', DSTO Aeronautical and Maritime Research Laboratory report, DSTO-TR-0609, Australia.

Zieher F, Langmayr F, Jelatancev A and Wieser K (2005), 'Thermal mechanical fatigue simulation of cast iron cylinder heads', SAE paper 2005-01-0796.

Zinner K (1971), 'Power increase and reliability of diesel engines', ASME paper 71-DGP-11.

Zoroufi M and Fatemi A (2005), 'A literature review on durability evaluation of crankshafts including comparisons of competing manufacturing processes and cost analysis', *Proceedings of the 26th Forging Industry Technical Conference*, Chicago, IL, November.

3
Optimization techniques in diesel engine system design

Abstract: This chapter discusses the fundamental concepts, advanced techniques, and theory in diesel engine system optimization to address the complex interactions between engine subsystems, from single- and multi-objective optimizations to deterministic and nondeterministic probabilistic optimizations. It begins with an overview of system optimization theory by introducing design of experiments (DoE), response surface methodology (RSM) and the Monte Carlo simulation to address the optimizations in design for target, design for variability and design for reliability. It then elaborates the theory of using RSM, advanced DoE optimization, and Monte Carlo simulation in engine system design. Finally, it details the optimizations of robust design for variability and reliability.

Key words: system optimization, single- and multi-objective optimizations, deterministic and nondeterministic probabilistic optimizations, design of experiments (DoE), response surface methodology (RSM), Monte Carlo simulation, Taguchi method, variability- and reliability-based optimizations, robust design, statistical probability distribution.

3.1 Overview of system optimization theory

3.1.1 Introduction to optimization

Optimization is a process to search for solutions to reach the best objective under certain constraints. Optimization techniques have been widely used in all engineering areas to select the best design. Assume the total number of factors is k. A single-objective optimization topic can be described as follows.

Select factors from a k-dimensional vector $X = [X_1, X_2, \ldots, X_k]^T$ to minimize the objective function $f(X) = f(X_1, X_2, \ldots, X_k)$, subject to the inequality constraint functions $g_j(X) \le 0$, $j = 1, 2, \ldots, m$, and the equality constraint functions $m_u(X) = 0$, $u = 1, 2, \ldots, p$ ($p < k$) within the factor range $X_i^L \le X_i \le X_i^U$, $i = 1, 2, \ldots, k$, where X_i^L and X_i^U represent the lower and upper limits of the factor X_i, respectively.

Factors, objective functions, and constraints are three key elements in any optimization topic. The k-dimensional space formed by the k factors is called factor space or design space. When $k > 3$, the design space cannot be graphically displayed. The larger the number of factors, the more complex the optimization topic.

The objective and constraint functions are mathematical models that describe the relationships between the input factors and the output responses. The function can be as simple as an explicit formula (e.g., engine cylinder heat transfer area as a function of bore and stroke) or as complex as a surface-fit formula obtained by the regression of implicit numerical simulation results (e.g., BSFC as a function of fuel injection timing and turbine area as in the case of engine cycle simulation).

Usually, minimizing an objective function is equivalent to maximizing the negative of the function. The objective can be either a single objective function or multi-objective functions, or a weighted combination of multiple objective functions in the form of

$$f(X) = w_1 \cdot f_1(X_1, X_2, \ldots, X_k) + w_2 \cdot f_2(X_1, X_2, \ldots, X_k) + \ldots + w_q \cdot f_q(X_1, X_2, \ldots, X_k) \quad 3.1$$

where $w_1, w_2, \ldots, w_q$ are weighting factors ($0 < w_i < 1$) and $\sum_{i=1}^{q} w_i = 1$.

The inequality and equality constraints are optional in the formulation of an optimization topic. The number of equality constraints should be less than the number of factors or the number of unknowns in the equality constraint equations. Otherwise, all the factors will have a single set of deterministic solutions. Then the optimization topic will become meaningless. However, there is no limitation on the number of inequality constraints.

If all the objectives and constraints are linear functions, it is called a linear optimization topic. Otherwise, it is a nonlinear one. The optimization topics in diesel engine system design are usually multi-factor and multi-objective nonlinear optimization. However, in many cases the topic is simplified as a single-objective optimization to minimize BSFC by converting other objectives to constraints.

The objective function may have multiple local optima (i.e., minimum or maximum), especially when the number of factors is large and the objective function is complex. The optimization seeks the global optima within the design space or solution domain enclosed by the factor range and design constraints (including the boundary). This can often be achieved by trying many different and distant initial values of the factors during the optimization search. In particular, it is worth noting that genetic algorithms have been largely used in engine optimization recently as a robust method can avoid premature convergence to a local minimum or maximum of the objective function. Details on the advantages and drawbacks of genetic algorithms are provided by Parmee and Watson (2000), Thiel *et al.* (2002), and Zottin *et al.* (2008).

The optimization problem can be solved graphically or analytically. A graphical example is to use a spreadsheet to plot the result and visually select the optimum. An analytical example is to compute the gradient of the objective function and search along the direction of the maximum gradient in order to find the optima. Detailed elaborations on the solution algorithms in optimization are provided in numerous books and they are beyond the scope of this book. The reader is referred to Siddall (1982), Montgomery (1991), Myers and Montgomery (2002), Eriksson *et al*. (1999) and Edwards *et al*. (2000) for more information. A comprehensive summary of the statistics for engine optimization is provided by Edwards *et al*. (2000).

3.1.2 Design of experiments (DoE), response surface methodology (RSM), and Monte Carlo simulation

Each factor in a design of experiments (DoE) is an input parameter. It can be a control factor or a noise factor. It can be any design parameter (e.g., turbine area), calibration parameter (e.g., EGR valve opening) or attribute parameter (e.g., intake manifold pressure). Each factor has two or more discrete deterministic levels of value in the DoE. If each factor is viewed as one 'dimension', multiple factors form a multi-dimensional factor space or design space in the DoE. Each DoE run is a combination of factor levels or a point in the factor space. The full list of all the DoE runs is called a design matrix (or array). Each response in a DoE is an output parameter (e.g., BSFC, EGR rate, stress). The factors have their main effect and interaction effect on a given response. The interaction effect refers to the fact that the behavior of the response to one factor with respect to the factor's levels depends on the level of another factor.

DoE is a statistical design technique to design experiments or simulations. It has been widely used to efficiently identify design solutions in engine development. Because the number of possible factor combinations as design options increases exponentially with the number of factors, it is important to use DoE to screen and identify the key factors. DoE is a systematic approach to reduce the cost and time in finding the design solution. Instead of using a trial-and-error or one-factor-at-a-time approach, DoE constructs a statistical plan of the combination of factors. A DoE simultaneously varies the levels of the factors by using fractional factorial designs instead of the full factorial design to greatly reduce the number of runs by orders of magnitude without losing the quality. The advantage of using DoE becomes very prominent when the number of factors is much greater than three. The DoE design methods include principally the Taguchi method and the response surface methodology (RSM).

The counterpart of the DoE approach in the nondeterministic world is the Monte Carlo simulation. The Monte Carlo simulation is a statistical probabilistic technique to design experiments or simulations to study the nondeterministic probability distribution of the factors and the responses.

In the Monte Carlo simulation, each factor may have 1,000 random levels. Each Monte Carlo run is a random combination of all the random factors. For example, if there are eight factors, the number of Monte Carlo runs can be anywhere from 1,000 to $1{,}000^8$. The response of this 1,000 or $1{,}000^8$ runs forms a statistical distribution. Usually, 1,000–10,000 runs can give a sufficiently accurate result. A greater number of runs does not change the result significantly.

After the DoE or Monte Carlo data are obtained from either experiment or simulation, optimization is usually needed to search for the optimal or robust design. There are numerous solution algorithms to search for the optima. However, they all require the continuous mathematical functions that link the input factors to the output response. These continuous mathematical functions are called emulators (e.g., a polynomial built by surface fit), and they are used as the objective function or constraint function in the optimization formulation introduced earlier. The response surface methodology (RSM) is the method to generate these emulators.

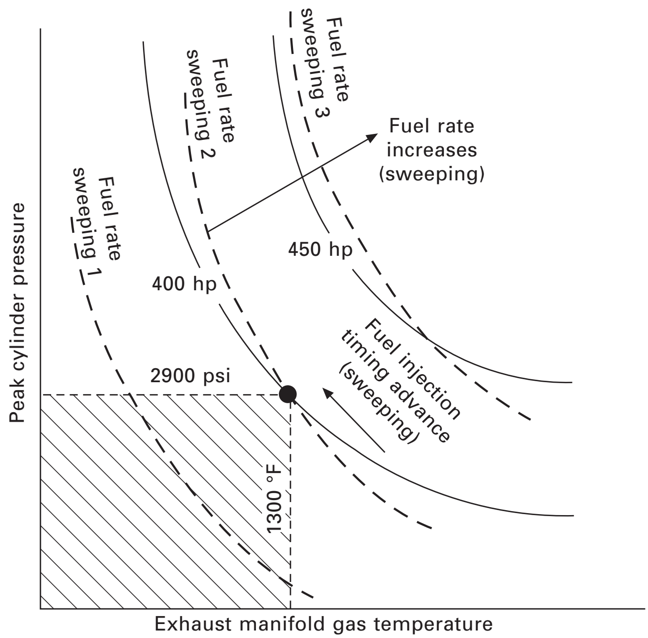
RSM is a statistical design and surface-fit approach. It uses a mathematical function (emulator) to describe the relationship between any input and output. The emulator can be used to study the main effect and the interaction effect, predict the outcome which has not been run, produce sensitivity maps, and conduct optimization. The order of surface fit in RSM is usually lower than a third. Therefore, it is not suitable for applications that are highly nonlinear. For example, in the fast changing and highly nonlinear transient emission cycles, neural network is a better tool to establish the relationship between the input and output parameters.

3.1.3 On-target simulation versus optimization

There are three types of optimization in diesel engine system design:

- parametric sweeping to calculate objective functions to find optima
- RSM DoE optimization
- RSM Monte Carlo optimization.

Not all the engine system design data are produced by the optimization formulation. On-target simulation is not an optimization. However, it is a frequently used technique to compute a design solution once all the design constraints are determined as equality constraints. Recall the formulation in Section 3.1.1 for the single set of deterministic solutions when the number of unknowns is equal to the number of equations of equality constraints. One example of on-target simulation is to find the engine power subject to the constraints of a given peak cylinder pressure and exhaust manifold gas temperature, as shown by the circle in Fig. 3.1. The mathematical formulation of this problem at a given engine speed is described as follows:



3.1 On-target simulation, parametric sweeping, and optimization.

$$\begin{cases} \dot{W}_E = f_{power}(\dot{m}_{fuel}, \phi_{inj}) & 3.2 \\ T_{exh} = f_{Texh}(\dot{m}_{fuel}, \phi_{inj}) = 1200\ °\mathrm{F} & 3.3 \\ p_{cyl} = f_{pcyl}(\dot{m}_{fuel}, \phi_{inj}) = 2900\ \mathrm{psi} & 3.4 \end{cases}$$

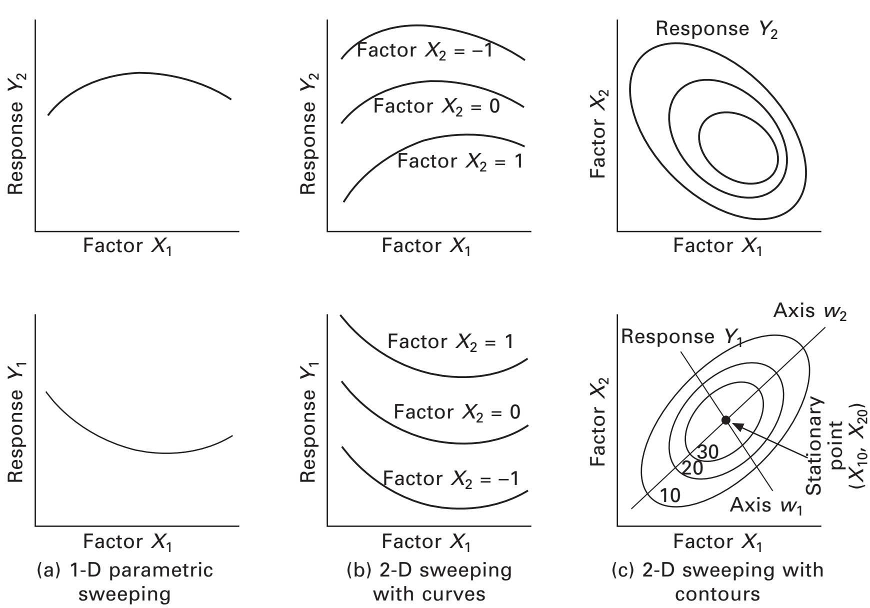
where $\dot{W}_E$ is the engine power, $\dot{m}_{fuel}$ is the fueling rate, ϕ_{inj} is the fuel injection timing, T_{exh} is the exhaust manifold gas temperature, and p_{cyl} is the peak cylinder pressure. The functions f are essentially implicit functions given by the engine cycle simulation. Note that there are two unknowns ($\dot{m}_{fuel}$ and ϕ_{inj}) and two equality constraints in the formulation of 3.2–3.4. Therefore, it gives a single set of deterministic solutions without optimization. After solving equations 3.3 and 3.4 for $\dot{m}_{fuel}$ and ϕ_{inj}, the engine power can be calculated by using equation 3.2. It should be noted that on-target simulation here is not the same as design for target. Design for target does not preclude using optimization techniques.

Parametric sweeping to calculate objective functions to find optima is an optimization. It computes the objective function by directly sweeping the factors and then finds the optima by checking the sweeping results against the constraints. If the above on-target problem is phrased differently, it becomes an optimization problem as follows. Find the maximum engine power subject to the constraints of peak cylinder pressure less than 2900 psi and exhaust manifold gas temperature lower than 1300°F. The formulation is given by:

$$\begin{cases} \text{Maximize} & f_{power}(\dot{m}_{fuel}, \phi_{inj}) \\ \text{subject to} & f_{Texh}(\dot{m}_{fuel}, \phi_{inj}) < 1300\ °\text{F} \\ & f_{pcyl}(\dot{m}_{fuel}, \phi_{inj}) < 2900\ \text{psi} \end{cases} \qquad 3.5$$

By plotting the two constraints on the two axes in Fig. 3.1, the solution domain of equation 3.5 is shown in the crossed area that is enclosed by the two design constraints. The two-dimensional sweeping data of $\dot{m}_{fuel}$ and ϕ_{inj} are shown by the dashed family curves in Fig. 3.1. It can be visually identified that the optimal solution is still the same circle, located at the boundary of the design space (crossed area). Note that this optimization problem can be solved graphically by using parametric sweeping because it is only a two-dimensional problem.

Parametric sweeping is a simple but very effective optimization technique. As shown in Fig. 3.2, sometimes several responses need to be plotted and checked as constraints before the optimal solution is selected. A large number of routine optimization topics in engine system design can be handled with this method, for example:



3.2 Illustration of optimization with parametric sweeping.

- Find the optimum fuel injection timing to minimize BSFC.
- Find the optimum exhaust valve opening to minimize engine BSFC.
- Find the optimum EGR circuit flow restriction and intake throttle opening to reach a target EGR rate while minimizing BSFC at peak torque.
- Find the optimum turbine wastegate opening and intake throttle opening to match a target air–fuel ratio while minimizing BSFC.
- Find the optimum turbine wastegate opening and intake throttle opening to match a target turbine outlet temperature for aftertreatment regeneration while minimizing BSFC.

The parametric sweeping method is usually restricted to only one or two design factors because it is difficult to plot the parametric data of more than three factors. An optimization topic involving three or more factors normally requires a RSM DoE approach by using emulators in the multi-dimensional design space. Some examples of the RSM DoE approach are as follows:

- Find the optimum fuel injection timing, injection pressure, EGR valve opening, and turbine area in emissions calibration to minimize BSFC while meeting NO_x and soot emissions.
- Find the optimum exhaust valve opening, exhaust cam acceleration, and maximum exhaust valve lift to minimize BSFC while maintaining acceptable valvetrain dynamics.
- Find the optimum EGR circuit flow restriction, intake throttle opening, and turbine wastegate opening to reach a target EGR rate while minimizing BSFC at peak torque.
- Find the optimum turbine areas of the high-pressure stage turbine and the low-pressure stage turbine, and EGR valve opening to match the target air–fuel ratio and EGR rate while minimizing BSFC.
- Find the optimum turbine wastegate opening, intake throttle opening, and the dosing fuel rate in DPF regeneration to maintain sufficient turbine outlet temperature while optimizing overall engine fuel consumption.
- Find the minimum BSFC subject to the durability and emission constraints by optimizing the following system design factors: stroke-to-bore ratio, engine compression ratio, valve timing, fuel injection timing, turbine area, and EGR circuit flow restriction.
- Find the optimum engine configuration subject to the constraints given by a combination of several potential technologies such as variable valve timing, waste heat recovery, turbocompounding, and VGT.

3.1.4 Methodology of diesel engine system optimization

During the optimization for a system, the control factors are system-level design parameters, and the objectives and constraints are system-level functions that reflect subsystem interaction and attribute interaction. The key elements in diesel engine system optimization are shown in Fig. 3.3. They are:

- system attributes
- system models
- response surface methodology (RSM) or neural network (NN) models.

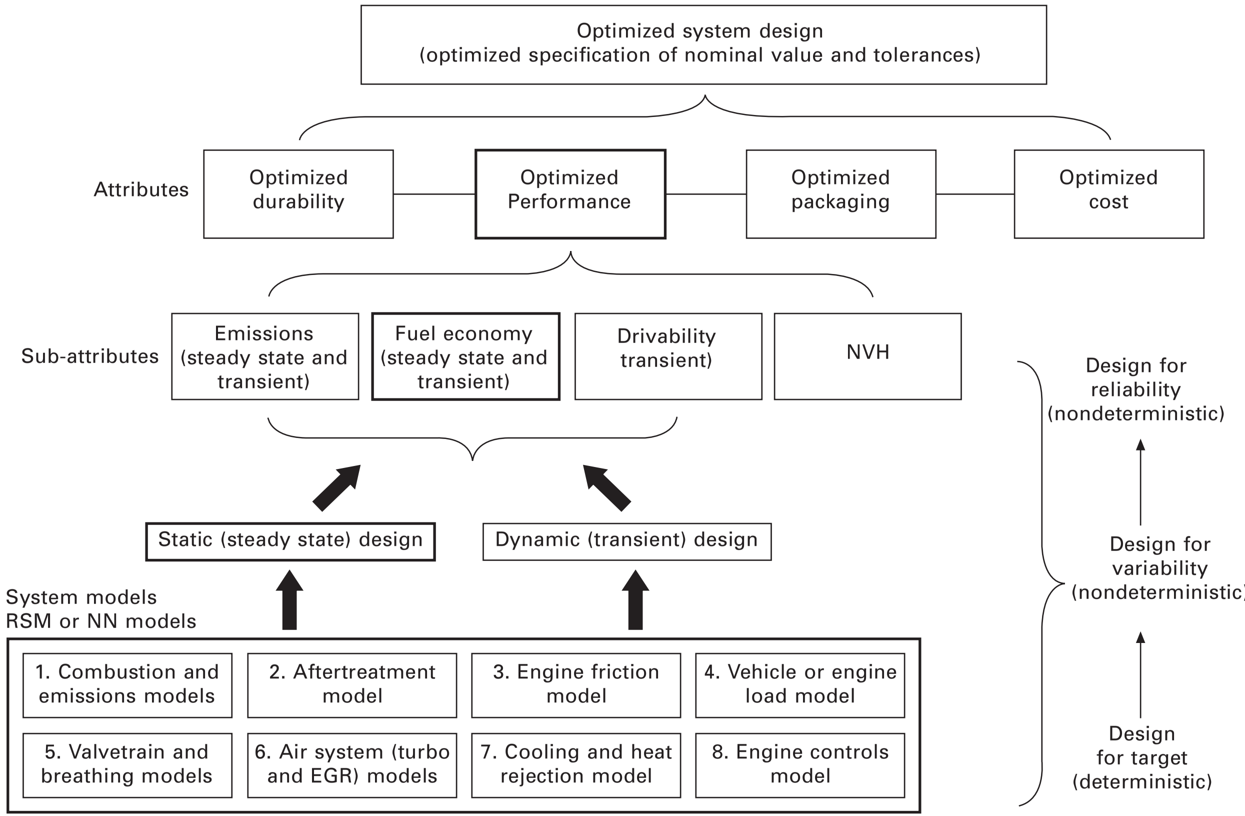
Simultaneously meeting the requirements of different attributes requires multi-objective optimization. Many attributes are conflicting in nature. Trade-offs among them are inevitable. In many cases, simplifying the multi-objective optimization is necessary by choosing the most important attribute or sub-attribute as the single objective. For example, fuel economy is the most important objective for commercial trucks, while power or drivability may be the most important for a racing car.

The scope of engine system optimization can be classified as follows: (1) hardware vs. software (engine controls); (2) steady state vs. transient; and (3) deterministic vs. nondeterministic. A transient model is more complex than a steady-state model because engine control strategies and time-marching calculation are involved. A nondeterministic model is more complicated than a deterministic model since probability distribution by the Monte Carlo simulation is involved. A reliability-based optimization is more difficult than a variability-based model because the time-dependent deterioration effects need to be modeled. A reliability-based nondeterministic system model covering all four attributes is ideal but very difficult to obtain in reality. Moreover, when one extra dimension of the complexity (i.e., transient time, probability, time-dependent deterioration) is added into the optimization scope, there is a big challenge on data display and the way to summarize the optimization result in a compact and comprehensible way.

The optimization for the category of deterministic 'design for target' (Fig. 3.3) is the foundation of diesel engine system design. It is very important. Most of the system optimization work is carried out in this category. It also provides pre-screened sub-optimal results for further optimization in the categories of nondeterministic 'design for variability' and 'design for reliability'.

Figure 3.3 also illustrates the need for future research in engine system optimization. This includes:

1. High-quality system models for each attribute that reflect subsystem interactions.
2. System models to reconcile the four attributes (performance, durability, packaging, and cost) to form multi-objective functions or a properly weighted objective function in the optimization.
3. High-quality transient performance models that include engine control software strategies.



3.3 Scope of diesel engine system design optimization.

4. High-quality transient thermo-mechanical durability models.
5. Reliability models.
6. High-quality response surface models and neural network models used in the optimization.

3.1.5 The Taguchi method

Dr Genichi Taguchi (1986) introduced an approach where standardized orthogonal arrays and linear graphs are used to simplify or 'shortcut' the DoE without going through the sophisticated steps of statistical data processing and mathematical solutions. In the Taguchi method, the control factors and noise (uncontrollable) factors are handled separately in two orthogonal arrays of the DoE matrix. The following needs to be first specified before a specially designed orthogonal array is selected to compose the DoE runs: (1) the factors; (2) the interactions between the control factors; and (3) the number of levels for each factor. The factor levels can be mixed in an array, for example, two levels for one factor and three levels for another factor. The control factors are placed in an inner array (also called the main array), and the noise factors (e.g., the tolerances in the control factors around the nominal values) are placed in an outer array, if any. The two arrays are crossed in such a way that the runs in the inner array are simply repeated for each run of the outer array. The procedure of the Taguchi method is summarized in Fig. 3.4. Note that the empty columns can be used in the array if the factors cannot fill all the columns.

In the Taguchi DoE method, interactions are often treated like any other factors in the array (e.g., the interaction between the factors *A* and *B*, $A \times B$, is treated as another factor). However, their presence is ignored for the preliminary determination of the optimum design. Alternatively, as shown in the example of L_{18} array in Fig. 3.4, an interaction is built in between the first two columns without sacrificing any other column. The interactions between the three-level factors are distributed uniformly to all the three-level columns so that the main effects can be estimated without the bias from the effect of interaction. Therefore, when the interaction effects exist, proper array selection is very important. The significance of the interaction can be analyzed by analysis of variance (ANOVA, a diagnostic tool in regression analysis). The determination of the interaction effect requires a separate study and is not convenient. Such a study may suggest a different optimum design from that suggested by the main effects analysis. When several interactions are included in an experiment, the optimum factor level selection may become extremely complex (Roy, 1990). In fact, most diesel engine system optimization problems involve strong interaction effects.

The Taguchi optimization method is often tied to robust design whose purpose is to provide on-target performance and maintain it in the face of

Taguchi orthogonal array L_{18} ($2^1 \times 3^7$, for a mixed array with two-level and three-level factors, 1 factor having 2 levels, 7 factors having 3 levels, 18 DoE runs)

DoE run number	Xc1	Xc2	Xc3	Xc4	Xc5	Xc6	Xc7	Xc8
1	–1	–1	–1	–1	–1	–1	–1	–1
2	–1	–1	0	0	0	0	0	0
3	–1	–1	1	1	1	1	1	1
4	–1	0	–1	–1	0	0	1	1
5	–1	0	0	0	1	1	–1	–1
6	–1	0	1	1	–1	–1	0	0
7	–1	1	–1	0	–1	1	0	1
8	–1	1	0	1	0	–1	1	–1
9	–1	1	1	–1	1	0	–1	0
10	0	–1	–1	1	1	0	0	–1
11	0	–1	0	–1	-1	1	1	0
12	0	–1	1	0	0	–1	–1	1
13	0	0	–1	0	1	–1	1	0
14	0	0	0	1	–1	0	–1	1
15	0	0	1	–1	0	1	0	–1
16	0	1	–1	1	0	1	–1	0
17	0	1	0	–1	1	–1	0	1
18	0	1	1	0	–1	0	1	–1

Control factor array (inner array)

Response 1
The 'S2-3-1' stands for the sample value of the 2nd run of array L_{18} with 3rd run of array L_9 for Response 1.

L_9 run/ L_{18} run	1	2	3	4	5	6	7	8	9	Mean value	Signal-to-noise ratio
1	S1-1-1	S1-2-1	S1-3-1	S1-4-1	S1-5-1	S1-6-1	S1-7-1	S1-8-1	S1-9-1	M1-1	SN1-1
2	S2-1-1	S2-2-1	S2-3-1	S2-4-1	S2-5-1	S2-6-1	S2-7-1	S2-8-1	S2-9-1	M2-1	SN2-1
3	S3-1-1	S3-2-1	S3-3-1	S3-4-1	S3-5-1	S3-6-1	S3-7-1	S3-8-1	S3-9-1	M3-1	SN3-1
4	S4-1-1	S4-2-1	S4-3-1	S4-4-1	S4-5-1	S4-6-1	S4-7-1	S4-8-1	S4-9-1	M4-1	SN4-1
5	S5-1-1	S5-2-1	S5-3-1	S5-4-1	S5-5-1	S5-6-1	S5-7-1	S5-8-1	S5-9-1	M5-1	SN5-1
6	S6-1-1	S6-2-1	S6-3-1	S6-4-1	S6-5-1	S6-6-1	S6-7-1	S6-8-1	S6-9-1	M6-1	SN6-1
7	S7-1-1	S7-2-1	S7-3-1	S7-4-1	S7-5-1	S7-6-1	S7-7-1	S7-8-1	S7-9-1	M7-1	SN7-1
8	S8-1-1	S8-2-1	S8-3-1	S8-4-1	S8-5-1	S8-6-1	S8-7-1	S8-8-1	S8-9-1	M8-1	SN8-1
9	S9-1-1	S9-2-1	S9-3-1	S9-4-1	S9-5-1	S9-6-1	S9-7-1	S9-8-1	S9-9-1	M9-1	SN9-1
10	S10-1-1	S10-2-1	S10-3-1	S10-4-1	S10-5-1	S10-6-1	S10-7-1	S10-8-1	S10-9-1	M10-1	SN10-1
11	S11-1-1	S11-2-1	S11-3-1	S11-4-1	S11-5-1	S11-6-1	S11-7-1	S11-8-1	S11-9-1	M11-1	SN11-1
12	S12-1-1	S12-2-1	S12-3-1	S12-4-1	S12-5-1	S12-6-1	S12-7-1	S13-8-1	S12-9-1	M12-1	SN12-1
13	S13-1-1	S13-2-1	S13-3-1	S13-4-1	S13-5-1	S13-6-1	S13-7-1	S14-8-1	S13-9-1	M13-1	SN13-1
14	S14-1-1	S14-2-1	S14-3-1	S14-4-1	S14-5-1	S14-6-1	S14-7-1	S15-8-1	S14-9-1	M14-1	SN14-1
15	S15-1-1	S15-2-1	S15-3-1	S15-4-1	S15-5-1	S15-6-1	S15-7-1	S16-8-1	S15-9-1	M15-1	SN15-1
16	S16-1-1	S16-2-1	S16-3-1	S16-4-1	S16-5-1	S16-6-1	S16-7-1	S17-8-1	S16-9-1	M16-1	SN16-1
17	S17-1-1	S17-2-1	S17-3-1	S17-4-1	S17-5-1	S17-6-1	S17-7-1	S18-8-1	S17-9-1	M17-1	SN17-1
18	S18-1-1	S18-2-1	S18-3-1	S18-4-1	S18-5-1	S18-6-1	S18-7-1	S19-8-1	S18-9-1	M18-1	SN18-1

Gray area: optional data

Taguchi orthogonal array L_9 (3^4, for an array with 4 factors having 3 levels, 9 DoE runs)

DoE run number	Xn1	Xn2	Xn3	Xn4
1	–1	–1	–1	–1
2	–1	0	0	0
3	–1	1	1	1
4	0	–1	0	1
5	0	0	1	–1
6	0	1	–1	0
7	1	–1	1	0
8	1	0	–1	1
9	1	1	0	–1

Noise factor array (outer array)

Response 1

Control factor	Control factor level	Main effect of mean value	Main effect of signal-to noise ratio
Xc1	–1	...	...
Xc1	–1	...	...
Xc2	–1	...	...
Xc2	0	...	...
Xc2	1	...	...
Xc3	–1	...	...
Xc3	0	...	...
Xc3	1	...	...

Control factor	Control factor level	Main effect of mean value	Main effect of signal-to noise ratio
Xc4	–1	...	...
Xc4	0	...	...
Xc4	1	...	...
Xc5	–1	...	...
Xc5	0	...	...
Xc5	1	...	...
Xc6	–1	...	...
Xc6	0	...	...
Xc6	1	...	...

Control factor	Control factor level	Main effect of mean value	Main effect of signal-to noise ratio
Xc7	–1	...	...
Xc7	0	...	...
Xc7	1	...	...
Xc8	–1	...	...
Xc8	0	...	...
Xc8	1	...	...

Plot main effects for all responses. Select optimum control factor levels based on the trade-offs between mean (performance) and $f_{S/N}$ ratio (robustness), and the trade-offs between different responses. Noise factors are not controllable. Run confirmation test.

3.4 Process of Taguchi's DoE method.

variability or noise factors. One key feature of the Taguchi method is the signal-to-noise ratio ($f_{S/N}$), which takes both the mean and variation into account. A signal-to-noise ratio in engineering used to mean a power ratio between a signal (i.e., meaningful information) and the background noise. Because many signals have a wide dynamic range, a signal-to-noise ratio is usually expressed in decibels (dB) and defined as 10 times the logarithm of the power ratio. Taguchi proposed using a signal-to-noise ratio to optimize the robustness of products or processes. Taguchi's signal-to-noise ratios are shown in Table 3.1. It should be noted that only the last signal-to-noise ratio has a true meaning and is dimensionless. If the noise outer array does not exist, the $f_{S/N}$ can be obtained by repeating several experimental runs at each DoE factor level. If there is a noise outer array, the sample points for calculating the $f_{S/N}$ will be the DoE runs in the outer array at a given factor level setting in the inner DoE. Thus, an $f_{S/N}$ exists for each DoE run of inner array. In order to achieve variance reduction to improve quality, the values in control factors are sought to maximize the $f_{S/N}$. Because the $f_{S/N}$ mixes both mean (location effect) and variance (dispersion effect), both mean and $f_{S/N}$ are used simultaneously in optimization. The levels of some control factors

Table 3.1 Taguchi's signal-to-noise ratios

Parameter name	Parameter symbol	Formula
Taguchi's signal-to-noise ratio (for smaller-the-better or to minimize the response)	$f_{S/N1}$	$f_{S/N1} = -10 \cdot \log_{10}\left(\frac{\sum_{i=1}^{n} Y_i^2}{n}\right)$
Taguchi's signal-to-noise ratio (for larger-the-better or to maximize the response)	$f_{S/N2}$	$f_{S/N2} = -10 \cdot \log_{10}\left(\frac{\sum_{i=1}^{n} \frac{1}{Y_i^2}}{n}\right)$
Taguchi's signal-to-noise ratio (for nominal-the-best or to achieve a target with minimum deviation, if the response mean and variance can be altered independently)	$f_{S/N3}$	$f_{S/N3} = -10 \cdot \log_{10} \sigma_{SV}^2$ or $f_{S/N3} = -10 \cdot \log_{10} \sigma^2$
Taguchi's signal-to-noise ratio (for nominal-the-best or to achieve a target with minimum deviation, if the response mean and variance are related)	$f_{S/N4}$	$f_{S/N4} = 10 \cdot \log_{10}\left(\frac{\bar{Y}^2}{\sigma_{SV}^2}\right)$ or $f_{S/N4} = 10 \cdot \log_{10}\left(\frac{\mu^2}{\sigma^2}\right)$

Note: These signal-to-noise ratios need to be maximized in optimization in robust design. Only $f_{S/N4}$ is a true dimensionless signal-to-noise ratio.

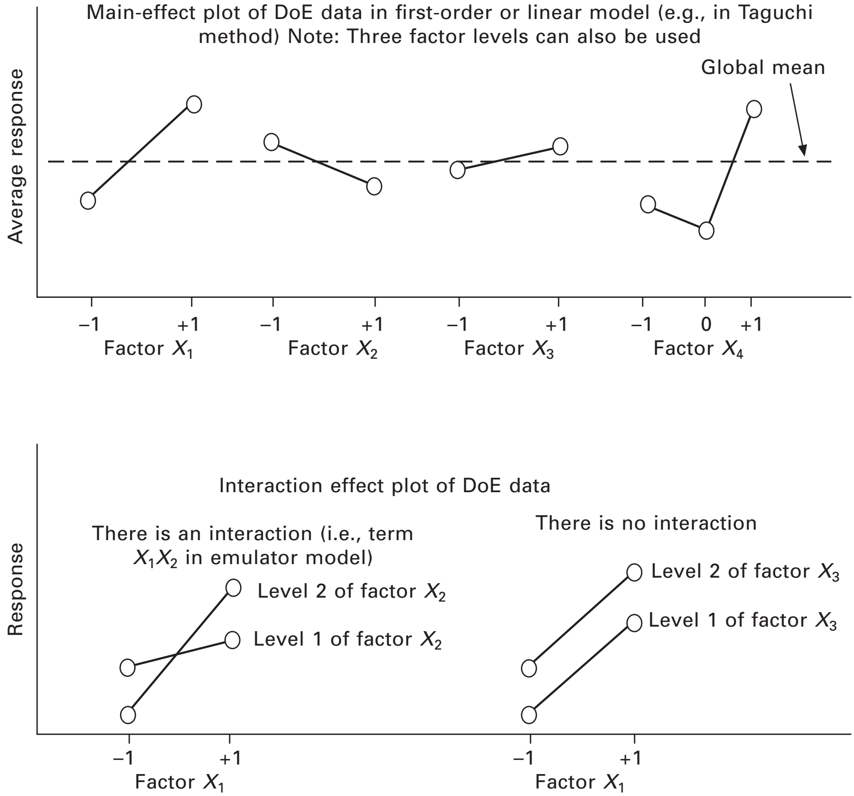
are optimized to bring the mean to the target, and the levels of some other control factors that affect $f_{S/N}$ are optimized in order to maximize $f_{S/N}$ and equivalently to minimize variance.

Standard ANOVA techniques are often applied on the mean response and the signal-to-noise ratio in order to identify the control factors that affect them. ANOVA can also quantify the effects and relative contribution of each factor to the total variation of the responses.

The summary statistics of the responses of each DoE run include mean, variance, and signal-to-noise ratio. Depending on optimization criteria, the output responses can be separated into three categories: smaller-the-better, larger-the-better, and nominal-the-best. Variable screening can be conducted using the main effect plot or linear regression. The response in the main effect plot of an independent factor is the average of all the responses over all the levels of other factors. It should be noted that the unit or the range is usually very different from one factor to the next. The main effect plot allows an intuitive visualization of the trends between the inputs and the outputs and identifies strong and weak factors as well as the optimal factor levels (Fig. 3.5). The greater the departure of the points from the global mean of all the response data, the more important the factor is. The optimal levels are chosen for each control factor from the main effect plot. In the interaction plot, parallel lines (strictly speaking, non-crossing lines within the factor range) indicate no interaction. It should be noted that the main effect plot can be highly misleading when there are strong interactions between the factors such as in many engine applications. Also note that the main effect plot can be biased if the experimental design is not properly balanced. The linear regression method computes the correlation coefficients for the factors and also outputs the p-values that indicate the statistical significance between the factors and the response. Although the linear regression method does not provide a visualization of the plot, it does not produce biased conclusions. After the optimum combination of the control factor level is selected from the main effects plot, the optimum signal-to-noise ratio and mean can be predicted. At the end, a confirmation run is conducted as a final check to confirm the predicted optimum outcome.

As summarized in Fig. 3.4, the Taguchi method consists of the following nine steps:

1. defining the goal or target;
2. brainstorming for possible factors;
3. designing a DoE (selecting factors, levels, ranges and orthogonal arrays);
4. conducting DoE runs to collect data;
5. analyzing and plotting the results with statistical analysis (e.g., ANOVA, regression analysis, main effect plot, and interaction plot);



3.5 Main effect and interaction plots.

6. identifying the optimum factor levels;
7. predicting the optimum responses by using a signal-to-noise ratio and confidence interval;
8. checking the predicted optimum against design constraints; and
9. conducting the confirmation run.

It should be noted that Taguchi's signal-to-noise ratio method does not always give the best results, especially when the objective is to achieve a target response rather than to minimize or maximize the response. As pointed out by Myers and Montgomery (2002), maximizing a signal-to-noise ratio does not distinguish which control factors are the location effects (affecting the mean) and which are the dispersion effects (affecting the variance). In other words, the signal-to-noise ratios confound the location effect and the dispersion effect so that they are not necessarily unique and probably should not be used as an optimization parameter. Taguchi's signal-to-noise ratios were criticized by some statisticians and its utility seems questionable. In fact, a robust design can be achieved by using RSM to optimize the response and reduce the variability due to noise factors, without using a signal-to-noise ratio.

The Taguchi method can work well for determining a robust design if the interaction effect between control factors is not important (Myers and Montgomery, 2002). The method provides abundant information about the interaction between the control factors and the noise factors due to crossed inner and outer arrays, but the method provides only little information about the interaction between the control factors or the interaction between the noise factors because the orthogonal arrays are heavily fractionated. The Taguchi method offers a quick and simple approach during product optimization and robust design in order to identify which factor contributes the most, the optimal factor level, or the impact of noise. However, it seems several weaknesses exist in the Taguchi method. First of all, the Taguchi method does not have an empirical surface-fit model which includes the higher-order or interaction effects. Second, Taguchi arrays usually only allow for two or three levels for each factor, and the preparation of orthogonal arrays for more levels is not convenient or economical. These make the Taguchi method not sufficiently accurate or convenient for engine system design applications. Third, only the main effects are emphasized in the Taguchi method, and the interaction effects of control or noise factors are not given enough attention in the DoE design arrays, especially the interaction effects between the control factors. Fourth, the usefulness of optimizing the signal-to-noise ratios seems questionable. Finally, the way that the noise factors are handled in an outer array is not efficient in minimizing the number of DoE runs because the entire main (inner) array needs to be repeated at each run of the outer array.

The Taguchi DoE method was introduced in detail by Roy (1990), Carey (1992), Fowlkes and Creveling (1995) and Goh (1994), and will not be elaborated in this book. In general, although the Taguchi DoE method can be used in engine system design, its use is usually restricted to preliminary concept screening or qualitative evaluation (e.g., by using the main effect plot). For precise system design the response surface methodology with surface-fit formula is much more powerful.

3.1.6 Response surface methodology (RSM)

As introduced earlier, optimization usually requires objective functions and constraint functions that are continuous mathematical functions relating the DoE input factors to output responses. The response functions (emulators) are constructed by the surface-fit techniques using least-squares regression to fit the DoE runs. The emulator model consists of the coefficients reflecting the influence of the factors. The usefulness of the emulator functions in terms of predicting the main effect and the interaction effect of the DoE factors and the accuracy of the functions are determined by the structure of the emulators (i.e., the terms constructed into the emulators) and the statistical design of the DoE matrix. Response surface methodology (RSM) is an advanced DoE method that systematically addresses the statistical design and data processing techniques of DoE.

The emulator model in RSM is usually a polynomial function containing the terms of the input factors up to third order and the interaction terms, e.g.,

$$
\begin{aligned}
Y_1 = C_0 + C_{11}X_1 + C_{12}X_1^2 + C_{13}X_1^3 + \\
C_{21}X_2 + C_{22}X_2^2 + C_{23}X_2^3 + \\
C_{31}X_3 + C_{32}X_3^2 + C_{33}X_3^3 + \\
C_{1121}X_1X_2 + C_{1131}X_1X_3 + C_{2131}X_2X_3
\end{aligned}
\qquad 3.6
$$

where Y_1 is a response and X_1, X_2 and X_3 are the factors. Both the control factors and the noise factors can be mixed together in the RSM DoE design. The coefficients C_i are determined with least-squares fit. RSM is a better and more formal method than the Taguchi method for solving DoE optimization and robust design problems. The comparison between the Taguchi method and RSM is given in Table 3.2. Detailed discussions about the DoE statistical matrix design and the data processing techniques of RSM are provided in

Table 3.2 Comparison between Taguchi method and response surface methodology

	The Taguchi method	Response surface methodology (RSM)
Statistical DoE design	• Crossed orthogonal arrays of fractional factorial design with control factors in inner array and noise factors in outer array • Primarily used to screen factors	• Control factors and noise factors are mixed in more economical DoE designs with fewer runs • Both linear and higher-order analysis • Primarily used for both screening and regression modeling
Graphical display	Main effect plot, interaction plot	Contour plot for main and interaction effects
Analytical evaluation	• Not easy to handle interaction effects • Analysis of variance (ANOVA) • Signal-to-noise ratio analysis	• Easy to handle interaction • Surface-fit emulator regression models (first-, second- and third-order) including interaction terms built with least squares • Analysis of variance (ANOVA) for regression models • Canonical analysis • Ridge analysis

Section 3.2. More elaborated theories of RSM are introduced by Montgomery (1991), Eriksson *et al*. (1999), and Myers and Montgomery (2002).

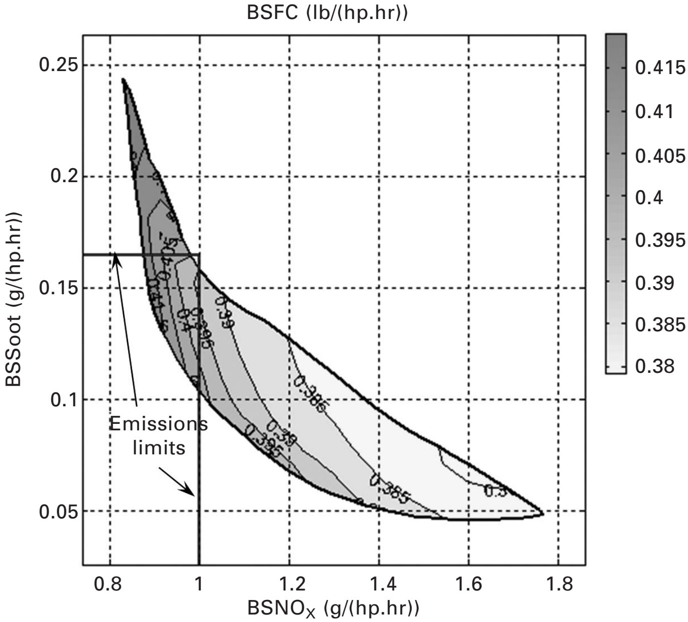
3.1.7 From single point to two-dimensional optimization map

Diesel engine system design has two characteristics in output data. First, a single 'design point' needs to be selected after comparing various solutions. For example, a large amount of optimization can be conducted at the rated power condition by comparing different hardware configurations and calibration settings. However, in the end, one nominal design point needs to be selected and reported as a final summary in the system design. Secondly, parametric maps are required during the course of system design to understand the sensitivity and the data trend. The techniques of constructing the parametric maps are important.

Diesel engine system design often handles a large number of factors. Parametric analysis and its graphical display is a very important fundamental technique every system design engineer must master. Graphical display of the optimization results is very important for the following reasons. First of all, during optimization not only is the final optimal result needed, but also the parameter sensitivity needs to be visually examined in order to obtain good understanding. The parametric data also need to be documented for future use or reference. Secondly, the design target is often a moving target during the development stage. A complete graphical map covering the uncertainties and the moving target is very convenient for system design to reference. Therefore, running and reporting one single point of data is not sufficient in engine system design. One- or two-dimensional maps are required. As one more dimension is added in graphical display, the amount of information packed increases exponentially. A useful graphical display usually has to be limited to either a two-dimensional contour map (preferred) or an equivalent three-dimensional cubic plot.

The one-dimensional map refers to plotting one factor (or response) on the horizontal axis of the map and plotting one response on the vertical axis. The two-dimensional map refers to plotting two factors or two responses on the horizontal and vertical axes, respectively, and plotting another response as constant-value contours.

A more powerful approach is to produce 'each point' in the one- or two-dimensional map through optimization. The 'optimization' here is different from the parametric sweeping introduced earlier. In the parametric sweeping (Fig. 3.2), each point is obtained without optimization. One example of two-dimensional optimization is shown in Fig. 3.6. It presents the DoE optimization result with respect to several factors (i.e., VGT vane opening, EGR valve opening, fuel injection timing, and injection pressure). Two



3.6 Two-dimensional optimization with minimum BSFC contour map.

responses, the brake specific NO_x and the brake specific soot, are placed on the horizontal and vertical axes of the map, respectively. The objective function, BSFC, is minimized at 'each' point in such a two-dimensional map subject to certain constraints (e.g., peak cylinder pressure, exhaust temperature, plus the two additional constraints of the NO_x and soot values at that point). As a result, the BSFC contour map obtained as such is the 'minimum BSFC' map in an optimized sense within the design space and bounded by the DoE factor range and optimization constraints. Such an optimized contour map is certainly much more complex and computational intensive than the simple parametric sweeping map shown in Fig. 3.2. The two-dimensional optimization map is a very powerful graphical display and a concise summary of the sensitivity and optimization results. The two-dimensional optimization technique proposed here is especially important for diesel engine system design because there are several critical trade-offs that often need to be used in a two-dimensional domain. These trade-offs include: (1) soot vs. NO_x for emissions compliance (Fig. 3.6); (2) air–fuel ratio vs. EGR rate for evaluating the air system capability; and (3) hydrocarbons vs. exhaust temperature for aftertreatment design. Normally, minimizing BSFC is appropriate in these two-dimensional optimizations.

In summary, there are five fundamental techniques in the optimization of diesel engine system design:

- parametric sweeping
- DoE main effect plot
- DoE emulator
- one-dimensional optimization by curves
- two-dimensional optimization by contour maps.

3.1.8 From single-objective to multi-objective optimization

Complete engine system optimization often requires meeting the requirements of multiple objectives, such as performance, durability, packaging and cost. The objectives of different attributes are usually conflicting and have trade-offs, for example, NO_x vs. PM, fuel economy vs. emissions or engine noise. Moreover, engine hardware design for the same attribute (e.g., volumetric efficiency, BSFC) also presents a trade-off between high speed and low speed, or between high load and low load. Typical examples include turbocharger matching and cam timing. Furthermore, engine operation in customer usage occurs in a wide range of environmental conditions such as hot and cold weather, low and high altitude. There is also a trade-off for a given design attribute between those different conditions. Although single-objective optimization is still a commonly used tool in engine system design, more advanced multi-objective optimization is needed to address these trade-offs.

A multi-objective optimization topic can be formulated as follows:

$$\begin{cases} \text{Minimize } f(X) \\ \text{subject to } g(X) \le 0 \\ \text{where } X = [X_1, X_2, \cdots, X_k]^T \\ \qquad f(X) = [f_1(X), f_2(X), \cdots, f_q(X)]^T \\ \qquad g(X) = [g_1(X), g_2(X), \cdots, g_m(X)]^T \end{cases} \tag{3.7}$$

Optimization with multiple conflicting objectives in a multi-dimensional design space is difficult. Multi-objective optimization may not have a single optimum. Instead, there is a set of alternative trade-offs generally known as the Pareto-optimal solutions. In any optimization study, it is necessary to define an objective function, which can be either a single response parameter or a predetermined combination of several parameters. Converting a multi-objective optimization to a single-objective optimization can be problematic because the optimization result can be influenced by a bias in the selected single objective function. Therefore, a direct exploration in a multi-objective design space would be a better alternative. The solution methods for multi-objective optimization topics usually include the following.

Single objective function with constraints

Select one criterion as primary and the remaining criteria as secondary. Use the primary criterion as the objective function in the optimization, and assign acceptable minimum or maximum values to the secondary criteria and use them as constraints.

Non-normalized weighted single objective function

When the objectives are the same parameter or attribute with the same unit but at different working conditions (e.g., BSFC or NO_x emissions at different engine speeds/loads or different ambient temperatures), apply weighting factors on all the objective functions and then construct a single objective function as follows:

$$F(X) = \sum_{j=1}^{q} w_j f_j(X) \quad \text{where } w_j \geq 0, \ \sum_{i=1}^{q} w_i = 1 \tag{3.8}$$

or

$$F(X) = \sum_{j=1}^{q} w_j [f_j(X) - f_j^{target}]^2 \tag{3.9}$$

where a target value is denoted with the superscript 'target'.

Normalized weighted single objective function

When the objectives are different parameters or attributes with different units (e.g., emissions, BSFC, noise), construct a merit function as a single objective by normalizing each parameter with their own reference values or target values as follows:

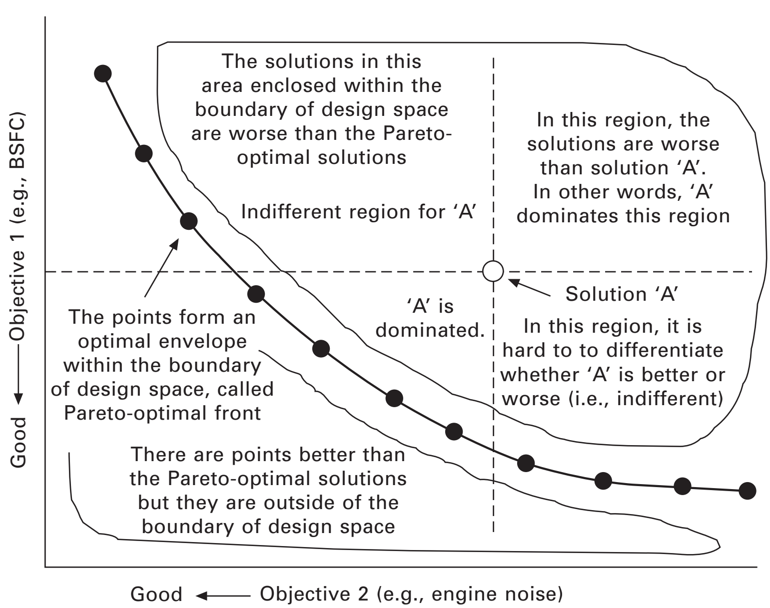
$$F(X) = C_5 \left[\begin{array}{l} w_1 \left(\dfrac{\psi_{NO_x}}{\psi_{NO_x}^{target}} \right)^{C_1} + w_2 \left(\dfrac{\psi_{PM}}{\psi_{PM}^{target}} \right)^{C_2} + w_3 \left(\dfrac{\psi_{HC}}{\psi_{HC}^{target}} \right)^{C_3} + \\ w_4 \left(\dfrac{\psi_{CO}}{\psi_{CO}^{target}} \right)^{C_4} + w_5 \dfrac{\eta_{BSFC}}{\eta_{BSFC}^{target}} + w_6 \cdot 10^{\left(\frac{p_{SPL} - p_{SPL}^{target}}{20} \right)} + \cdots \end{array} \right]^{C_6} \tag{3.10}$$

where ψ_{NO_x} is brake specific NO_x emissions, η_{BSFC} is brake specific fuel consumption, and p_{SPL} is the sound pressure level of engine noise in dB. The $w_1, w_2, \ldots, w_6$ are weighting factors. The $C_1, C_2, \ldots, C_4$ are the exponent constants to amplify the effects of certain normalized terms. The C_5 and C_6

are also arbitrary coefficients. For example, $C_5 = 1000$, $C_6 = -1$ and $C_1 = C_2 = \ldots = 2$ were used by Montgomery and Reitz (2000); $C_5 = 1$, $C_6 = 1$, $C_1 = C_2 = \ldots = 2$ and $w_1 = 1$, $w_2 = 1$, $w_3 = 0$, $w_4 = 0$, $w_5 = 1$, $w_6 = 0.5$ were used by Mallamo *et al.* (2004).

Pareto-optimal solutions

The concept of Pareto optimality is illustrated in Fig. 3.7. A Pareto-optimal front curve or surface can be obtained by optimization algorithms such as the multi-objective genetic algorithm (MOGA). In a single-objective optimization, the optimal solution is usually clearly defined. However, this is not the case for a multi-objective problem where the objectives can be conflicting. A single solution is hardly the best for all the objectives simultaneously. Instead of a single optimum, there is a set of trade-off solutions, generally known as Pareto-optimal solutions (also called non-dominated solutions). These solutions are optimal in the sense that no other solutions in the design space are better than them or can 'dominate' them when all the objectives are considered. In other words, all other solutions are 'dominated' solutions that are worse than the Pareto-optimal solutions. The method of Pareto-optimal solutions has been widely used in engine optimization (Li *et al.*, 2004; Courteille *et al.*, 2005; Zottin *et al.*, 2008). A concept of Pareto front quality index (PFQI) was developed by Kazancioglu *et al.* (2003) to objectively assess the



3.7 Illustration of Pareto optimality in two-objective optimization.

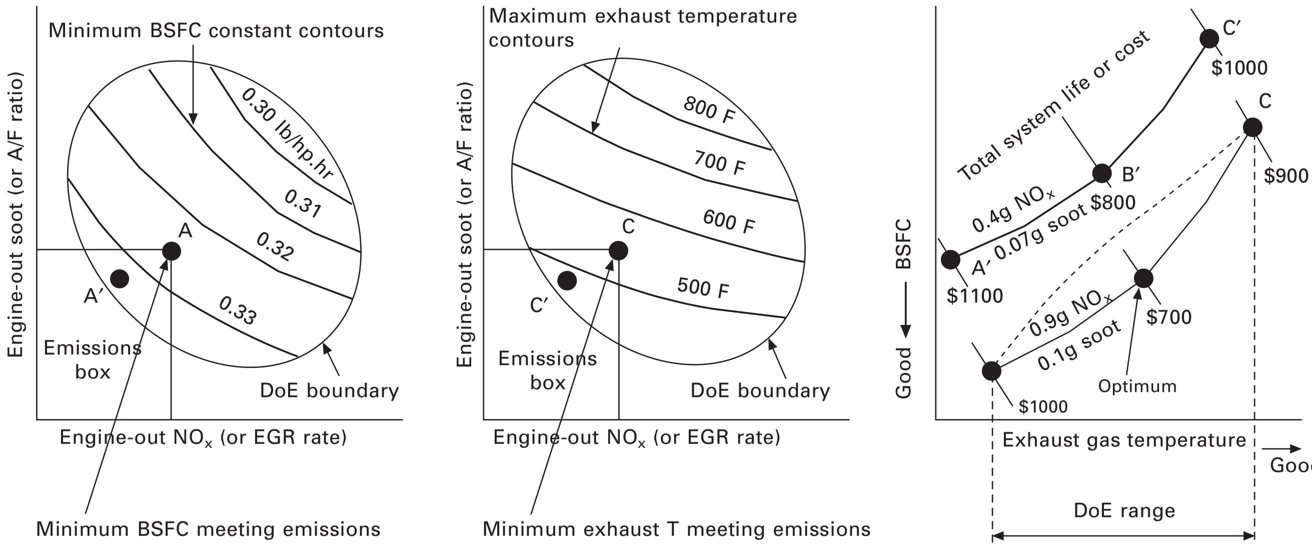
quality of Pareto optimal fronts; quantify the closeness to the utopia point of the Pareto-optimal solutions; and quantify the range and the evenness of the spread of the Pareto solutions.

Minimum-contour-based multiple Pareto-optimal solutions

In the past a Pareto optimal front was obtained by running an optimization routine. Using the two-dimensional optimization (minimization) contour map discussed in the previous section, the Pareto optimal can be generated more easily and flexibly. Taking an example of the DoE optimization shown in Fig. 3.8, the method of minimum-contour-based multiple Pareto-optimal solutions is explained as follows. The objective is to find the optimal solution of emissions level, BSFC and cost. A DoE can be constructed at a representative engine operating condition in speed and load. The DoE factors include air system control factors (e.g., VGT vane opening, EGR valve opening, intake throttle opening, and wastegate opening) and fuel injection control factors (e.g., main injection timing, post injection timing and quantity, fuel injection pressure). The DoE responses include air–fuel ratio, EGR rate, engine-out NO_x, engine-out soot, BSFC, and turbine outlet exhaust gas temperature. Aftertreatment operation (e.g., DPF regeneration) prefers hotter exhaust gas temperature in order to save cost from precious metal loading inside the DPF. But hotter exhaust gas may indicate more energy is lost or wasted and hence worse BSFC. Therefore, there is a trade-off between exhaust temperature and BSFC. Moreover, the emissions level affects such a trade-off. The total cost of the system for the customer is related to the emissions level the engine is designed for, the aftertreatment cost, and the cost saving from the fuel saving due to BSFC reduction. There may be an optimal (the lowest) cost within the design space.

Obviously, optimizing such a multi-dimensional design topic and presenting the result in a concise way is challenging. The minimum-contour-based method is characterized by the following three steps.

- Step 1 – minimum-contour optimization for the first objective (Fig. 3.8a). The method starts from conducting constrained single-objective optimization to minimize BSFC in the entire emissions domain (i.e., the soot vs. NO_x domain). The range of the DoE factors forms a boundary in this response domain. Every data point on the minimum BSFC contour map has the minimum achievable BSFC within the factor space. In other words, if a third axis were created to represent BSFC pointing vertically out of the paper from point A, there would be multiple BSFC values which are higher than the minimum value shown in the contour map at that given emission value. Once the minimum BSFC map is obtained, several points such as A and A′ at different emissions levels can be identified.



(a) Step 1: Minimum-contour optimization for the first objective

(b) Step 2: Maximum-contour optimization for the second objective

(c) Step 3: multiple Pareto-optimal curves at different constraints, and optimization

3.8 Illustration of minimum-contour-based multiple Pareto optimal solution.

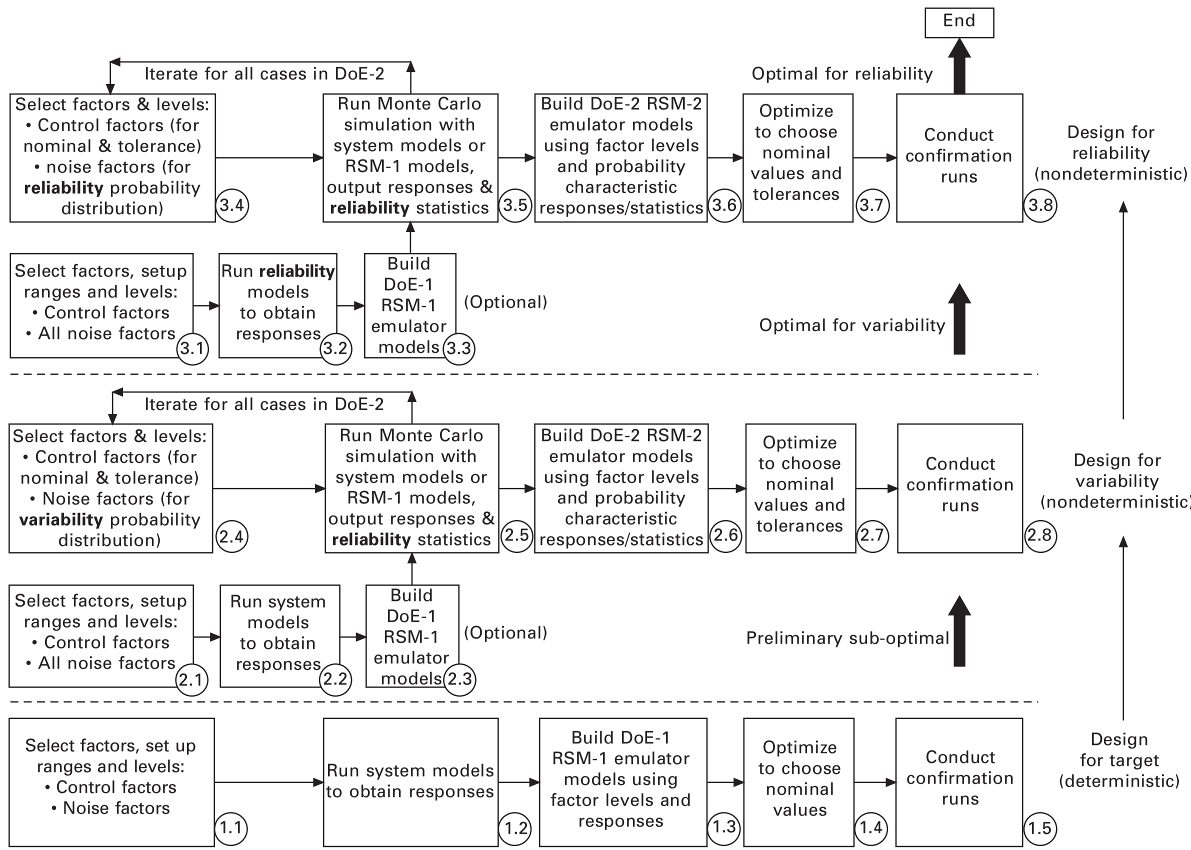
- Step 2 – minimum- or maximum-contour optimization for the second objective (Fig. 3.8b). The method continues by conducting another constrained single-objective optimization to maximize the exhaust temperature in the entire emissions domain. The DoE factors form the same boundary mapped in this domain. It should be noted that the maximum exhaust temperature point C, subject to the same emissions constraints as for point A, has a higher BSFC than point A.
- Step 3 – multiple Pareto-optimal curves at different constraints (Fig. 3.8c). The two extreme points (A and C) of the Pareto-optimal curve have been determined for a selected emissions level. Point B can be generated by running a similar optimization as in Step 2 to maximize the exhaust temperature while subject to a constraint at a certain BSFC level that is selected between the BSFC of point A and the BSFC of point C. Then, the Pareto curve can be constructed by linking points A, B, and C to display the trade-off. It should be noted that the dotted curve linking point A directly to point C is not a part of the Pareto optimal front because the points along the dotted curve are not optimized in any sense. Another Pareto-optimal curve at another emissions level A′–B′–C′ can be obtained similarly. On these Pareto-optimal points, the values of another objective function (e.g., the system cost) can be calculated and marked on the chart to form certain constant-value contours. It is noted that the system cost may reach a minimum along the Pareto-optimal curve. A final optimum solution can be selected from this chart.

Such a chart consisting of multiple Pareto-optimal curves is very convenient and powerful for selecting the final optimal solution in diesel engine system design. The chart is featured by two objectives (BSFC vs. exhaust temperature here, placed on the horizontal and vertical axes of the chart, respectively) at several levels of constraints (emissions here) with the values of a third objective (cost here) marked on the Pareto-optimal curves. The final design decision can be made based on such a concise summary of 'four dimensions' of information (i.e., three objectives of response/output plus one constraint) that are distilled from many more dimensions of DoE input factors.

Other optimization methods such as desirability functions are introduced in Edwards *et al.* (2000) and Myers and Montgomery (2002).

3.1.9 Optimization in design for target, design for variability, and design for reliability

Diesel engine system design requires an optimized specification of both nominal target value and tolerance. Steady-state engine optimization with a large number of factors usually requires a DoE technique. Figure 3.9 illustrates the processes of the optimization for diesel engine system design. The processes consist of three layers of work:



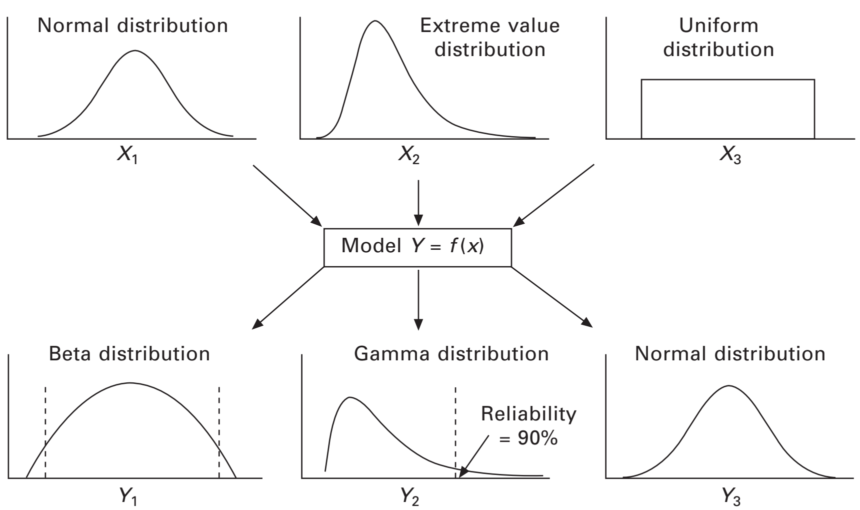
3.9 DoE optimization process for steady-state diesel engine system design.

- a deterministic 'design for target' process to screen for preliminary sub-optimal of the nominal value of the design specification
- a nondeterministic 'design for variability' process to reach the optimal design – both nominal value and tolerance of the design specification, subject to variability
- a nondeterministic 'design for reliability' process to reach the optimal design – both nominal value and tolerance of the design specification, subject to reliability.

The difference between variability and reliability is that reliability analysis includes the effect of the time-dependent noise factors (e.g., deterioration). The design for variability uses probabilistic objective functions to control both nominal value and tolerance range in order to make the design insensitive to noise factors.

The contents in steps 1.1–1.5 described in Fig. 3.9 for the layer of design for target are explained in detail in Section 3.2. The RSM-1 model mentioned in step 1.3 refers to the surface-fit emulator model that relates the nominal value of the response to factors. There is no emulator model for the tolerance in this layer.

The optimization in design for variability is illustrated in steps 2.4–2.5 in Fig. 3.9. The associated Monte Carlo simulation is shown in Fig. 3.10. Basically, the Monte Carlo simulation is a probability calculation by using random combinations of random samples selected from the probabilistic distributions of several input factors. The probabilistic distribution of the output response can be predicted along with an estimation of the failure rate or reliability. To make the estimation accurate, the amount of random



3.10 Statistical uncertainty propagation and design for variability.

samples needs to be very large. The details of the Monte Carlo simulation are provided in Section 3.4.

The noise factors mentioned in step 2.1 in Fig. 3.9 refer to all the noise factors covered by the variability analysis. The steps 2.1–2.3 compose DoE-1, and they are similar to steps 1.1–1.3 in nature. The level setting of the noise factors in step 2.1 is handled in the same manner as in step 1.1 (i.e., only for levels of mean values). The DoE-1 RSM-1 emulator surface-fit models are often needed as surrogate models to replace the computationally intensive engine cycle simulation models because the Monte Carlo simulation in step 2.5 requires thousands of runs. The thousands of Monte Carlo runs need to be iterated for each case in DoE-2. It should be noted that the level setting of the noise factors in the DoE-2 in step 2.4 is different from that in step 2.1 (or step 1.1). The noise factors in step 2.4 need to be described by several distribution factors (e.g., mean, standard deviation; scale parameter and shape parameter) to reflect its particular probabilistic distribution shape. These factors are called probability distribution factors. Each probability distribution factor is a factor in DoE-2. Each noise factor in step 2.4 needs to have several factor levels for each probability distribution factor within a reasonable range for the shape of the given type of probability function. For example, for a noise factor of turbine efficiency, its 'mean value' factor needs to have five levels of setting to cover a range of possible mean values of the probabilistic distribution of the turbine efficiency, for example at 58%, 59%, 60%, 61%, and 62%. Its 'standard deviation' factor also needs to have five levels of setting to cover a range of possible different shapes of the probabilistic distribution of the turbine efficiency, for example at 0.3%, 0.6%, 0.9%, 1.2%, and 1.5%. Obviously, the DoE size in step 2.4 is usually larger than that in step 2.1. For example, assuming the DoE-2 in step 2.4 has 10 factors (i.e., 4 control factors, and 3 noise factors which give 6 noise probability distribution factors) and 210 cases (runs), for each case the Monte Carlo simulation needs to be executed 1000 times by taking 1000 random probability sample combinations. Such a huge amount of computation usually cannot be handled by using the original detailed system models. Therefore, the RSM-1 model described in step 2.3 is needed here as the fast surrogate model.

The output of step 2.5 in Fig. 3.9 includes all the engine responses in the form of probabilistic distribution shapes, their statistical properties for a selected fit of probability distribution function, and probability statistics (i.e., failure rate for variability). The statistical properties of the responses may include the following: minimum, maximum, mean, standard deviation, skewness, excess kurtosis, and mode. (For the definition of these probability distribution parameters, see Tables A.1 and A.2 in the Appendix.) Suspected outliers in the probability distribution of the simulated responses are not uncommon. Outliers are not necessarily bad data points. They should be handled carefully instead of simply being removed automatically. The RSM-2 emulator models are described in step 2.6 by linking the factors of DoE-2 to the probability distribution responses and probability statistics. The emulator models allow an evaluation of the sensitivity of the output's probability distributions to all the input factors by using the analysis techniques introduced earlier (e.g., parametric sweeping, two-dimensional optimization with contour maps).

Step 2.7 is critical in robust optimization. In the traditional theory of robust design, Dr Taguchi used a 'two-step optimization' approach (Fowlkes and Creveling, 1995). In this approach, the product's tolerance is first reduced to a desirable probability distribution shape, then, the entire probability distribution curve is shifted to the desirable target by adjusting the nominal design value. Such a two-step approach has certain disadvantages. For example, the nominal target design and tolerance design are separated and their interactions are difficult to be handled efficiently. In the robust optimization theory for diesel engine system design here, those disadvantages are overcome by using a one-step simultaneous optimization of both nominal design and tolerance design. The mathematical formulation of the optimization by using the DoE-2 RSM-2 emulator models in step 2.7 enables such a simultaneous optimization because the models include all the statistical properties (nominal or mean, tolerance or deviation) for constrained optimization (e.g., subject to a constraint of failure rate at or less than certain prescribed target value). It should be noted that such an advantage of the proposed 'design for variability' approach over the traditional 'two-step optimization' approach can only be achieved by introducing the RSM into the robust design area.

The last layer of the system optimization is design for reliability. It is similar to the design for variability (Fig. 3.9), but still different. The reliability-related system models, probability distributions, and output statistics should be used in steps 3.2, 3.4, and 3.5 shown in Fig. 3.9, respectively. As a comparison, the variability-related items should be used in steps 2.2, 2.4, and 2.5.

3.2 Response surface methodology (RSM)

Response surface methodology (RSM) is a DoE approach in which the output variables are called 'responses' and the independent input variables are called 'factors'. The response function forms a surface or hyper-surface (if there are more than two factors) in the factor space. Optimization is conducted to search for a global optimal point (minimum or maximum) on a surface defined by the objective function. Detailed discussions on RSM are provided by Eriksson *et al.* (1999) and Myers and Montgomery (2002). A brief overview of the most important concepts and procedures in the RSM theory related to diesel engine system design is summarized in this section.

3.2.1 Overview of the RSM process

Engine optimization requires a continuous quantitative mathematical model to link the factors to responses (either steady state or transient) and solve for global optima. In engine applications the relationship between the design or operating factors and the responses is highly nonlinear and complex. For example, the impact of turbine area and EGR valve opening on air flow rate is governed by a large nonlinear system of equations of the engine in-cylinder cycle process, engine gas flow network, and turbocharging. The underlying mechanisms are so complex that they cannot be condensed to 'one equation' that still reflects the physical laws and can be used as the objective or constraint function. Moreover, sometimes the true functional relationships or physical models are not even understood or available, and the governing equations of the processes are simply lacking. The engineer must approximate the underlying processes with an appropriate empirical model in the form of:

$$Y_{obs} = Y + \upsilon = f(X_1, X_2, \ldots, X_k) + \upsilon \tag{3.11}$$

which links the input factors X_i to an approximation (Y) of the observed output response Y_{obs}. It should be noted that the terms in the empirical model usually do not contain any physical meaning. Here υ is an error term that reflects the sum of experimental errors and the 'lack-of-fit' model errors not accounted for in the function f. The experimental errors are usually assumed to have a normal distribution with a mean at zero and a constant variance. In the RSM, such a model is built with the linear regression technique by fitting a set of sample data to minimize the errors. Although more complicated mathematical functions (e.g., spatial stochastic process models or sophisticated weighted linear interpolation models, neural network, radial basis function network, and spline-cubic function) may be used to fit the data, polynomial models are most widely used to approximate the true functions. Usually, for a RSM problem with multiple input factors the function f is a first-, second- or third-order polynomial. Such an empirical model is called a response surface model or surface-fit emulator.

The RSM optimization process usually consists of the following six steps:

1. Prepare factors and their levels in a statistical DoE design matrix.
2. Generate responses by either experimental testing or numerical simulation (for instance, using the engine cycle simulation software to produce performance output data).
3. Build emulators to link the factors and the responses by using polynomials or other continuous functions with surface fit.
4. Validate the emulator models by checking the surface-fit accuracy with spare runs (i.e., use the DoE runs that are not used during emulator model building to test the accuracy of model prediction).

5. Conduct optimization by using the emulator and advanced searching algorithms to search for the global optima under given constraints.
6. Produce confirmation runs for the optimum result.

3.2.2 Building emulator models

Appropriate statistical DoE design is dependent upon the model structure of the emulator. Unlike the curve fit for a response with respect to only a single factor where a polynomial of very high order (e.g., 6th order) can be used, a lower-order polynomial (usually less than 3rd) is generally used in RSM to handle multiple factors. For a complex process, lower order or fewer terms in the polynomial model generally makes the model deviate more from the true physical mechanism, but allows fewer DoE runs to build the model.

For a simple linear system without interactions, a first-order polynomial model, also called main effect model, can be used. It is given by

$$Y_{obs} = Y + \upsilon = \left(C_0 + \sum_{i=1}^{k} C_i X_i \right) + \upsilon \tag{3.12}$$

where Y_{obs} is the observation values of the DoE test run, and Y is the fitted value by the polynomial model. If there is a 'curvature' on the surface of the response function, interaction terms or a higher-order polynomial must be used to model the 'curvature' in such a more complex system. For example, a linear-with-interaction model is given by

$$Y_{obs} = Y + \upsilon = \left(C_0 + \sum_{i=1}^{k} C_i X_i + \sum_{i=1,\, i<j}^{k-1} \sum_{j=2}^{k} C_{ij} X_i X_j \right) + \upsilon \tag{3.13}$$

where X_iX_j is the interaction term. Another example is a second-order (quadratic) model, which is a second-order Taylor series approximation:

$$\begin{aligned} Y_{obs} &= Y + \upsilon \\ &= \left(C_0 + \sum_{i=1}^{k} C_i X_i + \sum_{i=1}^{k} C_{ii} X_i^2 + \sum_{i=1,\, i<j}^{k-1} \sum_{j=2}^{k} C_{ij} X_i X_j \right) + \upsilon \end{aligned} \tag{3.14}$$

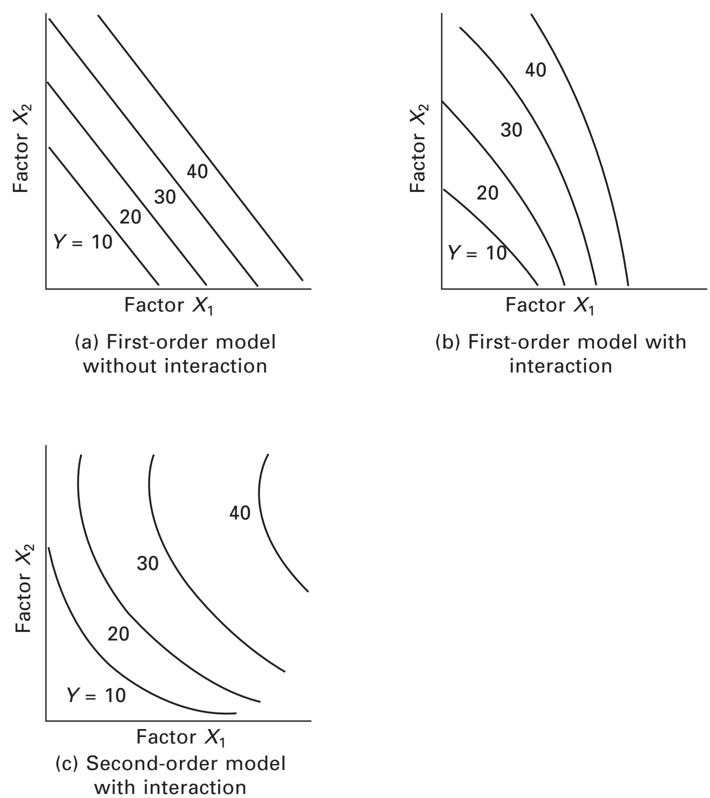
Note that equations 3.12–3.14 contain p terms of polynomial or regression coefficients, where

$$\begin{cases} p = 1 + k & \text{for first-order without interaction} \\ p = 1 + k + k(k-1)/2 & \text{for first-order with interaction} \\ p = 1 + 2k + k(k-1)/2 & \text{for second-order with interaction} \end{cases} \tag{3.15}$$

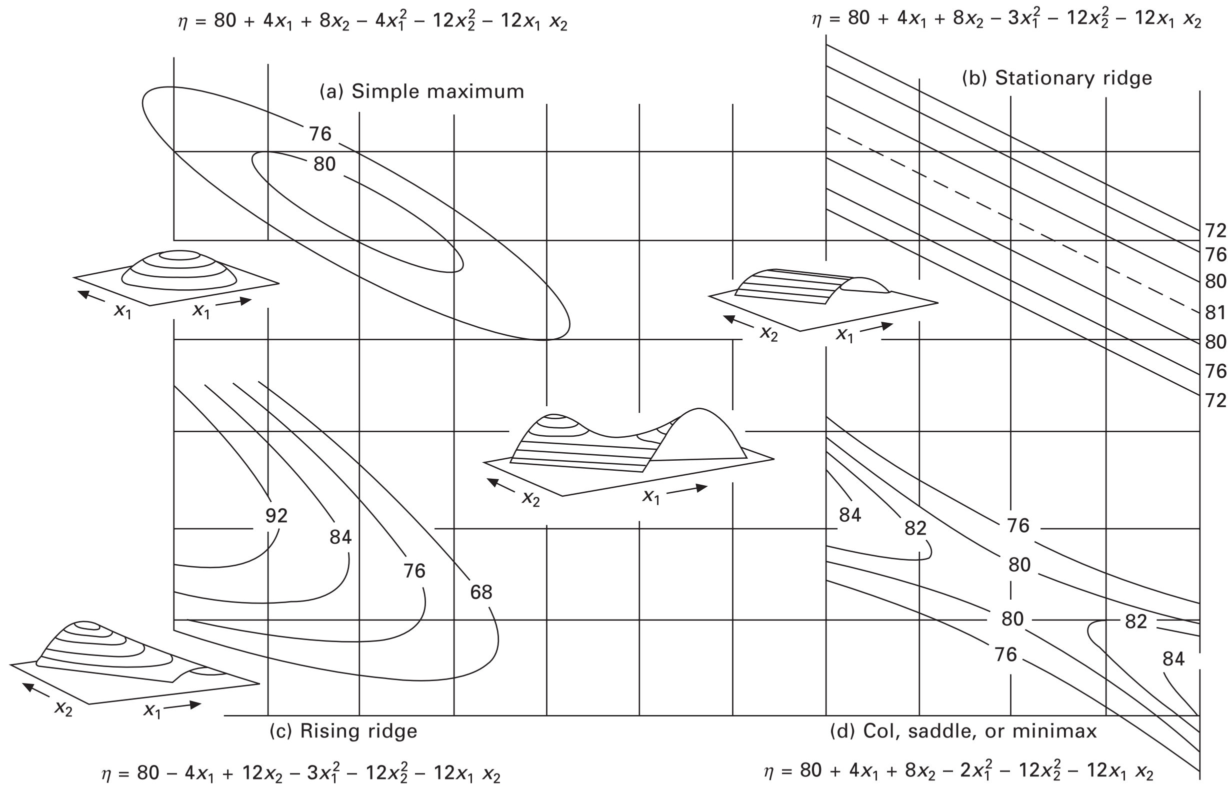
and k is the number of factors. One advantage of using polynomial in the emulator is that the polynomials are everywhere differentiable so that the derivative-based optimization algorithms can be used to search for the optima. A second-order model requires more DoE runs and also more factor levels to construct than a first-order model.

A first-order model of $Y = C_0 + C_1X_1$ shows a straight line in the domain of response vs. factor (Y vs. X_1). A first-order model of $Y = C_0 + C_1X_1 + C_2X_2$ without interactions gives a tilted plane in the three-dimensional space domain of response vs. factor (Y vs. X_1 and X_2). It is reflected as a contour map in the two-dimensional domain of X_1 vs. X_2 for the response that has parallel straight lines with constant intervals (Fig. 3.11a). A higher-order model including the interaction terms is characterized by the non-parallel variable-interval contours (Fig. 3.11b and Fig. 3.11c), or the shapes of stationary ridges, rising ridges, mountains or saddle surfaces if viewed in a three-dimensional space (Fig. 3.12).

Second-order polynomials have been widely used because they offer a good compromise between model flexibility and DoE size. Usually a second-order



3.11 Interaction effects.



3.12 Some examples of types of surfaces defined by the second-order model in two variables x_1 and x_2 (from Box and Draper, 1987; also adapted by Myers and Montgomery, 2002).

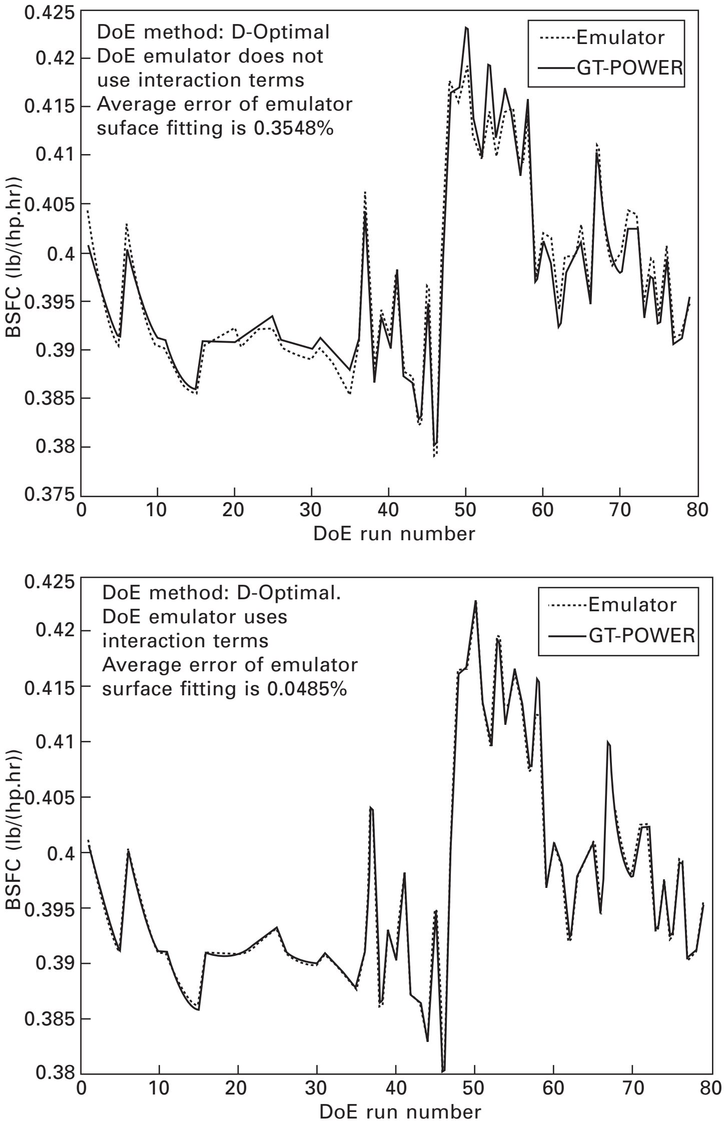
model is necessary and sufficient to approximate the true response surface in most applications of engine system design except for some extremely nonlinear scenarios (e.g., soot sensitivity, very complex interactions among the factors, or very large range of the DoE factors). In these cases, a third-order model becomes necessary. For example, the cubic terms X_i^3 and the interaction terms $X_i^2 X_j$ or $X_i X_j X_u$ need to be used for a highly nonlinear exhaust restriction flow coefficient, EGR circuit restriction flow coefficient, or VGT vane opening, also taking their interactions with other factors into account (Figs 3.13 and 3.14).

Polynomial models are 'parametric' models which provide the regression coefficients to represent the model structure and individual factor effects. In statistics, the regression coefficients are called 'parameters'. Non-parametric or non-polynomial models cannot output such descriptions of the model structure or the model coefficients. The non-parametric models include spatial stochastic process models, radial basis function, and neural network. They offer flexibility to adapt to the characteristics of highly nonlinear responses or to build accurate models in the presence of noisy or uncertain experimental or simulation data. The validity of higher-order polynomial models compared to other non-polynomial models has been proved by Bates *et al.* (2000). The details of stochastic process models are explained by Edwards *et al.* (2000). Neural network models are discussed by Shayler *et al.* (2000) and He and Rutland (2004).

Although sometimes it cannot match over the entire factor space, a polynomial model usually can approximate the true functional relationship quite well for a relatively small region. Assuming the number of DoE runs is n for each response parameter shown in equation 3.14, the sum of the squares of the errors of all n runs is given by:

$$\sum_{r=1}^{n} \upsilon_r^2 = \sum_{r=1}^{n}\left(Y_{obs,r} - C_0 - \sum_{i=1}^{k} C_i X_{i,r} - \sum_{i=1}^{k} C_{ii} X_{i,r}^2 - \sum_{i=1,\, i<j}^{k-1} \sum_{j=2}^{k} C_{ij} X_{i,r} X_{j,r}\right)^2 \qquad 3.16$$

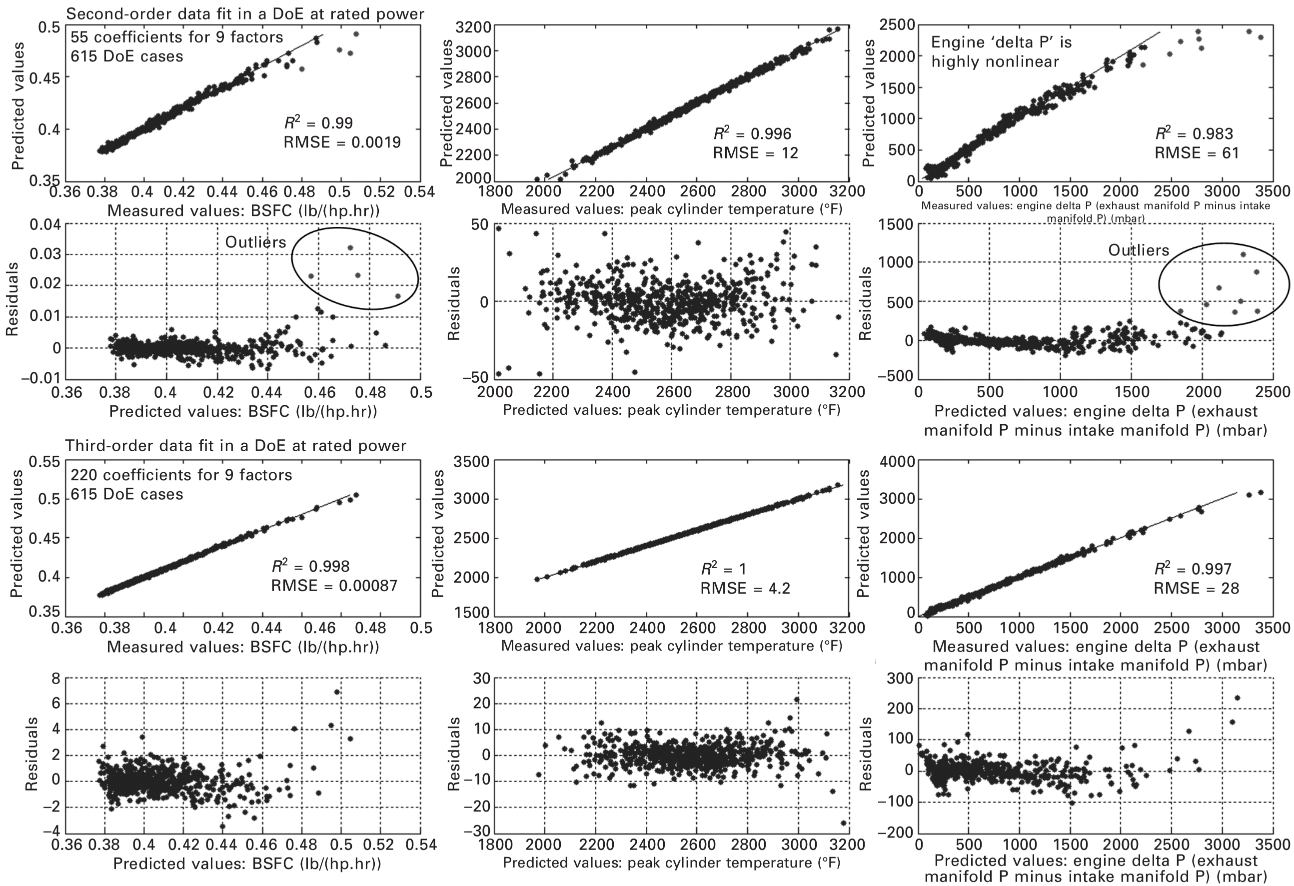
The method of least-squares is used to solve for the regression coefficients in the polynomial model by minimizing $\sum_{r=1}^{n} \upsilon_r^2$. The model parameters or regression coefficients C_i, C_{ii} and C_{ij}, $i, j = 1, \ldots k$, in equations 3.12–3.14 are solved by multiple linear regression analysis with the method of least-squares (also called surface fit). The coefficients are affected by the experimental design of the DoE matrix that is used to collect the data. The difference between the observation value Y_{obs} and the fitted value Y is called residual. The details of the mathematical solution of least-squares and analysis of variance (ANOVA) are provided by Eriksson *et al.* (1999) and Myers and



3.13 RSM emulator model accuracy – effect of interaction terms.

Montgomery (2002). Note that the number of DoE runs needs to be no less than the number of regression coefficients in equation 3.15, i.e., $n \geq p$.

A coefficient of determination R_d^2 (also called explained variance or goodness of fit) is defined as the ratio of the regression sum of squares to the total sum of squares. The total here refers to the regression sum of squares plus the residual sum of squares, as shown in equation 3.17:



3.14 RSM emulator model accuracy – effect of surface fit order.

$$R_d^2 = \frac{S_{SR}}{S_{ST}} = 1 - \frac{S_{SE}}{S_{ST}} = 1 - \frac{\sum_{r=1}^{n}(Y_{obs,r} - Y_r)^2}{\sum_{r=1}^{n}(Y_{obs,r} - \bar{Y}_{obs})^2} \text{ and } 0 \leq R_d^2 \leq 1 \quad 3.17$$

As pointed out by Myers and Montgomery (2002), a value of R_d^2 close to 1 does not necessarily imply that the regression model is good for accurate prediction. Adding a polynomial term to the model always increases R_d^2 because the least-squares solution for equation 3.16 approaches a deterministic solution with $R_d^2 = 1$, regardless of whether the additional term is statistically significant or not. If adding an additional term in the emulator results in a small increase in R_d^2, it suggests that the term does not really improve the model. It is possible to have a model with R_d^2 close to 1, but the prediction of the response has poor accuracy due to the negative impact of irrelevant terms in the model. More regression terms do not necessarily mean better model accuracy or predictability because the irrelevant terms which do not reflect the physical nature may interfere with the model's predictability. This is especially true for the interaction terms. Model accuracy checking and model pruning to delete irrelevant terms are always necessary to ensure the fitted model represents the true system behavior adequately.

To overcome the problem with R_d^2, an adjusted R_{adj}^2 (also called explained variance) is used in the literature (e.g., Myers and Montgomery, 2002):

$$R_{adj}^2 = 1 - \frac{\left(\dfrac{S_{SE}}{n-p}\right)}{\left(\dfrac{S_{ST}}{n-1}\right)} = 1 - \frac{n-1}{n-p}(1 - R_d^2) \text{ and } R_{adj}^2 < R_d^2 \quad 3.18$$

where n is the number of DoE runs or observations, and p is the number of regression coefficients in the model. In general, if unnecessary terms are added, the value of R_{adj}^2 will often decrease. When R_d^2 and R_{adj}^2 differ dramatically, there is a good chance that insignificant terms have been used in the emulator model (Myers and Montgomery, 2002). When less useful model terms are removed, R_d^2 decreases and R_{adj}^2 normally remains unchanged. This is a method to test or judge which polynomial terms are important in the model. The details of S_{SE}, S_{ST} and S_{SR} functions are given in Myers and Montgomery (2002). A parameter of goodness of prediction is the predicted variation $R_{prediction}^2$, given by

$$\begin{cases} R_{prediction}^2 = 1 - \dfrac{\upsilon_{PRESS}}{S_{ST}} = 1 - \dfrac{\sum_{r=1}^{n}(Y_{obs,r} - \hat{Y}_{(r)})^2}{\sum_{r=1}^{n}(Y_{obs,r} - \bar{Y}_{obs})^2} \\ -\infty \leq R_{prediction}^2 \leq 1 \end{cases} \quad 3.19$$

where υ_{PRESS} is the prediction error sum of squares, and $\hat{Y}_{(r)}$ is the predicted response at rth observation by using the emulator regression model fitted just without using the rth observation run. Such an approach is also called 'leave-one-out' cross validation (Bates *et al.*, 2000). The cross validation predicts each DoE run in turn when that run point is left out of the emulator equation. $R^2_{prediction}$ indicates the predictive capability of the original regression model that is fitted with all the runs from n observations. $R^2_{prediction}$ is more realistic and useful than R^2_{adj} and R^2_d to judge the quality of the regression models. Generally speaking, $R^2_{prediction}$ should be greater than 0.8–0.9 in order for a model to be acceptable. Moreover, $R^2_{prediction}$ and R^2_d should not be separated by more than 0.2–0.3. Figure 3.13 shows an example of the importance of a second-order interaction term in an air system DoE simulation. Figure 3.14 shows an example of the importance of the third-order terms in the emulator.

Another parameter to characterize the goodness of prediction or accuracy of the emulator model is the estimated root-mean-square error (RMSE). RMSE allows an estimation of the confidence intervals for model prediction. Larger error bars indicate less confidence in the emulator. According to Bates *et al.* (2000), an index of the inaccuracy of the emulator is defined as a RMSE percentage of the range of the response, given by

$$f_{RMSE\%} = 100 \times \frac{\upsilon_{RMSE}}{Y_{obs,max} - Y_{obs,min}} = 100 \times \frac{\sqrt{\frac{1}{n}\sum_{r=1}^{n} (\hat{Y}_{(r)} - Y_{obs,r})^2}}{Y_{obs,max} - Y_{obs,min}} \qquad 3.20$$

For example, $f_{RMSE\%} = 5\%$ means that if the emulator is used to predict the response at a new factor setting, the error of prediction can be expected to be roughly less than 5% when compared with the true value. A very narrow bar of confidence error indicates good statistical confidence in the prediction. A perfect emulator would produce predicted response values that overlap the true values with zero-length error bars.

Residual analysis and testing for lack-of-fit are common techniques for model accuracy checking. Choosing the right polynomial terms and deleting irrelevant terms in the model is a key step to ensure model accuracy. Adding DoE runs by using more factor levels and factor interactions usually can improve model accuracy and predictability but at the expense of longer time or higher cost for the runs. Good RSM DoE designs allow the estimation of the model coefficients with low uncertainty. The low uncertainty is measured by narrow error bars of the model's regression coefficients. A confidence interval refers to the interval between an upper bound and lower bound for an estimate at a selected degree of certainty (e.g., 95%). A prediction interval refers to the bounds for a predicted value of the response by using the model. A good example of the RSM model regression and optimization for diesel engine performance was given in the Chapter 15 in the book by Eriksson *et al*. (1999). Moreover, modern optimization software packages (e.g., MODDE, iSIGHT, Minitab) usually outputs the calculations of ANOVA, main effect, and interaction effect as standard post-processing of the DoE.

Lastly, it also should be noted that the transformation of factor and/or response variables from their natural unit to special forms (e.g., natural log, exponential, or trigonometric) may better reflect the physical mechanisms of the system hence improve the fit of the model to the DoE data runs. For example, a new response Y^* can be obtained by the transformation of the original response $Y^* = Y^{0.5}$ or $Y^* = \ln Y$ before surface fit is performed. Such a transformation may replace the higher-order polynomial terms for highly nonlinear phenomena or special curvatures. The transformation may also eliminate certain interaction terms from the emulator model so that the model can be simplified with better accuracy. Note that the transformed model is still classified as a polynomial model.

3.2.3 Statistical DoE design for RSM

The RSM technique can be applied to both experimental testing and simulation. There are three major differences between experimental tests and simulation:

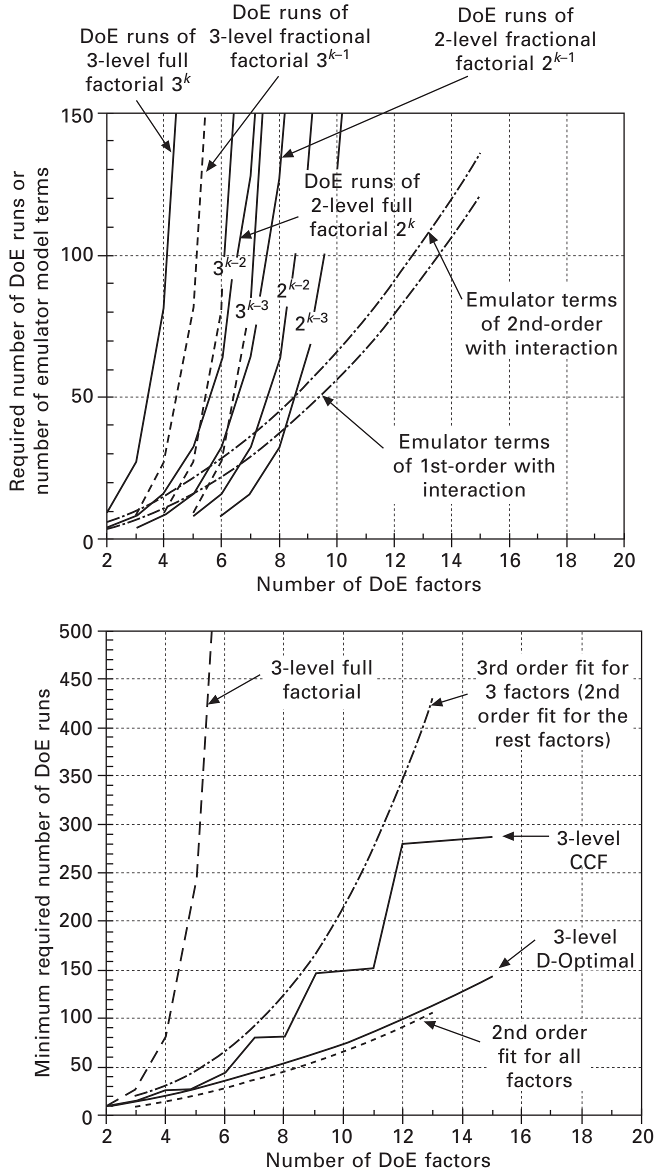
1. Any experimental work always contains measurement errors or variability. The input or output of test runs are not exactly repeatable. Simulation does not have this problem. Therefore, a randomized order of DoE runs and the repeated center-point runs for replicate errors used in an experimental DoE are not necessary in a simulation DoE.
2. Hardware limitations (e.g., excessively high peak cylinder pressure) mean that some experimental runs cannot be obtained within the designed DoE factor space so that these data points are simply missing. Simulation is not limited by such real world constraints. Extreme but still meaningful data sometimes can be obtained easily in the simulation and are included as useful information to build the emulator.
3. Simulation is usually much faster and less expensive than experimental testing. For example, a computer model can run overnight automatically without human intervention for a large number of DoE runs.

Overall, simulation usually has less restrictive or less demanding requirements on the statistical DoE design than experimental testing. Due to the above three reasons, the statistical DoE design method used for simulation can be different from the DoE method used for experimental testing. Although experimental data analysis such as emissions calibration or hardware testing is an important part in diesel engine system design, the main body of the system design work is conducted by simulation. The focus of this section is on the statistical DoE design of RSM used for engine system simulations. Myers and Montgomery (2002) provide a thorough discussion of the principles of experimental DoE design. Interested readers can find theoretical analysis and practical guidelines there for designing experiments.

The two most important properties of DoE design for engine system simulation are good emulator accuracy and cost-effective computation. Between them, good emulator accuracy has paramount importance for the quality of engine system design. Therefore, unlike experimental DoE tests where a minimum number of DoE runs is often a primary objective of the test, simulation DoE work usually uses as many DoE runs as possible (at affordable computing time) in order to increase the accuracy of the emulator model.

The emulator model accuracy is affected by the following: (1) the order and the number of the terms in the model (discussed in the previous section); (2) the number of DoE runs; (3) the factors (factor types, factor range, and factor levels); and (4) the distribution pattern of the DoE runs in the factor space (i.e., DoE design properties, interactions). These aspects are discussed below.

Engine design and operating factors can be classified into control factors and noise factors. The number of factors built into a RSM model needs to be carefully planned because too many factors will result in an excessively large number of DoE runs (Fig. 3.15) and poor accuracy in surface fit. The 'fishbone' diagram introduced in Chapter 1 can be used to brainstorm all possible influential factors. Unlike the Taguchi DoE method where the control factors and noise factors are separated into two different DoE arrays, the control and noise factors in a RSM DoE can be mixed more efficiently in one DoE array. Factor screening and range screening can be conducted with a first DoE by using a first-order model and a large number of factors along with a large factor range in order to identify important main effects and interaction effects. Fewer important factors along with a relatively smaller factor range around the optimum can be identified and built into a second refined DoE by using a higher-order emulator model in order to achieve better resolution for optimization. The factor range needs to reflect the operation region that can be neither too small nor too large. A too small range may result in an important factor behaving insignificantly or like a constant in the emulator model so that it does not provide much value to reflect its sensitivity in the DoE. A too large range often results in an extreme combination of the factor settings that may give unreasonable or impossible simulation results. For example, when the exhaust restriction is excessively high and the turbine area is extremely large, it will result in an extremely low air–fuel ratio that is even below the stoichiometric value.



3.15 DoE factors and number of DoE runs.

Once the number of factors and the factor ranges are determined, a DoE design method needs to be selected to generate the factor levels and the DoE runs. Different design methods give a different size of the DoE and different accuracy and predictability of the emulator model. The selection of factor levels depends on the emulator model structure (i.e., the order and the interaction terms used) and the DoE design method. Two-level DoE designs are sufficient for fitting the models that contain the first-order main effects and the low-order interactions, such as those used in the screening analysis. If a second-order model is needed (e.g., for the purpose of optimization), the factor needs to have at least three levels in order to estimate the quadratic terms. If a third-order model is needed, the factor needs to have at least four levels. The minimum required number of DoE design points (or runs, denoted as n) also depends on the emulator model structure: $n \geq p$, where p is given by equation 3.15. Figure 3.15 shows the number of DoE runs required in some particular DoE designs. The number of runs depends on the number of factors, DoE design method, and the polynomial order of accuracy of the emulator. Moreover, it should be noted that in order to ensure good resolution in factor spacing in the experimental DoE design, the interval between the levels needs to be large enough to cover three times the standard deviation of the factor during the experimental test. However, such a restriction and the issue of statistical scattering do not exist in a simulation DoE.

The statistical DoE design method affects the efficiency and quality of RSM emulator model building. The theory of the statistical properties of factorial designs is very complex. Many conclusions depend on the physical nature of the problem or assumed regression model. Some general guidelines for statistical DoE design are summarized below.

The first guideline is to use full factorial and fractional factorial designs appropriately. Full factorial design provides the most complete information to describe the relationship among the factors. But a full factorial design with more than three levels is not practical for engine system simulations when the number of factors is more than five or six due to the excessively large number of DoE runs (Fig. 3.15). Moreover, compared to the much more efficient fractional factorial design, full factorial design is often not necessary because it provides only marginally more information at the expense of a huge increase in the number of DoE runs.

To further illustrate the role of factorial design on the emulator model structure, consider the 2^4 full factorial design with four factors X_1, X_2, X_3 and X_4 at two levels, –1 and +1 (Table 3.3) for a first-order emulator model. The number of DoE runs is $2^4 = 16$. It is possible to estimate at maximum 16 model coefficients for all the effects contained in these 16 runs. The coefficients are distributed as follows: 1 constant term, 4 first-order terms (without the second-order terms due to only two factor levels), 6 two-factor interactions, 4 three-factor interactions, and 1 four-factor interaction. The corresponding model structure of the emulator is:

$$\begin{aligned} Y = {} & C_0 + C_1X_1 + C_2X_2 + C_3X_3 + C_4X_4 + \\ & C_5X_1X_2 + C_6X_1X_3 + C_7X_1X_4 + C_8X_2X_3 + C_9X_2X_4 + C_{10}X_3X_4 + \\ & C_{11}X_1X_2X_3 + C_{12}X_1X_2X_4 + C_{13}X_2X_3X_4 + C_{14}X_3X_4X_1 + \\ & C_{15}X_1X_2X_3X_4 \end{aligned} \tag{3.21}$$

Table 3.3 Illustration of full factorial design and confounding effects in fractional factorial design

Full factorial design					Fractional factorial design				
DoE run number	X_1	X_2	X_3	X_4	DoE run number	X_1	X_2	X_3	X_4 (generator constructed by the sign of $X_1 \cdot X_2 \cdot X_3$)
1	−1	−1	−1	**−1**	1	−1	−1	−1	**−1**
2	+1	−1	−1	**−1**	10	+1	−1	−1	**+1**
3	−1	+1	−1	**−1**	11	−1	+1	−1	**+1**
4	+1	+1	−1	**−1**	4	+1	+1	−1	**−1**
5	−1	−1	+1	**−1**	13	−1	−1	+1	**+1**
6	+1	−1	+1	**−1**	6	+1	−1	+1	**−1**
7	−1	+1	+1	**−1**	7	−1	+1	+1	**−1**
8	+1	+1	+1	**−1**	16	+1	+1	+1	**+1**
9	−1	−1	−1	+1					
10	+1	−1	−1	+1	DoE run number	X_1	X_2	X_3	X_4 (generator constructed by the sign of $-X_1 \cdot X_2 \cdot X_3$)
11	−1	+1	−1	+1	9	−1	−1	−1	+1
12	+1	+1	−1	+1	2	+1	−1	−1	−1
13	−1	−1	+1	+1	3	−1	+1	−1	−1
14	+1	−1	+1	+1	12	+1	+1	−1	+1
15	−1	+1	+1	+1	5	−1	−1	+1	−1
16	+1	+1	+1	+1	14	+1	−1	+1	+1
					15	−1	+1	+1	+1
					8	+1	+1	+1	−1

It is found that usually the lower order terms are relatively more important than the higher order terms in the emulator, and the three-factor or four-factor interaction terms are negligible to the response. This means that there is usually a redundancy in such an emulator model as equation 3.21 for the full factorial design. Those redundant runs corresponding to the higher-order interaction terms can be eliminated without affecting the accuracy of the emulator model. Fractional factorial designs address how to reduce such a redundancy. Moreover, fractional factorial design is very useful in screening many factors with a minimum number of runs.

Confounding is the most important concept in fractional factorial design. Confounding means when the number of DoE runs and the number of terms in the emulator model are reduced from the full factorial design, the estimated coefficients of the remaining terms in the model actually represent the combined factor effects mixed up with each other (i.e., the effects cannot be estimated completely independently from each other). In the above example, if the number of DoE runs has to be reduced to 8 (i.e., a fractional factorial design $2^{4-1} = 8$) from 16 (i.e., the full factorial design $2^4 = 16$), the maximum allowable terms in the emulator has to be reduced to 8 from equation 3.21 accordingly. There are two questions here: (1) How to reduce the number of DoE runs? (2) Which terms in equation 3.21 must be deleted accordingly? These questions are answered as follows.

From Table 3.3, it is observed that if only the DoE runs 1–8 are used, although all the main and interaction effects of the factors X_1, X_2 and X_3 are completely captured, the effects of X_4 are completely missed. As a result, all the terms involving X_4 in the emulator in equation 3.21 must be deleted. A similar argument applies for other factors. Therefore, the factor levels of X_4 in the DoE runs 1–8 have to be changed differently compared with the levels used in the full factorial design in order to add certain effects of X_4. In the fractional factorial example in Table 3.3, the main effects of all the factors can still be clearly estimated even if X_4 changes its level settings. However, there is a penalty that certain interaction effects are mixed or cannot be estimated clearly due to the factor level change of X_4. Such an effect is called 'confounding' when a full factorial design is reduced to a fractional factorial design. Confounding is essentially the inability to separate DoE information about individual factors or interactions.

There are generally two types of confounding generator methods: one using the original sign of the product of other factors for the generator (e.g., $X_4 = X_1X_2X_3$); and the other using the negative sign (e.g., $X_4 = -X_1X_2X_3$). The two generator methods are shown in Table 3.3 for the two fractional factorial designs constructed for X_4 as the generator. Note that in this example the four factors are rotatable (i.e., they are equally switchable among the factors). However, in the scenarios with more factors, the confounding pattern may be regulated in a desired direction or for a desired factor by adjusting the fractional factorial design. Due to confounding some interaction terms have to be removed from the emulator model. In this example, a maximum of eight terms in the model is allowed for eight DoE runs, for example,

$$Y = C_0 + C_1X_1 + C_2X_2 + C_3X_3 + C_4X_4 + C_5X_1X_2 + C_7X_1X_4 + C_{10}X_3X_4 \tag{3.22}$$

Keeping which terms in the model depends on the importance of the term for a particular problem.

Confounding is measured by 'DoE design resolution'. Low resolution means strong (or complicated) confounding of effects and hence poor estimate of the affected model coefficients. On the other hand, high resolution means weak confounding and an increase in the number of DoE runs. According to Eriksson *et al*. (1999), the resolution III design corresponds to a design in which the main effects are confounded with the two-factor interactions, and the two-factor interactions confounded with each other. This is undesirable in factor screening. Therefore, the resolution III designs should be used with great care. The resolution IV designs have the main effects unconfounded with the two-factor interactions, but the two-factor interactions are still confounded with each other. The resolution V or higher has the two-factor interactions unconfounded with each other. This means that the resolutions V and above are almost as good as the full factorial designs with very little confounding effects, but they require a large number of runs. Eriksson *et al*. (1999) concluded that 'the resolution IV designs are recommended for screening, because they offer a suitable balance between the number of factors screened and the number of experiments needed'.

Different fractional reduction patterns yield different confounding patterns. Confounding pattern affects the selection of the terms in the emulator model structure. In other words, if the number of terms in the emulator model has been determined, an appropriate confounding pattern can be selected in order to make the number of DoE runs greater than the number of model terms. Table 3.4 summarizes such a relationship. For example, if a fractional DoE of 2^{k-1} is used, the eight DoE runs for $k = 4$ cannot fit a complete model that has eleven terms. Either some polynomial terms have to be dropped from the model or the number of DoE runs has to be increased. Table 3.5 summarizes the resolutions of fractional factorial designs. Compared with the full factorial design, fractional factorial designs may greatly reduce the number of DoE runs for 'factor screening' at a penalty of confounding effects and possibly associated loss of accuracy or predictability in the emulator models. Good fractional factorial designs can give low uncertainty in the model coefficients and small prediction errors with a small number of DoE runs.

The second guideline for statistical DoE design is to differentiate the use of orthogonal designs and non-standard designs. In fractional factorial design,

Table 3.4 Relationship between confounding in fractional factorial design and emulator model structure

Number of factors (k)	Required emulator terms for first-order with interaction	Required DoE runs of 2-level full factorial 2^k	Required DoE runs of 2-level fractional factorial 2^{k-1}	Required DoE runs of 2-level fractional factorial 2^{k-2}	Required DoE runs of 2-level fractional factorial 2^{k-3}	Required emulator terms for second-order with interaction	Required DoE runs of 3-level full factorial 3^k	Required DoE runs of 3-level fractional factorial 3^{k-1}	Required DoE runs of 3-level fractional factorial 3^{k-2}	Required DoE runs of 3-level fractional factorial 3^{k-3}
2	4	4				6	9			
3	7	8	4 (III)			10	27	9		
4	11	16	8 (IV)			15	81	27	9	
5	16	32	16 (V)	8 (III)		21	243	81	27	9
6	22	64	32 (VI)	16 (IV)	8 (III)	28	729	243	81	27
7	29	128	64	32 (IV)	16 (IV)	36	2,187	729	243	81
8	37	256	128	64	32 (IV)	45	6,561	2,187	729	243
9	46	512	256	128	64	55	19,683	6,561	2,187	729
10	56	1,024	512	256	128	66	59,049	19,683	6,561	2,187
11	67	2,048	1,024	512	256	78	177,147	59,049	19,683	6,561
12	79	4,096	2,048	1,024	512	91	531,441	177,147	59,049	19,683
13	92	8,192	4,096	2,048	1,024	105	1,594,323	531,441	177,147	59,049
14	106	16,384	8,192	4,096	2,048	120	4,782,969	1,594,323	531,441	177,147
15	121	32,768	16,384	8,192	4,096	136	14,348,907	4,782,969	1,594,323	531,441

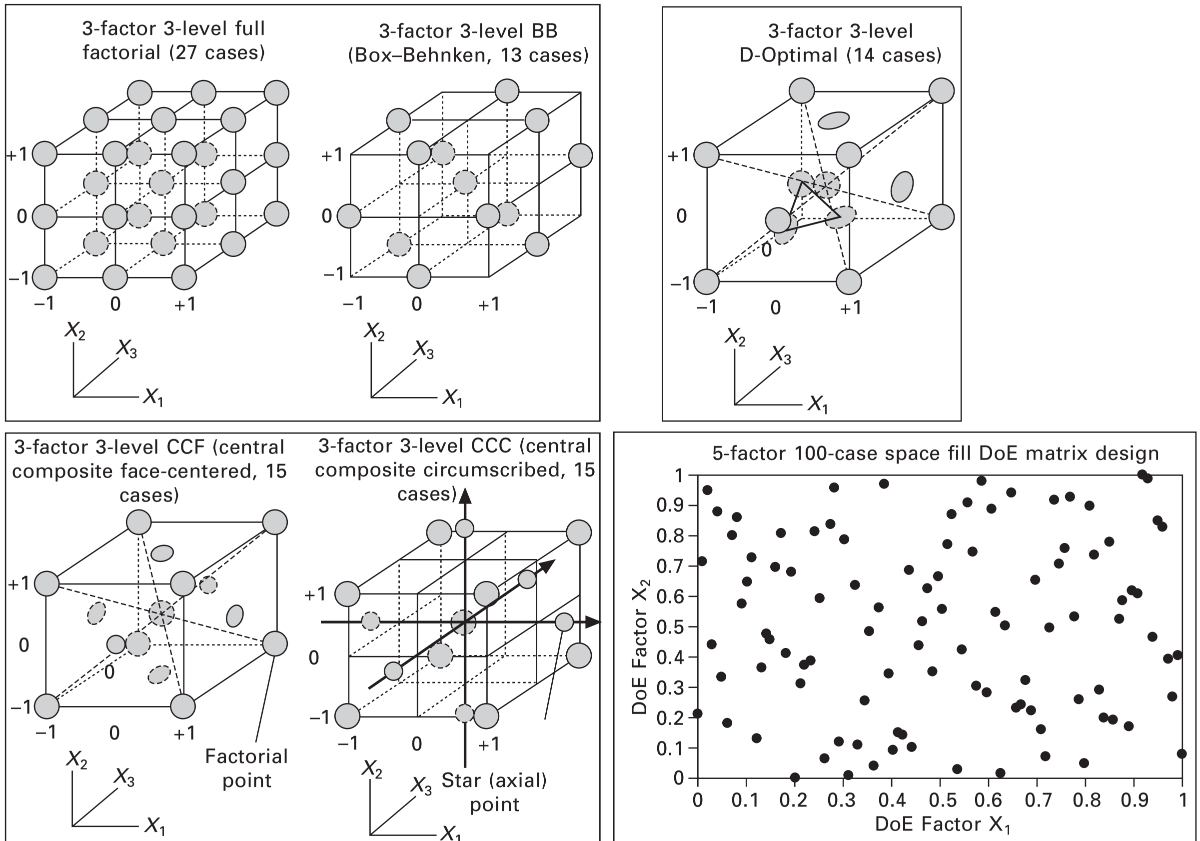
Note: The Roman numeral in parentheses indicate the DoE design resolution.

Table 3.5 Resolution of fractional factorial designs (based on the analysis by Eriksson *et al.*, 1999)

Number of DoE runs	Number of factors					
	3	4	5	6	7	8
4	2^{3-1} Resolution III	*Not available*	*Not available*	*Not available*	*Not available*	*Not available*
8	2^3 Full factorial	2^{4-1} Resolution IV	2^{5-2} Resolution III	2^{6-3} Resolution III	2^{7-4} Resolution III	*Not available*
16	*Not available*	2^4 Full factorial	2^{5-1} Resolution V	2^{6-2} Resolution IV	2^{7-3} Resolution IV	2^{8-4} Resolution IV
32	*Not available*	*Not available*	2^5 Full factorial	2^{6-1} Resolution VI	2^{7-2} Resolution IV	2^{8-3} Resolution IV

as pointed out by Myers and Montgomery (2002), standard designs, with a dominant preferable property of orthogonality, are the optimal design for first-order models or first-order-with-interaction models. But in the case of second-order models, the standard RSM designs are rarely optimal designs. The famous Plackett–Burman designs are special fractional factorial designs of resolution III (e.g., many used by the Taguchi DoE method in its inner array for control factors), and hence they can only be used to estimate main effects rather than interaction effects. They also noted that for the purposes of RSM analysis, the standard central composite (CCC/CCF) or the Box–Behnken (BB) designs should be used whenever possible, and the central composite design is the most popular class of second-order designs. However, it is found that the non-standard design of space-filling is also very effective for engine applications. The space-filling design is also known as Latin hypercube sampling, a random sampling of a prescribed number of DoE points (runs), n, to fill the factor space. The n points (runs) project onto n different levels in each factor, as shown in Fig. 3.16. The space-filling design is especially useful in the following situations: (1) a large design space; (2) large factor intervals; (3) a large number of factors; (4) many interactions; (5) the need to build highly accurate third-order emulator models. Figure 3.16 also shows several other DoE designs that have three factors and three levels. Note that the DoE design points (runs) of the CCC expand to five levels of factor settings residing on a sphere and hence the factor region is symmetrical. The CCF design has strictly three factor levels in a cube or hypercube. The CCC design is slightly better than the CCF design (Eriksson *et al.*, 1999).

The third guideline for statistical DoE design is to pay attention to rotatability



3.16 DoE matrix designs.

and assign factors in proper directions if needed. Rotatability means the statistical design provides equal precision of prediction in all directions. It is a very important property because the nature of the response surface is usually unknown prior to running the DoE cases. Rotatability in composite designs can be controlled by adjusting the distance of the 'star points' (axial points) to the center point. According to Montgomery (1991), any first-order orthogonal design is rotatable; the face-centered central composite designs are not rotatable, and this could be a serious disadvantage; Box–Behnken design is rotatable or nearly rotatable.

The fourth guideline for statistical DoE design is to tailor the non-reachable 'dead corners' in the DoE factor space with proper designs. In the DoE designs with an irregular factor space or polyhedron (e.g., corners missing in a cubic space of standard design, or the region is constrained by certain factor relationships), the Box–Behnken or non-standard D-Optimal design should be considered. The Box–Behnken design does not contain any points at the vertices of the cubic region that is created by the upper and lower limit levels of each factor (i.e., no corner points, Fig. 3.16). This feature could be advantageous when the points on the corners of the cube represent the factor-level combinations that are impossible to test or simulate due to certain constraints. The D-Optimal designs are constructed by including some of the extreme vertices of the constrained region, centers of edges between the hyper-cube corners, or the centers of the faces of the hyper-cube, and so on. The D-Optimal design makes efficient use of the entire factor space, and it can be a preferred choice when there are no classical designs that can well investigate the irregular region, and when the number of DoE runs that can be afforded is smaller than the number of runs of any available classical design. The D-Optimal method may tailor the corners of the factor space, and it also provides as much orthogonality as possible between the columns in the design matrix, hence maximizes the output information for a given number of DoE runs. In fact, the D-Optimal was regarded as the most suitable statistical design method by some authors for engine calibration due to its capability of handling the constraints on factor combinations (Roepke and Fischer, 2001). It should be noted that the irregular factor space occurs frequently in diesel engine system design. For example, a factor combination of a very high exhaust restriction, a very large VGT vane opening, and a very large EGR valve opening will result in an extremely low air–fuel ratio in some DoE runs, which belong to the impossible 'corners'.

Figure 3.17 shows an example of DoE input factor set-up used in diesel engine system design. In summary, the practical rules of RSM DoE design for diesel engine system design include the following:

1. The first-order model with two factor levels can be used to preliminarily screen the factors. The second-order or even third-order with more factor

DoE input factor range (Note: The bold data means the default values, i.e., 'L-def'.)

DoE input parameter index	Input parameter name	level 1 (= –1)	level 2 (= –0.333)	level 3 (= 0)	level 4 (= 0.333)	level 5 (= +1)
X_1	Overall turbine efficiency	57%	60%	63%	**66%**	69%
X_2	Overall compressor efficiency	64%	67%	70%	**73%**	76%
X_3	Ovberall turbine size (turbine area or VGT vane opening)	0.36	0.4	**0.44**	0.48	0.52
X_4	EGR circuit flow restriction (mm)	10	**13.23**	15.81	18.03	20
X_5	EGR cooler outlet gas temperature (K)	450	475	**500**	525	550
X_6	Charge air cooler cooling heat transfer coefficient (W/m^2.K)	550	650	**750**	850	950
X_7	Exhaust restriction (CDPF flow restriction parameter)	1.52	1.37	1.22	1.07	**0.92**

DoE input factor set-up with 'one-factor-at-a-time' method (Note: 'L1' means Level 1; 'L-def' means default value.)

Input	DoE simulation case →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
X_1	Overall turbine efficiency	L1	L2	L3	L4	L5	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	...
X_2	Overall compressor efficiency	L-def	L-def	L-def	L-def	L-def	L1	L2	L3	L4	L5	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	...
X_3	Turbine size	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L1	L2	L3	L4	L5	L-def	L-def	L-def	L-def	L-def	...
X_4	EGR circuit flow restriction (mm)	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L1	L2	L3	L4	L5	...
X_5	EGR cooler outlet temperature (K)	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	...
X_6	CAC cooling heat transfer coeffi (W/m^2.K)	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	...
X_7	Exhaust restriction (CDPF flow design)	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	L-def	...

DoE input factor set-up with the D-Optimal method (Note: –1 means Level 1; –0.3333 means Level 2; 1 means Level 5)

Input	DoE simulation case →	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
X_1	Overall turbine efficiency	1	–1	1	1	1	1	1	1	1	1	1	1	1	–1	–1	–1	–1	...
X_2	Overall compressor efficiency	1	1	1	1	1	–1	1	1	1	–1	1	1	–1	–1	–1	–1	–1	...
X_3	Turbine size	–1	–1	1	–1	–1	1	1	–1	1	–1	1	–1	–1	–1	–1	–1	–1	...
X_4	EGR circuit flow restriction	–1	1	1	1	1	1	1	1	–1	–1	–1	–1	1	–1	1	1	0.3333	...
X_5	EGR cooler outlet gas temperature	–1	–1	–1	1	–1	–1	1	–1	1	–1	–1	1	1	1	–1	0.3333	–1	...
X_6	CAC cooling heat transfer coefficient	–1	–1	–1	–1	1	1	1	–1	–1	1	1	1	1	–1	–1	1	1	...
X_7	Exhaust restriction (CDPF flow design)	–1	–1	–1	–1	–1	–1	–1	1	1	1	1	1	1	–0.3333	0.3333	–1	1	...

3.17 Illustration of DoE input factor set-up.

levels should be used to build the emulator regression model for the optimization in engine system design.
2. For a very small number of factors (e.g., two or three), the full factorial design may be used to conduct the DoE runs and build the emulator models.
3. For a large number of factors, design resolution of the confounding effect should be checked before a standard fractional factorial design and emulator model structure are selected.
4. Normally, the central composite designs with three to five factor levels should be used for the second-order emulator models. Sometimes, as many as ten factor levels can be used or realized by combining two five-level DoE that have two different factor ranges in order to build emulators.
5. The nonstandard space-fill design is very effective for the third-order emulator models.
6. For the irregular or constrained factor space where the 'corner' points are undesirable or impossible to run, or when the region is non-cubic, the Box–Behnken or the D-Optimal designs can be considered.
7. Adding complementary runs for unconfounding is a common technique to improve the model accuracy iteratively. Adding supplementary DoE runs to upgrade the emulator model (e.g., to cubic order) is also common, and this can be achieved with the D-Optimal supplementary design.

3.2.4 Analysis and optimization with response surface models

Canonical analysis

Once the emulator model is obtained by surface fit, it can be used to predict responses inside the factor space. It should be noted that the regression model should not be used for extrapolation outside the factor range. The model can also be used to analyze the sensitivity characteristics of the factor–response relationship. Most importantly, the model can be used to conduct optimization to search for the optima located on the response surface. A theoretical analysis of the optimization is reviewed below on the basis of canonical analysis.

The general form of a second-order response surface emulator model is given by

$$Y = C_0 + \sum_{i=1}^{k} C_i X_i + \sum_{i=1}^{k} C_{ii} X_i^2 + \sum_{i,i<j} \sum_{j=2}^{k} C_{ij} X_i X_j \qquad 3.23$$

In the optimization, the stationary point refers to the point of factor settings corresponding to zero partial derivatives of the response with respect to all the factors $\partial Y/\partial X_1 = \partial Y/\partial X_2 = \ldots = \partial Y/\partial X_k = 0$. The stationary point may represent a maximum response, a minimum response, or a saddle point (i.e., neither a maximum nor a minimum in a hyperbolic response). Denote Y as the approximation of Y_{obs} by ignoring the error υ. According to Montgomery (1991), equation 3.24 can be written in matrix notation as follows with the vectors $\hat{X}$ and $\hat{C}$, and a $k \times k$ matrix $\bar{C}$:

$$\begin{cases} Y = C_0 + \hat{X}'\hat{C} + \hat{X}'\bar{C}\hat{X} \\ \text{where } \hat{X} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_k \end{bmatrix} \quad \hat{C} = \begin{bmatrix} C_1 \\ C_2 \\ . \\ . \\ . \\ C_k \end{bmatrix} \quad \bar{C} = \begin{bmatrix} C_{11} & \frac{C_{12}}{2} & \cdots & \frac{C_{1k}}{2} \\ & C_{22} & \cdots & \frac{C_{2k}}{2} \\ & & \cdots & \\ \textit{symmetric} & & & C_{kk} \end{bmatrix} \end{cases} \tag{3.24}$$

The optima can be found by setting the derivative of Y with respect to the elements of the vector $\hat{X}$ equal to zero. This is expressed as:

$$\frac{\partial Y}{\partial \hat{X}} = \hat{C} + 2\bar{C}\hat{X} = 0 \tag{3.25}$$

The stationary point is the solution of equation 3.25, and given by

$$\hat{X}_0 = -\frac{1}{2}\bar{C}^{-1}\hat{C} \quad \text{and} \quad Y_0 = C_0 + \frac{1}{2}\hat{X}_0'\hat{C} \tag{3.26}$$

Canonical analysis is used to analyze the nature of the stationary point. Canonical analysis refers to a coordinate translation and rotation from the original coordinate system to be anchored at the stationary point, as shown in Fig. 3.2c earlier. The polynomial model can be transformed to a new coordinate system with the origin at the stationary point (X_{10}, X_{20}) and with the principal axes of the new coordinate system ($w_{c,1}$ and $w_{c,2}$) coming from the fitted response surface. This coordinate transformation results in the fitted emulator model to be transformed into the canonical form as follows (Montgomery, 1991):

$$Y = Y_0 + \lambda_{c,1}w_{c,1}^2 + \lambda_{c,2}w_{c,2}^2 + \ldots + \lambda_{c,k}w_{c,k}^2 = Y_0 + \sum_{i=1}^{k}\lambda_{c,i}w_{c,i}^2 \tag{3.27}$$

where $w_{c,1}$, $w_{c,2}$, . . ., $w_{c,k}$ are the transformed factors called canonical variables, and $\lambda_{c,1}$, $\lambda_{c,2}$, . . ., $\lambda_{c,k}$ are the eigenvalues or the characteristic roots of the matrix $\bar{C}$.

The nature of the stationary point on the response surface (maximum, minimum, or saddle point) can be determined from the signs and magnitudes of the eigenvalues. If the eigenvalues or the roots of the determinant equation of the canonical analysis are all positive, then the stationary point is a minimum. If the eigenvalues are all negative, then the stationary point is a maximum. If the eigenvalues have different signs, the stationary point is a saddle point. A very small (essentially zero) eigenvalue results in a ridge system where the response surface is elongated along the direction of that canonical eigenvalue. Ridge analysis of the response surface is very important, and is explained in detail by Myers and Montgomery (2002). All these theoretical analyses may help locate the optima (or stationary) point of the response surface, or point out the most efficient direction (along certain canonical axis) to revise the DoE factor range in statistical DoE design in order to approach the optima efficiently. These analyses may avoid the blind trial-and-error adjustment in the DoE factor range. The theoretical analysis may also replace the laborious graphical search for the global optima when the number of factors is large.

Optimization search algorithms

Many methods can be used to search for the optima, such as the simple but powerful method of gradient search. The gradient search includes steepest ascent search (in the direction of the maximum increase in the response to reach its highest peak) and steepest descent search (in the direction of the maximum decrease in the response to arrive at the bottom of its valley). Basically, the direction in gradient search is perpendicular to the parallel lines of the response contour plot. Gradient-based search algorithms sometimes have problems with numerical errors or oscillations around the local optima. Genetic algorithms (Parmee and Watson, 2000) or other search algorithms such as SIMPLEX do not require evaluating the derivatives of the model function (Zottin *et al*., 2008). Therefore, sometimes they can be more robust than the gradient-based algorithms, although they still can converge to the local optima rather than the global optima.

The required optimization technique for engine system design is the one which can handle the following: complex trade-offs; constraints and the responses having high nonlinearity (e.g., the behavior of engine soot and BSFC); or multiple local optima (e.g., badly behaved responses or oscillating numerical errors). For details of different optimization search algorithms, the reader is referred to Eriksson *et al*. (1999) and Thiel *et al*. (2002).

Advanced data display for sensitivity and optimization

In engine system design, due to the complexity of many different engine responses and the large amount of sensitivity data produced by the emulator models, presenting the analysis results of sensitivity study and optimization in an accessible and concise format is very important. In the RSM optimization, using a two-dimensional contour plot on a two-factor or two-response domain is a very effective method of graphical display for understanding system sensitivity and locating the optima.

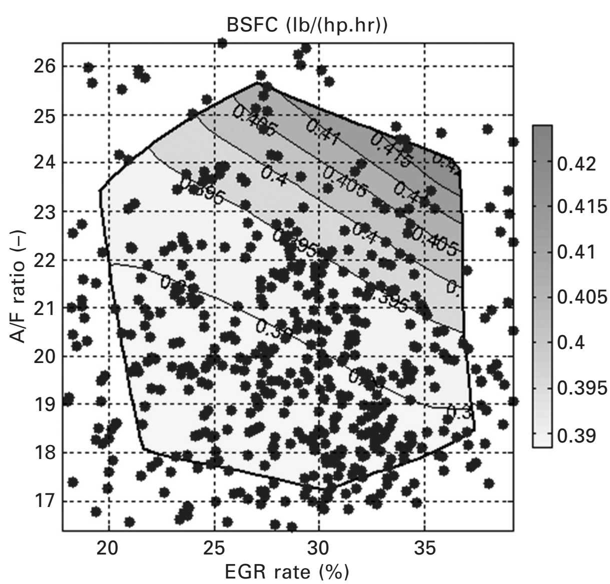
The most basic method to construct the contour plot is to compute response contours with the emulator model by using constant intervals in the X_1 vs. X_2 factor domain while holding all other factors constant. This method becomes awkward when the number of factors is much larger than three because many factors must be held constant in turns to construct such contours.

A four-dimensional response contour plotting method to handle four factors was proposed by Eriksson *et al.* (1999). They expanded a single two-dimensional plot to an array of grids of contour plots along the horizontal and vertical directions so that two more factors (dimensions) can be added as an outer array of plot. The two-dimensional contour plots for sweeping the first two factors are first created at each factor setting of the other two factors. For example, three two-dimensional plots are created side-by-side horizontally at three factor levels of the third factor. Then they are cloned vertically row-by-row at four factor levels of the fourth factor. In this way, a grid block of $3 \times 4 = 12$ two-dimensional contour plots can be created so that the parametric trends and the optimal design point can be observed conveniently for four factors in a one-factor-at-a-time parametric manner.

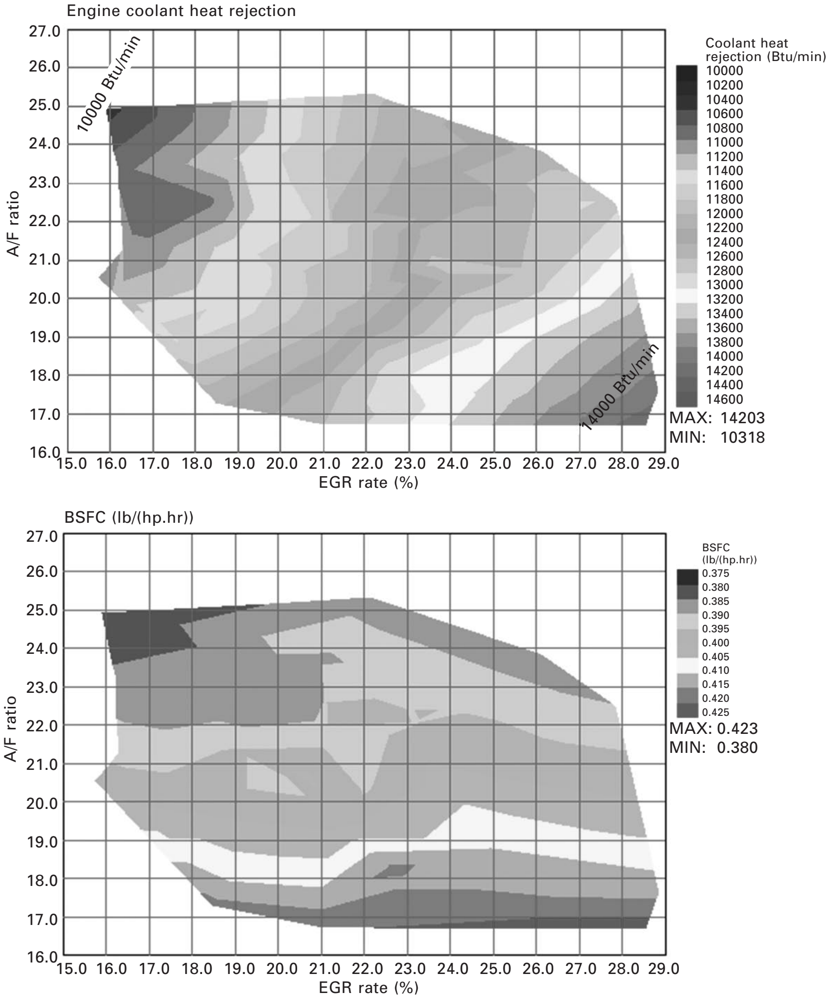
A much more powerful graphical approach proposed earlier (Fig. 3.6) is to generate the optimum contour maps by using constrained optimization where all the points in the factor domain are optimized, for example, in the sense of minimum fuel consumption. In such optimization using the emulator model, the factor settings at each point are no longer constant. Instead, they are optimized to correspond to the minimum fuel consumption subject to the constraints of each particular pair of X_1 and X_2 values. An even more powerful optimum contour map can be generated by mapping the factor domain (X_1 vs. X_2) to a response domain (Y_1 vs. Y_2). In this case, every point in the response domain is obtained by constrained optimization subject to the constraints of each particular pair of Y_1 and Y_2 values. The required factor settings for achieving the minimum fuel consumption map and the corresponding other responses can also be plotted as contour maps in the Y_1 vs. Y_2 response domain. The factor boundary can be computed by a search algorithm on such a response domain. The factor boundary is bounded by both the DoE factor range and constraints used in the optimization. The factor boundary in the response domain may give a Pareto frontier for the multi-objective optimization problem to optimize the trade-off between two responses on the horizontal and vertical axes of the map. In diesel engine system design, a frequently used response domain is the air system capability domain, which is represented by 'air–fuel ratio vs. EGR rate' (Fig. 3.18).

Figure 3.19 shows the DoE raw data of coolant heat rejection and BSFC plotted in the response domain without using the emulators or optimization. Figure 3.20 shows the optimized coolant heat rejection (in the sense of minimum BSFC) and BSFC by using the emulators. It is observed that Figs 3.19 and 3.20 give totally different data trends in heat rejection and BSFC. It shows that without optimization the raw DoE data can be misleading in the data trend because the factor settings used in a DoE matrix do not guarantee any trend in the raw data. The DoE factors used in this example include fuel injection timing, EGR valve opening, VGT vane opening, turbocharger efficiency, EGR cooler size, and charge air cooler size. This example shows the importance of using DoE emulator to process the DoE raw data and conduct optimization.

The information on parametric sensitivity of system response, generated either analytically or graphically, is often even more valuable than a single set of estimated optimum solutions. The theoretical advantages of ridge analysis, canonical eigenvalues, and the technique of optimum contour map can be combined in the future to generate a more powerful analysis tool for the RSM DoE design, optimization, and data display. This is an important research direction for optimization in diesel engine system design.



3.18 Illustration of raw DoE data and minimum BSFC optimization contours.

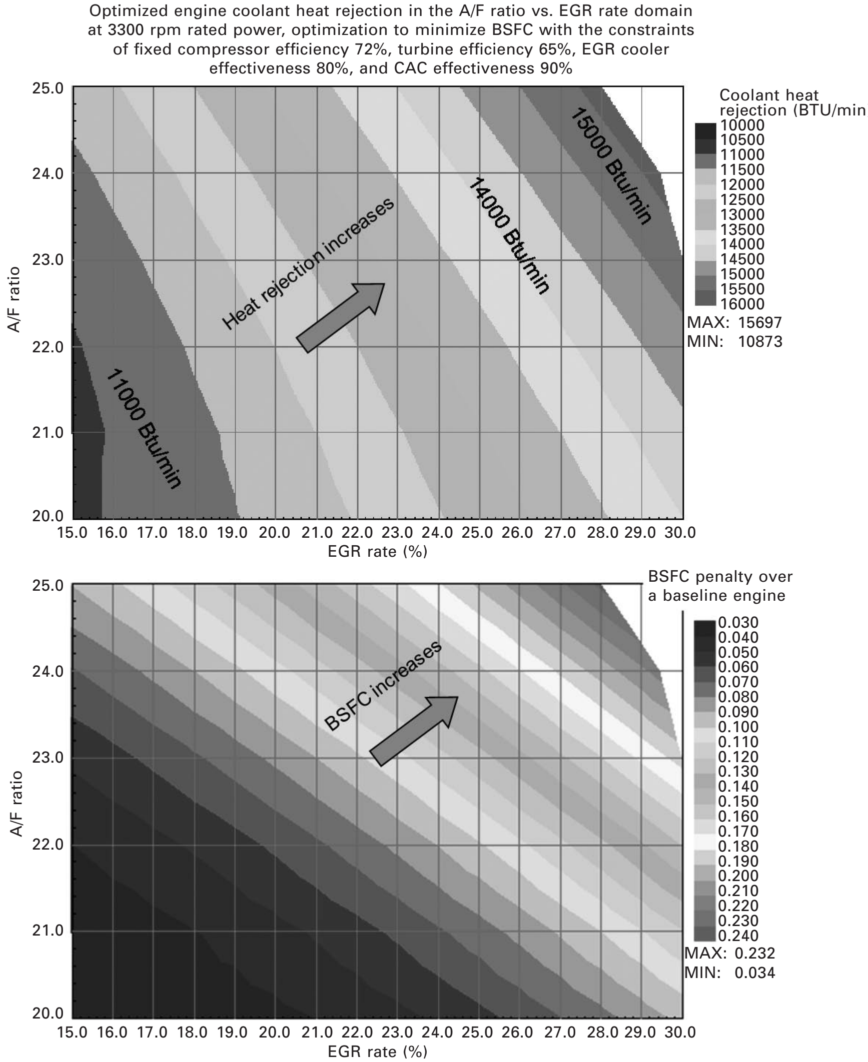


3.19 Raw DoE output data in response domain without optimization.

3.3 Advanced design of experiments (DoE) optimization in engine system design

3.3.1 Engine optimization with the Taguchi method

There are numerous studies using the Taguchi method in engine design and testing applications. The engine performance related design and calibration work is reviewed here. Baranescu *et al.* (1989) from Navistar used the Taguchi method and engine cycle simulation to analyze the effects of control factors and noise factors on diesel engine power. The analysis started with the cause-and-effect 'fishbone' diagram to brainstorm design factors. Five



3.20 Optimized DoE output data in response domain by using RSM emulators.

control factors (turbocharger build, intake flow area, exhaust flow area, intake cam, and exhaust cam) and four noise factors (ambient temperature, engine compression ratio, injection timing, and fueling rate) were selected. Most factors were set at three levels. They used an inner array for the control factors and an outer array for the noise factors. The signal-to-noise ratio of engine power was evaluated along with the mean value of engine power. They used the main effect plot to identify strong factors and weak factors. The interaction effect of two control factors, turbocharger build and intake flow area, was considered when a proper inner array L_{18} was selected. Possible trade-offs between the mean value and the signal-to-noise ratio of the engine power were pointed out and demonstrated by using their main effect plots side-by-side as a multi-objective optimization. Optimum control factor levels were selected for the best power (i.e., large mean) and robustness (i.e., large signal-to-noise ratio). The ANOVA study of the signal-to-noise ratio was provided including the interaction effect between the control factors. An estimate of cost savings due to the optimized design was also provided. This work is an important pioneering analysis of using the Taguchi DoE method on engine performance. It applied many important concepts of robust design and optimization in engine simulation applications.

Another important optimization work on diesel combustion and transient emissions using the Taguchi DoE method combined with the Monte Carlo simulation was reported by Navistar's Yan *et al.* (1993). Four important concepts were introduced to the area of optimization of diesel engine emissions and performance. They are: (1) measured signal-to-noise ratio for each DoE run; (2) transient optimization; (3) multi-objective optimization for multiple combustion parameters (PM, NO_x, HC, INSOL, SOF, and CBSFC); and (4) probability distribution sampling for analyzing the variability of production emissions with the Monte Carlo simulation. Understanding the statistical distribution of production emissions and reducing both the mean and the standard deviation of the transient emissions were the objectives of their analysis. Six control factors with three levels on each factor were selected in a single L_{18} array: piston-to-head clearance, nozzle flow, nozzle protrusion, spray cone angle, injection timing, and injection control pressure. Because it was an experimental work, they selected sufficiently wide intervals between the factor levels in order to cover three times the standard deviation of each design factor. The signal-to-noise ratio was used as an optimization response parameter to minimize the emissions variance by using the main effect plot. Unlike the signal-to-noise ratio calculation conducted by Baranescu *et al.* (1989) who used an inner L_{18} control factor array and an outer L_9 noise factor array, Yan *et al.* (1993) used engine transient measurements to obtain the mean, the variance, and the signal-to-noise ratio for each L_{18} DoE run. Each emission data in a DoE run is a summary point from the transient measurement. In contrast, Baranescu *et al.* (1989) used the outer array design points at a given inner array factor setting to compute the signal-to-noise ratio for each L_{18} run (as illustrated in Fig. 3.4) with steady-state engine cycle simulations. Yan *et al.* (1993) used the main effect plot to check the sensitivity of signal-to-noise ratio as a response with respect to all six control factors for six performance/emissions parameters. The data exhibited fairly nonlinear behavior of the signal-to-noise ratios with respect to the factor levels. The optimum factor setting was selected with the highest signal-to-noise ratio for each factor meanwhile minimizing the mean of PM and maintaining the mean of NO_x at a constant level. They also illustrated the method of calculating the optimum signal-to-noise ratio at 95% confidence interval and the estimated optimum response (emissions) values. The 95% confidence interval level is usually a good compromise between model accuracy and complexity. A high signal-to-noise ratio does not necessarily mean the statistical distribution of the emissions variability is acceptable. The authors continued their investigation by using the Monte Carlo simulation along with the emissions regression models obtained from an earlier step and other relevant models to calculate the statistical distributions of the emissions parameters for ten factors. The ten factors were: nozzle cone angle, nozzle protrusion, piston-to-head clearance, head gasket thickness, stem seal leak rate, oil contribution from other sources, timing, injection command pressure, nozzle flow, and intake port swirl ratio. The statistical distributions of the input factors were determined by actual engine testing data. A large amount of samples were taken as the input for the Monte Carlo simulation. The Monte Carlo simulation was run 500 times to obtain the result of emissions variability/scattering. They quantified the contribution from each factor on the standard deviation of the emissions distributions. This work leaped from the Taguchi DoE method of signal-to-noise ratio optimization to probability distribution analysis for performance variability, and it is one of the pioneering works in the Monte Carlo simulation for diesel engines.

Hunter *et al.* (1990) applied the Taguchi method to simultaneously optimize several diesel engine design and operating parameters for low emissions in a single cylinder engine. The control factors included engine compression ratio, nozzle area, nozzle protrusion, boost pressure, start-of-combustion timing, indicated mean effective pressure, and engine speed. The interactions included: compression ratio vs. nozzle area, compression ratio vs. nozzle protrusion, and nozzle area vs. nozzle protrusion. The responses included the mean and the signal-to-noise ratio of particulate matter, NO_x, HC, and smoke. They provided a detailed description of each step in the Taguchi method, especially the calculation procedure and formula for predicting the optimum responses (with 90% confidence interval) when each factor is independent and no significant interactions exist. Gardner (1992) used the Taguchi method to investigate the effects of changes in fuel spray cone angle, number of spray holes, nozzle hole area, nozzle tip protrusion, compression ratio, swirl level, and fuel injection timing on diesel engine combustion and emissions. He pointed out that, although the Taguchi method is a powerful tool for factor screening and optimization, it should be used with caution to understand the confounding and interaction effects in order to choose an appropriate orthogonal array and to avoid erroneous conclusions drawn from the main effect study. Win *et al.* (2002) used the Taguchi method to conduct an experimental study on diesel engine noise, emissions, and fuel economy. They used the signal-to-noise ratio and ANOVA to quantify the influence of engine speed, load, and fuel injection timing on performance response parameters.

Interesting experimental work was carried out by Yamamoto *et al.* (2002) by combining the Taguchi method (orthogonal array design and linear graph) with RSM (polynomial surface-fit empirical model) to optimize the parameters in fuel injection, turbine and EGR systems for low NO_x, PM emissions, and BSFC for a heavy-duty diesel engine. Their work is important for the following four contributions: (1) expanding the DoE optimization from a single speed–load mode to balancing all 13 modes in a steady-state emissions cycle; (2) checking optimization results against design constraints; (3) combining Taguchi's orthogonal array design with a special polynomial model; and (4) using a Chebychev's orthogonal polynomial as an empirical model in the Taguchi method. In their study, six calibration or design factors with three levels were used in a Taguchi L_{27} orthogonal array: VGT vane opening, EGR valve lift, injection timing, common rail injection pressure, injection nozzle diameter, and injection nozzle cone angle. The DoE method was applied at each engine mode point of the Japanese 13-mode emissions test cycle. After the optimum trade-off between the emissions and the BSFC at each mode point was obtained, the optimization for the entire 13-mode composite was made by a balancing between the modes. They used ANOVA to quantify the interaction effects among VGT vane opening, EGR valve lift, and fuel injection timing. It should be noted that they used a Chebychev's orthogonal polynomial, rather than a regular second-order polynomial, to fit a second-order empirical model including the interaction terms. In most other researchers' work using the Taguchi method, the main effect plot, the signal-to-noise ratio and confidence interval were usually used to identify the optimum factor levels and estimate the optimum responses. Yamamoto *et al.* (2002) did not take that approach. They used Chebychev's empirical equation to conduct a numerical parametric study at each mode. They took this different approach probably in an attempt to overcome the weakness of the Taguchi method in regression models. Checking the predicted optimization solution against design constraints was particularly mentioned in their work. The limiting constraints included smoke, turbocharger over-speed, engine over-speed, exhaust gas temperature, and maximum cylinder pressure. The authors probably did not realize that their work actually took a hybrid approach between the Taguchi method and RSM. Although it is worthwhile to further investigate the validity of using a low-resolution Taguchi orthogonal array with polynomial surface-fit, their attempt of using a Chebychev's orthogonal polynomial, which is a special polynomial, did indicate an opportunity worth further exploration in engine optimization theory.

The above examples show that the Taguchi DoE method has been successfully used in diesel engine performance development. The previous efforts have laid out a good foundation for this area to move forward to adopt the more advanced DoE optimization theories and techniques.

3.3.2 Engine calibration with RSM

Engine performance and emissions calibration is probably the area where the RSM DoE technique has been most widely used compared with other engine application areas. Engine calibration refers to tuning mechanically or electronically the adjustable parameters on the engine to reach desirable performance. In contrast to hardware design where a fixed sizing has to be chosen from a few candidates (e.g., engine compression ratio, turbine area, cooler size), a tunable parameter offers the flexibility of varying within a range, depending on different engine speed and load operating conditions or ambient conditions. Examples of mechanical tuning device include mechanical governor, mechanical fuel injection system, and pneumatically controlled turbine wastegate. Since electronic controls were applied to diesel engines starting in the 1990s, the flexibility and performance of the engine have been greatly improved. However, the complexity of engine calibration and the associated complexity in electronic control strategies and software have increased dramatically. There are many electronically controlled calibration parameters available in the engine, for example fuel injection timing, injection pressure, turbine wastegate opening, VGT vane opening, EGR valve opening, and intake throttle opening. Other possible calibration parameters include cooling control valve setting, variable valve timing or cam phasing angle, variable swirl control, and cylinder deactivation. Without using a DoE, finding the optimum setting for these calibration factors is almost impossible.

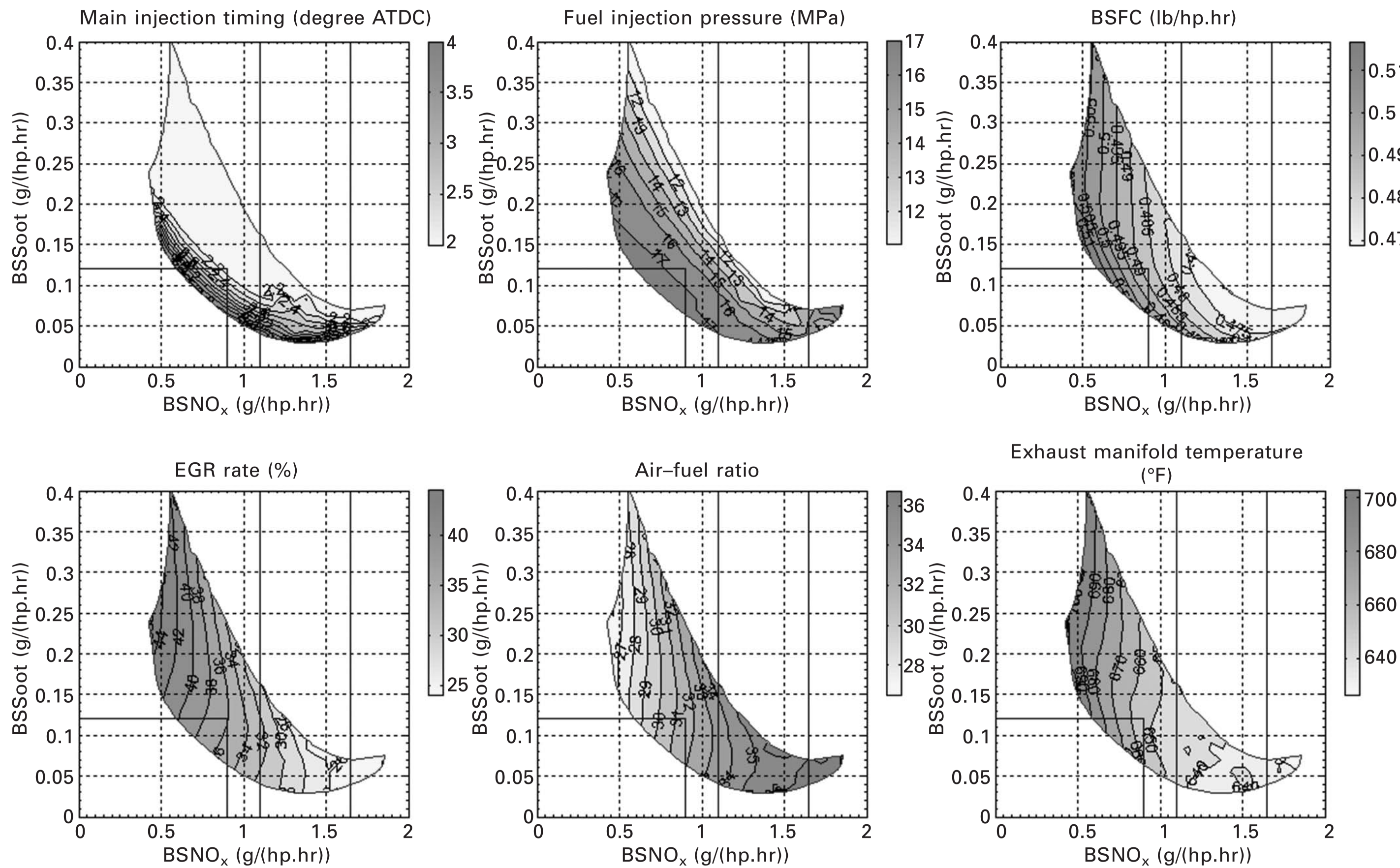
Roepke and Fischer (2001) introduced an efficient calibration method on variable valvetrain by using RSM for a gasoline engine. They pointed out that the D-Optimal is very suitable and efficient for engine calibration due to its capability of handling the constraints in factor combinations. The concept of model-based calibration was explained by Lumsden *et al.* (2004) in their RSM DoE work for stratified charge direct injection gasoline engines. They indicated that using a third-order model with the space-filling DoE design may yield better predictability than using the second-order model for hydrocarbons and BSFC. Mallamo *et al.* (2004) used RSM to optimize an off-road common-rail diesel engine performance with four calibration factors (main injection timing, pilot injection duration, pre-injection timing, and pre-injection duration) at three levels for each factor. The optimization was conducted for NO_x, PM, noise, and BSFC. It was proved that the model built based on the CCF design gave very similar results to the model built based on the full factorial design. Diesel engine calibration with DoE optimization techniques was also reported by Brooks *et al.* (2005).

Engine calibration has been moving in the direction of using RSM DoE with online automated testing, dynamic mapping, and offline optimization, as summarized by the MathWorks hosted SAE Panel Discussion (The MathWorks, 2007). Moreover, the powerful approach proposed earlier by using the two-dimensional minimum BSFC contour maps can greatly enhance the quality and efficiency of optimization work in both engine calibration and engine system design. Figure 3.21 shows an example of applying the approach of minimum BSFC contour maps to the DoE optimization of engine emissions calibration. Note that the factors and responses in the figure are all optimized in the entire map domain with the minimum BSFC. The boundary of the 'soot vs. NO_x' response domain is formed by the factor range of the DoE and the calibration constraints.

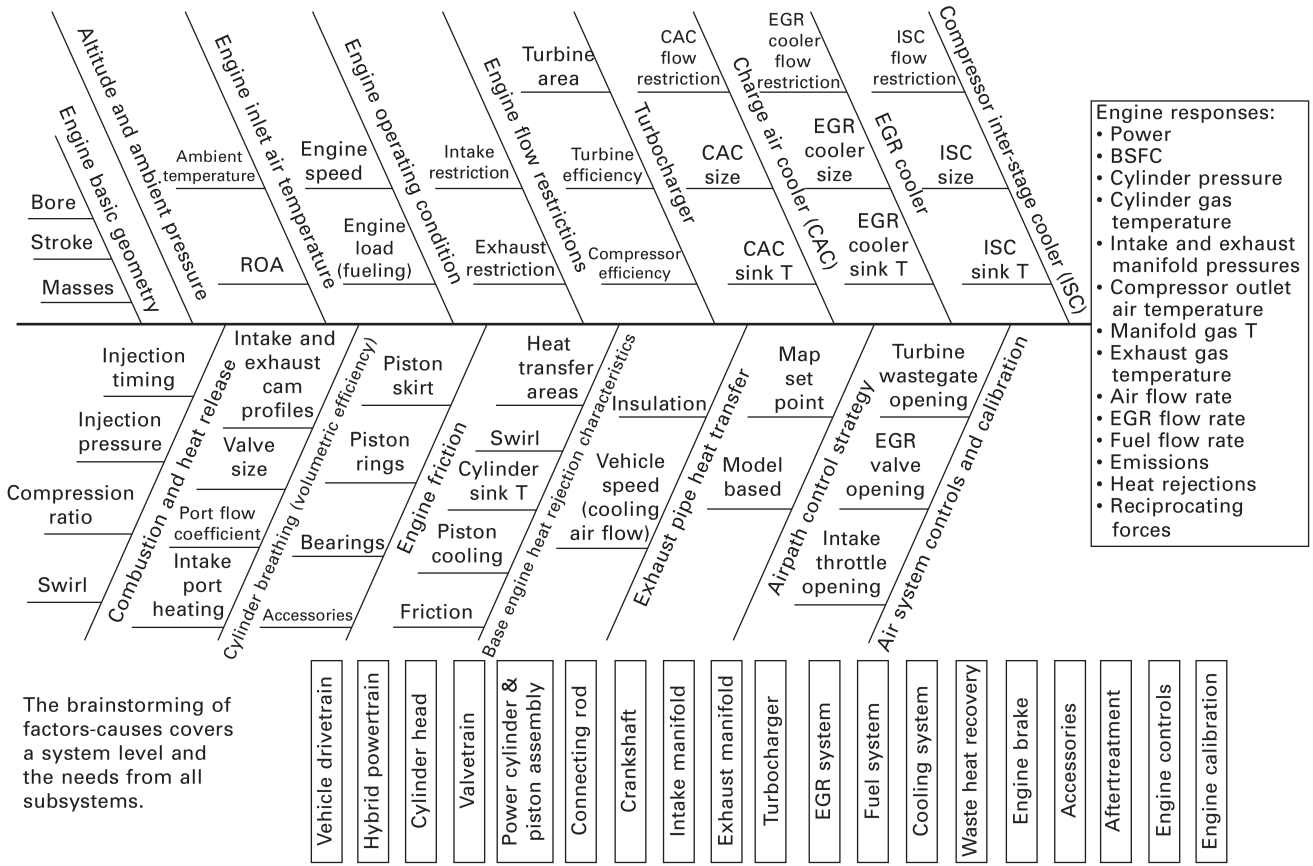
3.3.3 Engine system optimization with RSM

Diesel engine system design is conducted by using DoE optimization based on numerical simulation data since simulation can greatly reduce the cost and time of hardware testing during engine development. The control factors and noise factors can be brainstormed by using the cause-and-effect 'fishbone' diagram, as shown in Fig. 3.22 for engine performance simulation. Figure 3.22 shows all the important categories of factors in an engine performance model. The control factors may include hardware sizing or engine calibration parameters. The noise factors can be from piece-to-piece variation, environmental conditions, customer usage, or deterioration over time. A large number of DoE optimization examples in diesel engine system design are provided in Chapter 15 where the effects of subsystem interaction are thoroughly addressed.

It is worth noting that RSM was used by Dvorak and Hoekstra (1996) more than a decade ago for optimizing internal combustion engine performance. Recognizing the weakness of the Taguchi method in quantifying the interaction effects between engine performance factors, they used the central composite design in RSM and engine performance simulation to analyze the effects of eight design factors (intake valve diameter, exhaust valve diameter, intake runner length, intake beginning runner area, bore, stroke, intake cam duration, and exhaust cam duration) on engine power for a gasoline engine. A second-order regression model including the interaction effects were used. The optimization of engine power was conducted using the method of steepest ascent. It should be noted that a canonical analysis was performed to identify the canonical axis along which the greatest rate of change occurred for the engine power. A stationary point was derived as the optimum solution for the maximum engine power within the design space. It was pointed out by the authors that the principal axis information obtained from the canonical analysis was utilized to identify and construct the new factor range of an expanded DoE in order to further increase the



3.21 HD diesel engine steady-state emissions calibration optimization with DoE and RSM (mode B25).



3.22 Cause-and-effect diagram for engine system performance analysis.

engine power. This work is one of the earliest attempts that demonstrated that RSM techniques can be successfully applied to internal combustion engines.

RSM was also used by Rutter *et al*. (1996) in engine aftertreatment system optimization for a catalytic converter of a gasoline engine. They used a three-level CCF DoE design to study the optimum precious metal loading of platinum (Pt), palladium (Pd) and rhodium (Rh).

3.4 Optimization of robust design for variability and reliability

3.4.1 Overview of optimization for variability, reliability, and robustness

Many related terms in reliability engineering and robust engineering are used in the existing literature. Sometimes they are misused and this causes confusion. This section clarifies the differences among the key concepts related to optimization, uncertainty, variability, reliability, and robustness for advanced nondeterministic probabilistic optimization.

Nondeterministic probabilistic optimization

The earlier sections mainly discuss the deterministic optimization techniques to produce a single deterministic design solution for 'design for target' (either 'design for nominal' or 'design for limit'). Under-design or over-design can occur if the safety margin is not determined properly. The product may still not perform well in use when all sources of variation present even with an acceptable nominal design. Deterministic optimal design without considering variations is generally not reliable, not robust or not economic. The variation may cause the engine to operate at the conditions that violate the performance or durability constraints. The deterministic design generated by stacking up the worst-case tolerance is usually a non-economic over-design. Reliability-based design optimization (RBDO) and robustness-based optimization are required to handle variability. They are powerful optimization tools based on nondeterministic approaches for diesel engine system design.

There are three types of input factors in nondeterministic optimization:

- deterministic control factor (also called deterministic design variables in some other literatures)
- nondeterministic or random control factor (also called nondeterministic or random design variables)
- random noise factor (also called random design parameters).

Both the mean and the standard deviation of the control factors can be varied, while the settings of the noise factor cannot be controlled or changed. The output response is a function of the deterministic control factors, nondeterministic control factors and random noise factors. For example, fuel injection timing can be selected as a deterministic control factor; engine compression ratio can be selected as a nondeterministic control factor; and ambient temperature is a noise factor with random variations and it cannot be controlled. The objective function of RBDO and robust design can be any attribute measure such as performance, durability, packaging or cost. The constraints in the optimization usually include an evaluation of probabilistic reliability or failure rate. RBDO is used to optimize the system under the constraint of maximum allowable probability of failure (or conversely minimum acceptable reliability).

The need for nondeterministic probabilistic evaluation

The most efficient and economical way to handle the variations in design is to use a probabilistic analysis rather than stacking up the worst-case tolerances. The Monte Carlo simulation is the most widely used method in this area and plays a central role in RBDO and robust optimization. The Monte Carlo simulation uses random sampling to simulate the effect of the statistically distributed factors on the statistical distribution of the responses. The necessity of using a probabilistic analysis can be comprehended by using an example from an investigation conducted by Savage *et al.* (2007). They presented a methodology to analyze the effects of the variations from different vehicle cooling system design variables on the radiator inlet coolant temperature. A comprehensive list of all the influential variables and the range of variation were presented. The statistical distribution of the radiator inlet temperature was predicted with the Monte Carlo simulation as a result of the variations in nine factors (i.e., powertrain heat rejection, radiator heat rejection, condenser heat rejection, coolant flow rate, ram airflow rate, condenser air-side pressure drop, radiator air-side pressure drop, motor speed, and motor voltage). The radiator inlet temperature had a normal distribution from 239.8°F to 249.6°F with a range of 9.8°F by the Monte Carlo simulation. The worst-case stacking-up method gave a distribution of 235.0°F to 255.8°F with a range of 20.8°F. They found that the Monte Carlo simulation provided a more accurate variation range of the radiator coolant temperature.

Definition of variability-based optimization

Variability-based optimization applies optimization techniques in design for variability to optimize the design in the presence of a range of variability in order to make the design reliable or robust. Variability and uncertainty are different. There are two types of uncertainties: (1) random uncertainty, which is due to variations related to any attribute in the population and/or with respect to time; and (2) epistemic uncertainty (Farizal and Nikolaidis, 2007; Donders *et al.*, 2007), which is due to the deficiencies in deterministic simulation models caused by lack of knowledge.

Random uncertainty is also called variability. Strictly speaking, it further includes two types of variations: (1) variations within the population at a fixed moment in time during the product life, including the variations inherent to the physical system such as material characteristics, geometrical properties, and manufacturing tolerances, or the variations caused by changes in usage, test condition, and environment; and (2) variations with respect to time, i.e., a time-dependent degradation (for example, wear-out). For the first type, a design parameter may be used as the random variable in the statistical distribution to characterize the probability, such as the stress used in the stress–strength interference model. In the second type, time is the random variable.

Epistemic uncertainty usually produces systematic modeling errors and affects the entire population of a product in the same way. Random uncertainty can be handled by a probabilistic approach, while epistemic uncertainty cannot. RBDO and robust design are usually concerned with random uncertainty or variability.

Less variability does not mean robustness. A design is called robust if its response is insensitive to the variations in noise factors. Reliability refers to the chance of failure of a product. Variability is the cause of failure rate or reliability problems. Design for reliability (or RBDO) includes two distinct categories of analysis, namely (1) design for variability (or variability-based design optimization), which focuses on the variations at a given moment in time in the product life; and (2) design for time-dependent gradation. Most of the RBDO analysis in engine applications to date falls in the first category. It should be noted that the concepts of variability and reliability can refer to any performance or durability issues, not just durability alone.

A variability analysis may help define the engineering goal for the nominal target value and identify the major contributors to the response so that the variability can be reduced by changing the tolerances of the control factors. If the statistical distribution is too wide, design tolerances must be tightened to bring the variability within an acceptable range. On the other hand, if the distribution is narrower than necessary, design tolerances can be relaxed in order to reduce the design and manufacturing costs. In the optimization of robust design, the mean and its functional limit tolerance range (from LSL or lower specification limit to USL or upper specification limit) are optimized simultaneously. A target failure rate ($P_{failure}$) or reliability (R) is used in the variability-based optimization as a constraint. Note that the reliability is defined as $R = 1 - P_{failure}$.

Definition of reliability-based optimization

A design is called reliable if it satisfies all the requirements of performance and durability attributes within a specified period of mission time in the presence of variability. An example of unreliable design is as follows. The same unit of product may function unreliably at different times on the same day, or its performance or durability may deteriorate chronically over the next two years. Another example of an unreliable product is that 5% of 100 units of the same product do not function reliably after the product has been used for two years. Recall that if 5% of 100 units of the same product do not function well before the product is put in use (service), the problem is usually called a design quality or manufacturing quality problem.

A failure in terms of reliability is usually caused by one of the following: (1) an improper mean value in the control factor; (2) an improper tolerance in the variability control for the population in the control factor; or (3) the degradation with respect to time in the mean (nominal) or the tolerance. The failure can be either a performance attribute (e.g., emissions compliance, fuel economy) or durability attribute (e.g., crack due to fatigue, engine failures due to excessively hot radiator outlet coolant temperature).

In RBDO, the constraints are defined for the required reliability, and the design is optimized by changing the variable settings such that all the criteria (e.g., performance, durability, packaging, cost, and reliability) are met. The reliability estimation is usually expressed as a reliability value or failure rate, and sometimes is expressed in terms of tolerance as multiples of the standard deviation (σ). For example, a 6σ design has a higher reliability than a 3σ design, with a very small probability of failure in the order of 10^{-10}. However, the 6σ design is also more expensive. The standard deviation may apply to any variation such as the piece-to-piece variation among multiple units or time-to-time variation for a given unit. The mean value is designed at several multiples of the standard deviation (e.g., 3σ) away from the control limit with a prescribed failure rate. In fact, either a higher or lower failure rate is undesirable because neither under-design nor over-design is acceptable. More information on the RBDO theory can be found in Gu and Yang (2003), Mourelatos and Liang (2005, 2006), Rahman *et al.* (2007), and Donders *et al.* (2007).

Definition of robust optimization

As defined earlier, a design is called robust if its response is insensitive to the variations in noise factors. Robust design is to improve the quality of the product or process by minimizing the variation in the response without eliminating the sources of variation. The focus in robust design is usually to reduce the variation due to noise factors by changing the settings of the control factors. In design for robustness and robust optimization, the objective is to evaluate the 'variance' of a response parameter. Sometimes a separate objective or constraint is also formulated for the mean of the response. A reliability constraint does not necessarily present in this formulation. One example of robust design is to minimize the variation range of the radiator coolant temperature caused by the fluctuation of heat rejection and ambient temperature. Another example of robust design is to maintain a stable power output of a diesel generator to make it insensitive to various noise factors. More information on the robust design theory can be found in Gu and Yang (2003) and Mourelatos and Liang (2005, 2006).

Robustness and reliability need to be integrated in the early stage of the engine development process. Robustness should be included in engine optimization as an objective, along with the nominal values of fuel consumption, emissions, and NVH. A robust system exhibits low variation of its functions in the presence of noise factors. Robust design is to achieve the robust system by selecting the control factor levels that produce the least variation in functions. Equally important is tolerance control.

The relationship among attribute optimality, reliability, and robustness

Reliability (low failure rate at time *t*) and robustness (insensitivity to noise factors) are two distinct characteristics of a product. However, they share some common analysis and design methods. They are both closely related to variability. Therefore, they share a common probabilistic nature and require a statistical approach to address. High reliability or high robustness requires varying the mean and/or standard deviation of the control factors. Both reliability and robustness can be considered in an optimization topic. In summary, there are trade-offs among the three fundamental design principles of a product: attribute optimality (e.g., lowest BSFC or cost), reliability (e.g., long durability B10 life), and high robustness (e.g., small variation in radiator coolant temperature in response to large changes in ambient temperatures).

Adding a constraint in an optimization problem generally makes the attribute objective less optimal. For example, when reliability is added as a probabilistic constraint, the optimized engine BSFC in a nondeterministic nominal design will become worse compared to the deterministic case without any reliability constraint. If more margin is reserved for variability (e.g., demanding a higher reliability), the design for BSFC will become even less optimal. RBDO is the process to move from a deterministic optimum to the reliability optimum. Similarly, if reliability is formulated as a constraint in an optimization topic, it is impossible to satisfy robustness without sacrificing the attribute optimality because a reliability optimum may not be robust (i.e., insensitive to variation).

Although a RBDO topic and a robust optimization topic can be formulated separately to address different issues, in many scenarios they need to be formulated together as a multi-objective optimization topic in order to optimize a system mean attribute and simultaneously minimize its standard deviation (sensitivity to noise) subject to the same probabilistic reliability constraint. Such a formulation is called combined robust-RBDO. The two objectives (the mean attribute and the robustness) usually have a trade-off, i.e., improving one must sacrifice the other. The trade-off can be analyzed by a Parero optimality frontier calculation.

Mathematical formulation of deterministic optimization

For comparison purpose, the formulation of deterministic design optimization is given by:

$$\begin{cases} \text{Minimize } f(X_d, \bar{X}_r, \bar{X}_p) \\ \text{subject to } g_i(X_d) \leq 0, \, i = 1, \cdots, m \\ \qquad\qquad X_d^L \leq X_d \leq X_d^U \end{cases} \tag{3.28}$$

where X_d is a vector of k_d-dimensional deterministic control factors (deterministic design variables), $g_i(X_d)$ is a non-probabilistic constraint, X_r is a vector of k_r-dimensional nondeterministic random control factors (random design variables), $\bar{X}_r$ is the mean of X_r, X_p is a vector of k_p-dimensional nondeterministic random noise factors (random design parameters), $\bar{X}_p$ is the mean of X_p, and X_d^L and X_d^U are the lower and upper bounds of X_d.

Mathematical formulation of RBDO

A typical formulation of nondeterministic RBDO is expressed as follows:

$$\begin{cases} \text{Minimize } f(X_d, \bar{X}_r, \bar{X}_p) \\ \text{subject to } P(g_i(X_d, X_r, X_p) \leq 0) \geq R_i, \, i = 1, \cdots, m \\ \qquad\qquad X_d^L \leq X_d \leq X_d^U \\ \qquad\qquad \bar{X}_r^L \leq \bar{X}_r \leq \bar{X}_r^U \end{cases} \tag{3.29}$$

where $\bar{X}_r$ is a vector of random control factors (random design variables) whose mean and standard deviation can be varied, and $\bar{X}_p$ is a vector of random noise factors (random design parameters) whose mean and standard deviation cannot be changed. R_i is the desired reliability constraint, and $g_i(X_d, X_r, X_p) > 0$ indicates failure. Note that, as an example, this particular formulation only varies the deterministic control factors and the means of the random control factors (not their standard deviations) to reach a reliability-constrained optimum.

Mathematical formulation of robust optimization

A typical reliability-based robust design optimization formulation (i.e., including reliability as a constraint) is expressed as follows:

$$\begin{cases} \text{Minimize } f_v(X_d, X_r, X_p) \\ \text{subject to } P(g_i(X_d, X_r, X_p) \le 0) \ge R_i,\ i = 1, \cdots, m \\ \qquad X_d^L \le X_d \le X_d^U \\ \qquad \bar{X}_r^L \le \bar{X}_r \le \bar{X}_r^U \\ \qquad f(X_d, \bar{X}_r, \bar{X}_p) \le C_{target} \\ \qquad \text{(This is an optional constraint)} \end{cases} \quad 3.30$$

where $f_v(X_d, X_r, X_p)$ is the variance of objective function f, and C_{target} is a constant of the mean performance target. $f_v = 0$ indicates extremely robust, and it indicates the variations in the input factors do not create any change in the output response. This particular formulation is to reduce the attribute variation at the target mean for the attribute. The variation measure f_v in robust design is given by

$$f_v(X_d, X_r, X_p) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_{x_i}^2 \quad 3.31$$

The variation can also be expressed by the spread of the probability density function of the objective function, which is simply the percentile difference (Mourelatos and Liang, 2006):

$$\Delta R_f = f^{R_2} - f^{R_1} \quad 3.32$$

where f^{R_1} and f^{R_2} are a low and high percentile of f, respectively. For example, the 5th and 95th percentiles can be used.

It should be noted that in many formulations of robust optimization the mean is optimized (maximize or minimize) and the variance is minimized simultaneously with a trade-off in such a multi-objective formulation.

Mathematical formulation of combined robust-RBDO

The unified methodology of combining robust design with RBDO has been proposed by several authors (Gu *et al.*, 2004; Mourelatos and Liang, 2005, 2006). Essentially, it is the combination of equations 3.29 and 3.30. The topic then becomes a multi-objective optimization for two or more objectives (e.g., at least one for the mean of an attribute and the other for the variance of an attribute). For example, a combined robust-RBDO formulation can be as follows:

$$\begin{cases} \text{Minimize} & \mu_{f(X_d, X_r, X_p)} \text{ and } \sigma_{f(X_d, X_r, X_p)} \\ \text{subject to} & P(g_i(X_d, X_r, X_p) \le 0) \ge R_i,\ i = 1, \cdots, m \\ & X_d^L \le X_d \le X_d^U \\ & \bar{X}_r^L \le \bar{X}_r \le \bar{X}_r^U \end{cases} \tag{3.33}$$

where $\mu_{f(X_d, X_r, X_p)}$ is the mean of the objective function $f(X_d, X_r, X_p)$, and $\sigma_{f(X_d, X_r, X_p)}$ is the standard deviation of the objective function (however, not necessarily the same objective function or attribute). There is a trade-off between these two objectives. The response surface methodology (RSM) can be used to link the input to the output to evaluate the objective functions and the constraints. In fact, not only the mean can be changed, the standard deviation of the random control factors can also be changed in order to optimize the objective functions. There is a trend to regard robust design optimization as a subset of the generalized RBDO when a reliability constraint is required for the robust design problem.

Gu *et al*. (2004) introduced a formulation and its solution algorithm of RBDO combined with robust design. They also addressed the advantages of combining the Monte Carlo simulation with response surface models in RBDO. Mourelatos and Liang (2005, 2006) illustrated the design trade-offs between reliability and robustness in combined robust-RBDO optimization.

From two-step robust optimization to one-step optimization

In the conventional robust design theory, for example Taguchi's two-step optimization theory in quality and variability control (Fowlkes and Creveling, 1995), the mean and tolerance (deviation or variation) of the objective performance attribute are varied in two steps. In Taguchi's approach, the concept of RSM is not used. Because RSM allows more than one parameter to be modified at one time, using RSM to optimize both mean and tolerance simultaneously becomes feasible in robust design. By introducing the powerful RSM and Pareto optimality into robust design, the variability-control optimization problem can be handled with the above formulations in just one step, more effectively than the conventional two-step approach without RSM.

Solution algorithms of RBDO and robust optimization

There are many commercial software packages available for solving RBDO problems. Usually, RBDO problems require a dual-loop solution approach which includes an outer loop for optimization and an inner loop for calculating the reliability to evaluate the probabilistic constraint. The dual-loop approach is computationally intensive. In order to improve the computational efficiency of RBDO problems, many approximate solution methods have been developed to convert the dual-loop to a single-loop. Detailed discussion on the solution algorithms of RBDO and robust optimization are provided by Gu and Yang (2003), Mourelatos and Liang (2005, 2006) and Donders *et al.* (2007).

3.4.2 Basics of statistics and probability distribution selection

Statistical probability distributions

There is a statistical spread or variability in almost any value that can be measured in a population. A single number is often inadequate to represent an item. Instead, probability distributions are often more appropriate. There are many statistical distributions. The probability distribution of any random variable can be characterized by certain distribution parameters such as mean value, standard deviation, minimum and maximum values. Their definitions are introduced in Table A.1 in the Appendix. The standard deviation is a measure of the variance or scatter around the mean value. The larger the standard deviation, the wider the range of the variability.

The probability distribution of a random variable is described completely by its cumulative distribution function (CDF), whose value at each real number x is the probability that the variable is smaller than or equal to x. A probability distribution is called discrete if its cumulative distribution function only increases in discrete jumps. Discrete distributions are characterized by a probability mass function.

On the other hand, a probability distribution is called continuous if its cumulative distribution function is continuous. If the distribution of a real number x is continuous then x is called a continuous random variable. Diesel engine system design often handles continuous random variables, for example, the noise factor of piece-to-piece variation. The normal distribution, continuous uniform distribution, Weibull distribution, Beta distribution, and Gamma distribution are well-known continuous distributions (see Table A.2 in the Appendix). The continuous distributions are characterized by a probability density function (PDF). The relationship between CDF $f_{CDF}(x)$ and PDF $f_{PDF}(x)$ is given by

$$f_{CDF}(x) = P(X \le x) = \int_{-\infty}^{x} f_{PDF}(u)\mathrm{d}u \qquad 3.34$$

The probability that x is between the two points a and b is

$$P(a \le x \le b) = \int_{a}^{b} f_{PDF}(x)\mathrm{d}x \qquad 3.35$$

The integral of a continuous probability function is equal to one, given by

$$\int_{-\infty}^{\infty} f_{PDF}(x)\mathrm{d}x = 1 \quad \text{where} \quad f_{PDF}(x) \ge 0 \qquad 3.36$$

A histogram is a useful plot to show the frequency occurrence of the event values vs. the values distributed in different bins. Narrow bins will collect few data points and reduce the frequency occurrence or the population count. Large bins may lump together different ranges which are really different, thus distorting the real distribution of the data. Probability density can be interpreted as the ratio of frequency occurrence of the data in the bin to the bin size.

Reliability in statistics

The reliability function is the integral of the probability density function from x to infinity, given by

$$R(x) = P(X \ge x) = \int_{x}^{\infty} f_{PDF}(u)\mathrm{d}u = 1 - f_{CDF}(x) \qquad 3.37$$

where X is the mean time to failure. The hazard function in reliability engineering, an instantaneous failure rate representing the tendency to fail, is defined as

$$h_r(x) = \frac{f(x)}{R(x)} \qquad 3.38$$

Statistical distribution parameters

The population mean is the first moment-generating function of the probability density function about the origin (see Table A.1 in the Appendix). The variance of a distribution is equal to the second moment-generating function about the mean. The skewness of a distribution is equal to the third moment-generating function about the mean. Skewness is a measure of asymmetry of the probability distribution of a real-valued random variable. The kurtosis is equal to the fourth moment-generating function about the mean. Kurtosis is a measure of the sharpness of the peak and the size of the tail of the probability distribution. Higher kurtosis means the mean and extreme deviations occur with higher probability, and the distribution has a sharper peak and longer or fatter tails. Excess kurtosis is defined as kurtosis minus three, a correction to make the kurtosis value of the normal distribution equal to zero.

In general, the probability density function and the shape of a statistical distribution can be described by one or more of the following parameters: location parameter, scale parameter, shape parameter, or degrees of freedom parameter. The effect of the location parameter is to translate or shift the distribution curve horizontally. The effect of the scale parameter is usually to stretch or shrink the curve horizontally and vertically. The shrinking approaches a spike as the scale parameter goes to zero. Usually, non-positive scale parameters are not allowed. Any particular type of probability distribution is not a single distribution, but in fact a family of distributions. This is because the distribution has one or more shape parameters. Shape parameters allow a distribution to form a variety of shapes, depending on the value of the shape parameter. The combined use of the above statistical distribution parameters makes the probability distributions particularly useful in modeling applications since they are flexible enough to model a variety of data sets. The 'standard' form of any distribution is the form that has the location parameter equal to zero and the scale parameter equal to one. The statistical distribution parameters are extensively used in the nondeterministic optimization of diesel engine system design (Fig. 3.9).

Analytical relationships between probability distributions

The probability density function of the sum of two independent random variables is the convolution of each of their probability density functions. The probability density function of the difference of two independent random variables is the cross-correlation of each of their probability density functions. In diesel engine system design, the PDF of the engine response needs to be analyzed based on the PDF of different input factors. The relationships between the factors and responses in engine system design are usually not simple sum or difference but very nonlinear and complex. Therefore, a numerical simulation method such as the Monte Carlo simulation with random sampling is necessary in order to find out the PDF of the engine response.

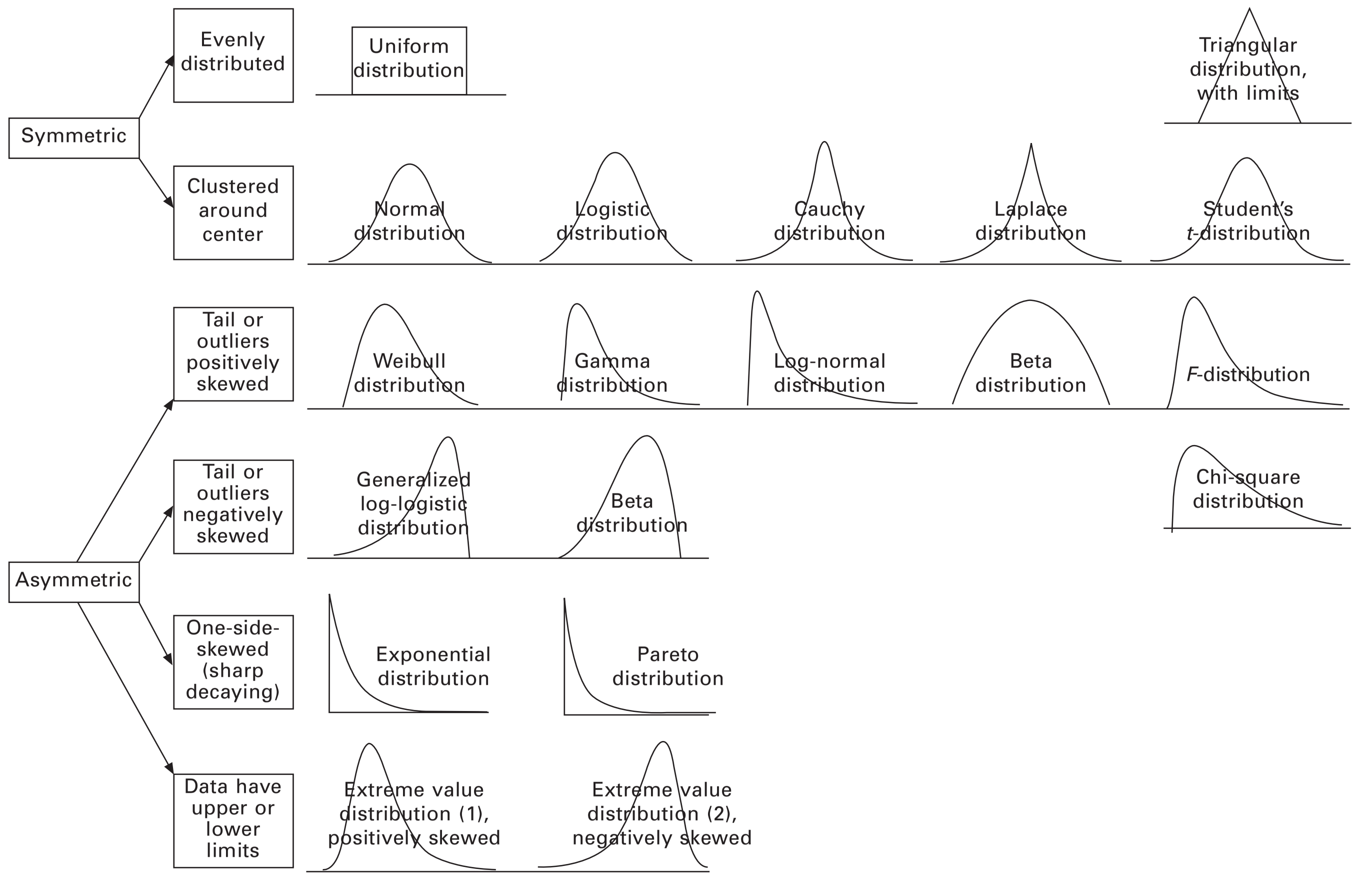
Probability distribution selection

There are many complex relationships between different probability distributions, and they are well developed in the theory of probability and statistics. There are numerous textbooks and handbooks in this area. Table A.2 (in the Appendix) is compiled to summarize the probability distribution functions that are commonly used in diesel engine system design. In the statistical distributions in Table A.2, a consistent set of symbols are used to facilitate the reader: x or y denotes the random variable value (real number), α denotes the location parameter, β denotes the scale parameter, and γ denotes the shape parameter. Note that most distributions and their shapes can be characterized by two or three parameters. For the basics of probability and statistics, the reader is referred to Law (2006). The details of probability distribution functions and their statistical properties are provided by Evans *et al*. (2000) and Dodson and Schwab (2006). Examples of various probability distribution curves can be found at the website of the US National Institute of Standards and Technology at www.itl.nist.gov/div898/handbook/secton3.

Design for variability and design for reliability make extensive use of probability distributions. In reliability engineering the distribution is usually fitted with respect to time (i.e., mean time to failure used as the random variable). The most commonly used distributions in reliability engineering are Weibull, normal, log-normal, and exponential distributions. On the other hand, in design for variability the statistical distribution is usually fitted with respect to the scattered values of the system factors or responses (i.e., the factors or responses used as the random variable).

The guidelines of selecting a probability distribution to fit univariate statistical data are summarized as follows:

1. It is important to plot and watch the histogram of the actual data to distinguish the data pattern according to the following aspects and then select an appropriate type of statistical distribution (Fig. 3.23):
   - continuous distribution vs. discrete distribution
   - symmetric distribution vs. asymmetric distribution
   - clustered around center vs. evenly distributed (i.e., classified by kurtosis)
   - asymmetric ‘tail’ – outliers skewed positively vs. skewed negatively or one-side skewed (i.e., classified by skewness)
   - no limits vs. having upper or lower limits (extremes).
2. It is necessary to calculate the statistics of the actual data distribution (such as mean, standard deviation, skewness and excess kurtosis (see Table A.1 in the Appendix). Check which distribution function matches them closely (see Table A.2 in the Appendix).
3. An easy-to-use statistical distribution is an important consideration in model selection. The selected distribution does not need to be the best-fit distribution for the data, but needs to be a sufficiently adequate model so that the statistical model may yield valid conclusions.
4. When fitting the probability data of an engine system response with a



3.23 Selecting continuous probability distribution functions.

selected distribution function, the nature of the relationship between the response and the factors in the engine processes needs to be considered in order to judge whether the selected distribution function is reasonable compared with the probability distribution functions used for the input factors.

5. The parameters of the selected distribution function for the factors or responses need to be estimated graphically with probability plotting or numerically with the methods of maximum likelihood estimation (Dodson and Schwab, 2006) or least-squares. An analysis of goodness of fit can be conducted to compare the distribution function of the actual data with that of the fitted distribution in order to accept or reject the fit. For details of selection, fitting and testing statistical distribution models, the reader is referred to Shapiro (1990).

Normal distribution and other commonly used distributions

When the true distribution of a variable is not known, a normal distribution (also known as the Gaussian distribution or the bell curve) is often assumed. Normal distribution is the most widely used distribution. It is ubiquitous in nature and statistics because of the well-known central limit theorem: every variable that can be modeled as a sum of many small independent variables is approximately normal. In other words, the sum or average of a large number of random independent variables is approximately normally distributed, regardless of the distribution of the variables being summed or averaged. The data in normal distribution have a strong tendency to cluster on a central value with symmetric deviations to both sides of the central mean value. The symmetry of deviation results in zero skewness. The low probability of large deviations from the mean value (i.e., small 'tail' size) gives zero excess kurtosis. The characteristics of the normal distribution are elaborated by Dodson and Schwab (2006).

In the family of symmetric distributions the shapes and kurtosis are different (Fig. 3.23). Some of them have a sharper peak and longer or fatter tails than the normal distribution (i.e., higher kurtosis). For example, the logistics distribution has a longer tail and higher kurtosis value than the normal distribution. Some others may have a lower, wider peak and shorter or thinner tails than the normal distribution (i.e., lower kurtosis). As an extreme case when the distribution flattens out it becomes the uniform distribution.

Some non-symmetric distributions have a peak in the distribution at low values and a gradual tapering tail toward higher values. These random variables can be modeled by a log-normal or Gamma distribution. Some distributions with the 'cliff' shape may be best modeled by an exponential distribution. In the family of positively skewed distributions (e.g., Weibull, Gamma, log-normal), increasing the shape parameter shifts the peak of the distribution to the left and the skewness increases. With a very high shape parameter, the distribution becomes one-side skewed and has a sharp decaying shape like in the exponential distribution.

When the statistical data are constrained by an upper or lower limit (e.g., cooler effectiveness limited to less than 100%, engine flow rate or pressure limited to greater than zero), using some of the above-mentioned distributions may create problems that unrealistic numbers exceeding the limits could be generated. Depending on the type of limit (upper only, lower only, or both), one of the following distributions that are limited by an upper or lower limit can be used to fit the data: uniform, triangular, positively skewed, negatively skewed, one-side skewed, or extreme value distributions. An alternative and approximate way is to use the symmetric clustered-around-center distributions (e.g., normal) to fit the data and then impose the limits to discard any data exceeding the limits. The penalty of using this method can be small if the tail in the distribution (i.e., the portion truncated off) is relatively small.

It should be noted that in statistics theory a few distributions are usually used in statistical inference analysis rather than modeling physical random variables, such as the chi-square, *F*- and student's *t*-distributions, because they are derived distributions from other basic distributions.

3.4.3 Introduction to Monte Carlo simulation

Probabilistic models are sometimes titled with the term 'Monte Carlo'. Monte Carlo simulation is a modeling tool for uncertainty and it has been used since the 1940s. Uncertainty cannot be simply replaced by a single average value. Otherwise, the estimate and design decisions based on that average will be way off in general. It is the engines at the extremes of the entire population that determine the success or failure of the design, rather than the nominal mean of the entire population. Probability distribution is a much more realistic way of describing uncertainty in variables subject to risk.

In general terms, the Monte Carlo method refers to any technique that approximates solutions to quantitative problems by statistical sampling. It is a general class of stochastic approach for analyzing uncertainty propagation from model input to model output. It uses random sampling of the probability distribution functions of the model inputs, often with an independent and random combination of several inputs at the same time, to produce outputs and estimate the probability distribution of the outputs. Independent sampling refers to the fact that there is no correlation between two or more input distributions. The calculation is usually conducted with several thousand random samples instead of a few discrete scenarios in order to satisfy the accuracy requirement of the probability evaluation. The term Monte Carlo was coined in the 1940s in reference to games of chance by the physicists working on the nuclear weapon projects in the Los Alamos National Laboratory. Its core idea is to use random samples as inputs to predict the behavior of a complex system or process.

A Monte Carlo simulation method consists of the following steps:

1. Create a deterministic parametric model linking the input with the output, $Y = f(X_1, X_2, \ldots, X_k)$.
2. Define a domain of input factors (number of factors, their sampling range of variation, and probability distribution).
3. Generate a set of inputs randomly and independently from the domain, denoted as the ith set of samples, $X_{1i}, X_{2i}, \ldots, X_{ki}$.
4. Perform a deterministic computation with the model for the ith iteration to obtain the outputs, $Y_{1i}, Y_{2i}, \ldots, Y_{ki}$, by using the ith set of inputs.
5. Repeat steps 3 and 4 for $i = 1$ to n (Note: n is the number of samples).
6. Aggregate the outputs of all individual computations to produce the probability distributions of the outputs (e.g., histograms or summary statistics such as frequency of occurrence, minimum, maximum, mean, median, standard deviation, variance, mean standard error, percentiles, error bars, tolerance zones, confidence intervals, probability or reliability predictions, skewness and excess kurtosis of the distribution).

Figure 3.10 illustrates the basic principle of the Monte Carlo simulation. Different types of probability distributions can be used for the inputs, for example, normal or Gaussian distribution, log-normal distribution, uniform distribution, exponential distribution, or triangular distribution. The output is not a single fixed value, but a range of possible outcomes with a probability distribution. The Monte Carlo method is powerful in handling multiple dimensions of uncertainties (i.e., many input factors). It can also handle any statistical distribution of random input factors with straightforward implementation. Good Monte Carlo simulation relies on the quality of random numbers. As more data points are sampled in the statistical probability distribution, the results of the Monte Carlo simulation converge to a better approximation. The large amount of sampling cases required may become a drawback of the Monte Carlo method when the function evaluation for each case is computationally expensive. The Monte Carlo method provides an estimate of the expected value of a random variable and also predicts the estimation error. The estimation error is given by

$$\upsilon \approx \frac{3\sigma}{\sqrt{n}} \qquad 3.39$$

where σ is the standard deviation of the random variable, and n is the number of samples. More in-depth discussions and advanced techniques related to the Monte Carlo method can be found in Hampson *et al.* (2002), Zou *et al.*

(2004), Daniels and Miazgowicz (2007), Donders *et al*. (2007), Farizal and Nikolaidis (2007) and Nikolaidis *et al*. (2008).

By exploring thousands of combinations of the full range of possible outcomes, using the Monte Carlo simulation may not only obtain more accurate results in the face of uncertainty but also reveal the sensitivity on which input has the biggest impact on the probability distribution of the outputs. The Monte Carlo method expands the engine sensitivity analysis from the level of deterministic single-point prediction used for design-for-target to a more advanced level of nondeterministic prediction used for design-for-reliability to assess the probability of the risks.

3.4.4 Previous research in reliability-based design optimization

Yan *et al*. (1993) analyzed the effects of statistical distribution of the design factors on emissions variability. The design factors included fuel injector nozzle flow, nozzle protrusion, spray cone angle, piston-to-head clearance, fuel injection timing and pressure, oil contribution, valve stem seal leak rate, and swirl ratio. The resulting mean value and the standard deviation of the emissions were used to set the emissions goal in engineering development to ensure the entire engine product population could meet the emissions requirement. A proper emissions deterioration factor was taken into consideration in defining the engineering emissions margins. They discovered that the injection timing variability was mainly responsible for NO_x and HC variations, while the variability from nozzle flow, injection pressure/timing and oil contribution controlled PM variation. Such sensitivity information could be used to adjust the tolerance of the control factors.

Dave and Hampson (2003) used the Monte Carlo simulation to investigate the effects of statistical distributions of four factors on the probability distributions of NO_x and BSFC. The four factors were injection timing, charge air cooler outlet temperature, cylinder wall temperature and intake valve opening timing. They used the RSM to optimize the mean values of the four design factors to shift the mean value of NO_x and reduce the standard deviation of NO_x so that the mean of NO_x was at 3σ below an upper control limit in order to ensure emissions compliance. Meanwhile, BSFC was minimized for a target failure rate of NO_x and BSFC.

Kokkolaras *et al*. (2005) analyzed the effects of variability on hybrid-hydraulic truck performance (e.g., fuel economy). The variability included many random design parameters in the fuel cell, the engine and the vehicle. The probability density function of the vehicle fuel economy was determined with a skewed distribution. In RBDO simulation, a large amount of Monte Carlo simulation was required. They had to use fast surrogate models to conduct the Monte Carlo simulation. They used a variable screening technique to identify the most important random factors in order to decrease the size of the RSM DoE in order to build accurate surrogate models. They came to the following findings:

- The most important factors for driving cycle fuel economy were fuel injection timing, vehicle frontal area, rolling resistance, and transmission efficiency.
- The most influential factors for vehicle acceleration performance were fuel injection timing, vehicle frontal area, intake boost pressure, transmission efficiency and engine compression ratio.
- The most important factors for silent watch fuel economy were fuel cell temperature, humidity ratio, and membrane thickness.

They used RBDO to quantify the trade-offs between fuel economy and reliability. They found the trade-off was highly nonlinear, and the fuel economy increased exponentially at very high reliability levels.

Catania *et al*. (2007) analyzed the variability of fuel consumption caused by driver driving style, vehicle weight, vehicle resistance, engine BSFC and transmission efficiency. They concluded that the probability distribution shape of fuel economy does not correlate with the distribution shape of the input factors.

The application of variability-based design optimization in piston assembly tribology was carried out by Hoffman *et al*. (2003), Ejakov *et al*. (2003) and Chan *et al*. (2004).

Rahman and Sun (2003) applied robust design and variability-based optimization to address the reliability problem of engine cooling system and the variation of top tank temperature. They assumed a normal distribution for each input factor in the Monte Carlo simulation. A probability distribution curve of the top tank temperature was produced and compared with a prescribed maximum design limit of the temperature. The failure rate was calculated as the probability of the top tank temperature exceeding the limit. As a design solution, they changed the mean and the standard deviation of the input factors in order to move the probability distribution curve of the top tank temperature to meet the reliability target.

Rahman *et al*. (2007) used multi-objective robust-RBDO to maximize the cooling system performance and simultaneously minimize its standard deviation under the probabilistic reliability constraints. The two objectives used were the mean performance and its variation. They achieved a balance (or trade-off) between reliability and robustness by simultaneously optimizing (maximizing) the mean performance and minimizing the performance variation. A trade-off occurred because it was impossible to improve one objective without sacrificing the other. The trade-off was computed by using the concept of Pareto efficiency. The Monte Carlo simulation with 10,000 samples was used to determine the statistics of an engine cooling system.

The input random factors included ambient temperature, engine speed, and heat rejection. The deterministic design control factors included cooling air flow rate and compressor pulley ratio. Three objectives were used in the optimization: (1) maximizing the mean of top tank temperature (for good BSFC); (2) minimizing the standard deviation of top tank temperature (for robustness); and (3) minimizing the front-end air flow rate (for good BSFC). By optimizing the control factors and controlling the variation of random control factors, an optimal balance between reliability and robustness of the engine cooling system was achieved. The probability of acceptable engine top tank coolant temperature was increased from 57% to 95%. Meanwhile, the standard deviation of the top tank temperature was reduced from 17.5°F to a much smaller value of 8.2°F.

3.4.5 Probabilistic simulation in diesel engine system design

A Monte Carlo simulation is conducted in Tables 3.6 and 3.7 and Figs 3.24–3.26 to investigate the impact of variability on the probability distributions of different engine performance parameters for a heavy-duty diesel engine at the rated power condition. The variability studied includes mainly the tolerances in engine design and control parameters. There are six cases in the simulation, corresponding to five different ambient conditions (Cases 1–5) and one sensitivity case to analyze the effect of exhaust restriction variation (Case 4S). Case 1 is at sea-level altitude (0 ft.) and 77°F (25°C) normal ambient in standard laboratory conditions. Case 2 is at sea-level altitude and 100°F (38°C) hot ambient for in-vehicle conditions with an increased air temperature at the compressor inlet by an amount of rise-over-ambient (ROA). Case 3 is at sea-level altitude, 122°F (50°C) hot ambient and in-vehicle. Case 4 is at 5500 feet (1676 meters) high altitude, 100°F (38°C) hot ambient and in-vehicle. Case 5 is at 10,000 feet (3048 meters) high altitude, 85°F (29°C) ambient and in-vehicle. Case 4S is the same as Case 4 except for using a 10% higher exhaust restriction flow coefficient that simulates a less restrictive aftertreatment system such as a clean DPF after soot regeneration. The standard deviation of the exhaust restriction flow coefficient of Case 4S is the same as that in Case 4. Table 3.6 shows the probability input data used in Case 4. There are in total 17 random input factors, all assumed in the normal distribution. Basically, the same coefficients of variation (i.e., the ratio of standard deviation to mean of the samples) are used for the other five cases. When the turbine wastegate is fully closed (as in Case 5), the standard deviation of the wastegate opening is assumed to be zero. The engine performance results are obtained with GT-POWER simulation for each sample of the Monte Carlo simulation.

Figure 3.24 shows the probability distributions of some input factors.

Table 3.6 Input data of Case 4 used in Monte Carlo simulation for probabilistic engine system design (an example at the rated fueling, 5500 feet high altitude and 100°F ambient temperature)

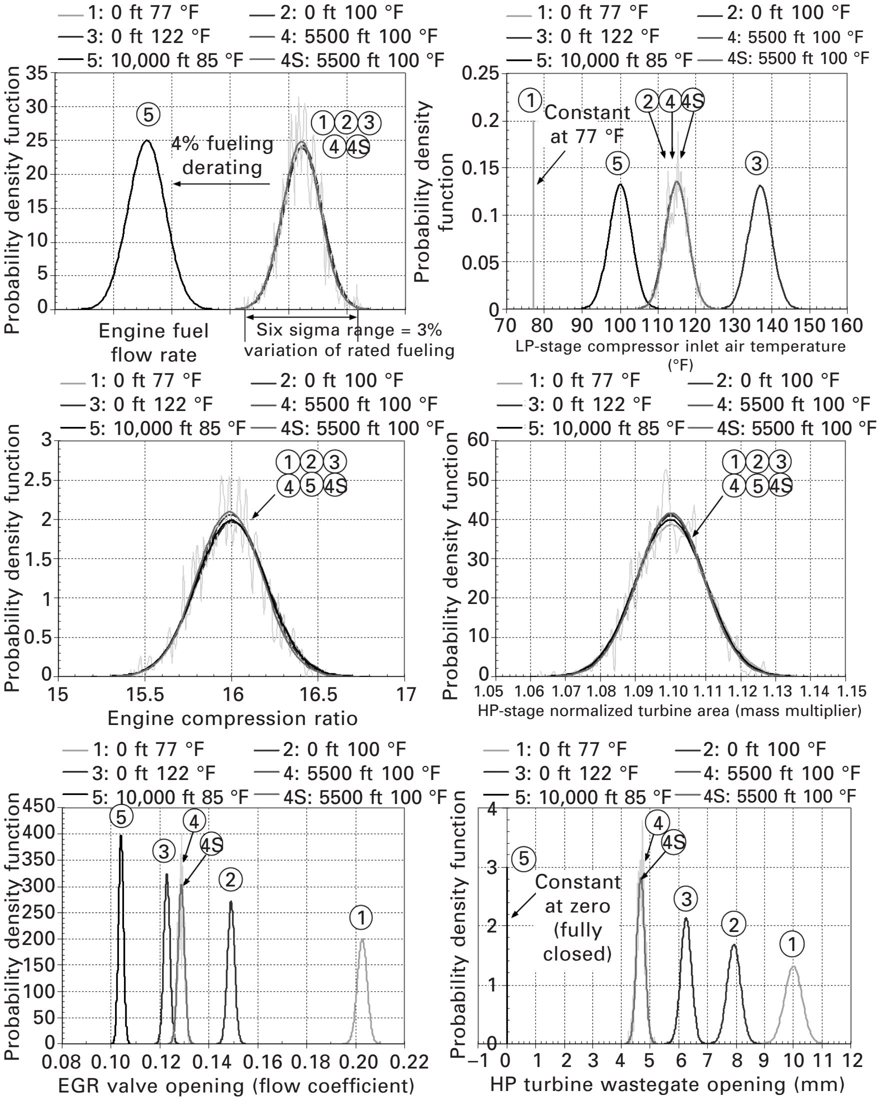
Input factor code	Parameter name	Unit	Type of factors	Baseline Mean	Baseline standard deviation	Coefficient of variation	Statistical distribution assumed in the model
X_1	Engine compression ratio	–	Random	16	0.2	1.25%	Normal distribution
X_2	HP turbine wastegate opening	mm	Random	4.68462	0.140539	3.00%	Normal distribution
X_3	EGR valve opening (flow coefficient)	–	Random	0.128757	0.001289	1.00%	Normal distribution
X_4	Fuel mass flow rate		Random	Baseline		0.5%	Normal distribution
X_5	Exhaust restriction flow coefficient	–	Random	0.39	0.02	5.13%	Normal distribution
X_6	HP compressor efficiency multiplier	–	Random	1	0.013	1.30%	Normal distribution
X_7	LP compressor efficiency multiplier	–	Random	1	0.013	1.30%	Normal distribution
X_8	HP turbine efficiency multiplier	–	Random	0.95	0.013	1.37%	Normal distribution
X_9	LP turbine efficiency multiplier	–	Random	0.95	0.013	1.37%	Normal distribution
X_{10}	Normalized HP turbine area (mass multiplier)	–	Random	1.1	0.01	0.91%	Normal distribution
X_{11}	Normalized LP turbine area (mass multiplier)	–	Random	1	0.01	1.00%	Normal distribution
X_{12}	Start-of-combustion timing	degree	Random	–10	0.1	1.00%	Normal distribution
X_{13}	Inter-stage cooler coolant inlet temperature	°F	Random	147.9	2	1.35%	Normal distribution
X_{14}	EGR cooler coolant inlet temperature	°F	Random	206.5	2	0.97%	Normal distribution
X_{15}	Engine coolant inlet temperature	°F	Random	216.8	2	0.92%	Normal distribution
X_{16}	Charge air cooler cooling air inlet temperature	°F	Random	113	2	1.77%	Normal distribution
X_{17}	LP-stage compressor inlet air temperature ($T_{AMB}+\Delta T_{ROA}$)	°F	Random	115	3	2.61%	Normal distribution

Notes:
(1) The coefficient of variation is calculated as the ratio of standard deviation to mean.
(2) X_{16} and X_{17} are only applicable for in-vehicle conditions rather than the standard lab engine condition at sea level (0 ft. altitude) 77°F ambient.
(3) In the case of sensitivity analysis on the effect of exhaust restriction variation, the mean value of the exhaust restriction flow coefficient is increased by 10% from the baseline mean 0.39 to 0.429.

Table 3.7 Coefficient of variation of the engine system response of the Monte Carlo simulation cases

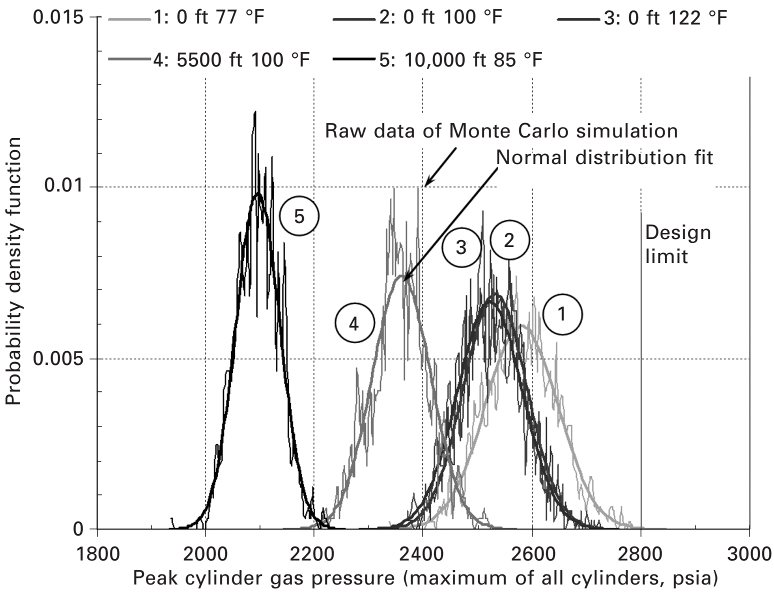
	Case 1	Case 2	Case 3	Case 4	Case 4S	Case 5	Average of five cases
Engine brake power	0.58%	0.58%	0.59%	0.57%	0.57%	0.58%	0.58%
BMEP	0.58%	0.58%	0.59%	0.57%	0.57%	0.58%	0.58%
Gross pumping loss PMEP	1.59%	1.48%	1.48%	1.40%	1.40%	1.30%	1.45%
360° gross IMEP	0.57%	0.55%	0.55%	0.54%	0.54%	0.51%	0.55%
Engine delta P	1.95%	1.91%	1.82%	1.78%	1.78%	1.63%	1.82%
BSFC	0.28%	0.27%	0.28%	0.30%	0.30%	0.31%	0.29%
Peak cylinder gas pressure (maximum of all cylinders)	2.60%	2.29%	2.38%	2.28%	2.28%	1.94%	2.30%
Peak cylinder gas temperature	0.96%	0.81%	0.76%	0.78%	0.78%	0.64%	0.79%
EGR rate	1.23%	1.16%	1.08%	1.10%	1.10%	1.02%	1.12%
Intake manifold gas temperature	0.68%	0.80%	0.73%	0.79%	0.79%	0.83%	0.77%
A/F ratio	2.11%	1.78%	1.65%	1.64%	1.64%	1.25%	1.69%
Intake manifold oxygen mass fraction	0.84%	0.65%	0.55%	0.63%	0.63%	0.50%	0.63%
Exhaust manifold gas temperature	1.27%	1.03%	0.97%	0.99%	0.99%	0.79%	1.01%
HP compressor outlet temperature	1.30%	1.18%	1.10%	1.12%	1.12%	1.21%	1.18%
Intake manifold boost pressure	1.80%	1.51%	1.48%	1.36%	1.36%	1.05%	1.44%
Exhaust manifold pressure	1.62%	1.36%	1.34%	1.23%	1.23%	1.04%	1.32%
Intake manifold mixture volumetric efficiency	0.07%	0.09%	0.10%	0.10%	0.10%	0.10%	0.09%
Total exhaust restriction (pressure drop)	10.95%	11.26%	11.25%	11.41%	11.41%	12.08%	11.39%
Engine coolant heat rejection	0.67%	0.71%	0.74%	0.72%	0.72%	0.75%	0.71%
Total engine coolant plus CAC heat rejection	0.81%	0.82%	0.86%	0.80%	0.80%	0.82%	0.82%
EGR cooler heat rejection	1.22%	1.17%	1.14%	1.14%	1.14%	1.22%	1.18%
Charge air cooler heat rejection	2.99%	2.75%	2.53%	2.31%	2.31%	2.12%	2.54%
HP-stage turbocharger actual speed	1.47%	1.05%	1.08%	1.16%	1.16%	1.23%	1.20%
LP-stage turbocharger actual speed	1.38%	1.26%	1.25%	1.39%	1.39%	1.43%	1.34%

Notes: The coefficient of variation is computed as the ratio of standard deviation to mean of the samples. The average of five cases includes Cases 1, 2, 3, 4 and 5.



3.24 Input data in Monte Carlo simulation for engine system design.

The oscillating data of the distribution in the figures are the raw data of the Monte Carlo simulation with 1000 random samples. The smooth curves of the distributions are the fitted results by using the normal distribution. The probability distribution of the raw data is obtained by using 100 bins that span over the entire range of the sample values for a given parameter. Fewer bins make the distribution appear less oscillating but may exhibit 'step changes' in the shape of the probability distribution curve because more samples



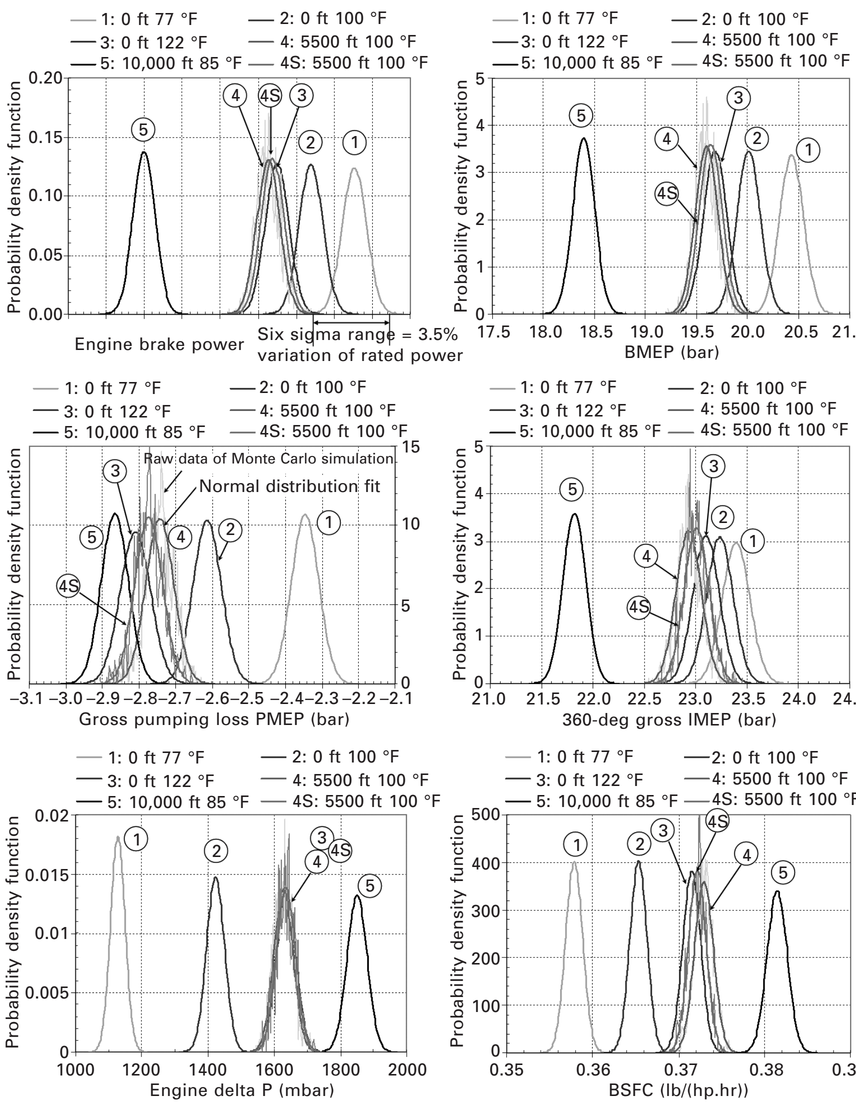
3.25 Output data illustration in Monte Carlo simulation – fitting the raw data with normal distribution.

will fall into each bin. In contrast, more bins make the distribution appear more oscillating and may even exhibit zero probability at certain parameter values because some bins may not have samples at all. The degree of the data oscillation is related to the number of bins used for data display, and does not indicate simulation accuracy. It is the number of samples in the Monte Carlo simulation that affects accuracy.

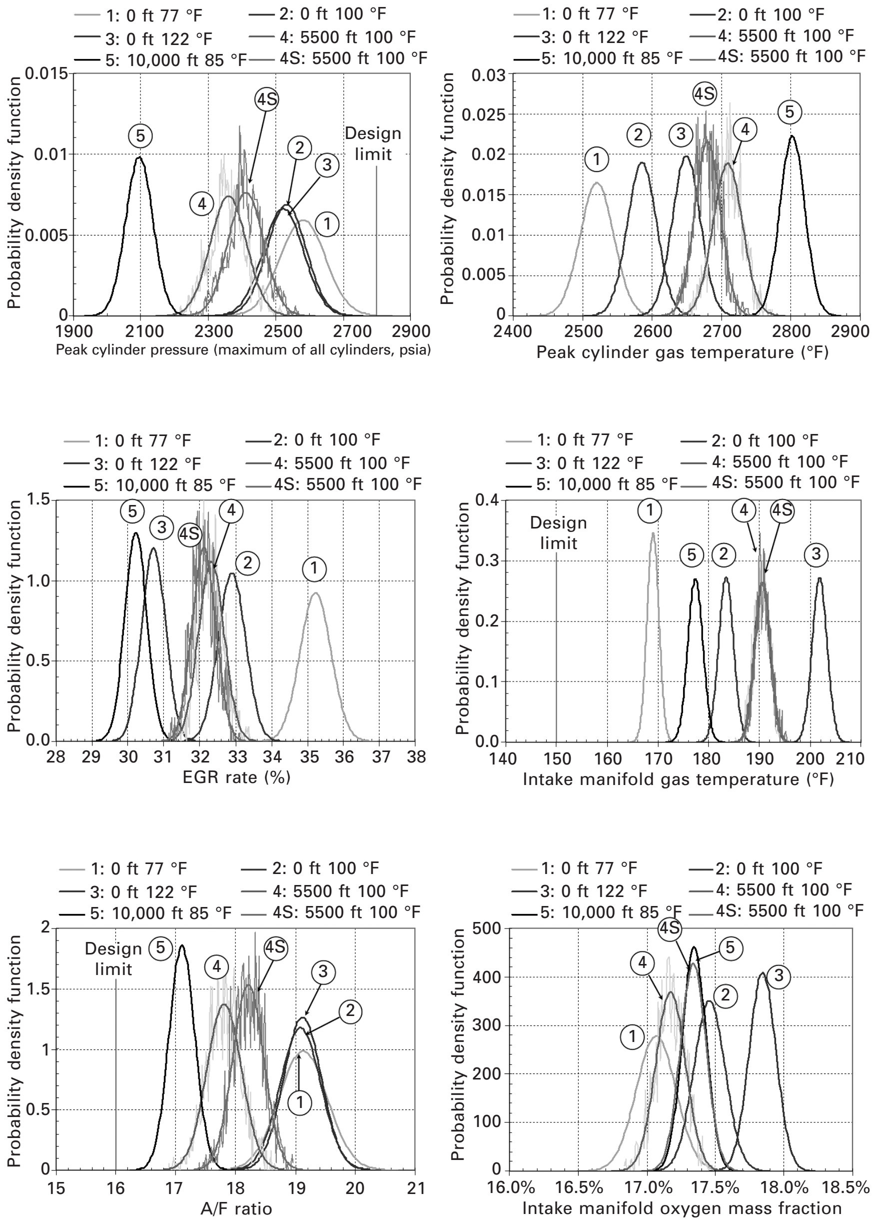
Figure 3.25 shows an example of the response parameters, displayed with both raw data of the Monte Carlo simulation and the fitted normal distribution curves. Figure 3.26 shows the probability distribution curves of all important response parameters in engine performance. Table 3.7 shows the calculated coefficients of variation for the responses of all six cases. A summary of this investigation is given as follows.

1. The variation range of each response parameter can be clearly observed from this study (Table 3.7 and Fig. 3.26). It should be noted that the variation range of the response is governed by the assumptions made in the variation ranges of the input factors shown in Table 3.6.
2. The shapes of the probability distributions of a given response parameter at different ambient conditions are different. This indicates the effect of the shapes of the probability distributions of the input factors (especially the turbine wastegate opening and the EGR valve opening), and the effect of complex nonlinear engine behavior at different ambient conditions.

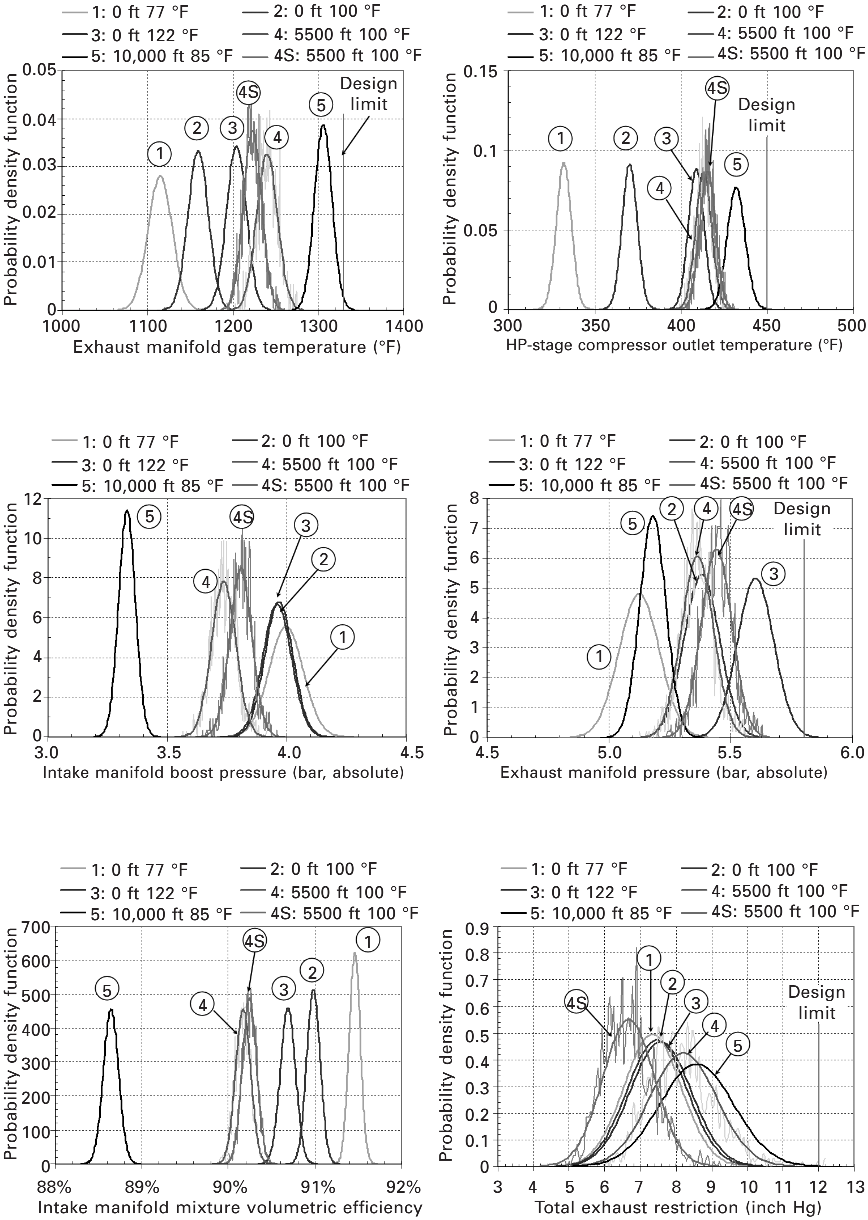
3. Different engine performance parameters exhibit their extreme values at different ambient conditions. For example, peak cylinder pressure has its worst (highest) value in Case 1 (sea level, 77°F), while exhaust manifold temperature reaches its worst (highest) value in Case 5 (10,000 feet, 85°F) even with 4% fueling derating. Different types of engine design constraints or limits are marked on Fig. 3.26 as examples. The probability distributions of the responses can be checked against these



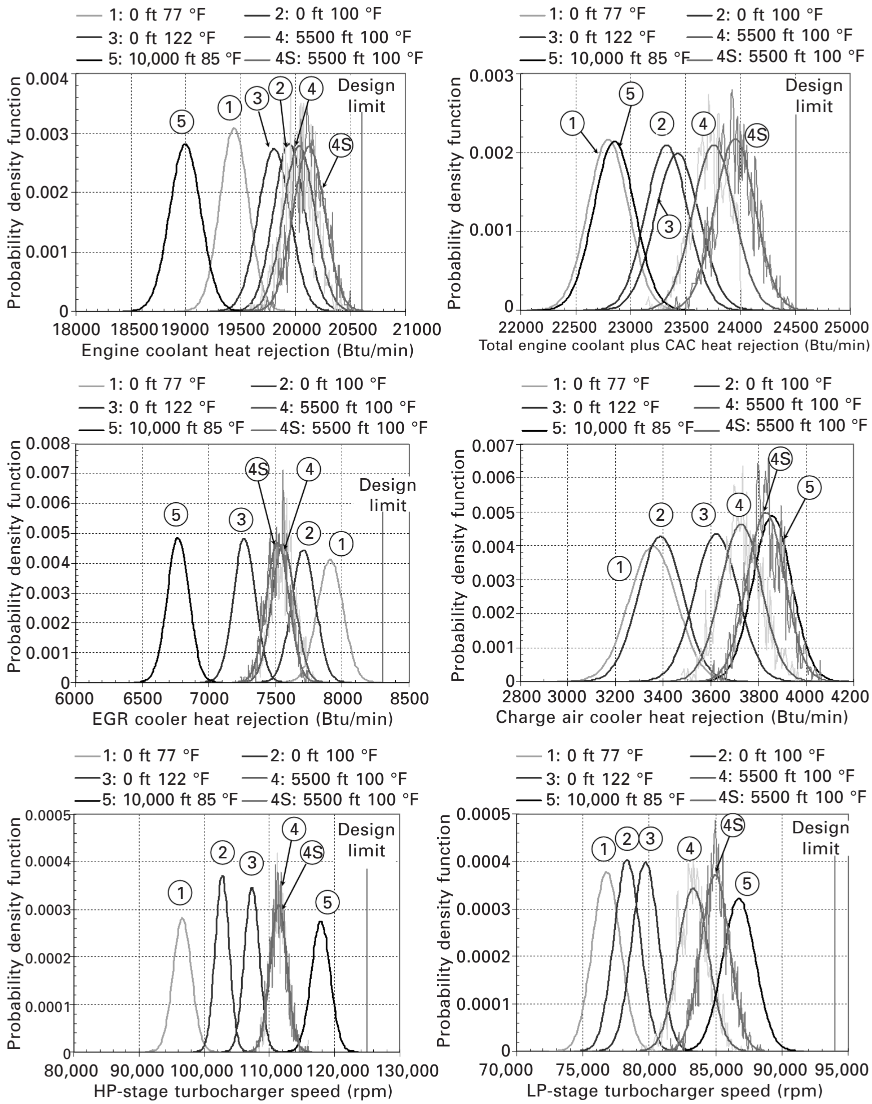
3.26 Output data in Monte Carlo simulation – probabilistic engine system design.



3.26 Continued



3.26 Continued



3.26 Continued

design limits to assess the probability of failure and reliability issues.

4. Case 4S illustrates a typical sensitivity analysis on the effect of design or calibration changes. Case 4S has a much lower exhaust restriction pressure drop compared with Case 4. This gives higher peak cylinder pressure in Case 4S due to its higher engine air flow rate. The sensitivity of the probability distribution shape of the engine response to the probability distributions of the input factors can be analyzed by this method.

5. This Monte Carlo simulation combines multiple engine operating conditions (i.e., at different ambient conditions) on one probability distribution chart to compare them conveniently. Similar plotting can be conducted by combining different engine speeds or loads on one probability chart.
6. Such a probabilistic analysis in engine system design provides much more information than the traditional deterministic approach in order to evaluate variability, reliability and safety margins in design.

3.5 References and bibliography

3.5.1 References and bibliography on DoE and optimization

Baker J and Brunson D (2000), 'Statistical modelling of engine systems', in Edwards S P, Grove D M and Wynn H P (eds), *Statistics for Engine Optimization*, Professional Engineering Publishing, Suffolk, pp. 61–82.

Baranescu R A, Desai S D and Ginder D A (1989), 'Optimization analysis of a diesel engine using cycle analysis and Taguchi methods', SAE paper 891895.

Bates R A, Fontana R, Randazzo C, Vaccarino E and Wynn H P (2000), 'Empirical modelling of diesel engine performance for robust engineering design', in Edwards S P, Grove D M and Wynn H P (eds), *Statistics for Engine Optimization*, Professional Engineering Publishing, Suffolk, pp. 163–173.

Box G E P and Draper N R (1987), *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, New York.

Brooks T, Lumsden G and Blaxill H (2005), 'Improving base engine calibrations for diesel vehicles through the use of DOE and optimization techniques', SAE paper 2005-01-3833.

Carey W R (1992), *Tools for Today's Engineer: Strategy for Achieving Engineering Excellence*, the 38th L. Ray Buckendale Lecture, SAE SP-913, SAE paper 920040, SAE International, Warrendale, PA.

Courteille E, Mortier F, Leotoing L and Ragneau E (2005), 'Multi-objective robust design optimization of an engine mounting system', SAE paper 2005-01-2412.

Dvorak T M and Hoekstra R L (1996), 'Optimizing internal combustion engine performance through response surface methodology', SAE paper 962525.

Edwards S P, Grove D M and Wynn HP (eds) (2000), *Statistics for Engine Optimization*, Professional Engineering Publishing, Suffolk.

Eriksson L, Johansson E, Kettaneh-Wold N, Wikstrom C and Wold S (1999), *Design of Experiments: Principles and Applications*, Umetrics AB, Sweden.

Fowlkes W Y and Creveling C M (1995), *Engineering Methods for Robust Product Design: Using Taguchi Methods in Technology and Product Development*, Addison-Wesley Publishing Company, Reading, MA.

Gardner T P (1992), 'Investigation of the effects of engine design parameters on diesel combustion and emissions using Taguchi methods', SAE paper 920116.

Goh T N (1994), 'Taguchi methods in practice: an analysis of Hoh's paradox', *Quality and Reliability Engineering*, 10(5), 417–421.

He Y and Rutland C (2004), 'Application of artificial neural networks in engine modelling', *International Journal of Engine Research*, 5(4), 281–296.

Hunter C E, Gardner T P and Zakrajsek C E (1990), 'Simultaneous optimization of diesel engine parameters for low emissions using Taguchi methods', SAE paper 902075.

Kazancioglu E, Wu G, Ko J, Bohac S, Filipi Z, Hu S J, Assanis D and Saitou K (2003), 'Robust optimization of an automobile valvetrain using a multi-objective genetic algorithm', *Proceedings of DETC2003/DAC-48714, ASME 2003 Design Engineering Technical Conference*, Chicago, IL, September 2–6.

Li Z, Kokkolaras M, Jung D, Papalambros P Y and Assanis D N (2004), 'An optimization study of manufacturing variation effects on diesel injector design with emphasis on emissions', SAE paper 2004-01-1560.

Lumsden G, Browett C, Taylor J and Kennedy G (2004), 'Mapping complex engines', SAE paper 2004-01-0038.

Mallamo F, Badami M and Millo (2004), 'Application of the design of experiments and objective functions for the optimization of multiple injection strategies for low emissions in CR diesel engines', SAE paper 2004-01-0123.

Montgomery D C (1991), *Design and Analysis of Experiments*, 3rd edn, John Wiley & Sons, New York.

Montgomery D C and Reitz R D (2000), 'Applying design of experiments to the optimization of heavy-duty diesel engine operating parameters', in Edwards S P, Grove D M and Wynn H P (eds), *Statistics for Engine Optimization*, Professional Engineering Publishing, Suffolk, pp. 83–98.

Myers R H and Montgomery D C (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd edn, John Wiley & Sons, New York.

Parmee I C and Watson A H (2000), 'An investigation of the utilization of genetic programming techniques for response curve modelling', in Edwards S P, Grove D M and Wynn H P (eds), *Statistics for Engine Optimization*, Professional Engineering Publishing, Suffolk, pp. 125–143.

Roepke K and Fischer M (2001), 'Efficient layout and calibration of variable valve trains', SAE paper 2001-01-0668.

Roy R (1990), *A Primer on the Taguchi Method*, Van Nostrand Reinhold, New York.

Rutter B, Hurley R, Eade D, Kisenyi J and Collingwood I (1996), 'Optimization of precious metal loadings in automotive catalysts using response surface methodology', SAE paper 961907.

Shayler P J, Dow P I, Hayden D J and Horn G (2000), 'Using neural networks in the characterization and manipulation of engine data', in Edwards S P, Grove D M and Wynn H P (eds), *Statistics for Engine Optimization*, Professional Engineering Publishing, Suffolk, pp. 145–161.

Siddall J N (1982), *Optimal Engineering Design: Principles and Applications*, Marcel Dekker, New York.

Taguchi G (1986), *Introduction to Quality Engineering*, Kraus International Publications, Whiter Plains, NY.

The MathWorks (2007), 'SAE 2007 SAE Commercial Vehicle Engineering Congress – MathWorks hosted panel on analytical calibration', SAE Panel Discussion. Available at: www.mathworks.com/mason/tag/proxy.html?dataid=9962&fileid...

Thiel M P, Klingbeil A E and Reitz R D (2002), 'Experimental optimization of a heavy-duty diesel engine using automated genetic algorithms', SAE paper 2002-01-0960.

Urano Y, Nakano Y, Takada H and Sugita M (2005), 'Optimization technique for transient emission reduction of heavy duty diesel engine', SAE paper 2005-01-1099.

Win Z, Gakkhar R P, Jain S C and Bhattacharya M (2002), 'Noise, emissions and fuel

economy investigation on a small DI diesel using Taguchi methods', SAE paper 2002-32-1793.
Yamamoto M, Yoneya S, Matsuguchi T and Kumagai Y (2002), 'Optimization of heavy duty diesel engine parameters for low exhaust emissions using the design of experiments', SAE paper 2002-01-1148.
Yan J, Rogalla R and Kramer T (1993), 'Diesel combustion and transient emissions optimization using Taguchi methods', SAE paper 930600.
Zottin W, Cuco A, Reis M and Silva R (2008), 'Application of optimization techniques in the design of engine components', SAE paper 2008-01-0219.

3.5.2 References and bibliography on probability, robust design, and reliability-based design optimization

Catania A E, d'Ambrosio S, Guenna G M, Petrolo D, Pilo L and Strazzullo L (2007), 'New modeling for reliable evaluation of parameter variability effects on vehicle fuel consumption', SAE paper 2007-01-0328.
Chan K Y, Kokkolaras M, Papalambros P, Skerlos S J and Mourelatoes Z (2004), 'Propagation of uncertainty in optimal design of multilevel systems: piston-ring/cylinder-liner case study', SAE paper 2004-01-1559.
Courteille E, Mortier F, Leotoing L and Ragneau E (2005), 'Multi-objective robust design optimization of an engine mounting system', SAE paper 2005-01-2412.
Daniels C and Miazgowicz K (2007), 'Incorporating design variation into a 1-D analytical model of a 4.6L-4V Ford engine for improving performance projections', SAE paper 2007-01-4098.
Dave A and Hampson G J (2003), 'Robust engine design using engine simulations', SAE paper 2003-01-0371.
Dinsmore M L and Unglenieks R (2005), 'Acoustical optimization using quasi-Monte Carlo methods and SEA modeling', SAE paper 2005-01-2431.
Dodson B and Schwab H (2006), *Accelerated Testing: A Practitioner's Guide to Accelerated and Reliability Testing*, SAE International, Warrendale, PA.
Donders S, d'Ippolito R, Auweraer H V, Hack M, Tzannetakis N, Farkas L and Desmet W (2007), 'Uncertainty-based design in automotive and aerospace engineering', SAE paper 2007-01-0355.
Ejakov M A, Yeager D A and Kayupov M (2003), 'Reliability and robustness analysis of engine ring-pack performance', SAE paper 2003-01-1221.
Evans M, Hastings N and Peacock B (2000), *Statistical Distributions*, 3rd edn, John Wiley & Sons, New York.
Farizal F and Nikolaidis E (2007), 'Assessment of imprecise reliability using efficient probabilistic reanalysis', SAE paper 2007-01-0552.
Fowlkes W Y and Creveling C M (1995), *Engineering Methods for Robust Product Design: Using Taguchi Methods in Technology and Product Development*, Addison-Wesley Publishing Company, Reading, MA.
Gale N F, Widener S K, Sui P C and Zhang H (1995), 'Analytical engine design methods: a review', SAE paper 950806.
Gogate S D and Duncan A E (2001), 'Digital prototype simulations to achieve vehicle level NVH targets in the presence of uncertainties', SAE paper 2001-01-1529.
Gu L and Yang R J (2003), 'Recent applications on reliability-based optimization of automotive structures', SAE paper 2003-01-0152.

Gu L, Li G and Yang R J (2004), 'An Excel based robust design tool for vehicle structural optimization', SAE paper 2004-01-1124.

Hampson G, Dave A, Tandel V and Smyth J (2002), 'AutoDOE optimization and direct execution of GT-Power engine simulations', *GT-Suite Users Conference*, Dearborn, MI, November 18.

Hoffman R M, Sudjianto A, Du X and Stout J (2003), 'Robust piston design and optimization using piston secondary motion analysis', SAE paper 2003-01-0148.

Kokkolaras M, Mourelatos Z, Louca L, Filipi Z, Delagrammatikas G, Stefanopoulou A, Papalambros P, Assanis D (2005), 'Design under uncertainty and assessment of performance reliability of a dual-use medium truck with hydraulic-hybrid powertrain and fuel cell auxiliary power unit', SAE paper 2005-01-1396.

Law A M (2006), *Simulation Modeling and Analysis*, 4th edn, McGraw-Hill, New York.

Li Z, Kokkolaras M, Jung D, Papalambros P Y and Assanis D N (2004), 'An optimization study of manufacturing variation effects on diesel injector design with emphasis on emissions', SAE paper 2004-01-1560.

Mourelatos Z P and Liang J (2005), 'A reliability-based robust design methodology', SAE paper 2005-01-0811.

Mourelatos Z P and Liang J (2006), 'A methodology for trading-off performance and robustness under uncertainty', *Transactions of the ASME, Journal of Mechanical Design*, 128, 856–863.

Nikolaidis E, Saleem S, Farizal F, Zhang G and Mourelatos Z P (2008), 'Probabilistic reanalysis using Monte Carlo simulation', SAE paper 2008-01-0215.

Rahman S and Sun R (2003), 'Robust engineering of engine cooling system', SAE paper 2003-01-0149.

Rahman S, Kayupov M, Li J and Mourelatos Z P (2007), 'Balance between reliability and robustness in engine cooling system optimal design', SAE paper 2007-01-0594.

Roy R (1990), *A Primer on the Taguchi Method*, Van Nostrand Reinhold, New York.

Savage J, Tang Y, Xie Z and Rahman S S (2007), 'Sensitivity analysis of powertrain cooling system performance', SAE paper 2007-01-0598.

Shapiro S S (1990), 'Selection, fitting, and testing statistical models', in Wadsworth H M (ed.), *Handbook of Statistical Methods for Engineers and Scientists*, McGraw-Hill, New York, pp. 6.1–6.34.

Tang Y, Grall N and Savage J (2008), 'Various impacting factors on a radiator top tank temperature', SAE paper 2008-01-1438.

Tao W, Yuan Y, Liu E A, Hill J, Zou Q and Barber G (2007), 'Robust optimization of engine lubrication system', SAE paper 2007-01-1568.

Yan J, Rogalla R and Kramer T (1993), 'Diesel combustion and transient emissions optimization using Taguchi methods', SAE paper 930600.

Zottin W, Cuco A, Reis M and Silva R (2008), 'Application of optimization techniques in the design of engine components', SAE paper 2008-01-0219.

Zou T, Mourelatos Z P and Mahadevan S (2004), 'Reliability analysis using Monte Carlo simulation and response surface methods', SAE paper 2004-01-0431.

4
Fundamentals of dynamic and static diesel engine system designs

Abstract: This chapter lays out the foundation of dynamic and static diesel engine system designs by linking the theoretical governing equations of the instantaneous engine in-cylinder cycle processes and the gas flow network of the air system. Engine manifold filling dynamics is discussed for dynamic system design. The chapter develops the theory of pumping loss and engine delta P, which are key design issues for modern high-EGR turbocharged diesel engines. The theory is used to predict engine hardware performance or determine hardware specifications to meet target performance. Four core equations for engine air system are proposed. Different theoretical options of engine air system design are summarized.

Key words: dynamic design, static design, engine thermodynamic cycle simulation, manifold filling dynamics, pumping loss, engine delta P, diesel engine air system.

4.1 Introduction to diesel engine performance characteristics

4.1.1 Engine performance maps

Engine performance is often characterized by the engine operating behavior in the speed–load domain, for example, the behavior of emissions, fuel consumption, noise, mechanical and thermal loading. Engine performance maps refer to the constant value contour plots of a given performance parameter in the speed–torque domain. A good understanding of engine performance maps is important to a system design engineer.

For an engine equipped with fixed-geometry hardware without flexible controls (e.g., fixed-geometry turbine, mechanical camshaft, mechanical water pump), optimum system performance is a large compromise between high-speed and low-speed operations, or between high-load and low-load operations. The trade-off effect is especially prominent when the range of the engine speed (or load) is wide. Mapping the operating characteristics of engine customer applications (e.g., vehicle driving cycles, marine engine load cycles) to the engine speed–load map and mapping the engine speed–load characteristics to the component characteristic map (e.g., a compressor map) are two important design techniques for a system engineer in conducting system integration. Owing to the paramount importance of engine speed–load characteristics, typical examples of engine performance maps are illustrated and discussed in this section.

For reference, engine glossary and abbreviations are provided in SAE J604 (1995). Engine testing is summarized in Martyr and Plint (2007). Engine power test codes can be found in SAE J1995 (1995) and J1349 (2008). Engine design details are provided in Haddad and Watson (1984a,b), Challen and Baranescu (1999) and Basshuysen and Schafer (2004).

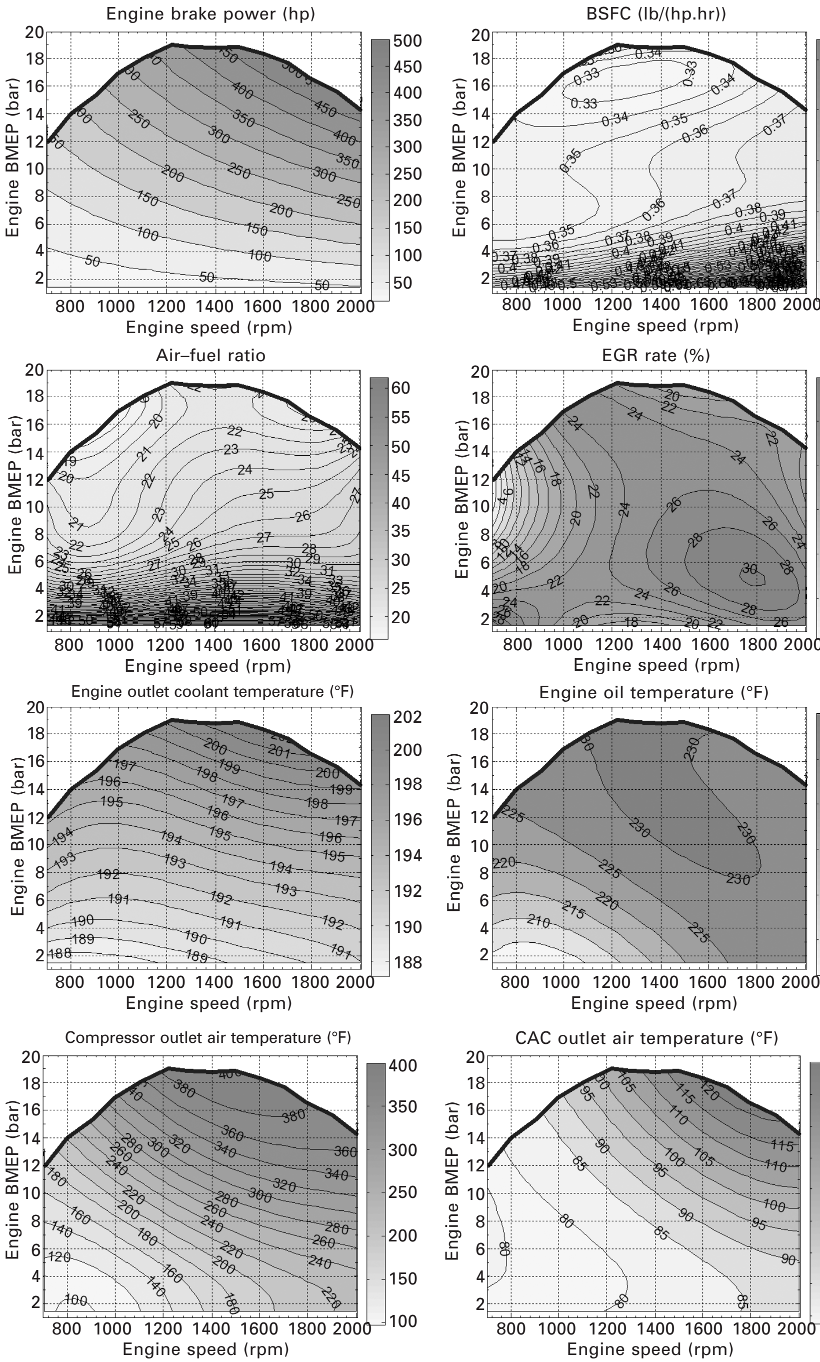
4.1.2 The characteristics of engine power and fuel consumption

Figure 4.1 presents the steady-state GT-POWER simulation data of a heavy-duty diesel engine at sea-level altitude and normal ambient temperature. The engine brake power map has second-order contours since power is equal to speed multiplied by torque. The operating points of a vehicle at a constant travelling speed but different transmission gear ratios are located along a constant-power curve.

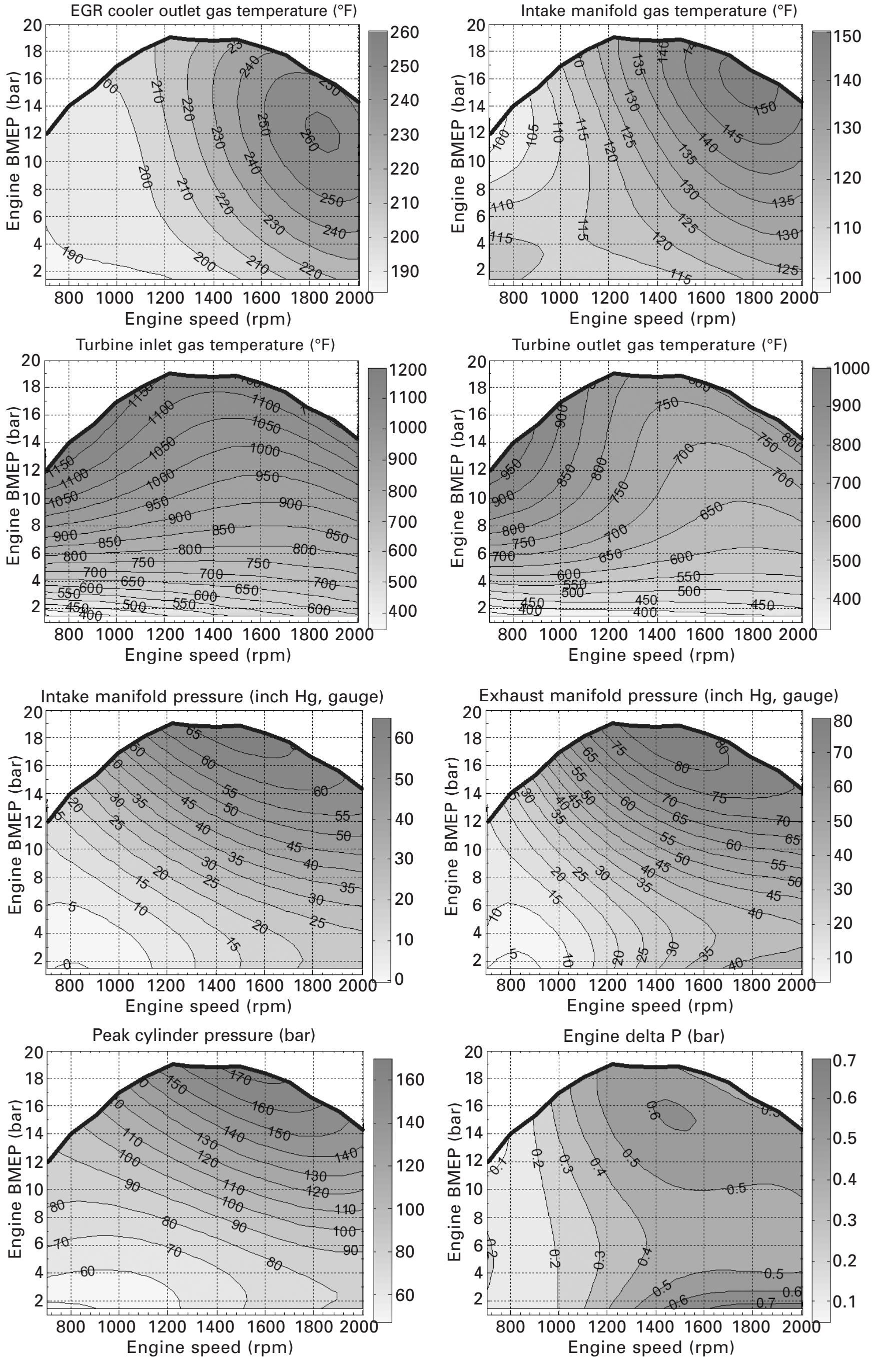
Engine BSFC contours usually exhibit a minimum (optimum) in the low-speed and high-load region such as at peak torque. BSFC is affected mainly by fuel injection timing, engine delta P, mechanical friction, and heat losses. The large increase in BSFC toward very low loads at a constant speed is primarily because the proportion of mechanical friction power relative to brake power is high, which means the engine mechanical efficiency is low. The increase of BSFC toward high engine speeds at a constant torque may be partly due to the increase in mechanical friction and engine delta P. The shape of the BSFC contours is also largely affected by the emissions calibration strategy, especially in the region above 25% load and between the peak-torque speed and rated speed.

4.1.3 The characteristics of engine air–fuel ratio and exhaust gas recirculation rate

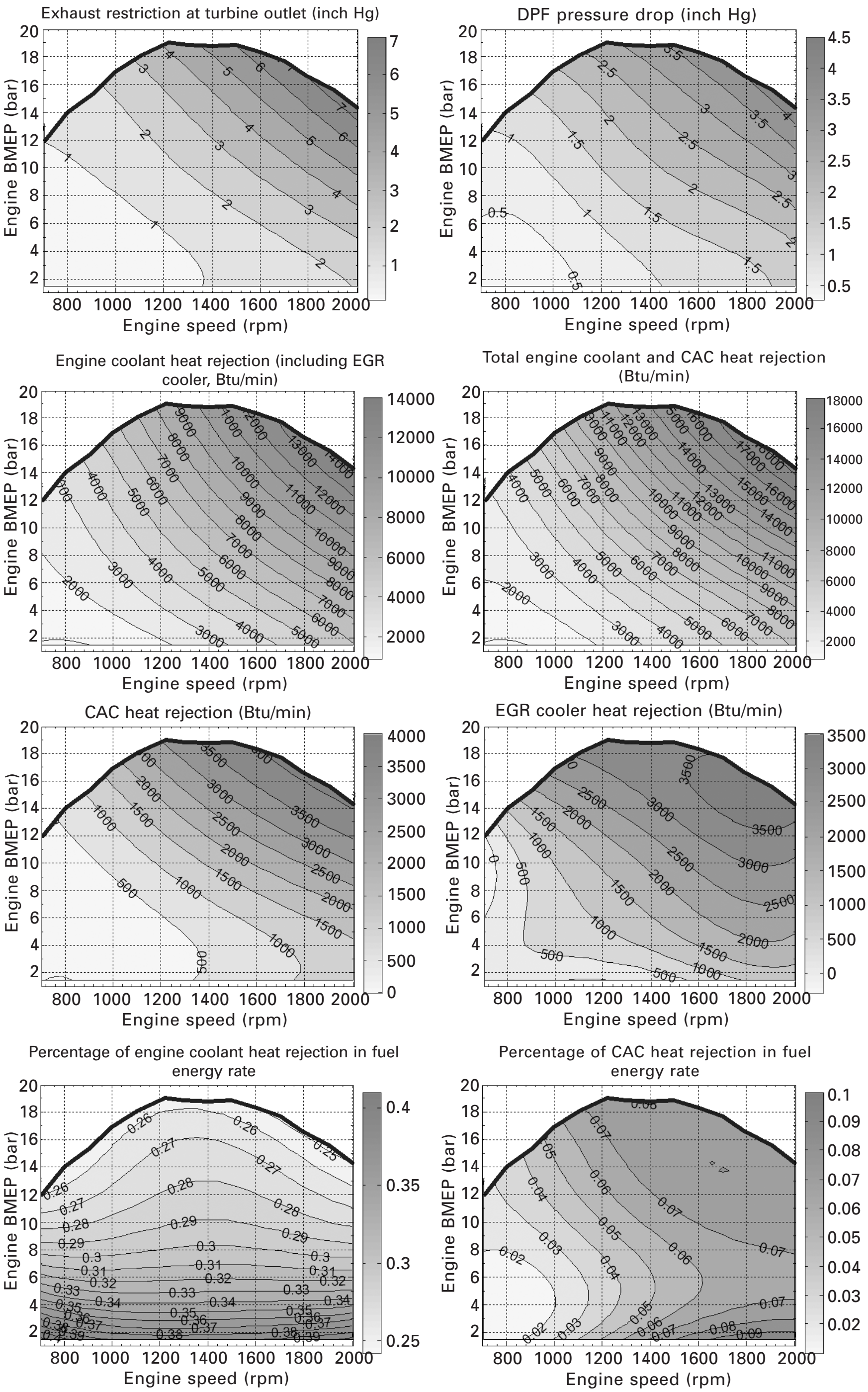
Engine air–fuel ratio is defined as the mass flow rate ratio of fresh air to fuel. The shape of the air–fuel ratio contours is largely affected by the type of turbocharger and EGR rate. An excessively high air–fuel ratio may produce high pumping loss, high peak cylinder pressure, and high compressor outlet temperature. An excessively low air–fuel ratio may produce the problems of deteriorated combustion efficiency, high smoke, and high exhaust gas temperature. Air–fuel ratio is affected by the engine air flow rate at a given engine speed and load mode, and the air flow rate is determined by the intake manifold boost pressure and engine volumetric efficiency. Unlike the fixed-geometry turbine or pneumatically controlled wastegated turbine, the electronically controlled wastegated turbine and VGT in modern diesel



4.1 Illustration of steady-state diesel engine system performance maps.



4.1 Continued



4.1 Continued

engines are able to flexibly regulate or reduce air–fuel ratio to a large extent or maintain it even at a nearly constant level, especially at medium or high loads.

EGR rate is defined as the mass flow rate ratio of EGR to the sum of fresh air and EGR. The shape of the EGR rate contours is determined primarily by emissions calibration. EGR rate directly affects coolant heat rejection, air–fuel ratio, and exhaust temperature.

4.1.4 Engine temperature characteristics

Engine outlet coolant temperature (or radiator inlet coolant temperature) is determined by coolant heat rejection and the characteristics of radiator and water pump. The coolant temperature usually reaches the highest value at full load but not necessarily at the rated speed. In fact, it often occurs at the peak torque or a medium speed. Engine oil temperature is affected by coolant flow rate and oil cooler heat rejection, which consists of the contributions from both mechanical friction heat and piston cooling. The highest oil temperature may occur at a lower engine speed than the rated speed.

Compressor outlet air temperature is affected mainly by compressor pressure ratio, flow rate, and efficiency. Charge air cooler (CAC) outlet air temperature (before mixing with EGR) is affected by cooler effectiveness and cooling medium temperature (also called cooler sink temperature). Excessively low CAC outlet temperature may generate water condensation problems. EGR cooler outlet temperature is affected by EGR cooler effectiveness, coolant temperature, and EGR flow rate. Again, excessively low EGR cooler outlet temperature may produce condensation problems of water and sulfuric acid. Intake manifold gas temperature (IMT) is the result of mixing between the fresh air at the CAC outlet and the EGR gas at the EGR cooler outlet. The required intake manifold gas temperature is determined by a compromise between the requirement to control NO_x emissions and the acceptable limit to control condensation for good durability. The highest IMT may not always occur at rated power.

Turbine inlet gas temperature (or exhaust manifold gas temperature) is determined mainly by a charge mass-to-fuel ratio, heat release rate and intake manifold gas temperature. The charge mass-to-fuel ratio here is defined as the mass flow rate ratio between the sum of fresh air and EGR to fuel. Turbine outlet temperature is affected by both turbine inlet temperature and turbine efficiency. It is important for aftertreatment performance and DPF regeneration, especially at low speeds/loads.

4.1.5 Engine pressure characteristics

Intake manifold pressure is determined by compressor pressure ratio or essentially turbine pressure ratio. It determines engine air flow rate through the relationship with volumetric efficiency. It also affects peak cylinder pressure and engine delta P. Exhaust manifold pressure is affected by turbine area and turbine flow rate. It affects turbine pressure ratio and engine delta P. The maximum peak cylinder pressure usually occurs at rated power. Engine delta P is affected mainly by engine air flow rate, turbine area, turbocharger efficiency and engine volumetric efficiency. Engine delta P greatly affects pumping loss and EGR driving. It is the most important parameter in the design of the air system of modern diesel engine. Engine delta P control is thoroughly elaborated in this book. Exhaust restriction usually refers to the total pressure drop from the turbine outlet to the ambient. Exhaust restriction is affected by engine air flow rate and the flow restriction characteristics of the exhaust and aftertreatment system. Exhaust restriction greatly affects turbine performance and hence air–fuel ratio. DPF pressure drop is variable with respect to time of operation, depending on the soot load in the DPF.

4.1.6 Engine heat rejection characteristics

Engine coolant heat rejection affects the coolant temperatures in the top tank, at the radiator inlet and engine inlet. It usually has the highest value near rated power or at a higher speed. Total engine coolant and CAC heat rejection presents the overall challenge for the vehicle cooling system design from the engine. Note that for engine manufacturers, the total heat rejection usually does not include transmission cooler heat rejection, which is normally estimated by the vehicle powertrain or drivetrain design team. However, in vehicle cooling system design, the heat rejections from the transmission cooler, the air conditioning and other miscellaneous sources all need to be taken into account. CAC or EGR cooler heat rejection is proportional to the air or gas flow rate and the gas temperature drop across the cooler. The percentage of fuel energy rate lost to engine coolant heat rejection is generally low at high loads and high at low loads. The reason is that at high loads the relative proportion of engine brake power in fuel energy is high, while the situation is opposite at low loads so that all the fuel energy primarily goes to coolant heat rejection and exhaust gas.

4.2 Theoretical formulae of in-cylinder thermodynamic cycle process

4.2.1 Overview of engine thermodynamic cycle calculations

Engine thermodynamic cycle analysis is the core of air system theory and the foundation for static (steady state) and dynamic (transient) engine system designs. An engine system model can be built with a modern commercial software package by putting the modeling elements together in a graphical programming environment. A large amount of engine performance simulation data can be produced by running the model to solve the differential equations involved. The system analysis engineers need to understand the cause-and-effect relationship between the relating parameters. They also need to interpret and present the computed results in a correct, intuitive, and simple fashion. For example, they need to explain what the reasons are for the change in pumping loss after the values of hardware design parameters and software calibration parameters are changed. The best way to understand the fundamentals is to examine the governing equations that describe the system behavior. Simplified forms of the equations are acceptable for this purpose. The parametric dependency can be revealed from the equations mathematically. This chapter presents the fundamental theory behind the cycle simulation software. Both the differential equations of the engine in-cylinder cycle process and the steady-state equations of the network system of the turbocharged EGR engine are described. Their links are discussed. The complementary sub-models of heat transfer, combustion and intake/exhaust valve flows are also provided.

Extensive reviews of internal combustion engine thermodynamic cycles have been provided by other researchers (Chow and Wyszynski, 1999; Ribeiro and Martins, 2007; Silva, 1993; Veiásquez and Milanez, 1995; Júnior *et al*., 2005; Primus, 1999; Proell, 1993; Makartchouk, 2002). The performance of internal combustion engines is introduced in several classical textbooks (Heywood, 1988; Stone, 1999; Pulkrabek, 2003; and Martyr and Plint, 2007) and SAE procedures (SAE J228, J604, J1312, J1829, J2548, and J2723). Turbocharging performance is reviewed in Watson (1999). System dynamics, which is the foundation for transient performance and model-based electronic controls, is introduced in Palm III (2005) and Ogata (2004). Diesel engine transient performance modeling is addressed in Rakopoulos and Giakoumis (2009).

4.2.2 Model assumptions of engine in-cylinder cycle process

The engine in-cylinder cycle process is determined by gas temperature (T), pressure (p) and mass (m). These parameters are solved by two ordinary differential equations for energy and mass conservations respectively and the ideal gas law equation. Initial conditions of in-cylinder p, T and m can be estimated from the intake manifold parameters, and the computation is conducted on a crank-angle basis with time-marching numerical integration until cyclic convergence is achieved for all the gas state parameters inside the entire engine. The assumptions for the in-cylinder thermodynamic model described in this chapter include the following:

- the in-cylinder single zone model (i.e., the burned and unburned mixture has a uniform temperature)
- zero-dimensional state parameters (i.e., at each moment, the pressure, temperature, or concentration in the cylinder are equal or homogeneous everywhere; no condensed species)
- ideal gases (i.e., the gases obey the equation of state $pV = mR_{gas}T$; the specific heat, internal energy, and enthalpy of the gases vary with temperature and constituents but not with pressure)
- the quasi-steady-state process for the gas flowing into and out of the cylinder (i.e., within a small time step in numerical integration the flow process is treated as steady state)
- ignoring the kinetic energy of the intake and exhaust gas flows.

4.2.3 Governing equations of the in-cylinder instantaneous cycle process

The energy conservation equation of the in-cylinder gas can be expressed as:

$$\frac{\mathrm{d}U}{\mathrm{d}\phi} = \frac{\mathrm{d}W}{\mathrm{d}\phi} + \sum_i \frac{\mathrm{d}Q_i}{\mathrm{d}\phi} + \sum_j h_j \cdot \frac{\mathrm{d}m_j}{\mathrm{d}\phi} \tag{4.1}$$

where ϕ is the crank angle (degrees), U is the internal energy of the in-cylinder gas, W is the mechanical work acting on the piston, Q_i is the heat exchanged through the system boundary and fuel combustion, h_j is the specific enthalpy, and $h_j m_j$ is the energy brought into and out of the cylinder by the intake and exhaust gas flow. The enthalpy can be calculated with a datum of zero enthalpy defined at 0 K or 298.15 K or any other arbitrary datum temperature. Note that a positive value for the heat or mass flow means flowing into the cylinder and a negative value means flowing out of the cylinder. Each term in equation 4.1 is further expressed as:

$$\frac{\mathrm{d}U}{\mathrm{d}\phi} = \frac{\mathrm{d}(m \cdot u)}{\mathrm{d}\phi} = u\frac{\mathrm{d}m}{\mathrm{d}\phi} + m\frac{\mathrm{d}u}{\mathrm{d}\phi} \tag{4.2}$$

$$\frac{\mathrm{d}W}{\mathrm{d}\phi} = -p\frac{\mathrm{d}V}{\mathrm{d}\phi} \tag{4.3}$$

$$\sum_i \frac{\mathrm{d}Q_i}{\mathrm{d}\phi} = \frac{\mathrm{d}Q_{fuel}}{\mathrm{d}\phi} + \frac{\mathrm{d}Q_{wall}}{\mathrm{d}\phi} \tag{4.4}$$

where V is the in-cylinder instantaneous volume, m is the in-cylinder gas mass, Q_{fuel} is the heat energy released from fuel combustion, and Q_{wall} is the heat transfer through the walls of the cylinder head, the piston and the liner. The energy of the intake and exhaust gas exchange mass flows is given by:

$$\sum_j h_j \cdot \frac{\mathrm{d}m_j}{\mathrm{d}\phi} = h_{in} \cdot \frac{\mathrm{d}m_{in}}{\mathrm{d}\phi} + h_{ex} \cdot \frac{\mathrm{d}m_{ex}}{\mathrm{d}\phi} \qquad 4.5$$

where m_{in} is the intake gas mass flowing into the cylinder, m_{ex} is the exhaust gas mass flowing out of the cylinder, and h_{in} and h_{ex} are the specific enthalpies of the gases at the intake and exhaust valves. Since the specific internal energy for ideal gases can be expressed as $u = u(T, \varsigma)$, where ς is the excess air–fuel ratio (defined as actual air–fuel ratio divided by stoichiometric air–fuel ratio, or the reciprocal of the 'equivalence ratio'), the following relationship is obtained:

$$\frac{\mathrm{d}u}{\mathrm{d}\phi} = \frac{\partial u}{\partial T} \cdot \frac{\mathrm{d}T}{\mathrm{d}\phi} + \frac{\partial u}{\partial \varsigma} \cdot \frac{\mathrm{d}\varsigma}{\mathrm{d}\phi} = c_v \cdot \frac{\mathrm{d}T}{\mathrm{d}\phi} + \frac{\partial u}{\partial \varsigma} \cdot \frac{\mathrm{d}\varsigma}{\mathrm{d}\phi} \qquad 4.6$$

Substituting these relationships into equation 4.1, the energy conservation equation is converted to the following form for solving the in-cylinder gas temperature T:

$$\frac{\mathrm{d}T}{\mathrm{d}\phi} = \frac{1}{m \cdot c_v} \times \left(\frac{\mathrm{d}Q_{fuel}}{\mathrm{d}\phi} + \frac{\mathrm{d}Q_{wall}}{\mathrm{d}\phi} - p\frac{\mathrm{d}V}{\mathrm{d}\phi} + h_{in}\frac{\mathrm{d}m_{in}}{\mathrm{d}\phi} + h_{ex}\frac{\mathrm{d}m_{ex}}{\mathrm{d}\phi} - u\frac{\mathrm{d}m}{\mathrm{d}\phi} - m\frac{\partial u}{\partial \varsigma} \cdot \frac{\mathrm{d}\varsigma}{\mathrm{d}\phi} \right) \qquad 4.7$$

The mass conservation equation can be expressed as

$$\frac{\mathrm{d}m}{\mathrm{d}\phi} = \frac{\mathrm{d}m_{in}}{\mathrm{d}\phi} + \frac{\mathrm{d}m_{ex}}{\mathrm{d}\phi} + \frac{\mathrm{d}m_{fuel\,B}}{\mathrm{d}\phi} \qquad 4.8$$

where m_{fuelB} is the fuel mass injected into the cylinder. If the total injected fuel mass per engine cycle is m_{fuelC} and the fraction of the fuel burnt is defined as $X_{fuel} = m_{fuelB}/m_{fuelC}$, then the following is obtained:

$$\frac{\mathrm{d}m_{fuelB}}{\mathrm{d}\phi} = m_{fuelC} \cdot \frac{\mathrm{d}X_{fuel}}{\mathrm{d}\phi} \qquad 4.9$$

The mass conservation equation is then converted to the following form:

$$\frac{\mathrm{d}m}{\mathrm{d}\phi} = \frac{\mathrm{d}m_{in}}{\mathrm{d}\phi} + \frac{\mathrm{d}m_{ex}}{\mathrm{d}\phi} + m_{fuelC}\frac{\mathrm{d}X_{fuel}}{\mathrm{d}\phi} \qquad 4.10$$

The heat release rate from the burnt fuel in equation 4.7 is

$$\frac{\mathrm{d}Q_{fuel}}{\mathrm{d}\phi} = \frac{\mathrm{d}m_{fuelB}}{\mathrm{d}\phi} \cdot q_{LHV} \cdot \eta_{com} = m_{fuelC} \cdot \frac{\mathrm{d}X_{fuel}}{\mathrm{d}\phi} \cdot q_{LHV} \cdot \eta_{com} \qquad 4.11$$

where q_{LHV} is the lower heating value of the fuel, and η_{com} is the combustion efficiency (η_{com} = 1 means complete combustion).

The ideal gas law applied to the in-cylinder gas gives the equation of state:

$$pV = mR_{gas}T \tag{4.12}$$

where R_{gas} is the gas constant. In the system of equations 4.7, 4.10 and 4.12, $\mathrm{d}Q_{wall}/\mathrm{d}\phi$, $h_{in}(\mathrm{d}m_{in}/\mathrm{d}\phi)$, and $h_{ex}(\mathrm{d}m_{ex}/\mathrm{d}\phi)$ are functions of gas temperature and pressure; $\partial u/\partial \varsigma$ and c_v are functions of gas temperature and constituents; $\mathrm{d}X_{fuel}/\mathrm{d}\phi$, $\mathrm{d}V/\mathrm{d}\phi$, m_{fuelC}, q_{LHV} and η_{com} are treated as known inputs. Therefore, the three unknowns, in-cylinder gas pressure p, temperature T, and mass m can be solved by a time-marching numerical integration. Equations 4.7 and 4.10 can be further simplified for different stages during an engine cycle, as explained below.

4.2.4 Simplification of governing equations for each stage of the cycle process

Compression stage (from intake valve closing to start of combustion)

In this stage, $\mathrm{d}m_{in}/\mathrm{d}\phi = 0$, $\mathrm{d}m_{ex}/\mathrm{d}\phi = 0$, $\mathrm{d}m_{fuel}/\mathrm{d}\phi = 0$, and $\mathrm{d}\varsigma/\mathrm{d}\phi = 0$. Therefore, $\mathrm{d}m/\mathrm{d}\phi = 0$ and $\mathrm{d}Q_{fuel}/\mathrm{d}\phi = 0$. The energy conservation equation 4.7 becomes

$$\frac{\mathrm{d}T}{\mathrm{d}\phi} = \frac{1}{m \cdot c_v}\left(\frac{\mathrm{d}Q_{wall}}{\mathrm{d}\phi} - p\frac{\mathrm{d}V}{\mathrm{d}\phi}\right) \tag{4.13}$$

Combustion stage (from start of combustion to end of combustion)

In this stage, $\mathrm{d}m_{in}/\mathrm{d}\phi = 0$ and $\mathrm{d}m_{ex}/\mathrm{d}\phi = 0$. Equation 4.3 becomes $\mathrm{d}m/\mathrm{d}\phi = \mathrm{d}m_{fuelB}/\mathrm{d}\phi = m_{fuelC}(\mathrm{d}X_{fuel}/\mathrm{d}\phi)$. Ignoring the impact of ς on u, the energy conservation equation 4.7 becomes

$$\frac{\mathrm{d}T}{\mathrm{d}\phi} = \frac{1}{m \cdot c_v}\left[m_{fuelC}(\eta_{com} \cdot q_{LHV} - u)\frac{\mathrm{d}X_{fuel}}{\mathrm{d}\phi} + \frac{\mathrm{d}Q_{wall}}{\mathrm{d}\phi} - p\frac{\mathrm{d}V}{\mathrm{d}\phi}\right] \tag{4.14}$$

Expansion stage (from end of combustion to exhaust valve opening)

In this stage, $\mathrm{d}m_{in}/\mathrm{d}\phi = 0$, $\mathrm{d}m_{ex}/\mathrm{d}\phi = 0$, $\mathrm{d}m_{fuelB}/\mathrm{d}\phi = 0$, and $\mathrm{d}\varsigma/\mathrm{d}\phi = 0$. Therefore, $\mathrm{d}m/\mathrm{d}\phi = 0$ and $\mathrm{d}Q_{fuel}/\mathrm{d}\phi = 0$. The energy conservation equation 4.7 becomes

$$\frac{\mathrm{d}T}{\mathrm{d}\phi} = \frac{1}{m \cdot c_v}\left(\frac{\mathrm{d}Q_{wall}}{\mathrm{d}\phi} - p\frac{\mathrm{d}V}{\mathrm{d}\phi}\right) \tag{4.15}$$

Exhaust stage (from exhaust valve opening to intake valve opening)

In this stage, $\mathrm{d}m_{in}/\mathrm{d}\phi = 0$, $\mathrm{d}m_{fuelB}/\mathrm{d}\phi = 0$ and $\mathrm{d}\varsigma/\mathrm{d}\phi = 0$. Equation 4.10 becomes $\mathrm{d}m/\mathrm{d}\phi = \mathrm{d}m_{ex}/\mathrm{d}\phi$. The energy conservation equation 4.7 becomes

$$\frac{\mathrm{d}T}{\mathrm{d}\phi} = \frac{1}{m \cdot c_v}\left[\frac{\mathrm{d}Q_{wall}}{\mathrm{d}\phi} - p\frac{\mathrm{d}V}{\mathrm{d}\phi} + (h_{ex} - u)\frac{\mathrm{d}m_{ex}}{\mathrm{d}\phi}\right] \tag{4.16}$$

Intake stage (from exhaust valve closing to intake valve closing)

In this stage, $\mathrm{d}m_{ex}/\mathrm{d}\phi = 0$ and $\mathrm{d}m_{fuelB}/\mathrm{d}\phi = 0$. Equation 4.10 becomes $\mathrm{d}m/\mathrm{d}\phi = \mathrm{d}m_{in}/\mathrm{d}\phi$. Ignoring the impact of ς on u, the energy conservation equation 4.7 becomes

$$\frac{\mathrm{d}T}{\mathrm{d}\phi} = \frac{1}{m \cdot c_v}\left[\frac{\mathrm{d}Q_{wall}}{\mathrm{d}\phi} - p\frac{\mathrm{d}V}{\mathrm{d}\phi} + (h_{in} - u)\frac{\mathrm{d}m_{in}}{\mathrm{d}\phi}\right] \tag{4.17}$$

Valve overlap stage (from intake valve opening to exhaust valve closing)

In this stage, $\mathrm{d}m_{fuelB}/\mathrm{d}\phi = 0$ and $\mathrm{d}Q_{fuel}/\mathrm{d}\phi = 0$. Equation 4.10 becomes $\mathrm{d}m/\mathrm{d}\phi = \mathrm{d}m_{in}/\mathrm{d}\phi + \mathrm{d}m_{ex}/\mathrm{d}\phi$. Ignoring the impact of ς on u, the energy conservation equation 4.7 becomes

$$\frac{\mathrm{d}T}{\mathrm{d}\phi} = \frac{1}{m \cdot c_v}\left[\frac{\mathrm{d}Q_{wall}}{\mathrm{d}\phi} - p\frac{\mathrm{d}V}{\mathrm{d}\phi} + (h_{in} - u)\frac{\mathrm{d}m_{in}}{\mathrm{d}\phi} + (h_{ex} - u)\frac{\mathrm{d}m_{ex}}{\mathrm{d}\phi}\right] \tag{4.18}$$

4.2.5 Key sub-models

Instantaneous in-cylinder volume

In equation 4.7, the instantaneous volume $\mathrm{d}V/\mathrm{d}\phi$ can be derived from engine geometry as

$$V = \frac{\pi B_E^2}{4}\left\{\frac{S_E}{\Omega - 1} + \frac{S_E}{2}\left[\left(1 + \frac{1}{f_{C\text{-}C}}\right) - \cos\phi - \frac{1}{f_{C\text{-}C}}\sqrt{1 - f_{C\text{-}C}^2 \cdot \sin^2\phi}\right]\right\} \tag{4.19}$$

$$\frac{dV}{d\phi} = \frac{\pi B_E^2 S_E}{8}\left[\sin\phi + \frac{f_{C\text{-}C}}{2}\frac{\sin(2\phi)}{\sqrt{1 - f_{C\text{-}C}^2 \cdot \sin^2\phi}}\right] \tag{4.20}$$

where Ω is the engine geometric compression ratio, B_E is the cylinder bore diameter, S_E is the engine stroke, and $f_{C\text{-}C}$ is the ratio between the connecting rod length (the 'center-to-center' length) and the crank radius.

Heat transfer through cylinder walls

The heat transfer terms in the energy conservation equation 4.7 can be derived in the following form:

$$\frac{\mathrm{d}Q_{wall}}{\mathrm{d}\phi} = \sum_{i=3} \frac{\mathrm{d}Q_{wall,i}}{\mathrm{d}\phi} = \frac{-1}{6N_E} \sum_{i=3} \alpha_g \cdot A_{wall,i}(T - T_{wall,i}) \qquad 4.21$$

where α_g is the instantaneous spatial-average heat transfer coefficient from the in-cylinder gas to the inner cylinder wall, N_E is the engine speed (rpm), A_{wall} is the heat transfer area, T_{wall} is the spatial-average temperature of the cylinder wall surface, and $i = 1, 2, 3$ refers to the cylinder head, the piston, and the liner, respectively. The cylinder liner heat transfer area is $A_{wall,3} = \pi B_E[L_{clear} + y(\phi)]$, where L_{clear} is the clearance height, and $y(\phi)$ is given by:

$$y(\phi) = \frac{S_E}{2}\left[\left(1 + \frac{1}{f_{C\text{-}C}}\right) - \cos\phi - \frac{1}{f_{C\text{-}C}}\sqrt{1 - f_{C\text{-}C}^2 \cdot \sin^2\phi}\right] \qquad 4.22$$

The α_g is critical for heat transfer calculation. The development history of the engine heat transfer theory has gone through three stages: (1) empirical modeling (pioneered by the Nusselt correlation in 1923); (2) semi-empirical similarity theory (represented by the Woschni correlation); and (3) the multi-zone CFD simulation since the 1980s. To date, the widely accepted and successful approach is still the Woschni correlation. It was developed in 1965 and was based on a correlation in the form of $Nu = 0.035Re^{0.8}Pr^{0.333}$, where Nu is the Nusselt number, Re is the Reynolds number, and Pr is the Prandtl number. It assumes that forced convection has a dominant effect on cylinder heat transfer as follows:

$$\alpha_g = K_1 B_E^{-0.214} (p v_{mp})^{0.786} \left(T^{-0.525} + K_2 \frac{T}{T_{ec}}\right) \qquad 4.23$$

In 1970 an improved formula was given by Woschni as:

$$\alpha_g = 820 B_E^{-0.2} p^{0.8} T^{-0.53} \left[C_1 v_{mp} + C_2 \frac{T_{bc} V_{cyl}}{p_{bc} V_{bc}} (p - p_{mot})\right]^{0.8} \qquad 4.24$$

in the unit of W/(m^2.K).

In the above formula, p is the in-cylinder gas pressure (MPa), T is the instantaneous in-cylinder bulk gas temperature (K), T_{ec} is the in-cylinder gas temperature at the end of compression, B_E is the cylinder diameter (m), and v_{mp} is the mean piston speed (m/s). Moreover, p_{bc}, T_{bc}, and V_{bc} are the in-cylinder pressure, temperature, and volume at the beginning of the compression stroke. V_{cyl} is the cylinder displacement (m^3), and p_{mot} is the in-cylinder pressure at the motoring condition without combustion. K_1 is a scavenging constant,

K_2 is a combustion constant, C_1 is a gas velocity coefficient, given by $C_1 = 6.18 + 0.2085(B_E\omega_p/v_{mp})$ for intake and exhaust strokes, and $C_1 = 2.28 + 0.154(B_E\omega_{pw}/v_{mp})$ for compression and expansion strokes, where ω_{pw} is the paddle wheel angular velocity (in radian per second) in the steady-flow swirl test. C_2 is a combustion chamber shape coefficient in the expansion stroke. C_2 = 0.00324 m/(K.s) for direct injection combustion chambers; C_2 = 0.00622 m/(K.s) for indirect injection combustion chambers; $C_2 = 0$ for intake, exhaust, and compression strokes. The last term in the Woschni correlation attempted to reflect the effects of radiation and combustion.

The Woschni correlation has several limitations, for example: (1) zero-dimensional homogeneous heat transfer was assumed; (2) the heat transfer equivalent diameter was treated as a constant B_E rather than an instantaneous parameter within an engine cycle; and (3) the modeling of radiation heat transfer was primitive. In fact, accurately determining α_g is very difficult, from either theoretical computation or experimental measurement. Fortunately, the impact of the accuracy of heat transfer modeling on the accuracy of computing engine power and gas flow rate is not extremely critical. For engine system-level cycle simulations, the main objective of accurately computing cylinder heat transfer is to predict the base engine heat rejection for cooling system design and exhaust manifold gas temperature. It is usually sufficient to apply a multiplier to the Woschni heat transfer coefficient α_g to calculate cylinder heat transfer more accurately. The multiplier is tuned based on the energy balance analysis of engine test data in order to capture the engine fundamental characteristics of the percentage of fuel energy lost to cylinder heat rejection. This topic is further explained in Chapter 12.

More information on engine cylinder heat transfer can be found in Woschni (1967), Hohenberg (1979), Jennings and Morel (1991), Hansen (1992), Imabeppu *et al.* (1993), Alkidas (1993), Shayler *et al.* (1996, 1997), Bohac *et al.* (1996), Wolff *et al.* (1997), Franco and Martorano (1999), Luján *et al.* (2003), Zeng and Assanis (2004), and Schubert *et al.* (2005).

Gas masses flowing into and out of the cylinder through engine valves

With the simplified assumptions of subsonic one-dimensional isentropic flow for intake valves, the intake mass flow rate entering the cylinder is given by:

$$\frac{\mathrm{d}m_{in}}{\mathrm{d}\phi} = \frac{C_{f,in}A_{in}p_{in}}{6N_E\sqrt{R_{in}T_{in}}} \cdot \sqrt{\frac{2\kappa_{in}}{\kappa_{in}-1}\left[\left(\frac{p}{p_{in}}\right)^{\frac{2}{\kappa_{in}}} - \left(\frac{p}{p_{in}}\right)^{\frac{\kappa_{in}+1}{\kappa_{in}}}\right]} \qquad 4.25$$

where N_E is the engine speed (rpm), $C_{f,in}$ is the intake valve flow coefficient, A_{in} is the intake valve instantaneous flow area, p_{in} and T_{in} are the pressure and temperature in the intake port just before the intake valve, respectively, R_{in} is the gas constant, κ_{in} is the ratio of specific heat capacities of the intake gas flow, and p is the in-cylinder pressure.

The exhaust mass flow rate out of the cylinder is given below. When

$$\frac{p_{ex}}{p} > \left(\frac{2}{\kappa_{ex}+1}\right)^{\frac{\kappa_{ex}}{\kappa_{ex}-1}} \qquad 4.26$$

the valve gas flow is subsonic and can be described as

$$\frac{\mathrm{d}m_{ex}}{\mathrm{d}\phi} = \frac{-C_{f,ex}A_{ex}p}{6N_E\sqrt{R_{ex}T}} \cdot \sqrt{\frac{2\kappa_{ex}}{\kappa_{ex}-1}\left[\left(\frac{p_{ex}}{p}\right)^{\frac{2}{\kappa_{ex}}} - \left(\frac{p_{ex}}{p}\right)^{\frac{\kappa_{ex}+1}{\kappa_{ex}}}\right]} \qquad 4.27$$

When

$$\frac{p_{ex}}{p} \le \left(\frac{2}{\kappa_{ex}+1}\right)^{\frac{\kappa_{ex}}{\kappa_{ex}-1}} \qquad 4.28$$

the valve gas flow is supersonic and can be described as

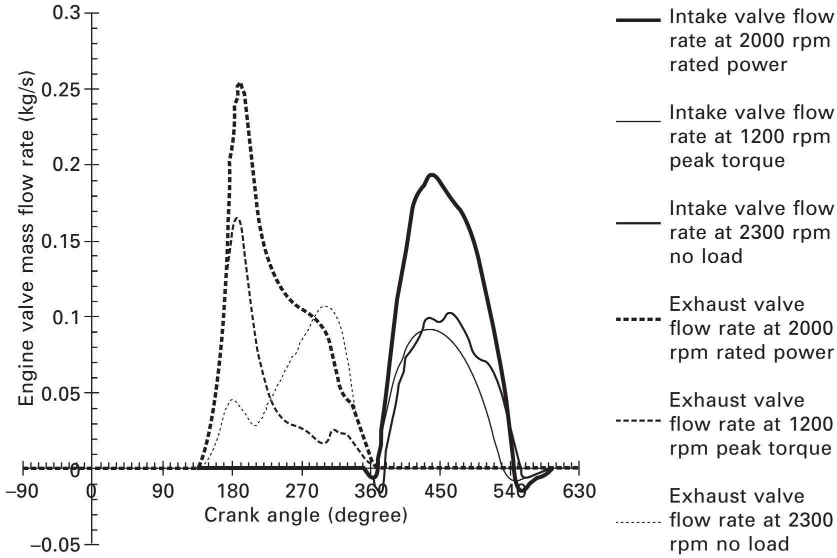
$$\frac{\mathrm{d}m_{ex}}{\mathrm{d}\phi} = \frac{-C_{f,ex}A_{ex}p}{6N_E\sqrt{R_{ex}T}}\left(\frac{2}{\kappa_{ex}+1}\right)^{\frac{1}{\kappa_{ex}-1}} \cdot \sqrt{\frac{2\kappa_{ex}}{\kappa_{ex}+1}} \qquad 4.29$$

where $C_{f,ex}$ is the exhaust valve flow coefficient, A_{ex} is the exhaust valve instantaneous flow area, and p_{ex} is the pressure in the exhaust port just behind the exhaust valve. Figure 4.2 shows typical engine valve flow rates in firing operation.

The valve and port flow coefficients $C_{f,in}$ and $C_{f,ex}$ can usually be obtained from the experimental flow bench test of the cylinder head. They are monotonic functions of normalized valve lift (Stone, 1999). The flow coefficients are determined by port design, valve diameter, valve seat angle, and flow recirculation effects between the valve and the cylinder bore. More detailed information about the valve/port flow coefficients can be found in Oldfield and Watson (1983), Agnew (1994), Danov (1997), Mattarelli and Valentini (2000), and Bohac and Landfahrer (1999).

Thermodynamic properties of in-cylinder gases

The in-cylinder specific internal energy is a function of temperature and species including the combustion products. For in-depth discussion of its calculation, the reader is referred to Stone (1999). Ferguson and Kirkpatrick



4.2 Illustration of engine valve flows in firing operation.

(2000) presented a method to calculate the thermodynamic properties of the combustion gas under chemical equilibrium.

Analysis of cylinder pressure and heat release rate for system design

The combustion heat release rate can be calculated by using a known measured cylinder pressure trace and the energy conservation equation:

$$\frac{dQ_{fuel}}{d\phi} = \frac{dU}{d\phi} - \frac{dW}{d\phi} - \frac{dQ_{wall}}{d\phi} \qquad 4.30$$

where $Q_{fuel} = m_{fuelC}X_{fuel}q_{LHV}\eta_{com}$. Or, alternatively the heat release rate may be specified as an input by using an empirical formula. The widely used Wiebe semi-empirical formula (also known as the Wiebe function) was developed based on the homogeneous chain reaction theory and gasoline engine test data. In the Wiebe function, the percentage of burned fuel X_{fuel} represents the combustion rate as follows:

$$X_{fuel} = 1 - e^{-C_{com}\left(\frac{\phi-\phi_{SOC}}{\Delta\phi_{com}}\right)^{\delta+1}} \qquad 4.31$$

$$\frac{dX_{fuel}}{d\phi} = C_{com}\frac{\delta + 1}{\Delta\phi_{com}}\left(\frac{\phi - \phi_{SOC}}{\Delta\phi_{com}}\right)^{\delta} \cdot e^{-C_{com}\left(\frac{\phi-\phi_{SOC}}{\Delta\phi_{com}}\right)^{\delta+1}} \qquad 4.32$$

where δ is a non-dimensional shape factor characterizing the instantaneous change of fuel concentration of the effective burning portions during the combustion process. The δ value is a function of engine type and speed, and affects the shape of the heat release rate. A smaller δ produces a faster burning rate, hence the peak of the normalized heat release rate occurs earlier. In the formula, $\Delta\phi_{com}$ is the combustion duration in the unit of degree crank angle, ϕ_{SOC} is the crank angle at the start of combustion (SOC), and C_{com} is a constant. If $X_{fuel} = 0.999$ (i.e., 99.9% of the fuel is burnt at the end of combustion), $C_{com} = 6.908$. The start of combustion is determined by the start of fuel injection and the ignition delay. The start of injection is usually defined as the moment when the injection needle has lifted a specified distance from its seat. Practically, the start of combustion is regarded as the moment when the heat release rate becomes zero or when the accumulated heat release rate reaches a minimum, or approximately as the moment when the first-order time derivative of the cylinder pressure trace reaches a minimum after the start of injection.

Ignition delay consists of both physical and chemical processes. In the physical processes, the fuel sprays break up, vaporize and mix with the air. In the chemical processes, pre-flame oxidation of the premixed fuel occurs, and localized ignition in multiple areas within the combustion chamber happens. Ignition delay is dependent upon in-cylinder pressure and temperature, mean piston speed, and fuel cetane number. A high cetane number reduces ignition delay and may prevent diesel knock. In the empirical modeling of diesel engine combustion, many efforts were made to predict the reaction rate of the fuel burning process (which is split into the premixed and the diffusion phases) based on ignition delay models or the more fundamental Arrhenius equation of reaction rate. Moreover, there are many models regarding fuel propagation rate and the diffusion of the oxygen into the fuel jet (Stone, 1999). More information on diesel engine ignition delay can be found in Ryan (1987), Rosseel and Sierens (1996), and Kamimoto *et al.* (1998).

Bypassing the complex details of the combustion processes, the Wiebe function provides a simple, convenient but still effective way to calculate the in-cylinder thermodynamic bulk pressure and temperature. The $\Delta\phi_{com}$, ϕ_{SOC} and δ in the Wiebe function can be easily determined by using experimental diagrams of the cylinder pressure trace and heat release rate analysis. These parameters can then be used as input data in cycle simulations. The formulae in equations 4.31 and 4.32 are based on a single Wiebe function that can simulate the heat release rate of low- and medium-speed diesel engines reasonably well. For more complex heat release rates which characterize the diesel premixed and diffusion combustion phases, two or three Wiebe functions can be linearly super-positioned together.

Predicting $\Delta\phi_{com}$, ϕ_{SOC}, and δ for other operating conditions based on a set of known $\Delta\phi_{com}$, ϕ_{SOC}, and δ values for a given condition was attempted by previous researchers, but still remains a significant challenge for modern EGR diesel engines equipped with advanced fuel systems. A practical approach is to rely on engine testing to obtain the experimental data for $\Delta\phi_{com}$, ϕ_{SOC}, and δ in the entire engine speed–load domain or at various ambient conditions, and then to empirically interpolate or surface-fit them to obtain the required values. Another approach is to rely on advanced combustion simulation (e.g., KIVA) to deduce certain general trends or phenomenological correlations of heat release rates.

Diesel engine combustion heat release rate has been extensively researched and modeled. Abundant information can be found in Meguerdichian and Watson (1978), Watson *et al.* (1980), Miyamoto *et al.* (1985), Grimm and Johnson (1990), Sierens *et al.* (1992), Tuccillo *et al.* (1993), Oppenheim *et al.* (1997), Homsy and Atreya (1997), Kamimoto *et al.* (1997), Ladommatos *et al.* (1998), Egnell (1999), Brunt and Platts (1999), Assanis *et al.* (2000), Nieuwstadt *et al.* (2000), Lakshminarayanan *et al.* (2002), Schihl *et al.* (2002), Hountalas *et al.* (2004), Cesario *et al.* (2004), Friedrich *et al.* (2006), Ponti *et al.* (2007), Manente *et al.* (2008), Nuszkowski and Thompson (2009), and Thor *et al.* (2009). Diesel engine combustion analysis is summarized in Hsu (2002) and Borman and Ragland (1998).

4.3 Engine manifold filling dynamics and dynamic engine system design

The thermodynamic processes of the transient performance of the intake and exhaust systems are introduced in this section. The engine manifold filling dynamics predicts the dynamic instantaneous values of the gas pressure, temperature, and flow rate in the intake manifold and the exhaust manifold as a function of time or crank angle. It is the foundation of dynamic engine system design in the time domain or crank angle domain, which is related to transient engine performance and engine controls. More information about the mathematical model formulation of dynamic engine system design can be found from the references provided in Chapter 14 (transient performance and electronic controls) in the sections regarding the mean-value models used by the engine controls community.

If the engine cylinder itself is the only subject of study for understanding the in-cylinder cycle process, a simplified approach to solve the above governing differential equations of the in-cylinder cycle process in the previous section is to assume the pressures in the intake port and the exhaust port are known constants as inputs. However, a more realistic and complex approach, often required by dynamic engine system design or high-fidelity static (steady state) engine system design, is to model them as instantaneous (i.e., dynamic) unknowns by using the equations that govern the gas flows in the intake and exhaust systems. Such an instantaneous calculation can be coupled with the equations of the in-cylinder cycle process if the purpose is for high-fidelity static or dynamic engine system design. The instantaneous manifold dynamics calculation can also be carried out independently, as normally practiced by the engine controls community, to decouple it from or bypass the equations of the in-cylinder cycle process, if the purpose is for approximate dynamic engine system design or electronic controls. When the crank-angle-based details of the in-cylinder cycle process are bypassed, the cylinder is treated as a cycle-average 'mean value' object, for example for the parameters of engine volumetric efficiency (representing the breathing or mass conservation feature of the engine), exhaust manifold gas temperature (representing the energy conservation feature of the engine) and engine brake power (derived from the p–V diagram and the mechanical friction of the engine). The mean-value model is a low-fidelity model in terms of the in-cylinder cycle process. The high-fidelity static engine system design refers to the simulation feature of including the manifold gas dynamics model to couple with the in-cylinder cycle process model. The high-fidelity dynamic engine system design refers to the simulation feature of including the in-cylinder cycle process to couple with the manifold gas dynamics. They share commonality in the logics and the methodology in diesel engine system design. Their difference is that the simulations in static system design can be conducted on a non-real-time basis, but the dynamic system design preferably needs to be conducted on a real-time basis. The details of the crank-angle-resolution real-time model used for high-fidelity dynamic engine system design are elaborated in Chapter 14.

If the gas properties in the intake and exhaust ducts are assumed to change only as a function of time or crank angle, and their variations along the duct dimension are ignored, then the zero-dimensional 'filling-and-emptying' method can be used as follows. Denoting p_3, T_3, m_3,V_3, and A_{3inl} as the gas pressure, the temperature, the mass, the volume, and the inbound effective flow area in the exhaust port and manifold, respectively; p, T, and m as the gas pressure, the temperature, and the mass, respectively, in the cylinder from which the gas is flowing into the exhaust control volume; and p_{3out}, T_{3out}, m_{3out}, and A_{3out} as the gas pressure, the temperature, the mass, and the outbound effective flow area at the downstream of the exhaust control volume (e.g., flowing to the turbine), respectively, the following equations are obtained:

$$\text{Mass conservation: } \frac{\mathrm{d}m_3}{\mathrm{d}\phi} = \frac{\mathrm{d}m}{\mathrm{d}\phi} + \frac{\mathrm{d}m_{3out}}{\mathrm{d}\phi} \tag{4.33}$$

$$\text{Energy conservation: } \frac{\mathrm{d}(m_3 u_3)}{\mathrm{d}\phi} = \frac{\mathrm{d}m}{\mathrm{d}\phi}h + \frac{\mathrm{d}m_{3out}}{\mathrm{d}\phi}h_3 + \frac{\mathrm{d}Q_{loss}}{\mathrm{d}\phi} \tag{4.34}$$

Substituting equation 4.33 into 4.34 gives

$$\frac{\mathrm{d}T_3}{\mathrm{d}\phi} = \frac{1}{m_3 c_{v3}}\left[\frac{\mathrm{d}m}{\mathrm{d}\phi}(h - u_3) + \frac{\mathrm{d}m_{3out}}{\mathrm{d}\phi}R_{ex}T_3 + \frac{\mathrm{d}Q_{loss}}{\mathrm{d}\phi}\right] \quad 4.35$$

where

$$\frac{\mathrm{d}m_{3out}}{\mathrm{d}\phi} = \frac{-A_{3out}\,p_3}{6N_E\sqrt{R_{ex}T_3}} \cdot \sqrt{\frac{2\kappa_{ex}}{\kappa_{ex}-1}\left[\left(\frac{p_{3out}}{p_3}\right)^{\frac{2}{\kappa_{ex}}} - \left(\frac{p_{3out}}{p_3}\right)^{\frac{\kappa_{ex}+1}{\kappa_{ex}}}\right]} \quad 4.36$$

Note that the above equation is valid for subsonic flow conditions, for example. If the flow is supersonic, the supersonic flow equation should be used accordingly. Applying the ideal gas law to the gas in the exhaust control volume (the port and the manifold) gives the following:

$$p_3V_3 = m_3R_{ex}T_3 \quad 4.37$$

The exhaust port heat loss can be modeled by forced convective water cooling. The heat transfer of the exhaust manifold is often modeled by natural convection heat transfer. T_3, m_3, and p_3 are solved by using the governing equations 4.33, 4.35, and 4.37 as a function of crank angle or time. The equations describing the gas dynamics in the intake port and manifold can be formulated in a similar way. The pressure and flow boundary conditions at the outlet of the exhaust manifold are modeled by the turbine flow characteristics (to be detailed in the next section). The pressure and flow boundary conditions at the inlet of the intake manifold are modeled by the compressor flow characteristics (to be detailed in the next section). In the mean-value model, the flow boundary conditions at the outlet of the intake manifold and at the inlet of the exhaust manifold are modeled by using the engine volumetric efficiency model which links the intake manifold pressure to the engine gas flow rate (to be detailed in the next section). In the crank-angle-based high fidelity model, the boundary conditions at the cylinder side for the intake/exhaust manifold control volumes are calculated by (or coupled with) the detailed in-cylinder cycle process model introduced in the previous section.

When the manifold pipe is short, the ‘filling-and-emptying’ method may give satisfactory simulation results. If the gas pressure wave propagation and reflection along the intake or exhaust pipe need to be calculated for manifold tuning design or valve timing analysis, the partial differential equations of one-dimensional gas pressure wave dynamics must be solved with the method of characteristics lines or finite volume or finite difference numerical methods (Benson, 1982; Horlock and Winterbone, 1986). Comprehensive theories of engine unsteady one-dimensional gas wave dynamics and the applications in manifold design were provided by Annand and Roe (1974) and Winterbone and Pearson (1999, 2000). Engine manifold unsteady gas wave dynamics was also researched by Blair (1991), Peters and Gosman (1993), Arias *et al*. (2000), Chalet *et al*. (2006) and Royo *et al*. (1994). The more complicated three-dimensional CFD modeling of manifold flows is usually a specialized component-level analysis, and is generally outside the scope of engine system design analysis.

4.4 Mathematical formulation of static engine system design

4.4.1 Predicting hardware performance

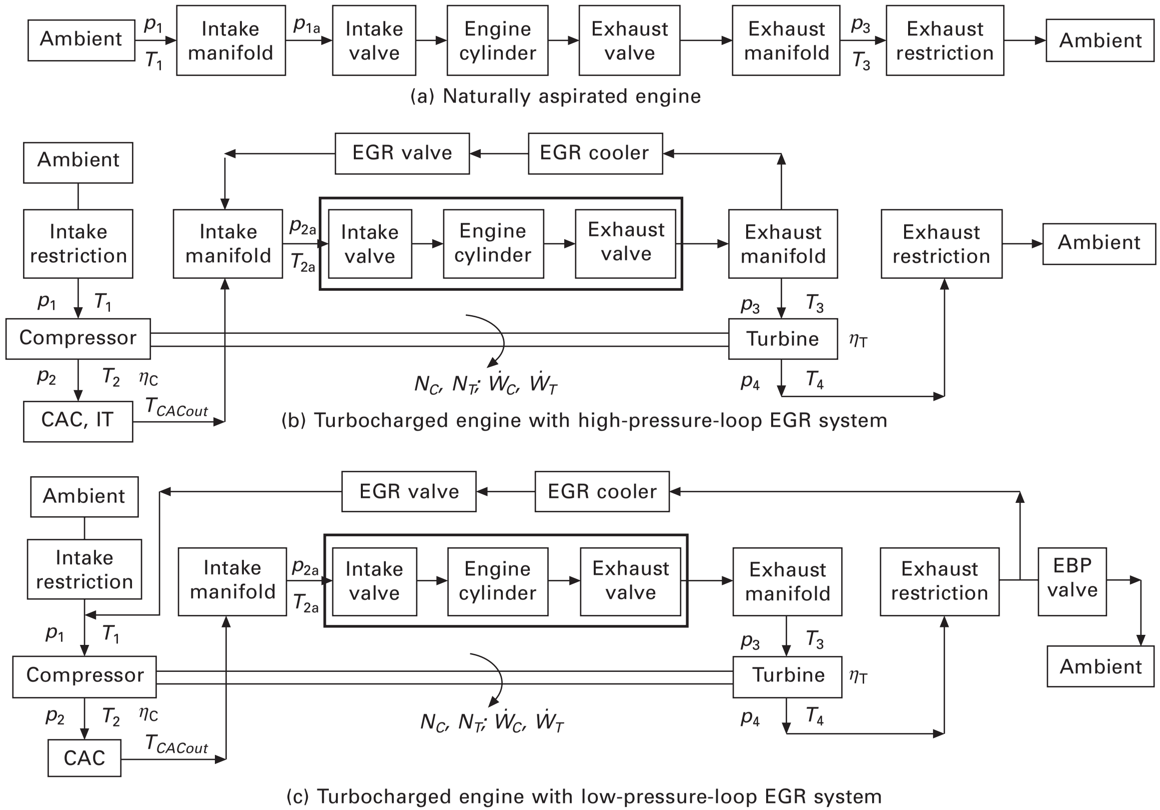
In order to understand engine air system design and the factors affecting pumping loss, the mathematical formulation of the entire engine gas flow network (Fig. 4.3) needs to be derived in a simple format so that a closed-form solution can be analyzed to facilitate an intuitive understanding.

After the instantaneous cylinder pressure p is solved by the in-cylinder cycle process, if the engine mechanical friction work $W_{f,E}$ is known (defined as negative value), the engine brake work W_E of total n_E cylinders can be calculated by using the following formula:

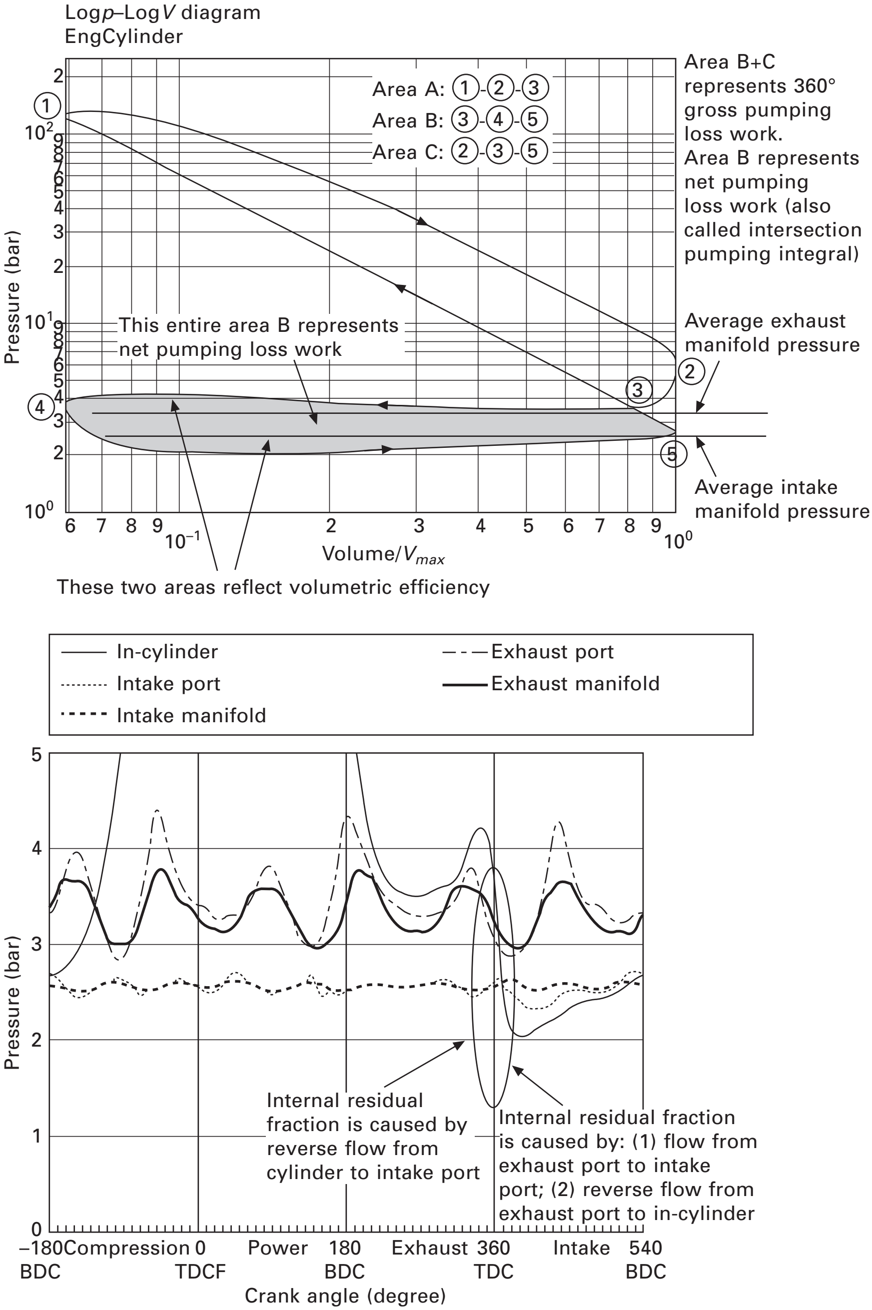
$$W_E = \sum_{j=1}^{n_E} \int p_j \cdot \mathrm{d}V_j + W_{f,E} \qquad 4.38$$

Equation 4.38 is the link between the in-cylinder cycle process and the performance of other subsystems in the engine gas flow network.

In the p–V diagram, the area encompassed by the curve of 'in-cylinder pressure vs. instantaneous volume' during intake and exhaust strokes is the work from pumping loss (Fig. 4.4). The PMEP in this book, unless especially mentioned, corresponds to the gross $p - V$ integral from 180° to 540° (i.e., including both areas of B and C in Fig. 4.4). However, the actual net pumping loss usually should be only the area B (Pierik and Burkhard, 2000). This distinction is important when evaluating variable valve actuation systems. The characteristic in-cylinder pressure differential for the pumping loss is $\Delta p_{cyl} = p_{exhaust} - p_{intake}$, where $p_{intake} = p_{IntakeManifold} - \Delta p_{in}$ and $p_{exhaust} = p_{ExhaustManifold} + \Delta p_{ex}$. The Δp_{in} is the pressure drop across the flow restrictions of the intake manifold, ports, and valves. The Δp_{ex} is the pressure drop across the flow restrictions of the exhaust valves, ports, and manifold. The pumping loss indicator Δp_{cyl} can be further derived as $\Delta p_{cyl} = (p_{ExhaustManifold} - p_{IntakeManifold}) + (\Delta p_{in} + \Delta p_{ex})$ = (engine delta P) + $(\Delta p_{in} + \Delta p_{ex})$. The engine delta P is defined as the exhaust manifold pressure minus the intake manifold pressure. It is noted that the pumping loss consists of two parts: the engine delta P, and the flow restrictions, which are related to volumetric efficiency. Both parts are equally important in diesel engine system design. Engine delta P is related to turbocharger and EGR circuit,



4.3 Schematic diagrams of engine system layout.



4.4 Pumping loss, engine delta P and manifold gas wave dynamics (GT-POWER simulation).

while volumetric efficiency is related primarily to the designs of cylinder head, valvetrain, and manifolds.

The system of equations governing the steady-state engine performance is derived below. Starting with a simpler example of the naturally aspirated non-EGR engine (Fig. 4.3(a)), at a given engine speed, fuel rate, and brake power, the four unknowns to be solved are the engine air flow rate $\dot{m}_{air}$, the exhaust manifold gas temperature T_3, the inlet pressure p_1, and the exhaust manifold pressure p_3. T_3 can be solved based on the engine energy balance of thermodynamic first law. The air flow rate $\dot{m}_{air}$ is solved based on the definition of engine volumetric efficiency by assuming the volumetric efficiency is a known value. The pressure p_1 is solved by using the intake flow restriction characteristics (i.e., a curve of pressure drop vs. volume flow rate). The pressure p_3 is solved by using the exhaust flow restriction characteristics.

For turbocharged EGR engines (Fig. 4.3(b) and (c)), the situation is much more complex. At a given engine speed, fuel rate and brake power, assuming the turbo characteristics (turbo maps) are known, there are 18 unknowns to be solved: $\dot{m}_{air}$, $\dot{m}_{EGR}$ T_3, p_4, $T_{1,ROA}$, p_1, p_2, T_2, p_3, T_4, N_C, N_T, η_C, η_T, p_{2a}, T_{2a} (for the high-pressure-loop EGR system) or T_1 (for the low-pressure-loop EGR system), T_{CACout}, and $T_{EGRcoolerGasOut}$. The detailed explanation of the high-pressure loop and low-pressure loop EGR systems is provided in Chapter 13 and Fig. 13.4. The engine delta P is $p_3 - p_{2a}$. In order to understand the system performance, a set of 18 equations is formed (equations 4.39–4.56) based on the thermal, flow or efficiency characteristics of each element or device. It should be noted that the formulated equations also serve as a foundation for developing future real-time model-based algorithms in advanced electronic controls.

Unlike using the detailed differential equations of the in-cylinder cycle process (equations 4.7, 4.10 and 4.12) for high-fidelity modeling, the engine in the air system circuit is treated as a single lumped element here in terms of energy balance so that the exhaust enthalpy or the exhaust manifold gas temperature can be calculated by the following:

$$\dot{Q}_{fuel} = \dot{W}_E + \Delta\dot{H}_{in-ex} + \dot{Q}_{base-coolant} + \dot{Q}_{miscellaneous} \qquad 4.39$$

where $\Delta\dot{H}_{in\text{-}ex}$ is the rise in gas enthalpy rate from the intake manifold to the exhaust manifold, $\dot{Q}_{base-coolant}$ is the base engine heat rejection to the coolant (if water cooled) or the cooling air (if air cooled), and $\dot{Q}_{miscellaneous}$, the miscellaneous heat losses, is treated as a known quantity here and is detailed in Chapter 12. Equation 4.39 is used to solve for the exhaust manifold gas temperature T_3 when the engine brake power $\dot{W}_E$ is assumed to be known or calculated from equation 4.38.

It should be noted that the exhaust temperature measured by a thermocouple is different from the time-averaged and the mass-averaged temperatures. This issue may cause large errors in engine cycle simulations. Proper care should be taken to ensure correct temperature input. More detailed discussions on this issue can be found in Caton (1982), Heywood (1988), Kar *et al.* (2004, 2006), and Son and Kolasa (2007).

The four-stroke engine may also be treated as a single lumped element in the air system circuit network for overall breathing performance by using the definition of intake manifold air–EGR mixture non-trapped volumetric efficiency shown as follows:

$$\eta_{vol} = \frac{2\dot{m}_{mixture}}{\rho_{2a} N_E V_E} = \frac{2(\dot{m}_{air} + \dot{m}_{EGR}) T_{2a} R_{gas}}{p_{2a} N_E V_E} \qquad 4.40$$

where η_{vol} is the engine volumetric efficiency with a reference gas density defined at the intake manifold, N_E is the engine crankshaft speed (revolution per second), V_E is the engine displacement, T_{2a} and p_{2a} are at the intake manifold, $p_{2a} = p_2 - \Delta p_{IntakeThrottle} - \Delta p_{CAC}$, $\dot{m}_{air}$ is the engine fresh air mass flow rate, and $\dot{m}_{EGR}$ is the EGR mass flow rate. The η_{vol} is related to valve size, valve timing, valve lift profile, port flow coefficient, and manifold design. The η_{vol} is also affected by intake manifold temperature, engine delta P and the consequent internal residual gas fraction. The trapped residue gas is affected by engine delta P and gas scavenging. In fact, it is very difficult to use hand calculations to compute η_{vol}. Note that equations 4.39 and 4.40 describe the engine macro behavior, and they are essentially the lumped simplified forms of the detailed in-cylinder cycle process described in equations 4.7, 4.10, and 4.12.

Volumetric efficiency is a very important system parameter. More information on engine volumetric efficiency can be found from Livengood *et al.* (1952), Fukutani and Watanabe (1979), Roussopoulos (1990), and Smith *et al.* (1999). Thorough elaborations on volumetric efficiency are provided in Chapters 9 and 13.

The lumped intake air flow restriction at the compressor inlet is given by:

$$\dot{m}_{air} = f_1(C_{d,int}, p_{ambient} - p_1, T_{ambient}) \qquad 4.41$$

where f_1 is a known function (e.g., second-order polynomial) and $C_{d,int}$ is a lumped flow restriction coefficient of the intake system at the compressor inlet, including the restrictions of the air filter and any regulating valves, for example, an intake throttle valve used in a low-pressure loop EGR system. Equation 4.41 is used to solve p_1. It should be noted that the intent here is not to provide a detailed precise form of the function *f*. The goal is rather to illustrate how to formulate the air system design problem mathematically to match the number of unknowns with the number of equations in order to facilitate a clear and intuitive understanding. The mathematical formulation of the design topic here cannot become over-constrained or under-constrained. Such a mathematical formulation is important for identifying various air system control 'knobs' and understanding their performance behavior.

The intake fresh air temperature at the compressor inlet is given by:

$$T_{1,ROA} = T_{ambient} + \Delta T_{ROA} \tag{4.42}$$

where ΔT_{ROA} is a 'rise-over-ambient' (ROA) value of the air temperature increase from the ambient to the compressor inlet. ΔT_{ROA} is related to intake pipe insulation and vehicle underhood thermal management. In a high-pressure loop EGR system, $T_1 = T_{1,ROA}$, where T_1 is the compressor inlet gas temperature. In a low-pressure EGR system, $T_1 > T_{1,ROA}$ due to EGR mixing.

The following equation of lumped exhaust flow restriction at the turbine outlet is used to solve the turbine outlet pressure p_4:

$$\dot{m}_{exh} = \dot{m}_C + \dot{m}_{fuel} + \dot{m}_{LubeOilCons} = f_2(C_{d,exh}, p_4 - p_{ambient}, T_4) \tag{4.43}$$

where $\dot{m}_{exh}$ is the exhaust mass flow rate, $\dot{m}_C$ is the compressor mass flow rate (note that in a high-pressure loop EGR system, $\dot{m}_C = \dot{m}_{air}$), $\dot{m}_{fuel}$ is the fuel flow rate, and $C_{d,exh}$ is a lumped exhaust restriction flow coefficient.

The lumped EGR circuit flow restriction is given by

$$\dot{m}_{EGR} = f_3(C_{d,EGR}, p_{EGRinl} - p_{EGRout}, T_{EGRcoolerOut}) \tag{4.44}$$

The pressure p_{EGRinl} refers to the gas pressure at the EGR circuit inlet, and p_{EGRout} refers to the gas pressure at the EGR circuit outlet. For example, in a high-pressure-loop (HPL) EGR system, $p_{EGRinl} = p_{ExhaustManifold}$. The $p_{ExhaustManifold}$ is equal to the turbine inlet pressure p_3 plus any pressure drop from the exhaust manifold or the EGR pickup location to the turbine inlet. In a HPL EGR system, $p_{EGRout} = p_{2a}$, and the engine delta P is the EGR driving force. In a low-pressure loop (LPL) EGR system, p_{EGRinl} is a pressure somewhere between p_4 and $p_{ambient}$, depending on where the EGR flow is picked up between different aftertreatment devices, and p_{EGRinl} is a function of $C_{d,exh}$ and p_4 or a function of $C_{d,int}$ and p_1. In a LPL EGR system, $p_{EGRout} = p_1$. $C_{d,EGR}$ is a lumped flow restriction coefficient including the flow restrictions of the EGR cooler and the tubing (i.e., fixed flow restrictions) as well as the flow restriction due to EGR valve opening (i.e., adjustable flow restriction). The discussion below is focused mainly on the HPL EGR system as an example.

The compressor outlet air temperature T_2 can be calculated based on the definition of compressor isentropic efficiency (η_C, defined on a total-to-total basis for temperatures and pressures) as below, assuming a single-stage compressor or two-stage compressor without inter-stage cooling:

$$\eta_C = \frac{(p_2/p_1)^{(\kappa_c-1)/\kappa_c} - 1}{(T_2/T_1) - 1} \tag{4.45}$$

The turbine outlet gas temperature T_4 can be calculated based on the definition of turbine isentropic efficiency (η_T, defined on a total-to-static basis) as below:

$$\eta_T = \frac{1 - (T_4/T_3)}{1 - (p_4/p_3)^{(\kappa_t - 1)/\kappa_t}} \tag{4.46}$$

The compressor power on a total-to-total basis is given by

$$\dot{W}_C = \frac{\dot{m}_C c_{p,C} T_1}{\eta_C}\left[\left(\frac{p_2}{p_1}\right)^{(\kappa_c - 1)/\kappa_c} - 1\right]$$

The turbine power is on a total-to-static basis given by

$$\dot{W}_T = \eta_T \dot{m}_T c_{p,T} T_3\left[1 - \left(\frac{p_4}{p_3}\right)^{(\kappa_t - 1)/\kappa_t}\right]$$

The turbocharger power balance is given by $\dot{W}_C = \dot{W}_T \eta_{TC,mech}$, where $\eta_{TC,mech}$ is the turbocharger mechanical efficiency if it has not been included in the turbine efficiency. The turbocharger power balance can be expanded as follows:

$$1 - \left(\frac{p_2}{p_1}\right)^{(\kappa_c - 1)/\kappa_c} + \eta_C \eta_T \eta_{TC,mech}\left(\frac{\dot{m}_T}{\dot{m}_C}\right)\left(\frac{c_{p,T}}{c_{p,C}}\right)\left(\frac{T_3}{T_1}\right)\left[1 - \left(\frac{p_4}{p_3}\right)^{(\kappa_t - 1)/\kappa_t}\right] = 0 \tag{4.47}$$

Note that the measured gas stand turbine efficiency is usually the product of the isentropic efficiency η_T multiplied by the mechanical efficiency of the turbocharger $\eta_{TC,mech}$. The overall turbocharger efficiency is defined as $\eta_{TC} = \eta_C \, \eta_T \, \eta_{TC,mech}$. Assuming the piston blow-by is negligible, the turbine flow rate and the compressor flow rate are related by $\dot{m}_T = \dot{m}_C + \dot{m}_{fuel} - \dot{m}_{WG}$, where $\dot{m}_{WG}$ is the turbine wastegate or bypass flow rate. It should be noted that there is a humidity effect on the c_p in equation 4.47 due to the water vapor in the atmosphere. The reader is referred to Israel and Hu (1993) for their discussions on the humidity effect on a composite gas constant and air density. The composite gas constant is a function of barometric pressure and relative humidity.

The compressor speed needs to be equal to the turbine speed, and this relationship is given as

$$N_C = N_T \tag{4.48}$$

The compressor efficiency map can be described as a function (e.g., a sixth-order polynomial) of corrected compressor flow rate and pressure ratio, as shown in any compressor map. This can be expressed as

$$\eta_C = f_4(\dot{m}_{C,corr}, p_2/p_1) \tag{4.49}$$

The turbine efficiency map with a given effective area can be described as a function (e.g., a sixth-order polynomial) of corrected turbine flow rate and pressure ratio, as shown in any turbine map. This can be expressed as

$$\eta_T = f_5(\dot{m}_{T,corr}, p_3/p_4) \tag{4.50}$$

The compressor speed map can also be described as a function of corrected compressor flow rate and pressure ratio as follows:

$$N_C = f_6(\dot{m}_{C,corr}, p_2/p_1) \tag{4.51}$$

The turbine speed map with a given effective area is also a function of corrected turbine flow rate and pressure ratio as follows:

$$N_T = f_7(\dot{m}_{T,corr}, p_3/p_4) \tag{4.52}$$

In fact, the turbine efficiency can also be described as a parabolic function of speed ratio based on its aerodynamic performance. Different turbine pressure ratios can form a family of parabolic curves. Denoting the subscript '*o*' for the total state and '*s*' for the static state, the speed ratio is defined by

$$\frac{v_T}{C_{T0}} = \frac{\pi d_T N_T}{\sqrt{2(h_{o3} - h_{s4})}}$$

The EGR and air mixing in the intake manifold for a HPL EGR system can be described by

$$\begin{aligned} &\dot{m}_{air} \int_0^{T_{CACout}} c_{p,air} \mathrm{d}T + \dot{m}_{EGR} \int_0^{T_{EGRcoolerGasOut}} c_{p,EGR} \mathrm{d}T \\ &= (\dot{m}_{air} + \dot{m}_{EGR}) \int_0^{T_{2a}} c_{p,mix} \mathrm{d}T \end{aligned} \tag{4.53a}$$

The EGR and air mixing at the compressor inlet for a LPL EGR system can be described by

$$\begin{aligned} &\dot{m}_{air} \int_0^{T_{1,ROA}} c_{p,air} \mathrm{d}T + \dot{m}_{EGR} \int_0^{T_{EGRcoolerGasOut}} c_{p,EGR} \mathrm{d}T \\ &= (\dot{m}_{air} + \dot{m}_{EGR}) \int_0^{T_1} c_{p,mixture} \mathrm{d}T \end{aligned} \tag{4.53b}$$

The effectiveness of the charge air cooler (CAC) is defined as

$$\varepsilon_{CAC} = \frac{T_2 - T_{CACout}}{T_2 - T_{CACcooling}} \tag{4.54}$$

The effectiveness of the EGR cooler is defined as

$$\varepsilon_{EGRcooler} = \frac{T_3 - T_{EGRcoolerGasOut}}{T_3 - T_{EGRcoolantInlet}} \tag{4.55}$$

A detailed discussion on cooler effectiveness and design parameters is provided in Chapter 12.

The lumped flow restriction of the charge air cooler and its downstream intake throttle valve (if any) is characterized by a lumped flow coefficient $C_{d,CAC\text{-}IT}$ as

$$\dot{m}_{air} = f_8(C_{d,CAC-IT}, p_2 - p_{2a}, T_{CACout}) \tag{4.56}$$

After the actual turbine mass flow rate $\dot{m}_T$ is solved by the above 18 equations, the turbine effective cross-sectional area A_T can be calculated using equation 4.57 for the compressible flow. This is provided as below for an axial flow turbine as a simplified illustration:

$$\dot{m}_T = A_T \cdot \frac{p_3}{\sqrt{R_{ex}T_3}} \cdot \sqrt{\frac{2\kappa_t}{\kappa_t - 1}} \cdot \sqrt{\left(\frac{p_4}{p_3}\right)^{2/\kappa_t} - \left(\frac{p_4}{p_3}\right)^{(\kappa_t+1)/\kappa_t}} \tag{4.57}$$

where p_3 and T_3 are defined in the total state. Note that equation 4.57 is only valid for subsonic flow conditions when

$$\frac{p_4}{p_3} > \left(\frac{2}{\kappa_t + 1}\right)^{\kappa_t/(\kappa_t - 1)} \tag{4.58}$$

However, when the following condition applies as

$$\frac{p_4}{p_3} \le \left(\frac{2}{\kappa_t + 1}\right)^{\kappa_t/(\kappa_t - 1)} \tag{4.59}$$

the gas flow becomes supersonic, and the turbine flow equation is described as

$$\dot{m}_T = A_T \cdot \frac{p_3}{\sqrt{R_{ex}T_3}} \left(\frac{2}{\kappa_t + 1}\right)^{1/(\kappa_t - 1)} \cdot \sqrt{\frac{2\kappa_t}{\kappa_t + 1}} \tag{4.60}$$

The flow equation for radial flow turbines, which are used in most automotive diesel engines, is much more complex and requires the modification of an enthalpy factor with respect to the term p_4/p_3 in equation 4.57. The turbine effective area A_T is the product of the following two contributing factors:

the turbine physical area, which is a constant related to the nozzle throat area and the exducer throat area; and the turbine flow coefficient, which is a variable as a function of turbine pressure ratio and turbine speed.

In equations 4.39–4.56, the parameters of gas pressure, temperature, and flow rate can be either cycle-average steady-state values or instantaneous quantities as a function of crank angle in a quasi steady-state. The earlier equation 4.7 in a transient differential form is essentially the same as the steady-state form equation 4.39 in nature. They are both based on the energy conservation that can solve for T_3. The earlier equation 4.10 is essentially the same as 4.40 in nature. They are both based on the mass conservation that can solve for $\dot{m}_{air}$. It is worth noting that the steady-state cycle-average integral of the in-cylinder pressure is essentially the engine power, reflected by equation 4.38. In the formulation of equations 4.39–4.56, the engine brake power is assumed to be a known input value. If the power is an unknown, the governing equation 4.12 of the in-cylinder instantaneous cycle process must be used to solve for the in-cylinder pressure, and then its steady-state counterpart equation 4.38 must be used to integrate over an engine cycle in order to solve for the brake power.

Because equations 4.51–4.52 are included to describe particular characteristics of the turbo hardware, the formulation of 4.39–4.56 can only be used to predict the engine performance of given hardware. Equations 4.39–4.56 consist of a nonlinear system that has to be solved iteratively by assuming an initial value for one of the 18 unknowns and iterating to convergence.

4.4.2 Hardware specification design for target performance

In the previous section, predicting hardware performance was discussed. In fact, in engine air system specification design, the objective is to find the required hardware (e.g., turbine effective area) to match with a given functional target of engine air flow rate (or gravimetric air–fuel ratio) and EGR rate. Recall that the EGR rate here is defined as the ratio between the EGR mass flow rate and the total mass flow rate of EGR and fresh air. If the desirable compressor and turbine efficiencies are assumed as fixed input, a system of 16 equations (4.39–4.48 and 4.51–4.56) is formed. The 16 unknowns are $\dot{m}_{air}$, $\dot{m}_{EGR}$, T_3, p_4, $T_{1,ROA}$, p_1, p_2, T_2, p_3, T_4, N_C, N_T, p_{2a}, T_{2a}, T_{CACout}, and $T_{EGRcoolerGasOut}$. The turbine speed can still be guaranteed to be equal to the compressor speed because equations 4.48, 4.51, and 4.52 are used. However, the turbocharger speed is usually not required in the output of engine system specification calculations. Instead, the turbine area is a required specification design parameter. Therefore, by deleting the turbocharger speed and the turbocharger map equations 4.48–4.52 that describe fixed turbo hardware characteristics, and by adding equation 4.57, a mathematical system of 14 equations (4.39–4.47 and 4.53–4.57) can be formed for engine system specification design. In fact, this system can be used for two purposes: (1) to calculate the required hardware for a given performance target; or (2) to calculate the performance for a given hardware configuration, although in a simplified form of 4.39–4.56. For example, the 14 unknowns can be $\dot{m}_{air}$ or A_T, $\dot{m}_{EGR}$, T_3, p_4, $T_{1,ROA}$, p_1, p_2, T_2, p_3, T_4, p_{2a}, T_{2a}, T_{CACout}, $T_{EGRcoolerGasOut}$ and it should be noted that this set of 14 equations can be reformulated to any smaller sets if the number of unknowns is reduced accordingly. In engine air system design, it is important to understand the engineering implications behind this mathematical formulation. When $\dot{m}_{air}$ is known and A_T is unknown, the required turbine area needs to be found to match the given air flow rate. In such a calculation without using turbocharger speed maps and turbocharger efficiency maps, the calculation does not guarantee that the turbine speed is equal to the compressor speed, although it still calculates the flow rate and the pressure ratio of the compressor and the turbine.

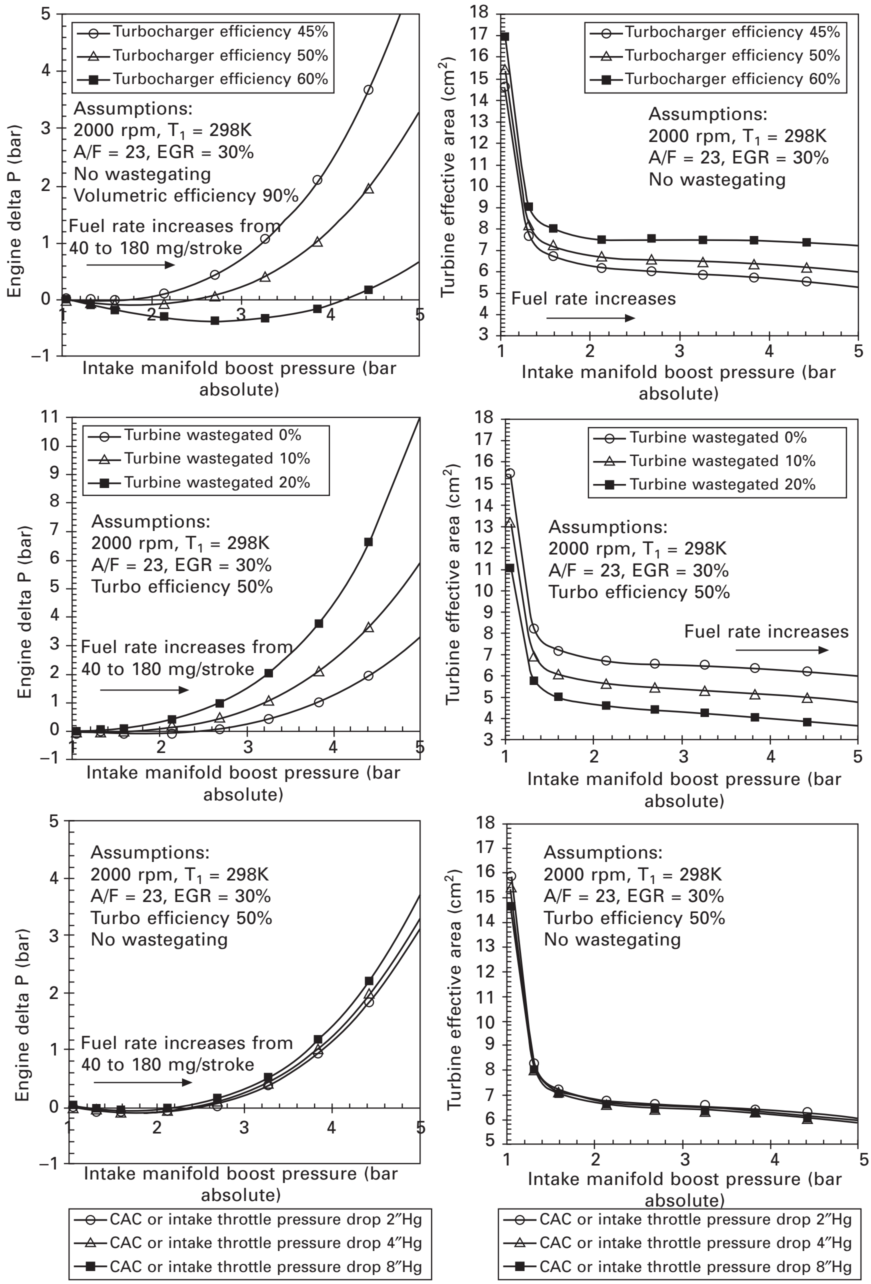
In system specification design analysis, in order to achieve the minimum engine delta P at a fixed target of air–fuel ratio and EGR rate, it is often convenient and necessary for the engine manufacturer to start by computing the required turbocharger efficiency and propose the desired turbocharger specification (i.e., turbocharger flow rate, pressure ratio, and a 'fixed' required efficiency) to the turbocharger suppliers. It should be noted that the assumption of imposed turbocharger efficiency sometimes might not be realistic for real turbocharger hardware, especially when the fixed efficiencies are specified at multiple engine speed and load modes. The turbocharger suppliers have the responsibility to select or design turbochargers to reach the efficiency requirements specified by the engine manufacturer. They are also responsible for matching a turbocharger with a proper rotating shaft speed to reach the desired flow rate and pressure ratio specified by the engine manufacturer as closely as possible. After obtaining the proposed turbocharger maps from the suppliers, the engine manufacturer can check turbocharger hardware performance with computer simulations (i.e., using the formulation of equations 4.39–4.56). The performance gap between the proposed turbocharger hardware and the system design specification can then be identified.

4.4.3 The characteristics of engine delta P

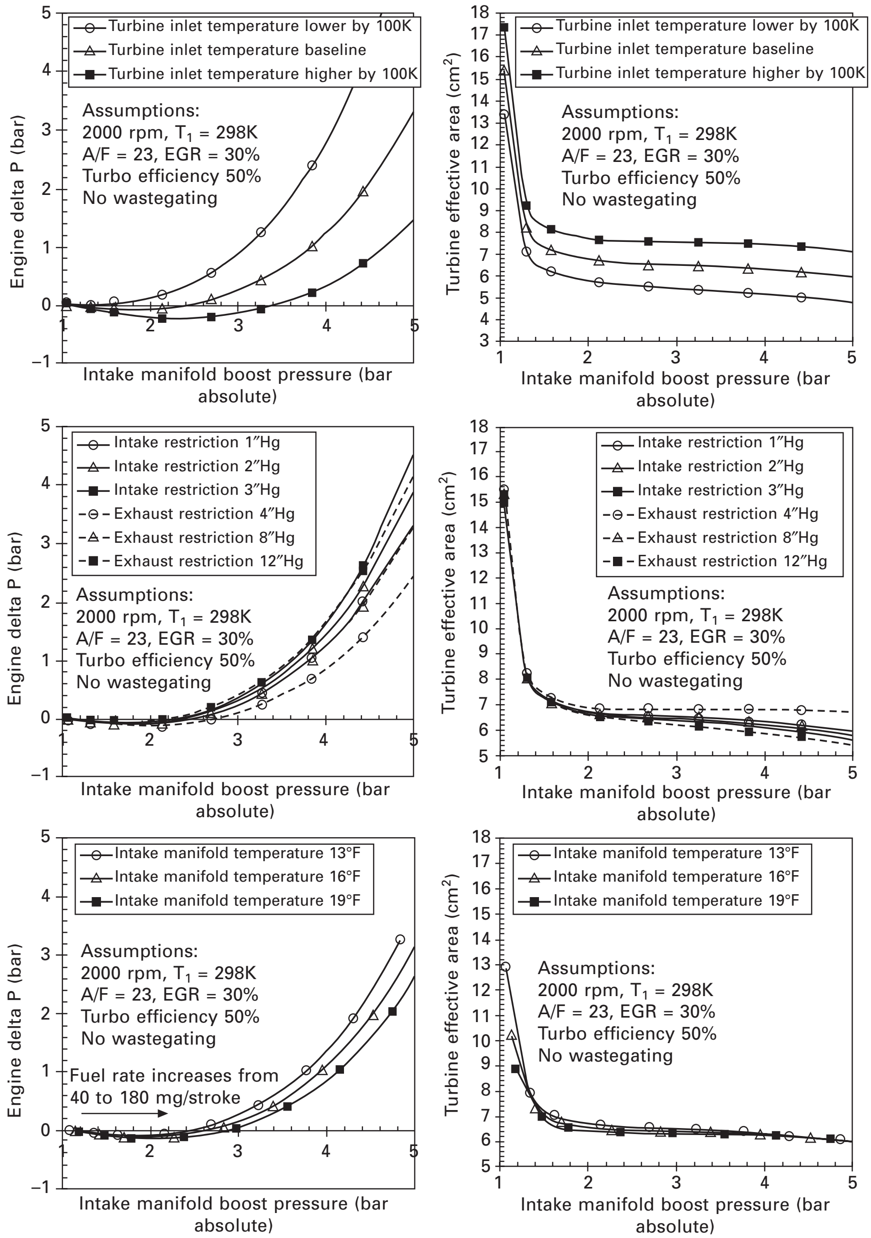
Engine delta P is a critical part of pumping loss. There are four core equations related to engine air system performance: equation 4.40 for engine–turbocharger coupled breathing characteristic, equation 4.44 for EGR circuit flow restriction, equation 4.47 for linking the intake manifold boost pressure to the exhaust manifold pressure, and equation 4.57 (or its equivalent counterpart of radial flow turbines) for generating the engine air flow by the turbine. Substituting 4.40 into 4.47, the parametric dependency of engine delta P ($p_3 - p_{2a}$) can be reflected by 4.47 and 4.57. With the engine air flow rate and the EGR rate as known target inputs (or equivalently the intake manifold pressure p_{2a} and the EGR rate as known targets by using 4.40), when T_3 and p_4 are constant, if the turbine area A_T varies, the exhaust manifold pressure p_3 changes according to 4.57. Thus, a change in p_2, $\eta_C \eta_T \eta_{TC,mech} (\dot{m}_T / \dot{m}_C)(T_3/T_1)$ or p_4 is required in order to balance the equation in 4.47. Therefore, the parameters that can reduce the engine delta P (i.e., minimize p_3 at a given p_{2a}) are shown directly in 4.47 and 4.57. These parameters are: p_1, p_2, the term $\eta_C \eta_T \eta_{TC,mech} (\dot{m}_T / \dot{m}_C)(T_3/T_1)$, p_4, and A_T. Note that the compressor inlet temperature T_1 is affected by ROA. The compressor inlet pressure p_1 is affected by the degree of flow throttling at the compressor inlet. The compressor outlet pressure p_2 is affected by engine speed, volumetric efficiency (especially valve timing), effective engine displacement and intake manifold temperature, as shown in equation 4.40. The pressure p_2 is also affected by charge air cooler flow restriction and intake throttle as shown in equation 4.56. The turbine inlet temperature T_3 is affected by cylinder cooling and heat losses from the exhaust port and the exhaust manifold, as shown in equation 4.39. T_3 is also affected by the air–fuel ratio. The turbine flow rate $\dot{m}_T$ is affected by turbine wastegate opening. The turbine outlet pressure p_4 is affected by aftertreatment exhaust restriction $C_{d,exh}$ as shown in equation 4.43. The turbocharger efficiency $\eta_C \eta_T \eta_{TC,mech}$ and the turbine area A_T are turbocharger design parameters.

EGR circuit flow restriction also affects engine delta P and other system design parameters. In a HPL EGR system, the EGR circuit flow restriction coefficient $C_{d,EGR}$ (or lumped EGR valve opening) can be solved using equation 4.44 when $p_3 - p_{2a}$ and the EGR flow rate are known. On the other hand, at a fixed air flow rate and EGR rate (or equivalently, at a fixed EGR mass flow rate), when the EGR circuit flow restriction is also fixed, p_3 must be calculated using equation 4.44. Then the turbine area A_T needs to change according to equation 4.57, and simultaneously another change is required in one of the following in order to balance equation 4.47: p_1, p_2, the term $\eta_C \eta_T \eta_{TC,mech} (\dot{m}_T / \dot{m}_C)(T_3/T_1)$, or p_4.

Using the model of exhaust manifold gas temperature as shown in equation 14.4 in Chapter 14, the factors affecting the cycle-average engine delta P are obtained by solving equations 4.40, 4.47, and 4.57. Figure 4.5 illustrates that the following design may reduce engine delta P at a given speed–load mode and fixed targets of air–fuel ratio and EGR rate: higher turbocharger efficiency, higher exhaust manifold gas temperature, less turbine wastegating, lower exhaust restriction pressure drop, and lower charge air cooler pressure drop. It is also observed that at higher engine load (fueling rate), engine delta P becomes higher. It is worth noting that in some engine designs with reed



4.5 Theoretical analysis of engine delta P.



4.5 Continued

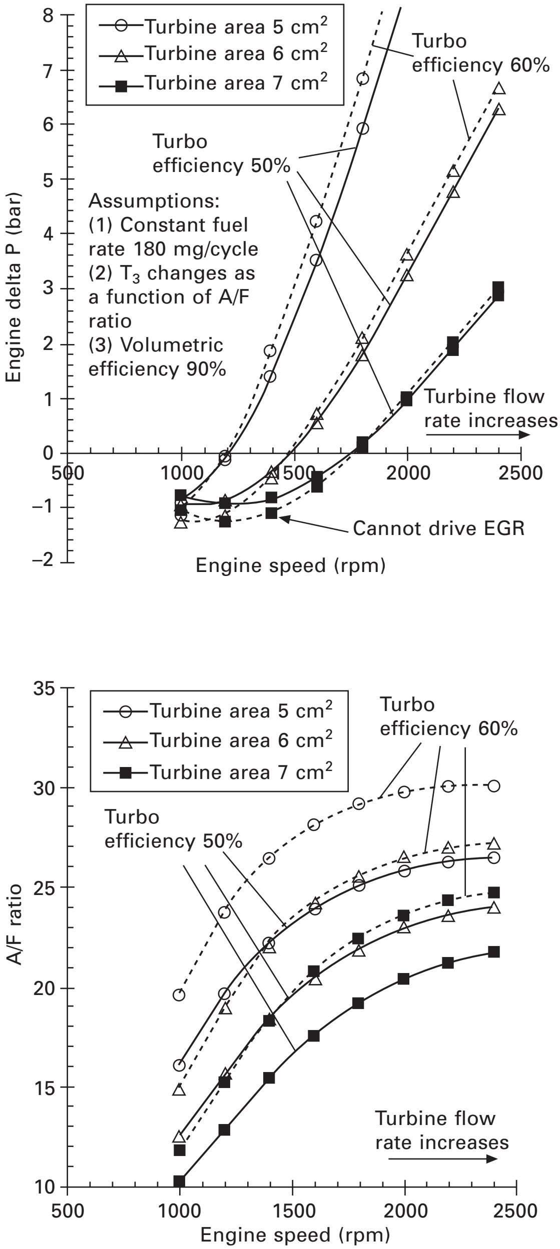
valves (check valves), EGR flow can be harvested into the cylinder with the pulsating pressure differential even if the cycle-average engine delta P becomes slightly negative.

The effect of turbine wastegating can be explained as follows. With wastegating the turbine flow loss must be compensated by an increase in the turbine pressure ratio in order to maintain a sufficient turbine power in order to deliver a target compressor boost pressure. The larger turbine pressure ratio results in an inevitable increase in both the exhaust manifold pressure and the engine delta P. Variable geometry turbine (VGT) does not have wastegate flow losses, hence it may produce lower pumping loss and BSFC than a wastegated turbine, provided both turbines have a similar efficiency. It should be noted that in some two-stage turbochargers, there is usually a high-pressure-stage turbine bypass valve or wastegate to divert the exhaust flow to the low-pressure-stage turbine inlet. The bypassed exhaust flow in a two-stage turbocharger is not all wasted because it is utilized by the low-pressure-stage turbine to deliver certain boost. But some flow throttling loss does occur at the wastegate when the valve is partially open.

In EGR engine design, insufficient engine delta P and inadequate capability to drive EGR flow often occur at low engine speeds, especially at the peak torque condition. In a turbocharged engine, the turbine effective area delivers and controls the exhaust manifold pressure and the compressor boost pressure. It acts as a restricting orifice and also a power-producing element to drive the compressor. The required turbine area can be determined using equation 4.57. It is the turbine area that primarily decides the engine delta P relationships described above. With a fixed turbine area, engine delta P reduces as the engine speed decreases at a fixed load, as shown in Fig. 4.6. The figure is obtained by using the four core equations (4.40, 4.44, 4.47, and 4.57). It is observed that a smaller turbine area produces higher engine delta P to enable EGR driving at low speeds. Figure 4.6 also shows that at a fixed target air–fuel ratio at a low speed, turbocharger efficiency should not be too high because otherwise the large turbine area would have produced very low or even negative engine delta P that cannot drive EGR flow (EGR rate 30% in Fig. 4.6).

4.4.4 Summary of theoretical options of engine air system design

From the set of 14 equations used for engine system specification design (4.39–4.47 and 4.53–4.57), any 14 parameters can be chosen as unknowns to formulate the equation system in order to evaluate the effects of engine hardware design or operating conditions. Sixteen typical air systems are listed in Table 4.1 to cover most air system design possibilities with HPL EGR. In Table 4.1 the second column lists possible unknowns, either a performance parameter or a hardware design parameter. One unknown needs to be chosen from each equation in order to formulate an air system. The parameters listed under each system number are the 14 unknowns. The parameters not shown



4.6 The effect of engine speed on engine delta P.

under each system number but shown in the second column are assumed as known input data. The **hardware** or **calibration** parameters are shown as **bold** in Table 4.1. For example, in system 1, η_C (compressor efficiency),

Table 4.1 Mathematical formulation of HPL EGR engine air systems at given engine speed, fueling rate, and power

Eq. no.	Possible unknowns	Engine system number															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4.39	T_3, $Q_{base\text{-}coolant}$	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3	T_3
4.40	η_{vol}, $\dot{m}_{air}$, $\dot{m}_{EGR}$, T_{2a}, p_{2a}, V_E	$\dot{m}_{air}$	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	p_{2a}	T_{2a}
4.41	$\dot{m}_{air}$, $C_{d,int}$, p_1, $p_{ambient}$	p_1	p_1	p_1	p_1	p_1	p_1	p_1	p_1	p_1	p_1	p_1	p_1	p_1	p_1	p_1	p_1
4.42	$T_{1,ROA}$, ΔT_{ROA}, $T_{ambient}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$	$T_{1,ROA}$
4.43	$C_{d,exh}$, p_4, $p_{ambient}$	p_4	p_4	p_4	$\boldsymbol{C_{d,exh}}$	p_4	p_4	p_4	$\boldsymbol{C_{d,exh}}$	p_4	p_4	$\boldsymbol{C_{d,exh}}$	p_4	p_4	$\boldsymbol{C_{d,exh}}$	$\boldsymbol{C_{d,exh}}$	p_4
4.44	$\dot{m}_{EGR}$, $C_{d,EGR}$, p_3, p_{2a}	$\dot{m}_{EGR}$	$\boldsymbol{C_{d,EGR}}$	$\boldsymbol{C_{d,EGR}}$	$\boldsymbol{C_{d,EGR}}$	$\boldsymbol{C_{d,EGR}}$	$\boldsymbol{C_{d,EGR}}$	p_3	p_3	p_3	p_3	p_3	p_3	p_3	p_3	p_3	p_{2a}
4.45	η_C, p_1, p_2, T_1, T_2	T_2	T_2	T_2	T_2	T_2	T_2	T_2	T_2	T_2	T_2	T_2	T_2	T_2	T_2	T_2	p_2
4.46	η_T, p_3, p_4, T_3, T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4	T_4
4.47	p_1, p_2, η_C, η_T, $\eta_{TC,mech}$, $\dot{m}_T$, T_1, T_3, p_3, p_4	p_2	p_3	p_3	p_4	p_2	$\boldsymbol{\eta_{T/C}}$	$\dot{m}_T$	p_4	p_2	$\boldsymbol{\eta_{T/C}}$	p_4	p_2	$\boldsymbol{\eta_{T/C}}$	p_2	$\boldsymbol{\eta_{T/C}}$	$\boldsymbol{\eta_{T/C}}$
4.53a	$\dot{m}_{air}$, $\dot{m}_{EGR}$, T_{CACout}, $T_{EGRcoolerGasOut}$, T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{2a}	T_{CACout}

Table 4.1 Continued

Eq. no.	Possible unknowns	Engine system number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4.54	E_{CAC}, T_2, T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_{CACout}	T_2
4.55	$\varepsilon_{EGRcooler}$, T_3, $T_{EGRcoolerGasOut}$ (= T_{EGR})	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}	T_{EGR}
4.56	$\dot{m}_{air}$, p_2, p_{2a}, $C_{d,CAC\text{-}IT}$ (= $C_{d,IT}$)	p_{2a}	p_2	p_2	p_2	$\boldsymbol{C_{d,IT}}$	p_2	p_2	p_2	$\boldsymbol{C_{d,IT}}$	p_2	p_2	$\boldsymbol{C_{d,IT}}$	p_2	$\boldsymbol{C_{d,IT}}$	p_2	$\boldsymbol{C_{d,IT}}$
4.57	$\dot{m}_T$, A_T, T_3, p_3, p_4	p_3	$\boldsymbol{A_T}$	$\dot{m}_T$	p_3	p_3	p_3	$\boldsymbol{A_T}$	$\boldsymbol{A_T}$	$\boldsymbol{A_T}$	$\boldsymbol{A_T}$	$\dot{m}_T$	$\dot{m}_T$	$\dot{m}_T$	p_4	p_4	p_3

$C_{d,EGR}$ (EGR valve opening), A_T (turbine effective area) and $\dot{m}_T$ (turbine flow rate, essentially representing turbine wastegate opening) are assumed to be known fixed hardware or calibration parameters. System 1 solves 14 performance parameters as unknowns with a given set of hardware and calibration. For a fixed pair of air–fuel ratio and EGR rate target, each of systems 2–16 needs to be solved for two hardware or calibration parameters. The two hardware selections have a certain flexibility of regulating the air–fuel ratio and the EGR rate within a range. Different systems of hardware formulation provide a different range of controllability. Systems 2–6 have high pumping loss when the EGR circuit flow coefficient $C_{d,EGR}$ is low (i.e., more restrictive). Systems 7–9, 11 and 12 are practical ones with a known turbocharger efficiency and a known maximum $C_{d,EGR}$. They can be used to seek hardware to reach a target air–fuel ratio and EGR rate with the minimum engine delta P ($p_3 - p_{2a}$). Systems 10 and 13 can be used to compute the required turbocharger efficiency, turbine area or wastegate opening with a given exhaust restriction flow coefficient $C_{d,exh}$ in order to achieve a target air–fuel ratio and EGR rate. System 14 may have some redundancy in functionality because the exhaust back pressure valve and the intake throttle valve have similar or related effects on engine performance. System 15 (or 16) can be used to compute the required turbocharger efficiency and exhaust (or intake) restriction with a given turbine. The mathematical formulations of LPL EGR and hybrid EGR systems can be derived similarly. In LPL EGR, although intake air throttling at the compressor inlet may be used to induct EGR flow, it is usually less effective than using an exhaust back pressure valve at the turbine outlet. It is worth noting that the non-EGR engine can be viewed as a simplified form of the above system formulation. The engine applications of Table 4.1 will be elaborated in Chapter 13.

4.5 Steady-state model tuning in engine cycle simulation

In engine cycle simulation, tuning the model is one of the most difficult tasks in EPSI analysis because of the complex interactions among the performance parameters of different subsystems. Model tuning refers to calibrating the model to adjust the hardware features to match known engine test data. The key to building a successful model is to identify and tune appropriate model parameters to reflect the true physics governing the engine performance test data. The theoretical foundation of model tuning can be found in Section 4.4.2. Those equations capture many hardware design parameters that can easily be tuned in the model (e.g., turbine area and efficiency). However, there are more complex physical processes that cannot be modeled very accurately by the zero-dimensional or one-dimensional engine cycle simulation software, such as the in-cylinder combustion, BSFC, and heat transfer. It should be noted that although the cycle simulation model is the best tool for engine system design due to its sufficient accuracy on macro engine performance parameters and fast computing time, it has certain limitations on predictability and it requires careful model tuning with experimental data to handle the uncertainties. Lumping some uncertainties together in subsystem models (e.g., within the boundary of the system of equations formulated earlier) is an effective approach in system-level analysis when detailed accurate experimental data of each component are not available and when certain complex physics cannot be modeled accurately. Over-complicated modeling on minor details or on highly uncertain physical processes (e.g., in-cylinder heat transfer and emissions) slows down the system analysis and provides only marginally added benefits in product design. The effectiveness of system simulation relies on a fine balance between model complexity and simplicity. It is worth noting that the quality of model tuning and model predictability must be judged by comparing with multiple data points, because the model could be 'purposely tuned' to match a single data point without good predictability at other points.

Some limitations of the engine cycle simulation model are listed as follows.

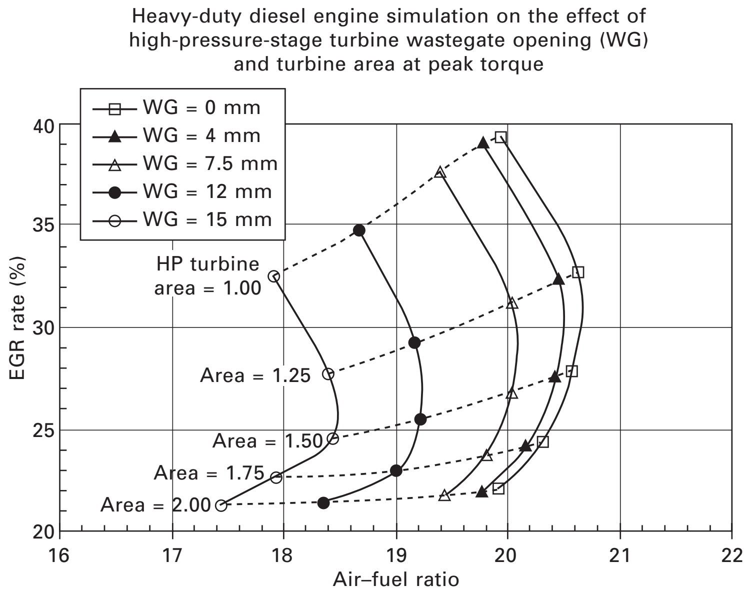
1. The model cannot accurately predict engine in-cylinder processes of heat transfer, combustion and emissions because it is difficult to simulate the three-dimensional distributions of the thermodynamic properties of the gases in the cylinder. It is also difficult to simulate the complex combustion processes as a function of both space and time for the in-cylinder heterogeneous mixture of air, fuel and combustion products. The model usually only uses a burn rate derived from a measured cylinder gas pressure trace, or at most it uses zero-dimensional phenomenological combustion models to compute a burn rate.
2. The model cannot capture all the micro details of thermal-fluid components such as the soot layer effect in EGR cooler fouling or the computational fluid dynamics (CFD) effects around the valves. Instead, the model uses a lumped approximate approach to simulate their global or macro performance characteristics.
3. The input data of engine hardware characteristics used in the model (e.g., turbocharger efficiency maps) are normally obtained from an off-engine test bench. They may not represent the true behavior of the devices under the realistic transient boundary conditions on the real engine.

All these difficulties cause simulation errors when compared to the engine test data.

In the engine cycle simulation model, the following eight key elements characterize the system: intake restriction, exhaust restriction, charge air cooler, inter-stage cooler (if any, characterized by cooling capacity and flow restriction), EGR cooler (characterized by cooling capacity and flow restriction), compressor (characterized by efficiency), turbine (characterized by efficiency and effective area), and engine cylinder (characterized by mechanical friction, base engine heat rejection, and volumetric efficiency). During model tuning at a given engine speed, it is important to match the following three types of system performance parameters with high-quality test data:

1. The five major design targets (brake power, BSFC, air–fuel ratio, EGR rate, and intake manifold gas temperature).
2. The parameters related to design constraints or design limits (e.g., peak cylinder pressure and the in-cylinder combustion pressure pattern, exhaust manifold gas temperature, compressor outlet air temperature, exhaust manifold pressure, engine delta P, coolant heat rejection, engine outlet coolant temperature, and turbocharger speed).
3. The rest of the 18 parameters mentioned earlier ($\dot{m}_{air}$, $\dot{m}_{EGR}$, T_3, p_4, $T_{1,ROA}$, p_1, p_2, T_2, p_3, T_4, N_C, N_T, η_C, η_T, p_{2a}, T_{2a} or T_1, T_{CACout}, and $T_{EGRcoolerGasOut}$). These parameters are not only related to model tuning, but also essential for engine system design specifications.

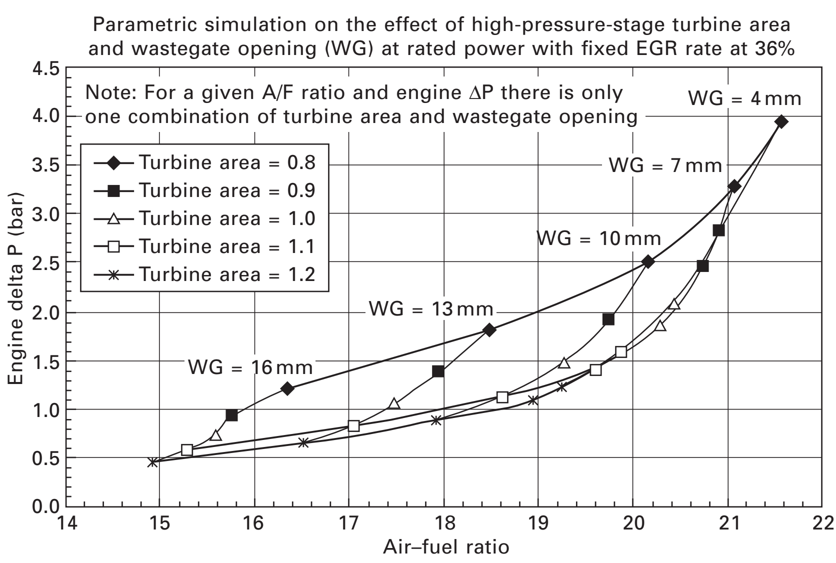
Figure 4.7 shows an example of model tuning to match a given EGR rate and air–fuel ratio at peak torque by adjusting the turbine wastegate opening and



4.7 Effect of turbine area and wastegate opening on EGR rate and air–fuel ratio (peak torque).

turbine effective area. It should be noted that usually in a good system design the turbine wastegate opening should be set fully closed at peak torque when selecting the turbine area. Note that in this air system capability domain of 'EGR rate vs. air–fuel ratio', the parametric sweeping of turbine wastegate opening and turbine area form a two-dimensional data area which can cover the known engine test data point (or a system design target point). Also note that the behavior of the data area in this two-dimensional sweeping is dependent upon the settings of two other critical air system parameters: the EGR valve opening (or essentially the EGR circuit flow restriction), and the intake valve opening (or essentially the flow restriction from the compressor inlet to the intake manifold). When the flow restriction in the EGR circuit increases, the data area in Fig. 4.7 will shift toward the lower right direction. When the intake throttle valve is closed, the data area in Fig. 4.7 will shift toward the upper left direction, with only a slight increase in EGR rate but a large reduction in air–fuel ratio. Therefore, when conducting such a two-dimensional sweeping of turbine wastegate opening and turbine area to try to match a given test data or a system design target at peak torque, it is important to determine the appropriate EGR valve opening (or EGR circuit flow restriction design) and intake throttle opening.

A similar model tuning technique can be applied to adjust the model to match a known engine delta P and air–fuel ratio at a given EGR rate at rated power, as shown in Fig. 4.8. A two-dimensional parametric sweeping of turbine area and wastegate opening is conducted to form a two-dimensional



4.8 Effect of turbine area and wastegate opening on engine delta P and air–fuel ratio (at rated power).

data area in the figure to cover the known engine test data point (or a system design target point). At rated power, usually the intake throttle valve needs to be set at fully open because there is a large engine delta P available to drive EGR so that there is no need to close the intake throttle valve to help drive EGR. Moreover, at rated power, unlike at peak torque where the EGR valve usually needs to be set at fully open, the EGR valve has to be partially closed in order to prevent overflowing too much EGR. Note that every data point in Fig. 4.8 is obtained by using a PID (proportional, integral, and differential) controller in the simulation model to automatically adjust the EGR valve opening to achieve 36% EGR rate. The behavior of the data area in Fig. 4.8 is dependent upon another critical air system design parameter: the turbocharger efficiency. When the turbocharger efficiency used in the model increases, the data area will shift toward the right. In contrast, when the turbocharger efficiency is reduced in the model, the data area will shift toward the left. Therefore, when conducting such a two-dimensional sweeping of turbine wastegate opening and turbine area to try to match a given test data or a system design target, it is important to determine the turbocharger efficiency required.

In the engine system design or model tuning parameters, the exhaust restriction is further discussed in the aftertreatment analysis in Chapter 8 and in Chapter 15 for the subsystem interaction effects. The coolers and engine cylinder heat rejection are discussed in Chapter 12. Engine volumetric efficiency and turbocharger performance are elaborated in Chapter 13. Emissions design targets are explained in Chapter 7. Engine durability design constraints are elaborated in Chapters 2 and 15.

It should be noted that using a PID feedback controller to automatically search and match the data is an important technique to accelerate the model tuning or design process. Examples of PID controller applications in diesel engine system design include using a constant-torque PID control, a target air–fuel ratio PID control, and a target EGR rate PID control.

Another challenge encountered in model tuning is to identify the root cause of the simulation errors and minimize the errors, especially for multiple speed/load mode points. For example, an error in the exhaust manifold gas temperature may be caused by several factors such as a wrong air–fuel ratio, EGR rate, BSFC, injection timing, or cylinder heat rejection. A DoE method can be used to speed up the process of error finding and minimize the uncertainties. Theoretically, there is only one unique set of model tuning results which can match the true physical behavior of the engine on all performance parameters. The challenge is to approximate that unique solution with a minimum weighted simulation error and reasonable adjustment in model tuning parameters for each subsystem.

A guideline for steady-state engine performance model tuning is summarized below in order to illustrate the logic and contents of EPSI simulation.

1. Set up key input data in the model (e.g., valve size, valve lift profile, port flow coefficient, cylinder and pipe geometry, pipe heat transfer, turbocharger maps).
2. Check the quality of engine performance test data with energy balance calculations and the turbocharger power balance calculation.
3. Tune engine volumetric efficiency by using either the engine motoring or firing test data.
4. Calculate mechanical friction power by subtracting the calculated pumping loss power from the measured total engine motoring power.
5. Tune intake restriction and exhaust restriction by matching the simulation data to the test characteristic curve of 'pressure drop vs. volume flow rate'.
6. Tune charge air cooler flow restriction and effectiveness by matching the simulation data to the test curves of 'charge air cooler pressure drop vs. volume flow rate' and 'cooler effectiveness vs. air mass flow rate'.
7. Tune EGR cooler flow restriction and effectiveness by matching the simulation data to the test curves of 'EGR cooler pressure drop vs. volume flow rate' and 'cooler effectiveness vs. EGR mass flow rate'.
8. Calculate the heat release rates at various engine speeds and loads by using engine cylinder pressure test data.
9. Set up the model properly to reflect the true heat rejection characteristics of the miscellaneous heat losses. Tune cylinder heat rejection by adjusting both the Woschni heat transfer multiplier and cylinder liner/head/piston heat transfer coefficients to match the simulation data to the computed test data of base engine heat rejection characteristics. Or, equivalently from the basis of energy balance, tune the Woschni multiplier and those heat transfer coefficients to match the exhaust manifold gas temperature.
10. Plot the engine test points on the compressor and turbine maps. Look up compressor and turbine efficiencies. Calculate turbocharger efficiencies with engine firing test data. Determine turbocharger efficiency multipliers.
11. At a given engine speed and load, use the DoE approach or a simpler parametric sweeping method to tune the DoE or sweeping factors (e.g., turbine effective area or VGT/wastegate opening, turbocharger efficiency multiplier, EGR valve opening, start-of-combustion timing, duration of heat release rate) to match the simulation results to the engine test data of engine delta P, air–fuel ratio, EGR rate, brake power (or BSFC), exhaust manifold gas temperature, in-cylinder combustion pressure pattern, and other system performance parameters. Use optimization software to process the DoE simulation data to compute the factor values for minimized weighted errors in BSFC, exhaust manifold gas temperature, or engine delta P at a target air–fuel ratio and EGR rate.

12. Repeat step 11 at other engine speed and load modes. For transient engine simulations, the tuned engine control modeling parameters need to be smooth functions of the engine speed and load.

4.6 References and bibliography

Agnew D (1994), 'What is limiting your engine air flow: using normalized steady air flow bench data', SAE paper 942477.

Alkidas A C (1993), 'Effects of operational parameters on structural temperatures ands coolant heat rejection of a S.I. engine', SAE paper 931124.

Annand W J and Roe G E (1974), *Gas Flow in the Internal Combustion Engine*, G.T. Foulis & Co. Ltd, Sparkford.

Arias J R, Moreno E, Navarro E and Varela E (2000), 'Using 1-D and 3-D models for the simulation of gas exchange processes', SAE paper 2000-01-0658.

Assanis D N, Filipi Z S, Fiveland S B and Syrimis M (2000), 'A methodology for cycle-by-cycle transient heat release analysis in a turbocharged direct injection diesel engine', SAE paper 2000-01-1185.

Basshuysen R and Schafer F (2004), *Internal Combustion Engine Handbook*, SAE International, Warrendale, PA.

Benson R S (edited by Horlock J H and Winterbone D E (1982), *The Thermodynamics and Gas Dynamics of Internal Combustion Engines*, Volume 1, Clarendon Press, Oxford.

Blair G P (1991), 'An alternative method for the prediction of unsteady gas flow through the internal combustion engine', SAE paper 911850.

Bohac S V and Landfahrer K (1999), 'Effects of pulsating flow on exhaust port flow coefficients', SAE paper 1999-01-0214.

Bohac S V, Baker D M and Assanis D N (1996), 'A global model for steady state and transient S.I. engine heat transfer studies', SAE paper 960073.

Borman G L and Ragland K W (1998), 'Direct-injection engine combustion', *Combustion Engineering*, McGraw-Hill, Boston, MA.

Brady R N (1996), *Modern Diesel Technology*, Prentice-Hall, Englewood Cliffs, NJ.

Brunt M J and Platts K C (1999), 'Calculation of heat release in direct injection diesel engines', SAE paper 1999-01-0187.

Caton J A (1982), 'Comparisons of thermocouple, time-averaged and mass-averaged exhaust gas temperatures for a spark-ignited engine', SAE paper 820050.

Cesario N, Muscio C, Farina M, Amato P and Lavorgna M (2004), 'Modelling the rate of heat release in common rail diesel engines: a soft computing approach', SAE paper 2004-01-2967.

Chalet D, Chesse P, Tauzia X and Hetet J-F (2006), 'Comparison of different methods for the determination of pressure wave in the inlet and exhaust systems of internal combustion engine', SAE paper 2006-01-1542.

Challen B and Baranescu R (eds) (1999), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA.

Chow A and Wyszynski M L (1999), 'Thermodynamic modeling of complete engine systems – a review', *Proc. IMechE*, *Part D: Journal of Automobile Engineering*, 213, 403–415.

Danov S (1997), 'Identification of discharge coefficients for flow through valves and ports of internal combustion engines', SAE paper 970642.

Egnell R (1999), 'A simple approach to studying the relation between fuel rate heat release rate and NO formation in diesel engines', SAE paper 1999-01-3548.

Ferguson C R and Kirkpatrick A T (2000), *Internal Combustion Engines: Applied Thermosciences*, 2nd edn, Wiley, New York.

Franco A and Martorano L (1999), 'Methods to evaluate in-cylinder heat transfer and thermal load in the small internal combustion engines', SAE paper 1999-01-1252.

Friedrich I, Pucher H and Offer T (2006), 'Automatic model calibration for engine-process simulation with heat-release prediction', SAE paper 2006-01-0655.

Fukutani I and Watanabe E (1979), 'An analysis of the volumetric efficiency characteristics of 4-stroke cycle engines using the mean inlet Mach number Mim', SAE paper 790484.

Grimm B M and Johnson R T (1990), 'Review of simple heat release computations', SAE paper 900445.

Haddad S D and Watson N (eds) (1984a), *Principles and Performance in Diesel Engineering*, Ellis Horwood Limited, Chichester.

Haddad S D and Watson N (eds) (1984b), *Design and Applications in Diesel Engineering*, Ellis Horwood Limited, Chichester.

Hansen A C (1992), 'A diagnostic quasi-dimensional model of heat transfer applied to a motored compression-ignition engine', SAE paper 920542.

Heywood J B (1988), *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.

Hohenberg G F (1979), 'Advanced approaches for heat transfer calculations', SAE paper 790825.

Homsy S C and Atreya A (1997), 'An experimental heat release rate analysis of a diesel engine operating under steady state conditions', SAE paper 970889.

Horlock J H and Winterbone D E (1986), *The Thermodynamics and Gas Dynamics of Internal Combustion Engines*, Volume 2, Clarendon Press, Oxford.

Hountalas D T, Kouremenos D A and Fiveland S B (2004), 'Some considerations on the estimation of the heat release of DI diesel engines using modelling techniques', SAE paper 2004-01-1405.

Hsu B D (2002), *Practical Diesel-Engine Combustion Analysis*, SAE International, Warrendale, PA.

Imabeppu S, Shimonosono H, Hirano Y, Fujigaya K and Inoue K (1993), 'Development of a method for predicting heat rejection to the engine coolant', SAE paper 931114.

Israel M A and Hu H (1993), 'Impact of ambient parameters on operating efficiency of compression release engine retarder systems on heavy duty diesel engines', SAE paper 932972.

Jennings M J and Morel T (1991), 'A computational study of wall temperature effects on engine heat transfer', SAE paper 910459.

Júnior J P P, Valle R M and Barros J E M (2005), 'Algebraic modeling for thermodynamic study of the compression-ignition engine', SAE paper 2005-01-4143.

Kamimoto T, Minagawa T and Kobori S (1997), 'A two-zone model analysis of heat release rate in diesel engines', SAE paper 972959.

Kamimoto T, Akiyoshi M and Kosaka H (1998), 'A numerical simulation of ignition delay in diesel engines', SAE paper 980501.

Kar K, Roberts S, Stone R, Oldfield M and French B (2004), 'Instantaneous exhaust temperature measurements using thermocouple compensation techniques', SAE paper 2004-01-1418.

Kar K, Swain A, Raine R, Roberts S and Stone R (2006), 'Cycle-by-cycle variations in exhaust temperatures using thermocouple compensation techniques', SAE paper 2006-01-1197.

Ladommatos N, Abdelhalim S M, Zhao H, Hu Z (1998), 'Effects of EGR on heat release in diesel combustion', SAE paper 980184.

Lakshminarayanan P A, Aghav Y V, Dani A D and Mehta P S (2002), 'Accurate prediction of the rate of heat release in a modern direct injection diesel engine', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 216, 663–675.

Livengood J C, Rogowski A R and Taylor C F (1952), 'The volumetric efficiency of four-stroke engines', SAE paper 520259.

Luján J M, Serrano J R, Arnau F and Dolz V (2003), 'Heat transfer model to calculate turbocharged HSDI diesel engines performance', SAE paper 2003-01-1066.

Makartchouk A (2002), *Diesel Engine Engineering: Thermodynamics, Dynamics, Design, and Control*, Marcel Dekker, New York.

Manente V, Vressner A, Tunestal P and Johansson B (2008), 'Validation of a self tuning gross heat release algorithm', SAE paper 2008-01-1672.

Martyr A J and Plint M A (2007), *Engine Testing*, 3rd edn, Elsevier Ltd., Oxford, UK.

Mattarelli E and Valentini A (2000), 'On the flow modeling through the valve assembly in engine cycle simulations', SAE paper 2000-01-0570.

Meguerdichian M and Watson N (1978), 'Prediction of mixture formation and heat release in diesel engines', SAE paper 780225.

Miyamoto N, Chikahisa T, Murayama T and Sawyer R (1985), 'Description and analysis of diesel engine rate of combustion and performance using Wiebe's functions', SAE paper 850107.

Morel T, Keribar R, Silvestri J and Wahiduzzaman S (1999), 'Integrated engine/vehicle simulation and control', SAE paper 1999-01-0907.

Naterer G F and Camberos J A (2008), *Entropy Based Design and Analysis of Fluids Engineering Systems*, CRC Press, Boca Raton, FL.

Neumeister J, Taylor J and Gurney D (2007), 'Virtual air path calibration of a multi cylinder high performance GDI engine using 1D cycle simulation', SAE paper 2007-01-0490.

Nieuwstadt M J, Kolmanovsky I V, Brehob D and Haghgooie M (2000), 'Heat release regressions for GDI engines', SAE paper 2000-01-0956.

Nuszkowski J and Thompson G J (2009), 'Application and repeatability of transient heat release analysis for heavy duty diesel engines', SAE paper 2009-01-1125.

Ogata K (2004), *System Dynamics*, 4th edn, Pearson Prentice-Hall, Upper Saddle River, NJ.

Oldfield S G and Watson N (1983), 'Exhaust valve geometry and its effect on gas velocity and turbulence in an exhaust port', SAE paper 830151.

Oppenheim A K, Barton J E, Kuhl A L and Johnson W P (1997), 'Refinement of heat release analysis', SAE paper 970538.

Palm III W J (2005), *System Dynamics*, McGraw-Hill, New York.

Peters B and Gosman A D (1993), 'Numerical simulation of unsteady flow in engine intake manifolds', SAE paper 930609.

Pierik R J and Burkhard J F (2000), 'Design and development of a mechanical variable valve actuation system', SAE paper 2000-01-1221.

Ponti F, Corti E, Serra G and Cesare M D (2007), 'Common rail multi-jet diesel engine combustion model development for control purposes', SAE paper 2007-01-0383.

Primus R J (1999), 'Visual thermodynamics: processes in log(p)-log(T) space', SAE paper 1999-01-0516.

Proell W A (1993), *The Thermodynamic Exploration for High Efficiency Internal Combustion Engines*, Cloud Hill Press, Las Vegas, NM.

Pulkrabek W W (2003), *Engineering Fundamentals of the Internal Combustion Engine*, 2nd edn, Prentice Hall, Upper Saddle River, NJ.

Rakopoulos C D and Giakoumis E G (2009), *Diesel Engine Transient Operation – Principles of Operation and Simulation Analysis*, Springer-Verlag London Limited, London.

Ribeiro B and Martins J (2007), 'Direct comparison of an engine working under Otto, Miller and Diesel cycles: thermodynamic analysis and real engine performance', SAE paper 2007-01-0261.

Rosseel E and Sierens R (1996), 'The physical and the chemical part of the ignition delay in diesel engines', SAE paper 961123.

Roussopoulos K (1990), 'A convenient technique for determining comparative volumetric efficiency', SAE paper 900352.

Royo R, Corberán J and Pérez A (1994), 'Optimal design of the intake system in 4-S I.C.E.', SAE paper 940210.

Ryan T W (1987), 'Ignition delay as determined in a variable-compression ratio direct-injection diesel engine', SAE paper 872036.

SAE J1004 (2004), 'Glossary of engine cooling system terms'.

SAE Surface Vehicle Recommended Practice J228 (1995), 'Airflow reference standards'.

SAE Surface Vehicle Recommended Practice J604 (1995), 'Engine terminology and nomenclature – general'.

SAE Surface Vehicle Standard J1312 (1995), 'Procedure for mapping performance – spark ignition and compression ignition engines'.

SAE Surface Vehicle Recommended Practice J1829 (2002), 'Stoichiometric air–fuel ratios of automotive fuels'.

SAE Surface Vehicle Recommended Practice J2548 (2001), 'Test method to obtain lug-up torque data for heavy-duty truck and bus engines'.

SAE Surface Vehicle Standard J1995 (1995), 'Engine power test code – spark ignition and compression ignition – gross power rating'.

SAE Surface Vehicle Standard J2723 (2007), 'Engine power test code – engine power and torque certification'.

SAE Surface Vehicle Standard J1349 (2008), 'Engine power test code – spark ignition and compression ignition – net power rating'.

Schihl P, Tasdemir J, Schwarz E and Bryzik W (2002), 'Development of a zero-dimensional heat release model for application to small bore diesel engines', SAE paper 2002-01-0073.

Schubert C, Wimmer A and Chmela F (2005), 'Advanced heat transfer model for CI engines', SAE paper 2005-01-0695.

Shayler P J, Chick J P and Ma T (1996), 'Effect of coolant mixture composition on engine heat rejection rate', SAE paper 960275.

Shayler P J, Chick J P and Ma T (1997), 'Correlation of engine heat transfer for heat rejection and warm-up modeling', SAE paper 971851.

Sierens R, Hove W V and Riemslagh K (1992), 'Heat release analysis and power cycle calculation of the combustion in a medium-speed diesel engine', *ASME ICE – Vol. 17, Diesel Engine Processes: Turbocharging, Combustion and Emission*, 41–51, presented at the Energy-Sources Technology Conference and Exhibition, Houston, TX, January.

Silva L L C (1993), 'Simulation of the thermodynamic processes in diesel cycle internal combustion engines', SAE paper 931899.

Silvestri J, Morel T, Lang O, Schernus C and Rauscher M (2000), 'Advanced engine/drivetrain/vehicle modeling techniques', *9th Aachener Kolloquium Fahrzeug- und Motorentechnik*.

Smith L A, Fickenscher T and Osborne R P (1999), 'Engine breathing – steady speed volumetric efficiency and its validity under transient engine operation', SAE paper 1999-01-0212.

Son S and Kolasa A E (2007), 'Estimating actual exhaust gas temperature from raw thermocouple measurements acquired during transient and steady state engine dynamometer tests', SAE paper 2007-01-0335.

Stone R (1999), *Introduction to Internal Combustion Engines*, 3rd edn, SAE International, Warrendale, PA.

Thor M, Andersson I and McKelvey T (2009), 'Parameterized diesel engine heat release modeling for combustion phasing analysis', SAE paper 2009-01-0368.

Tuccillo R, Arnone L, Bozza F, Nocera R and Senatore A (1993), 'Experimental correlations for heat release and mechanical losses in turbocharged diesel engines', SAE paper 932459.

Veiásquez J A and Milanez L F (1995), 'Computational model for simulation of processes in diesel engines', SAE paper 952304.

Watson N (1999), *The Theory of Turbocharging*, in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 27–72.

Watson N, Pilley A D and Marzouk M (1980), 'A combustion correlation for diesel engine simulation', SAE paper 800029.

Winterbone D E and Pearson R J (1999), *Design Techniques for Engine Manifolds*, SAE International, Warrendale, PA.

Winterbone D E and Pearson R J (2000), *Theory of Engine Manifold Design*, Professional Engineering Publishing, London and Bury St Edmunds.

Wolff A, Boulouchos K and Mueller R (1997), 'A computational investigation of unsteady heat flux through an I.C. engine wall including soot layer dynamics', SAE paper 970063.

Woschni G (1967), 'Universally applicable equation for the instantaneous heat transfer coefficient in the internal combustion engine', SAE paper 670931.

Zeng P and Assanis D N (2004), 'Cylinder pressure reconstruction and its application to heat transfer analysis', SAE paper 2004-01-0922.

5 Engine–vehicle matching analysis in diesel powertrain system design

Abstract: This chapter discusses vehicle powertrain performance and the impact on diesel engine system design in firing operation. It begins by summarizing the formulae of vehicle performance analysis, followed by analytical engine–vehicle matching methods with engine characteristic maps for optimum fuel economy and drivability. It then describes the transient powertrain performance simulations of vehicle driving cycles and acceleration. It emphasizes the concept of 'top–down' powertrain design approach to optimize the interface between the vehicle and the engine. Hybrid powertrain performance analysis is also briefly introduced.

Key words: vehicle performance, drivetrain, powertrain dynamics, engine–vehicle matching, driving cycle, fuel economy, drivability, transient performance, hybrid.

5.1 The theory of vehicle performance analysis

5.1.1 Introduction of vehicle analysis areas

Automotive engines are designed for real-world vehicle operating needs. Engine development is focused on improving engine cycle processes and torque curve shape. At the same time, it strives to reduce parasitic losses and fuel consumption. Drivetrain design is focused on increasing drivetrain efficiency, reducing resistance forces, and selecting proper transmission ratios to match the engine. Engine–vehicle system integration for the diesel powertrain requires a close collaboration between the two. Among the four major areas of vehicle dynamics (powertrain transmission, braking, suspension, and steering systems), the first two areas are directly related to engine system design.

The following topics are important in vehicle performance analysis for engine system design: longitudinal dynamics and vehicle force distribution, tire–road rolling resistance, aerodynamic drag resistance, vehicle inertia force, transmission and torque converter modeling, vehicle acceleration calculation, and vehicle fuel economy calculation.

Heavy-duty vehicle drivetrain design fundamentals are introduced in Jones (1991). Powertrain design process is discussed by Dopson *et al.* (1995) and Ferraz *et al.* (2001). Vehicle powertrain simulation methodologies are presented by Morel *et al.* (1999), Ciesla *et al.* (2000), and Guzzella and

Sciarretta (2005). Heavy-duty vehicle test cycles are investigated by Bata *et al.* (1994) and Clark *et al.* (2003). The coastdown method is discussed by Petrushov (1997) and Miller (2004). Overviews on vehicle fuel economy are provided by Ehlbeck and Mayenburg (1991), Simner (1996), Patton *et al.* (2002) and Sovran and Blaser (2003). Transmission efficiency has been researched by Kluger and Greenbaum (1993). Torque converter design and performance are reviewed by Mercure (1979). The SAE procedures listed in the section of references at the end of this chapter also provide valuable fundamental information on vehicle, powertrain and drivetrain.

5.1.2 Formulae of vehicle performance analysis

Heavy-duty trucks are usually front-engined and rear-wheel-driven in order to obtain effective engine cooling and better tractive force during acceleration and upgrade climbing. In order for a vehicle to move, the tractive force provided by the engine and the transmission needs to be greater than the total of all static resistance forces. However, the maximum tractive force acting on the vehicle driving wheels is limited by the tire–road adhesion force, even if the engine and the transmission can produce a higher tractive force.

The vehicle force balance at any steady-state or transient condition can be written as

$$F_t + F_{er} + F_{acc} + F_{br} + F_{rf} + F_a + F_i + F_{gl} + F_{dr} = 0 \qquad 5.1$$

When the vehicle moves at a constant speed, the inertia force $F_i = 0$. If F_i is not equal to zero, the vehicle either accelerates or decelerates in transients. F_t is the tractive force acting on the vehicle wheels and is delivered from engine firing operation. Note that the tractive force due to engine firing is defined as a positive value, and all resistance forces are defined as negative values. During engine firing, the engine brake retarding force $F_{er} = 0$. F_{acc} is a vehicle equivalent resistance force acting on the wheels and caused by accessory loads such as the cooling fan, the air conditioning and the power steering. If the driver does not use the service brake (wheel brake), the service brake force $F_{br} = 0$. F_{rf} is the tire–road rolling friction force. F_a is the aerodynamic resistance force. F_{gl} is the gravity force on a gradient. F_{dr} is the resistance force of drivetrain retarders such as the hydrodynamic or electromagnetic retarder. During engine firing without wheel braking or drivetrain retarder braking, the tractive force acting on the vehicle wheels is given by

$$F_t = -(F_{acc} + F_{rf} + F_a + F_i + F_{gl}). \qquad 5.2$$

On the other hand, the maximum allowable tractive force is limited by the tire–road adhesion force, which is equal to the normal load multiplied by the road adhesion coefficient. The adhesion force is determined by the axle load, the road surface condition and the tires. For example, an icy road has a very low adhesion force, and the vehicle wheels may slip on the ice. It should be noted that each term in the vehicle force balance equation 5.1 is for the total vehicle mass m_V (including trailers, if any).

The gravity force is given by $F_{gl} = m_V g \sin\theta$, and the road grade is defined as $G_r = \tan\theta$, where θ is the road slope angle (positive for downhill, negative for uphill), and g is the acceleration due to gravity. The rolling friction resistance is calculated by $F_{rf} = -m_V g \cdot f_{rf} \cos\theta$ if the aerodynamic lift force is ignored in the net normal load. The tire rolling friction coefficient f_{rf} increases when the vehicle speed, the tractive force, or the tire tilting angle increases. The friction coefficient decreases when the tire pressure or the tire temperature increases. The tire rolling friction coefficient is also affected by tire structure, tire material and road surface condition. The friction coefficient is independent of the vertical load. Usually, the rolling friction coefficient values are between 0.005 and 0.01 on a concrete road surface. Moreover, for a radial-ply truck tire, when $N_V < 100$ km/hour, previous experimental data gave $f_{rf} = 0.006 + 0.23(0.001N_V)^2$, where N_V is vehicle speed; and for a bias-ply truck tire, $f_{rf} = 0.007 + 0.45(0.001N_V)^2$ (Wong, 1993).

The aerodynamic resistance force is calculated by $F_a = -0.5\rho_{AMB} f_a A_V N_{Vw}^2$, where ρ_{AMB} is the ambient air density, f_a is the aerodynamic resistance coefficient, A_V is the projected frontal area of the vehicle in the direction of travel, and N_{Vw} is the vehicle speed relative to the wind.

The transient inertia force is given by $F_i = -\xi m_V a_V$, where m_V is the total effective vehicle mass including the payload, and a_V is the vehicle's linear acceleration. The ξ is the rotational mass coefficient which is defined as the ratio of the total vehicle inertia force to the linear inertia force. The total refers to the sum of the inertia force caused by both the linear motion of the vehicle mass and the equivalent inertia force caused by all the rotating masses. The ξ can be derived as follows:

$$\xi = 1 + \frac{I_{drive}}{m_V r_{tire}^2} + \frac{I_E C_{tr} i_{gr}^2 i_{ax}^2 \eta_t}{m_V r_{tire}^2} + \frac{I_E C_{tr} i_{gr} i_{ax}^2 \eta_t N_V}{m_V r_{tire}^2 a_V}\left(\frac{\mathrm{d}i_{gr}}{\mathrm{d}t}\right) \qquad 5.3$$

where r_{tire} is the dynamic tire radius, I_{drive} is the total equivalent mass moment of inertia of all the driveline components including vehicle wheels, I_E is the moment of inertia of the engine rotating components connected to the driveline such as the flywheel, i_{gr} is the transmission gear ratio, i_{ax} is the overall drive axle gear ratio, t is time, and η_t is the drivetrain efficiency representing the frictional power losses of the entire drivetrain (from the engine crankshaft to the vehicle wheels, including the clutch or the torque converter, the transmission, the universal joints, the differential, the drive axles, the final drive gear, etc.). Note that vehicle accessory power is defined as a resistance power rather than a frictional power loss.

The drivetrain efficiency of manual transmissions can be around 95% at lower gears and 97–98% at the 1:1 direct transmission top gear. The efficiency of automatic transmissions is roughly 10% lower. The overall drivetrain efficiency of heavy-duty trucks and buses usually peaks around 80–90%. The efficiency varies widely at different engine speeds, loads and gear numbers (Kluger and Greenbaum, 1993). C_{tr} in equation 5.3 is the torque ratio of the torque converter used with an automatic transmission (for manual transmissions without the torque converter, C_{tr} can be set to 1). Torque converter efficiency is $\eta_{TV} = C_{sr}C_{tr}$, where the speed ratio C_{sr} is defined as the output speed divided by the input speed, and the torque ratio C_{tr} is defined as the output torque divided by the input torque. When locked up without hydraulic coupling, the torque converter efficiency reaches its highest value. C_{sr} and C_{tr} are obtained from the characteristic chart of the torque converter when the input capacity factor c_{TV} is known. Note that $c_{TV} = N_{TV}/J_{TV}^{0.5}$, where N_{TV} is speed and J_{TV} is torque. The torque converter input capacity factor c_{TV} is equal to the engine capacity factor c_E, which is calculated using $c_E = N_E/J_E{}^{0.5}$, where N_E is the engine speed and J_E is the engine torque. In the last term of equation 5.3, the transient gear ratio change $\mathrm{d}i_{gr}/\mathrm{d}t$ can produce a significant resistance inertia force for continuously variable transmissions. The ξ can be approximated by an empirical formula $\xi = 1 + 0.04 + 0.0025 i_{gr}^2 i_{ax}^2$ (Wong, 1993). The ξ in equation 5.3 is a very important parameter that includes the engine moment of inertia. Based on equation 5.3 the transient powertrain equation 5.20 is derived later.

The vehicle tractive force acting on the wheels F_t, the tractive torque J_t, and the engine brake torque J_E are related by the following relationship:

$$F_t = \frac{J_t}{r_{tire}} = \frac{J_E C_{tr} i_{gr} i_{ax} \eta_t}{r_{tire}} \tag{5.4}$$

The engine firing brake torque at the crankshaft can be calculated by

$$J_E = \frac{-(F_{acc} + F_{rf} + F_a + F_i + F_{gl}) r_{tire}}{C_{tr} i_{gr} i_{ax} \eta_t} \tag{5.5}$$

The engine speed can be calculated by

$$N_E = \frac{N_V i_{gr} i_{ax}}{2\pi r_{tire} C_{sr}(1 - f_{slip})} \tag{5.6}$$

where N_E is the engine speed (revolution per second), f_{slip} is the slip of the vehicle running gear, $f_{slip} = 2$–5% (Wong, 1993), C_{sr} is the speed ratio of the torque converter used with an automatic transmission (for manual transmissions without torque converter, C_{sr} can be set to 1). Engine brake power can be calculated by

$$\dot{W}_E = J_E N_E \tag{5.7}$$

Note that the term 'brake power' derives from the engine dynamometer that absorbs power by the action of a brake (e.g., a friction brake). The engine 'brake power' refers to the crankshaft net power which is input to the vehicle drivetrain in either firing or braking operation, and it is not the same as the engine 'braking power or retarding power' in engine brake operation. Figure 5.1 illustrates the vehicle tractive force characteristics with a seven-speed transmission.

Denoting t as time, vehicle acceleration a_V is calculated by:

$$a_V = \frac{dN_V}{dt} = \frac{1}{\xi m_V}(F_{acc} + F_t + F_{rf} + F_a + F_{gl}) \qquad 5.8$$

Vehicle speed is calculated by

$$N_{V,t_2} = N_{V,t_1} + \int_{t_1}^{t_2} a_V \, dt \qquad 5.9$$

Vehicle travel distance is given by

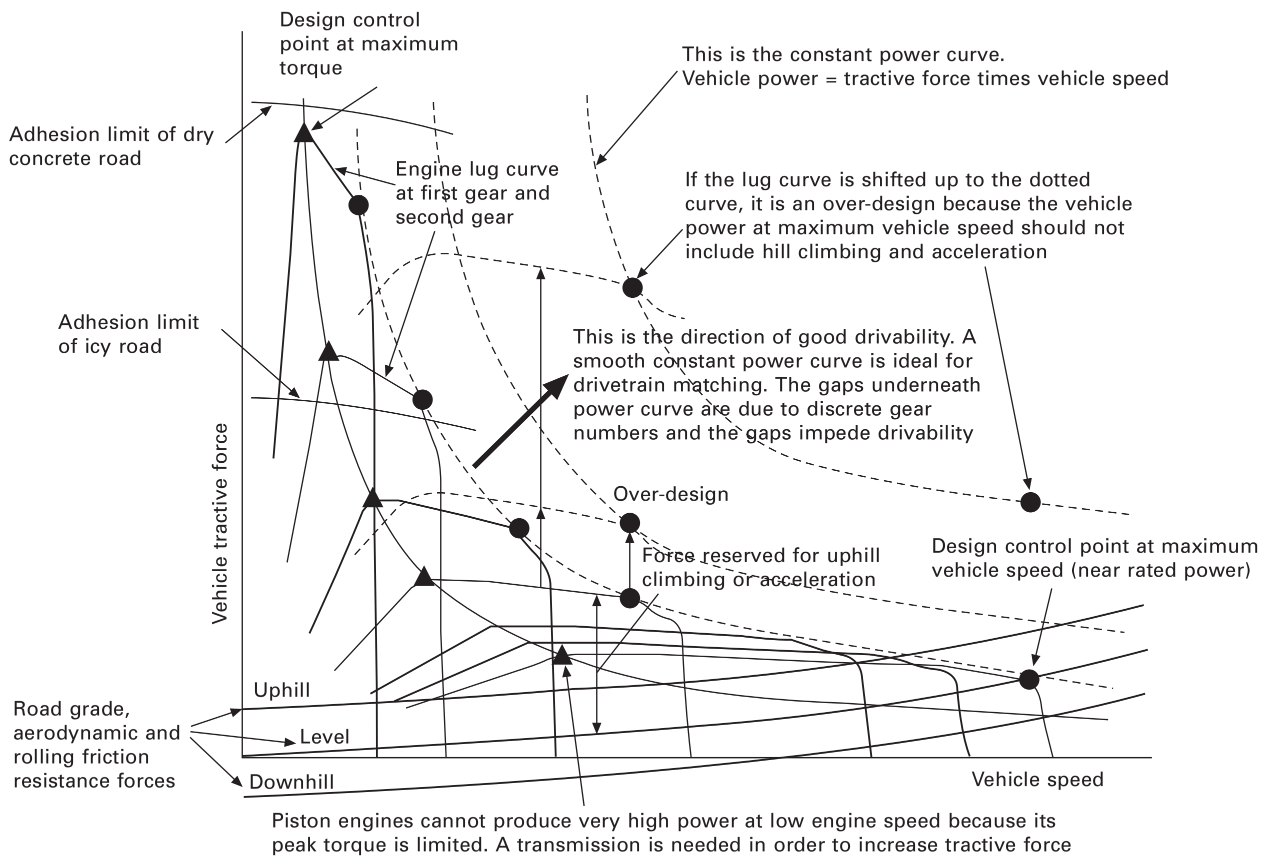
$$l_{V,t_2} = l_{V,t_1} + \int_{t_1}^{t_2} N_V \, dt \qquad 5.10$$

Vehicle acceleration time is calculated by

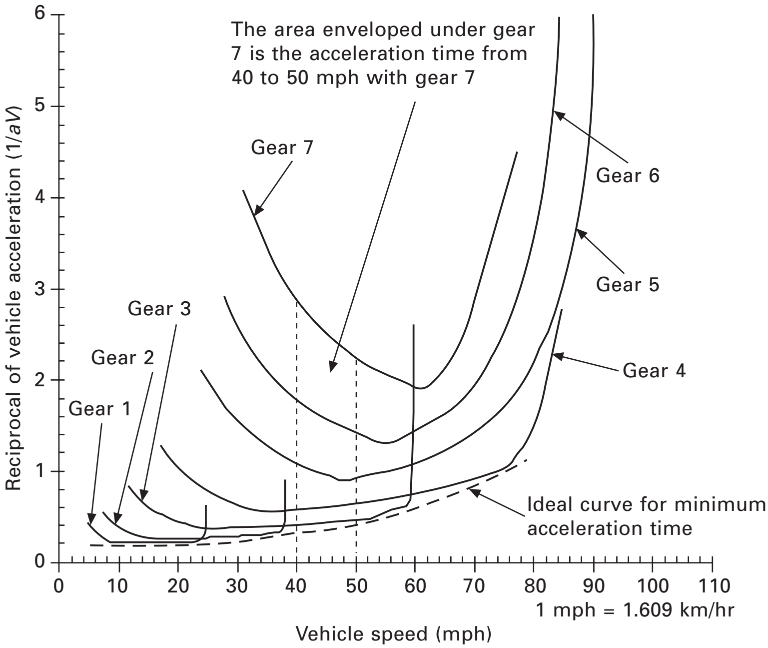
$$\begin{aligned} \Delta t_{ac} &= \int_{t_1}^{t_2} dt = \int_{N_{V1}}^{N_{V2}} \left(\frac{1}{a_V}\right) dN_V \\ &= \int_{N_{V1}}^{N_{V2}} \left(\frac{\xi \cdot m_V}{F_{acc} + F_t + F_{rf} + F_a + F_{gl}}\right) dN_V \end{aligned} \qquad 5.11$$

where F_t is the transient vehicle tractive force resulting from the available engine transient torque during acceleration in each gear. It should be noted that under the same fueling rate the transient engine torque is usually lower than the steady-state engine torque due to losses in combustion, pumping loss, turbocharger lag, and thermal inertia. In some cases the transient power can be 5–8% lower than the steady-state power. It is observed that the vehicle acceleration time from the speed N_{V1} to N_{V2} is the area under the curve of 'the reciprocal of acceleration vs. vehicle speed', plus the gear-shifting time (typically 0.2–0.6 s for each gear change). The area under the curve depends on where the transmission gear is shifted. Figure 5.2 shows the vehicle acceleration characteristics with the full load engine torque curve for a seven-speed transmission calculated using equation 5.8.

Vehicle dynamics can also be expressed by a power balance as follows:



5.1 Vehicle tractive force characteristics.



5.2 Vehicle acceleration characteristics.

$$\dot{W}_t = -(\dot{W}_{acc} + \dot{W}_{rf} + \dot{W}_a + \dot{W}_i + \dot{W}_g) \qquad 5.12$$

where

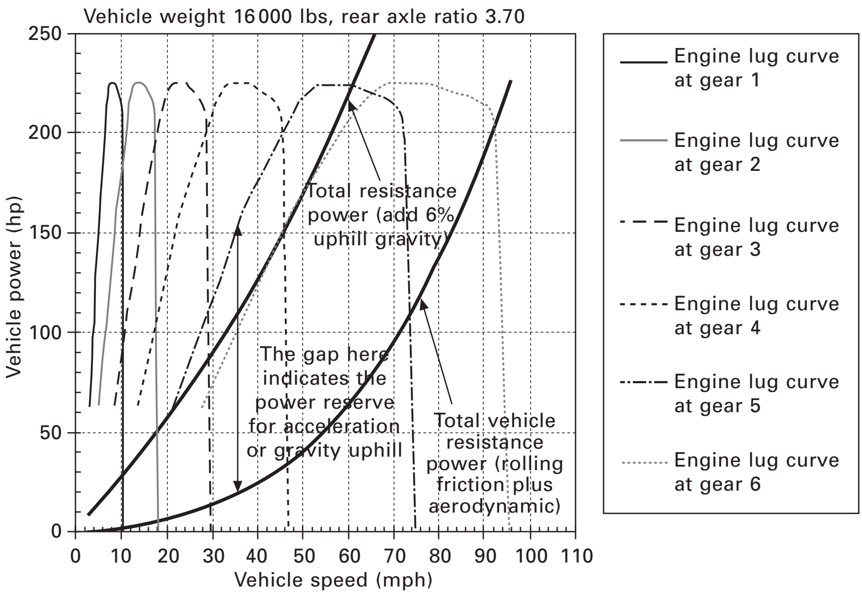
$$\dot{W}_{acc} = \dot{W}_{E,acc}(2 - \eta_t) \qquad 5.13$$

$$\dot{W}_t = \dot{W}_E \eta_t = F_t N_V \qquad 5.14$$

The engine brake power can be calculated by

$$\dot{W}_E = \frac{-(F_{acc} + F_{rf} + F_a + F_i + F_{gl})N_V}{\eta_t} \qquad 5.15$$

Figure 5.3 illustrates the vehicle tractive power characteristics with a six-speed transmission. Unlike the torque calculation, the calculation of the required engine power in equation 5.15 does not relate to the drivetrain transmission ratio. The difference between the engine brake power and the vehicle accessory power is the net available power at the inlet of the clutch or the torque converter to drive the vehicle. Figure 5.4 shows the calculated vehicle power requirements under different running conditions.

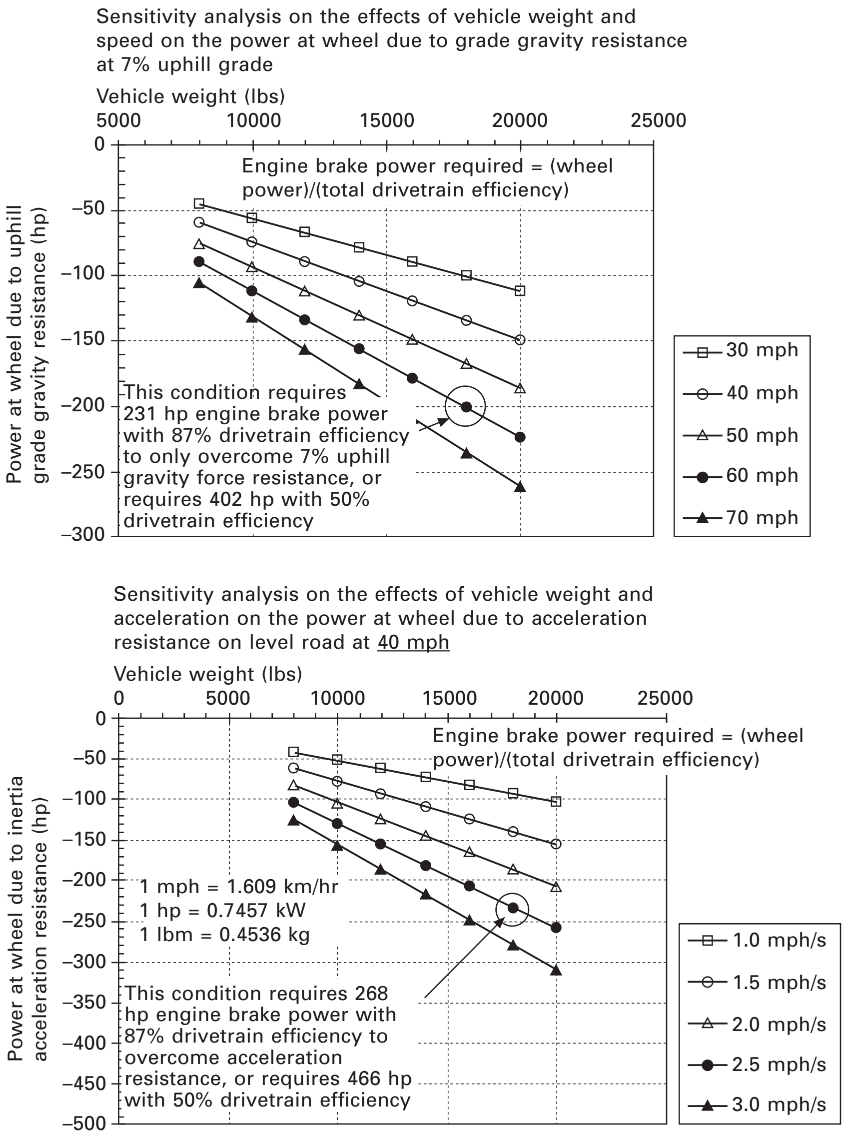


5.3 Vehicle tractive power characteristics.

5.2 Engine–vehicle steady-state matching in engine firing operation

5.2.1 Overview of matching criteria

Drivability refers primarily to a wide range of physical experience felt while driving including acceleration, deceleration, gradeability, transition between engine operating modes and gear shifts, as well as powertrain NVH during the events of tip-in, tip-out, takeoff, gearshift, etc. Drivability evaluation is related to the entire vehicle dynamics engineering and controls, including the engine, the engine mount stiffness and damping, the powertrain command and the management of torque-demand coordination, the gearbox stiffness and inertia, the clutch, the drivetrain, the wheels, the tires, the suspension, the vehicle body, etc. However, in engine system design, the vehicle driving performance is usually determined by three aspects: (1) the maximum vehicle speed on level ground; (2) acceleration time; and (3) the maximum gradeability under a combined vehicle weight rating at constant speed in first gear. Overall vehicle transient acceleration performance is usually characterized by the time taken to accelerate from 0 to 60 mph, shifting from first gear to near the highest gear. Alternatively, it can be characterized by the time spent to cover a distance of 400 meters. Vehicle launch capability is measured by the time to accelerate from 5 to 35 mph. Vehicle overtaking capability is measured by the time to accelerate from 35 to 60 mph or 60 to 80 mph during



5.4 Vehicle power requirements under different operating conditions.

acceleration in a high gear. The maximum desirable gradeability is usually less than 8% uphill grade on highway for heavy-duty commercial vehicles, and approximately 12–16% in mountains. The maximum gradeability is also related to the vehicle's startability.

Vehicle fuel economy is measured by miles per gallon of fuel (mpg) or liters of fuel per 100 km of travel. Several methods exist to evaluate fuel economy.

1. A curve of the fuel amount consumed in liters per 100 km during cruising at each constant vehicle speed plotted versus the vehicle speed.
2. The fuel amount consumed in mpg or liters per 100 km for a specific driving cycle.
3. The weighted average fuel consumption, for example, the US EPA's CAFE composite fuel economy indicator, defined as

$$\eta_{FC,composite,\text{mpg}} = \frac{1}{\dfrac{0.55}{\eta_{FC,urban}} + \dfrac{0.45}{\eta_{FC,highway}}} \quad 5.16$$

   where $\eta_{FC,urban}$ is the fuel economy of the city (urban) driving cycle in mpg, and $\eta_{FC,highway}$ is the fuel economy of the highway (suburban) driving cycle in mpg.
4. Fuel economy during full-pedal acceleration.

The total fuel mass consumed during a driving cycle can be calculated by

$$m_{fuel} = \int_{t_1}^{t_2} \dot{m}_{fuel}\,\mathrm{d}t = \int_{t_1}^{t_2} (\eta_{BSFC}\dot{W}_E)\mathrm{d}t \quad 5.17$$

where $\dot{m}_{fuel}$ is the fuel flow rate and η_{BSFC} is the engine BSFC. If treated as a quasi steady-state approximation, the BSFC at any non-idle condition (i.e., engine brake power greater than zero) can be determined at each engine speed–load point of the driving cycle from the engine BSFC map. The real transient BSFC must be computed by a high-fidelity crank-angle-resolution engine cycle simulation model that reflects the transient combustion efficiency and heat losses, as well as the transient pumping loss during turbo lag which is related to engine controls. A similar approach can be used to compute the total emissions over the driving cycle if the engine emissions maps are available. Engine brake power can be computed by vehicle powertrain transient simulation considering the acceleration inertia effect (equation 5.3). It should be noted that during vehicle deceleration, the fuel rate $\dot{m}_{fuel}$ is either the idle fueling amount or zero (especially during engine braking), depending on the vehicle fueling calibration strategy and the use of engine brake. If the vehicle cruises at a constant speed N_V to travel a distance L_V, the fuel mass consumed can be calculated in simplified form as:

$$m_{fuel} = \eta_{BSFC}\dot{W}_E\left(\frac{l_V}{N_V}\right) \quad 5.18$$

Engine–vehicle matching analysis was discussed by Chana *et al.* (1977), Wong and Clemens (1979), Thring (1981), Phillips *et al.* (1990), Fluga (1993), Watson (1995), Jawad *et al.* (1999), Mikulec and Li (2000), Walker and Ford

(2000), Millo *et al.* (2000), McManus and Anderson (2001), Regner *et al.* (2002), Callahan *et al.* (2003), Giannelli *et al.* (2005), Montazeri-Gh *et al.* (2005), Wehrwein and Mourelatos (2005), and Pagerit *et al.* (2006).

5.2.2 Drivability and drivetrain design parameters

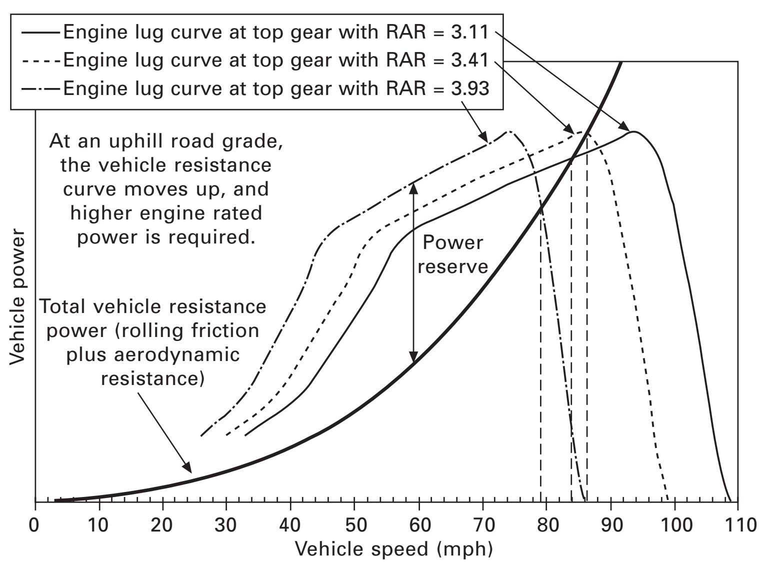
The application scope of the vehicle application determines the design limits of the drivetrain. The vehicle tractive power requirements ($\dot{W}_t = F_t N_V$) are best explained in Fig. 5.1. The entire domain is formed by a family of constant-power curves. The ideal characteristic is a constant power over the full vehicle speed range because it provides the vehicle with a high tractive force at a low speed where the demands for grade climbing and acceleration are high. To achieve superior vehicle performance, fuel economy, and emissions, the drivetrain parameters must be optimized together with the engine torque curve.

Final drive gears are used in driven axles and transaxles to transmit torque from the gearbox output shaft to the driven wheels through a constant gear reduction in order to keep the wheel size practical. The final drive ratio (also known as the drive axle ratio) is defined as the ratio of the number of teeth on the crown wheel of the bevel drive to the number of teeth on the bevel pinion. The drive axle ratio is determined based on the best trade-off between acceleration time and vehicle fuel economy. The minimum drivetrain transfer ratio $(i_{gr}i_{ax})_{min}$ at the highest (top) transmission gear and the engine torque J_E should be designed together to ensure the combined term $J_E i_{gr} i_{ax} \eta_t$ can overcome the minimum resistance forces at the maximum vehicle design speed. It should be noted that only rolling friction and aerodynamic resistance are considered when the vehicle is on level ground. The $(i_{gr}i_{ax})_{min}$ is determined mainly by the drive axle ratio. From a power perspective, once the maximum vehicle speed target is specified, the required engine power can be calculated using equation 5.15. The selection of the minimum drivetrain ratio is based on an optimized compromise between the engine speed and the torque at a fixed power ($\dot{W}_E = J_E N_E$). The choice of engine speed depends on engine design factors such as mechanical friction and diesel combustion quality at high speeds. The maximum vehicle speed target should be high enough to ensure the vehicle can have certain power in reserve to climb an upgrade or accelerate at a lower normal driving speed in the highest gear. Another important analysis in the selection of the rear axle ratio is to conduct a sensitivity calculation to analyze its effect on vehicle fuel economy and acceleration ability, then choose the best trade-off from the so-called 'C shape' curve (Chana *et al.*, 1977; Wong and Clemens, 1979).

The engine speed at the maximum vehicle speed depends on the drivetrain matching strategy. The traditional strategy is to match the engine speed at the right side of the engine rated speed (i.e., where the rated or the maximum power occurs, Fig. 5.5) in order to favor a power reserve at the expense of fuel consumption. Alternatively, the engine speed can be matched to locate at the left side of the engine rated speed in order to reduce fuel consumption but sacrifice the power reserve. A third option is to match the engine speed just at the rated speed. Engine rated power is determined from the above drivetrain-matching considerations. The diesel engine's maximum speed is determined by three factors: (1) mechanical limitations such as valvetrain separation and the high reciprocating inertia forces of the heavy components; (2) the rapidly increasing friction at high speeds; and (3) the time required to complete the combustion process of the diesel fuel. High-speed governors are used to rapidly reduce the fueling amount in order to steeply reduce the brake torque from a speed slightly above the rated speed to the high-idle speed.

Engine rated power requirement is determined by the maximum vehicle speed at an acceptable uphill road grade. If the road grade is set at zero, it is the special case of level ground mentioned above. In the drivetrain matching for light-weight cars, level ground analysis is commonly used. However, for heavy trucks, certain uphill road grade target is usually used to determine the engine rated power requirement.

Engine rated speed is determined by the following considerations. A higher rated speed provides a wider span of the engine speed range from peak torque to rated power so that fewer transmission gears can be used to cover a vehicle speed range for better drivability. Vehicle acceleration can also be

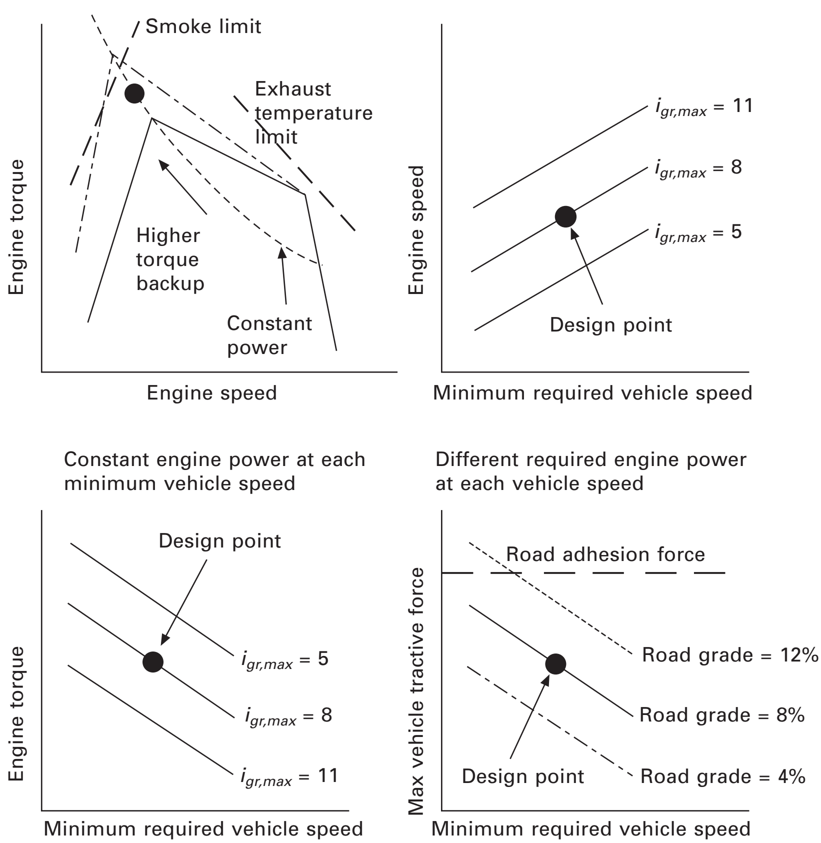


5.5 Integrated design of engine and drivetrain for rated power/speed matching.

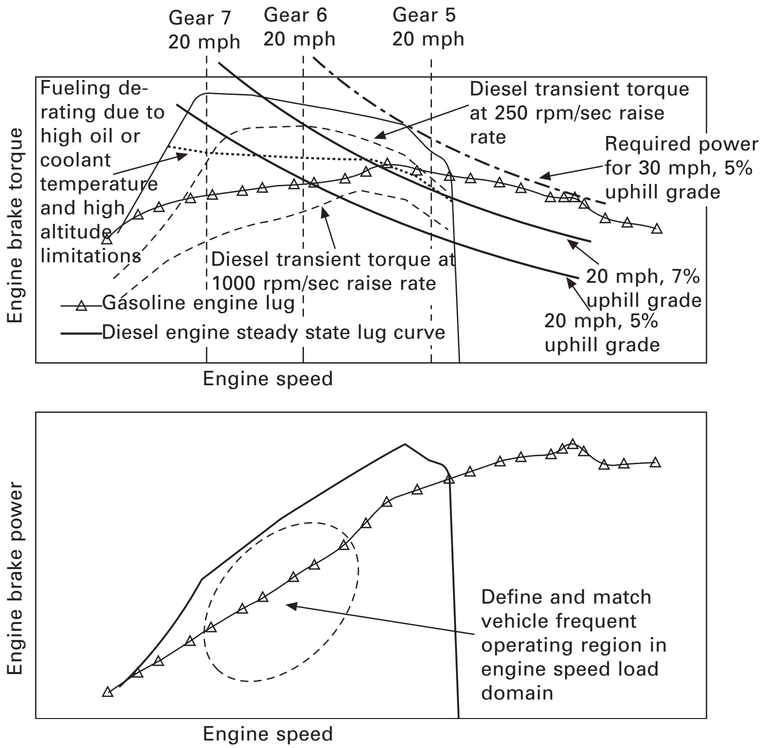
faster, as reflected by a smaller area enveloped under the '1/*a* curve' shown in Fig. 5.2. On the other hand, a lower rated speed generally enables lower emissions on the heavy-duty FTP transient cycle, lower BSFC and lower engine noise at high speeds. Engine brake specific soot emission is usually much higher and more difficult to control at the 'C speed' (the higher speed in the 13-mode test, see Fig. 5.12 later) than the A and B speeds (the lower speeds) in the US SET emission modes. The engine dynamometer speed schedule in some heavy-duty emissions certifications is practically proportional to the declared rated speed. For example, a higher rated speed may center the FTP cycle close to the C speed rather than the B speed compared to the case of using a lower rated speed. A C-speed centered FTP cycle usually produces higher emissions in both NO_x and PM than a B-speed centered cycle. Although lowering the declared rated speed for emissions certification in heavy-duty FTP transient cycle can effectively lower emissions results, the allowable declared rated speed is dependent on the shape of the curve of engine power vs. engine speed. New emission regulations may change the way engine dynamometer speed schedule is calculated for an emissions transient test cycle, for example, changing from using the rated speed to a 'maximum test speed' encountered in the cycle. This may impact the design decisions related to the shape of the engine torque curve and the governed engine speed from an emissions perspective.

The maximum drivetrain transfer ratio $(i_{gr}i_{ax})_{max}$ and the engine torque J_E should be selected together to ensure the combined term $J_E i_{gr} i_{ax} \eta_t$ can overcome the maximum resistance forces (including uphill climbing or acceleration) at a minimum required vehicle speed. The $(i_{gr}i_{ax})_{max}$ is achieved by combining the final drive ratio and the maximum transmission gear ratio. Again, from a power perspective, once the minimum required vehicle speed target is specified, the required engine power can be calculated using equation 5.15. The $(i_{gr}i_{ax})_{max}$ can be calculated using equation 5.6. The selection of the maximum drivetrain ratio should be based on an optimized trade-off between the engine speed and the peak torque at a fixed power requirement. The choice of engine peak torque speed depends on the desirable speed range between peak torque and rated power for torque backup and acceleration. The peak torque speed is limited by the engine's ability to breathe sufficient air in order to produce high peak torque. Usually, it is desirable to have a low peak-torque speed and a high torque level for better drivability and faster acceleration. Note again that the acceleration time is related to the area under the curve of '$1/a_V$ vs. vehicle speed' as shown in Fig. 5.2. Moreover, the engine peak torque should not exceed the durability limit of the drivetrain. Also note that the maximum tractive force should not exceed the tire–road adhesion force in order to avoid tire spinning or slipping. The traction control system provides flexible firing or braking torque to prevent the tire from spinning during acceleration or on a slippery road. A design approach to determine the minimum required vehicle speed, the maximum transmission gear ratio, the engine speed and the torque is illustrated in Fig. 5.6. A proper maximum drivetrain ratio is also required to ensure sufficient startability for vehicles on level ground.

Diesel engines are characterized by high torque due to turbocharging. During fast transient acceleration the steady-state full-load condition cannot be achieved because of turbocharger lag and a minimum required air–fuel ratio to avoid excessive smoke. Vehicle acceleration is limited by engine transient torque characteristics (the curves shown in Fig. 5.7). Figure 5.7 also shows the vehicle driving characteristic curves and the engine derating curve. It is worth noting that the design decision on the shape of the modern engine torque curve is also affected by the requirement to meet emission regulations for heavy-duty diesel engines, as explained earlier in the discussion about the rated speed.



5.6 Integrated design of maximum drivetrain transfer ratio for engine peak torque.



5.7 Diesel full-load lug curve and transient torque.

In the earlier discussions, torque curve shape and torque backup were mentioned. The advantages of a large torque backup coefficient, which is defined as the ratio of peak torque to the torque at rated power, can be summarized as follows.

1. A high torque backup makes the 'tractive force vs. vehicle speed' curve closer to the ideal constant power curve. It allows the vehicle to accelerate or climb a steep upgrade from a low speed with a high torque. It may reduce the number of gears and gear changes required during driving. High torque backup allows the engine to stably respond to a load increase (e.g., climbing a hill) when operating under the full-load condition. The higher the torque backup, the less rapidly the engine speed decreases when going uphill. In contrast, if the engine runs along the lug curve below the peak torque speed where engine torque decreases with a falling speed, a sudden load increase may stall the engine. A lower peak torque speed is desirable because the gear-changing speed needs to be higher than the peak torque speed in order to prevent the engine from stalling. When operating in the region of low engine speed

and high brake torque, a large torque backup requires no or less gear down-shifting when climbing hills or accelerating. This provides driving convenience.

2. A high torque backup provides less vehicle acceleration time as shown by the smaller area under the curve of '$1/a_V$ vs. vehicle speed' in Fig. 5.2. This results in a higher overall vehicle speed and associated better fuel economy in driving cycles because of the higher acceleration and faster attainment of the higher speed.
3. As a result of less downshifting to low gears, there is less chance to operate in the region of high engine speed, low brake torque and high BSFC. This results in better fuel economy.
4. A high coefficient of speed reserve, which is defined as the ratio of rated power speed to peak torque speed, can increase gradeability and defer the required gear downshift. The coefficient of adaptability (defined as the torque backup coefficient multiplied by the coefficient of speed reserve) and gear ratios determine how often the driver must shift gears.

A variation of drivetrain transfer ratio between the minimum and the maximum enables the vehicle to be driven properly at various conditions of load, speed, and road surface. The ratio change can be a discrete change in several gear ratios such as in a traditional transmission, or it can be a continuous change with variable ratios. The number of transmission gears depends on the ratio span between $(i_{gr}i_{ax})_{max}$ and $(i_{gr}i_{ax})_{min}$, and directly affects the vehicle's acceleration performance and fuel economy. A larger number of gears enables the vehicle to operate near the maximum power and enhances the acceleration and hill-climbing capabilities. With a continuously variable transmission (CVT) the vehicle can operate in the region of low fuel consumption (i.e., the region of high load and low speed) on the engine speed–load map when vehicle power requirements change during driving. The gear ratios are designed based on the usage frequency of each gear and also based on a rule that the engine should be operated within a similar speed range in each gear. The ratios, usually less than 1.7–1.8, generally follow a pattern similar to a geometric progression:

$$\frac{i_{gr1}}{i_{gr2}} \geq \frac{i_{gr2}}{i_{gr3}} \geq \cdots \geq \frac{i_{gr(n-1)}}{i_{grn}} \qquad 5.19$$

Properly selecting the transmission gear number and ratios can make the curves of vehicle tractive force close to the ideal constant-power shape, hence maximizing the vehicle acceleration capability. A detailed introduction to drivetrain mechanical structure was provided by Nunney (1998).

Drivability is comprehensively discussed by Everett (1971), Yonekawa *et al.* (1984), Naruse *et al.* (1993), Ciesla and Jennings (1995), List and Schoeggl (1998), Dorey and Holmes (1999), Balfour *et al.* (2000), Schoeggl *et al*. (2001), Sutton *et al*. (2001), Dorey *et al*. (2001), Hayat *et al*. (2003), and D'Anna *et al*. (2005). The reader may find useful information in these studies.

For CVT, the reader is referred to SAE J1618 (2000), Kluger and Fussner (1997), Kluger and Long (1999), Soltic and Guzzella (2001), Osamura *et al*. (2001), Frijlink *et al*. (2001), Min *et al*. (2003), and Lee *et al*. (2004).

5.2.3 Powertrain matching and vehicle fuel economy

The goal of system integration analysis is to seek the optimum combination of the engine and the drivetrain through sensitivity analysis. The optimum is usually the best compromise between vehicle fuel economy and acceleration time. The design options to improve vehicle fuel economy include the following:

1. Design proper engine displacement and higher BMEP level or higher load factor for driving cycles, possibly by downsizing the engine. High BMEP operation provides higher thermal efficiency for the engine and a lower proportion of mechanical friction. But the requirements for vehicle launch or off-idle responsiveness (i.e., basically starting with naturally aspirated conditions) and engine rated power may set limitations on displacement reduction.
2. Reduce engine BSFC and extend the low BSFC region as much as possible in the engine speed–load map. The desirable shape of the constant low BSFC contours from a vehicle driving perspective can be determined by matching (plotting) the frequently encountered vehicle driving cycles on the BSFC map.
3. Increase torque backup and the coefficient of speed reserve to improve the shape of the engine torque curve.
4. Reduce engine and vehicle weight, aerodynamic resistance, tire rolling friction, and accessory power loss. The aerodynamic resistance force is proportional to the square of the vehicle speed. Excessively reducing the rolling friction coefficient in tire design may result in losses in road adhesion and driving smoothness, especially on wet roads. A detailed discussion of fuel consumption and design trade-offs is provided by Steinberg and Goblau (2004).
5. Improve drivetrain efficiency (e.g., using torque converter lockup to prevent slip and efficiency loss in automatic transmissions). Increase the number of transmission gears to make the engine operate in the low BSFC region more often. Identify the frequent operating conditions of vehicle speed and resistance power, and then map these conditions to a region in the engine speed–load domain and try to match the region close to the low BSFC area by adjusting the rear axle ratio, the transmission gear ratios and the shift schedule.

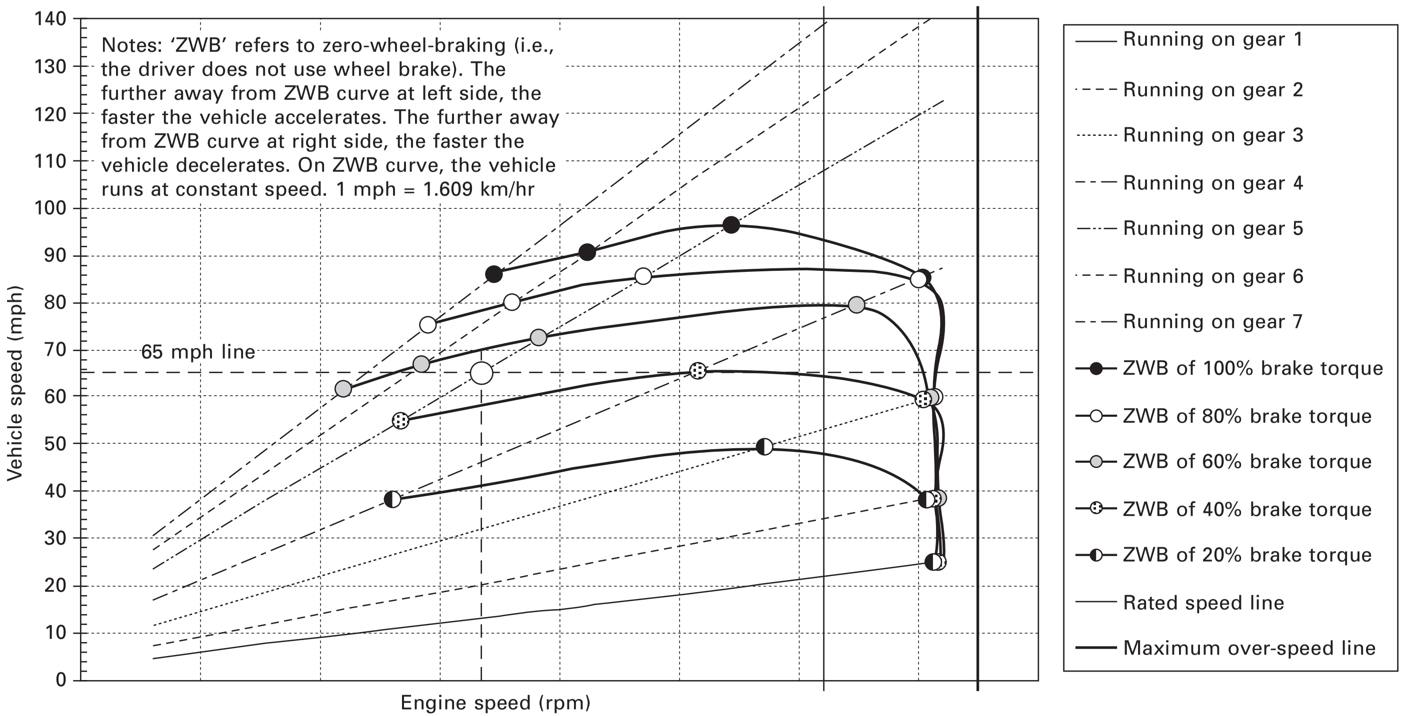
6. As shown in equation 5.17, the fuel rate consumed depends on BSFC, engine power and vehicle speed. At a low vehicle speed, although the power consumption is low, the BSFC is high. At a high vehicle speed, a high engine power is needed to overcome the fast increasing aerodynamic resistance, but the BSFC is low due to the high BMEP level. In fact, the lowest fuel consumption occurs at a medium vehicle speed.
7. Steady-state cruising at constant vehicle speed (or essentially constant engine power consumption) by using a higher gear gives a lower engine speed, higher engine torque, higher load factor and hence lower BSFC. An extreme case of this type of driving is to use the overdrive gear which has an even smaller gear ratio than the top gear. However, the torque reserve or power reserve for uphill climbing and acceleration becomes less.

5.2.4 Analytical engine–drivetrain matching with engine characteristic maps

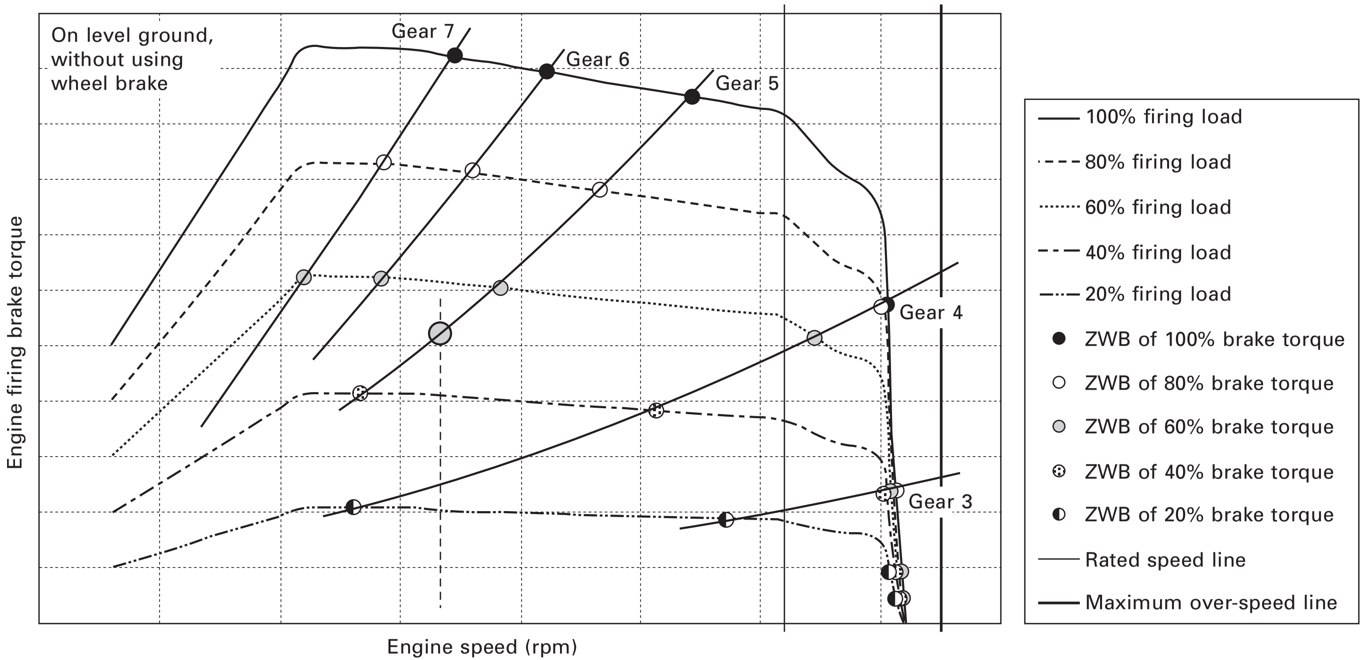
In analytical engine–vehicle matching, the vehicle characteristics are mapped on engine performance maps. The vehicle characteristics include vehicle speed contours, gear lines, road grade lines, and typical vehicle acceleration contours. The engine performance characteristics consist of the contours of firing power, BSFC, NO_x, soot, air–fuel ratio, EGR rate, transient torque smoke limit, etc., as discussed in Chapter 4 and Section 5.2.2. The matching quality and parametric sensitivity can be viewed conveniently, and the best trade-off between drivability and fuel economy can be selected. The analytical formulas presented in Section 5.1.2 are used to construct the vehicle characteristic curves graphically. The scope of the analytical matching approach includes the following:

- Conduct a vehicle force balance to predict the engine power required and the load factor for various vehicle operating conditions in order to better design the engine and its air system for real-world driving scenarios.
- Select the drive axle ratio, the transmission gear number, the gear ratios, and design the transmission shift schedule based on the best trade-offs.
- Simulate transient driving cycles and calculate the cycle composite fuel economy and emissions.

An analytical matching approach is proposed in Fig. 5.8 to map the vehicle operating point to the engine speed–load domain. A concept of 'ZWB' (zero-wheel-braking) is introduced in the approach. The 'ZWB' points on each gear are calculated by using the vehicle force balance for a given vehicle weight, drivetrain ratio and road condition in both the 'engine speed vs. vehicle speed' domain and the 'engine speed vs. torque' domain. The 'ZWB'



5.8 Illustration of engine–vehicle matching method with the 'ZWB' concept.



5.8 Continued

anchor points are connected to form the characteristic curves such as the gear lines in the engine speed–load domain. Then, at a given vehicle speed, the vehicle operating point can be visually selected at a preferred gear, and the corresponding engine speed can be graphically determined. The selected engine speed is used in the engine speed–load domain to intersect with the selected gear line. Then the intersection point is the mapped condition of the running vehicle on the engine map. It should be noted that when the drivetrain condition changes, the 'ZWB' points also change.

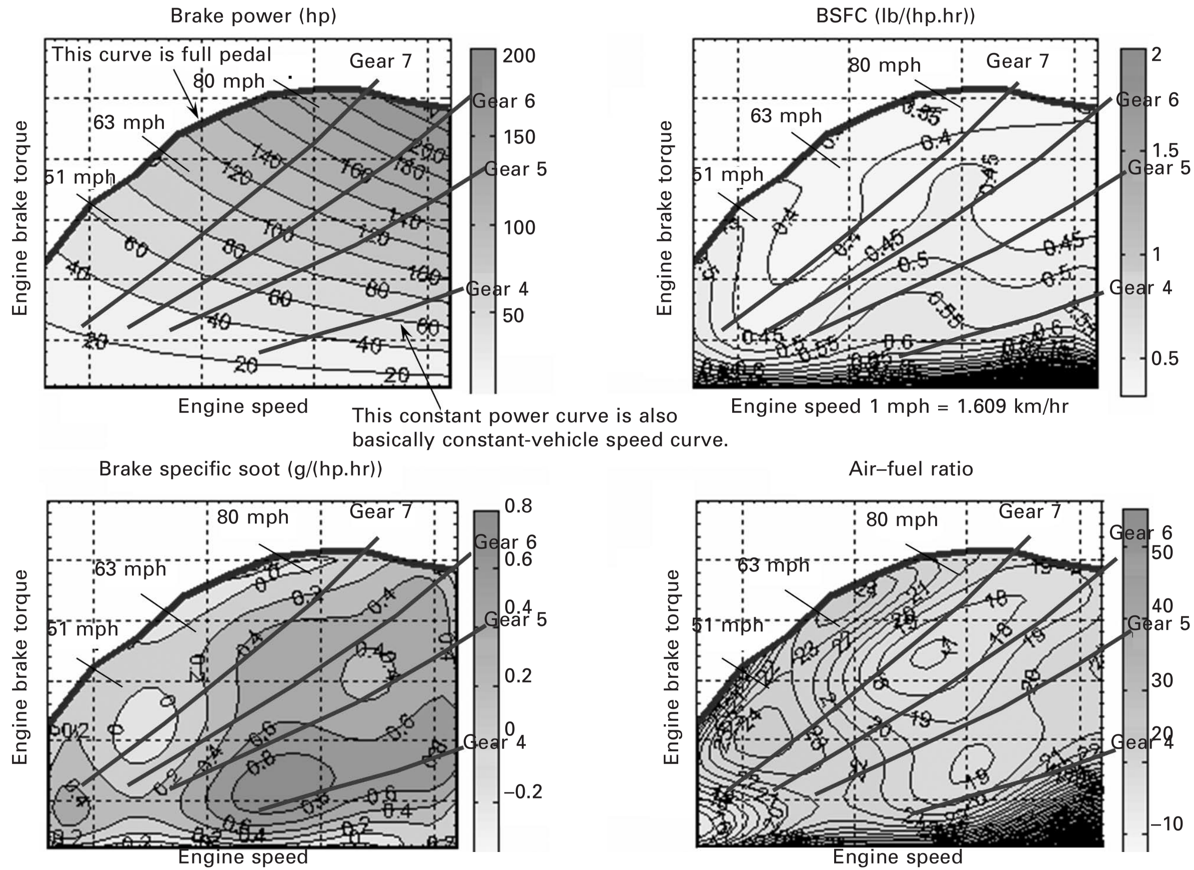
A more powerful matching approach is shown in Fig. 5.9 where the vehicle speed curves, gear lines and for a selected drivetrain or road condition are computed and superpositioned on various engine performance maps. Figure 5.10 shows the rear axle lines and road grade lines on the engine BSFC map. The driving points can be conveniently selected on the engine maps by considering all the performance characteristics, such as BSFC, air–fuel ratio smoke limit, power reserve for acceleration, emissions, etc. Moreover, it is noted that the engine map domain may be viewed as a 'distorted coordinate' domain of the transmission shift map (i.e., accelerator pedal position vs. vehicle speed in Fig. 5.11, compared to Fig. 5.9). Once the transmission shift schedules are visually located on the engine maps based on all the trade-offs, the transmission shift map can be directly 'computed' with mathematical coordinate transformation conveniently. Such an analytical method of engine–transmission matching is more advanced than the traditional 'trial-and-error' approach of generating the transmission shift map.

With this advanced analytical technique of engine–vehicle matching, drivetrain design parameters and engine–transmission matching can be evaluated as a part of engine system design. The calculated vehicle operating characteristics marked on the engine maps provide guidance from vehicle driving requirements to engine design and testing.

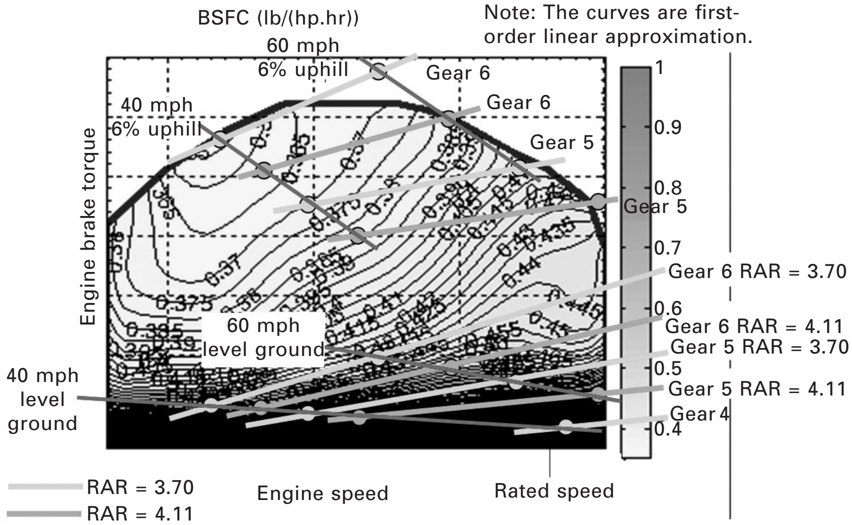
5.3 Powertrain/drivetrain dynamics and transient performance simulation

5.3.1 Theoretical analysis of transient powertrain dynamics

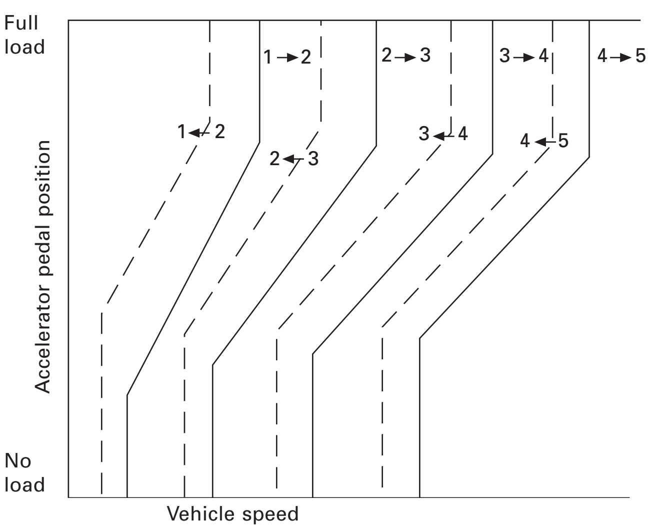
Transient powertrain simulation is important for both understanding engine transient behavior and calculating the cycle-composite fuel economy and emissions of vehicle driving cycles or engine speed–load certification test cycles. Taking the time derivative of equation 5.6, the engine acceleration (i.e., $\mathrm{d}N_E/\mathrm{d}t$) can be obtained as a function of vehicle acceleration (i.e., $\mathrm{d}N_V/\mathrm{d}t$) and a time derivative of the transient gear ratio (i.e., $\mathrm{d}i_{gr}/\mathrm{d}t$). In the derivation below, the transient gear ratio effect is neglected for simplicity. By substituting equations 5.3, 5.8 and the time derivative of 5.6 into 5.1,



5.9 Engine–vehicle matching method with engine performance characteristic maps.



5.10 Engine–vehicle matching for different road grades.



5.11 Illustration of transmission shift schedule map.

equation 5.20 is derived as an ordinary differential equation to solve for the transient engine speed N_E when a transient engine tractive torque J_E is produced (see equation 5.3 on how to relate engine brake torque to vehicle tractive torque):

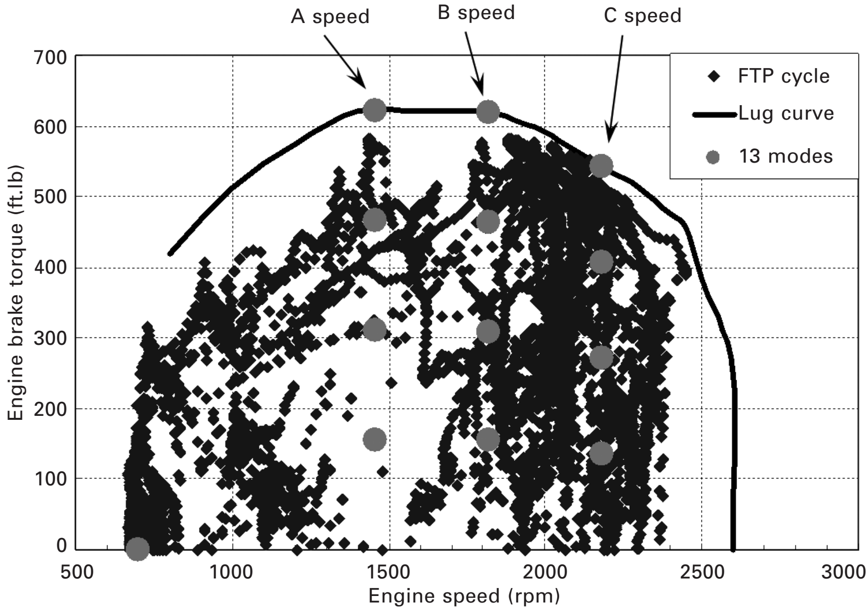
$$I_E \frac{dN_E}{dt} = \frac{r_{tire}\left(F_t + F_{er} + F_{acc} + F_{br} + F_{rf} + F_a + F_{gl} + F_{dr} - m_V a_V - \frac{I_{drive}}{r_{tire}^2} a_V\right)}{2\pi i_{gr} i_{ax} \eta_t \eta_{TV} (1 - f_{slip})} \tag{5.20}$$

All the terms in equation 5.20 are in the transient state and can be modeled in further detail. The tractive force F_t is produced by engine torque. Transient engine torque can be modeled in one of the following methods: (1) empirical formula; (2) model-based formula; (3) mapped engine model; (4) mean-value cylinder model; or (5) high-fidelity crank-angle-resolution engine cycle simulation model. The last two methods may contain a smoke limiter, turbo lag effect, and transient engine control strategies. Diesel engine transient torque characteristics are shown in Fig. 5.7. The available transient torque level depends strongly on how fast the transient event is.

5.3.2 Vehicle driving cycle simulation

In vehicle driving cycle simulations, a vehicle speed schedule as a function of time is imposed as input, and transient engine torque (or retarding torque or vehicle braking torque) and engine speed are computed to try to meet the vehicle speed and load requirements imposed by the vehicle speed schedule and the gear-shifting schedule, unless the imposed transient change of the vehicle speed is unrealistic or too fast. Once the engine's speed and load state is obtained, fuel consumption and emissions can be calculated either by using a high-fidelity engine cycle simulation model or by using the engine BSFC and emission maps.

Figure 5.12 shows the US EPA FTP emissions certification schedule mapped on the engine map for a heavy-duty diesel engine. It is observed that for heavy-duty engine certifications the high speed and high load engine operation occupies a major portion of the FTP transient cycle. Figure 5.13 shows the vehicle speed schedules of the highway fuel economy test (HWFET) driving cycle and the HD-UDDS driving cycle. Figure 5.13 also shows the GT-DRIVE simulation results of a heavy-duty Class-8 truck in these two driving cycles. It is observed that the truck stays much more time in the region of medium/high loads in the HWFET cycle than in the lighter duty HD-UDDS cycle. Obviously, the engine design and calibration strategies for these two different vehicle applications should be optimized differently. Figure 5.14 illustrates some other commonly used driving cycles for heavy-duty and light-duty applications. Figures 5.15–5.17 illustrate the GT-DRIVE simulation results of a medium-duty truck operating in three different driving cycles: the city route, the suburban route, and the highway route. It is observed that the city route application gives the lowest engine load factors, and the highway route demands the engine operate at high speed



5.12 Illustration of SET and FTP operating points on HD diesel engine speed–load map.

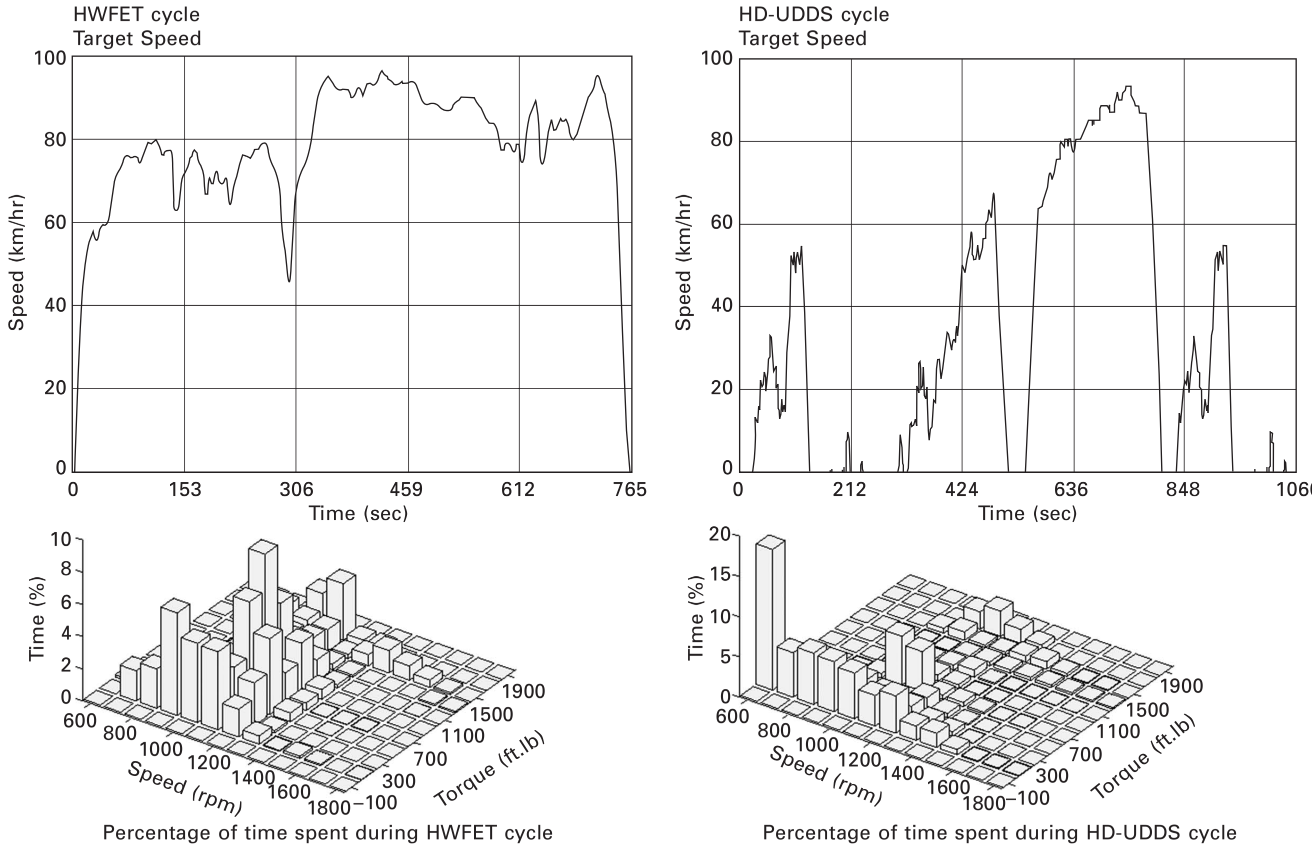
and high load for a long period of time. Again, engine design and calibration should be optimized differently to suit these different applications for this medium-duty truck. Figure 5.18 shows the driving cycle transient simulation of a light-duty truck in the engine speed–load domain. It is observed that, unlike the heavy-duty applications, the light-duty engine mainly operates at low speeds/loads due to the low vehicle weight. The rated power condition is only encountered during fast acceleration or high-road-grade operation in light-duty applications.

5.3.3 Vehicle acceleration simulation

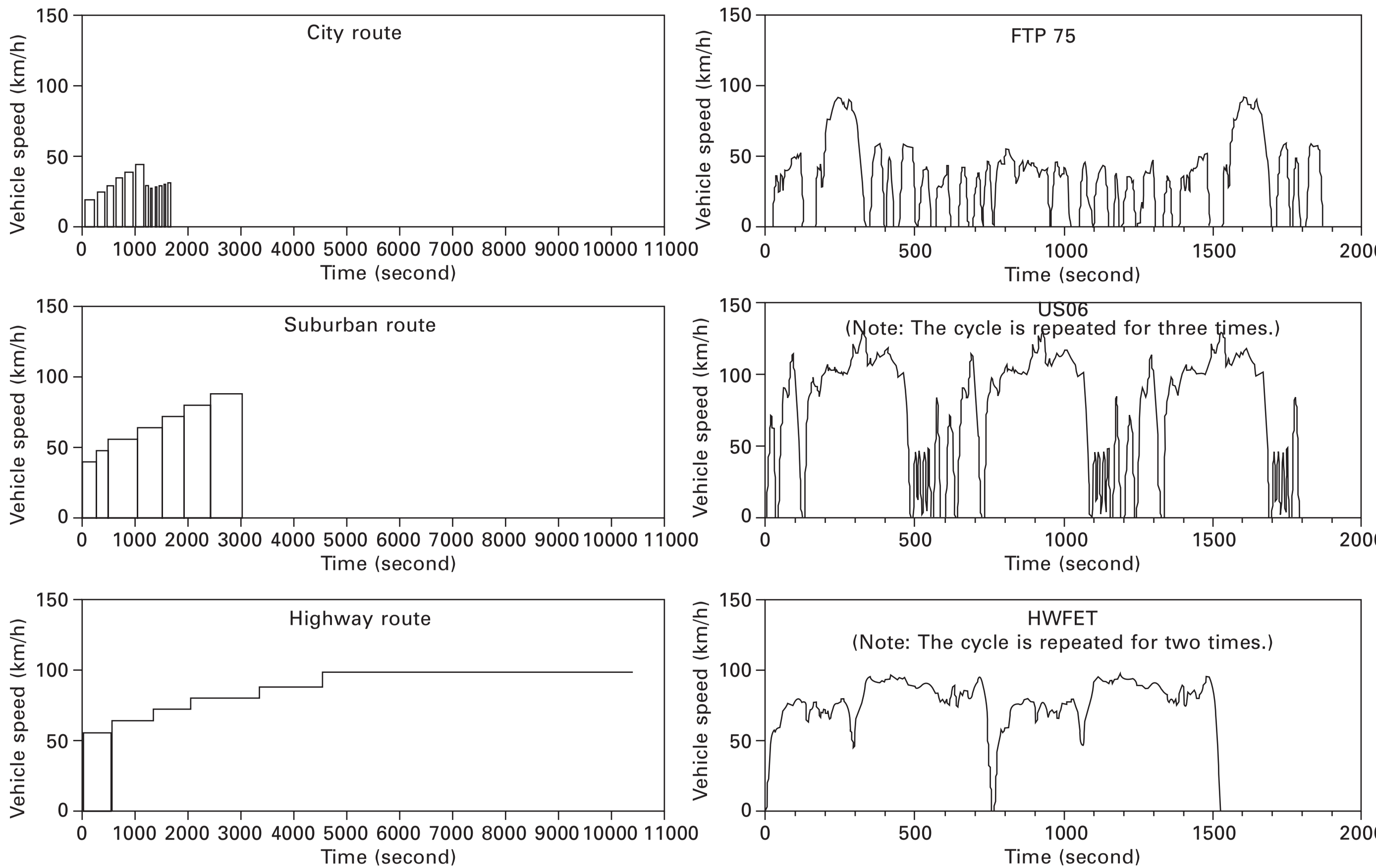
In vehicle full-pedal acceleration simulations, step-fueling and a transmission gear shift schedule can be imposed as input. Transient engine speed and vehicle speed can be computed as a function of time.

Figure 5.19 shows a GT-DRIVE simulation of a light-duty truck during transient full-pedal acceleration from standstill. The rear axle ratio study in Fig. 5.19 shows that the larger rear axle ratio gives faster vehicle acceleration. The vehicle weight effect study in Fig. 5.19 illustrates that the vehicle weight has a paramount impact on acceleration time.

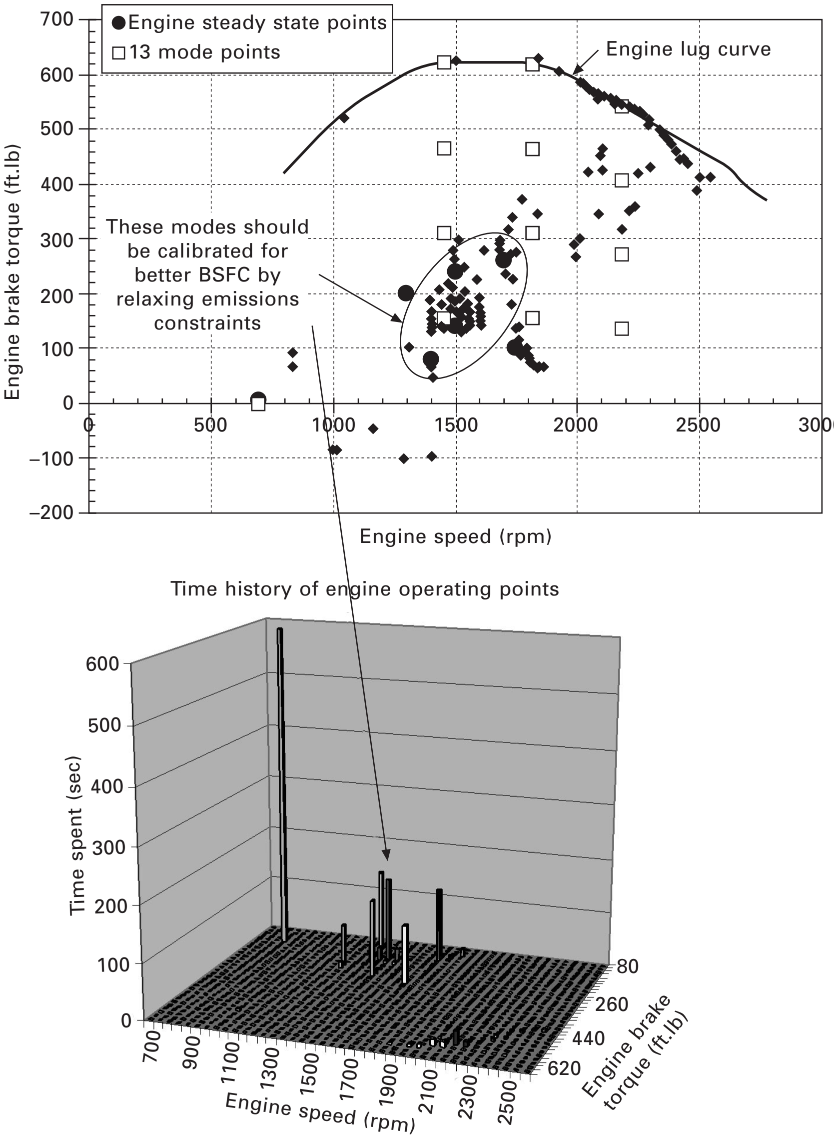
The effect of engine displacement on powertrain transient acceleration is illustrated in Fig. 5.20. It is observed that in this light-duty truck example



5.13 Heavy-duty Class-8 truck driving cycle simulation.

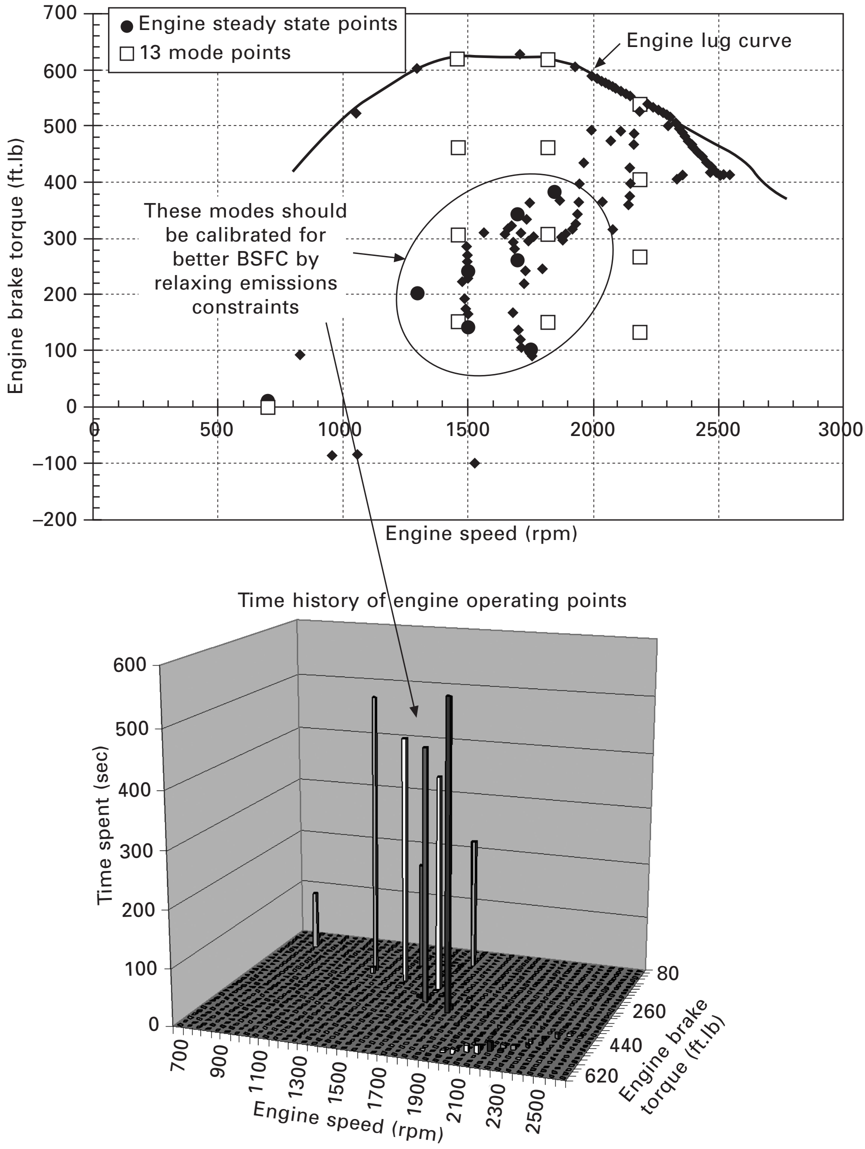


5.14 Illustration of commonly used driving cycles.



5.15 Medium-duty truck driving cycle simulation – city route.

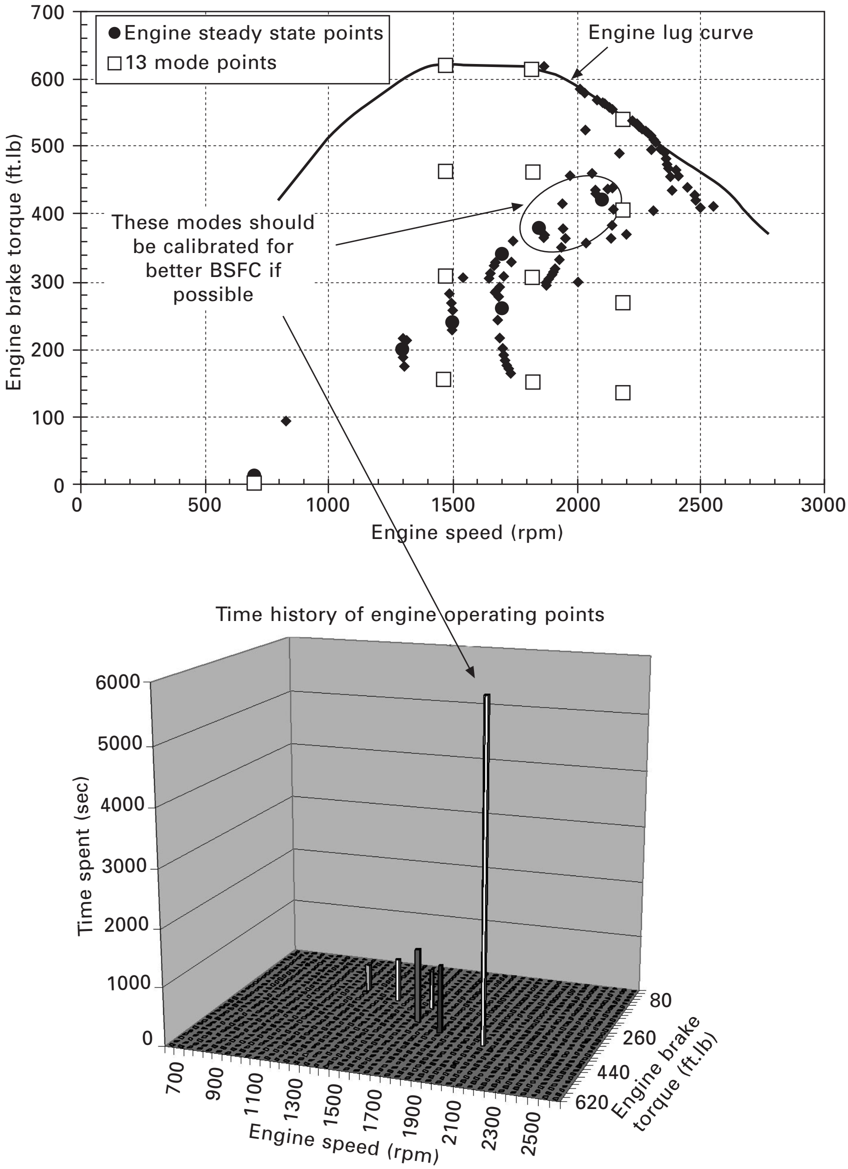
the 8.3% reduction in engine displacement from 4.8 L to 4.4 L makes a significant difference on the vehicle acceleration trajectory mapped in the engine speed–load domain. The 4.4 L diesel engine takes about 0.3 seconds longer than the 4.8 L engine to reach 30 mph from standstill (i.e., 3.7 seconds compared to 3.4 seconds) due to its inferior natural aspiration capability at the beginning of the acceleration.



5.16 Medium-duty truck driving cycle simulation – suburban route.

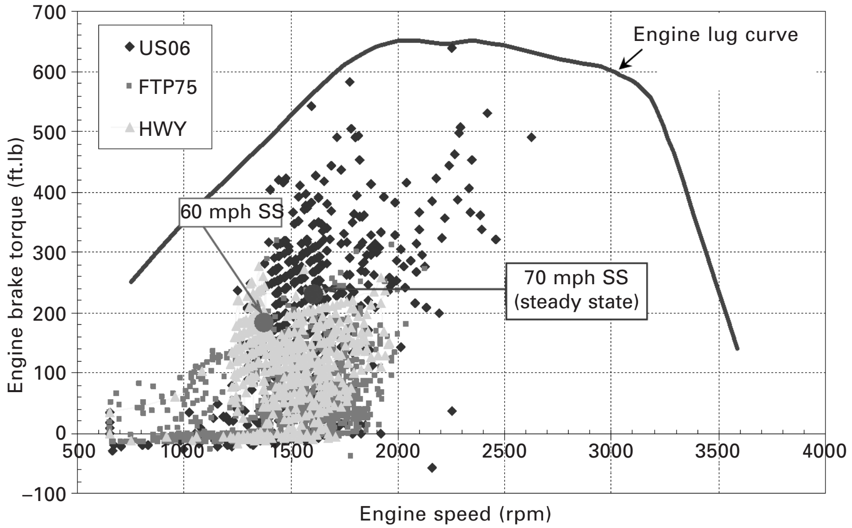
5.3.4 Integrated vehicle–engine driving simulation for system design

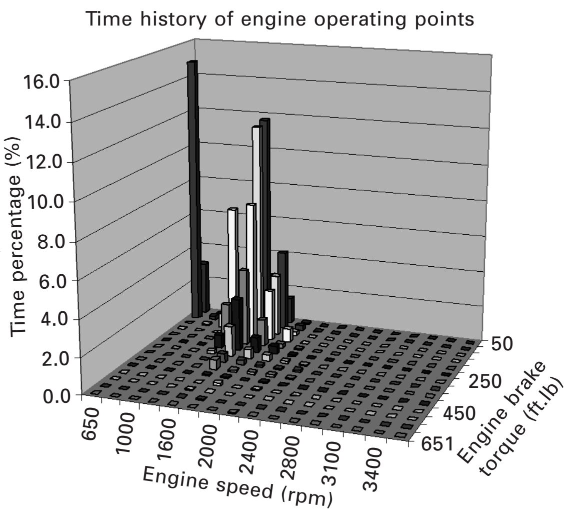
Figure 5.21 demonstrates a simulation capability for a long and slow transient driving route at high altitude uphill climbing coupled with engine coolant-induced power derating for a heavy-duty truck. Such a vehicle driving



5.17 Medium-duty truck driving cycle simulation – highway route.

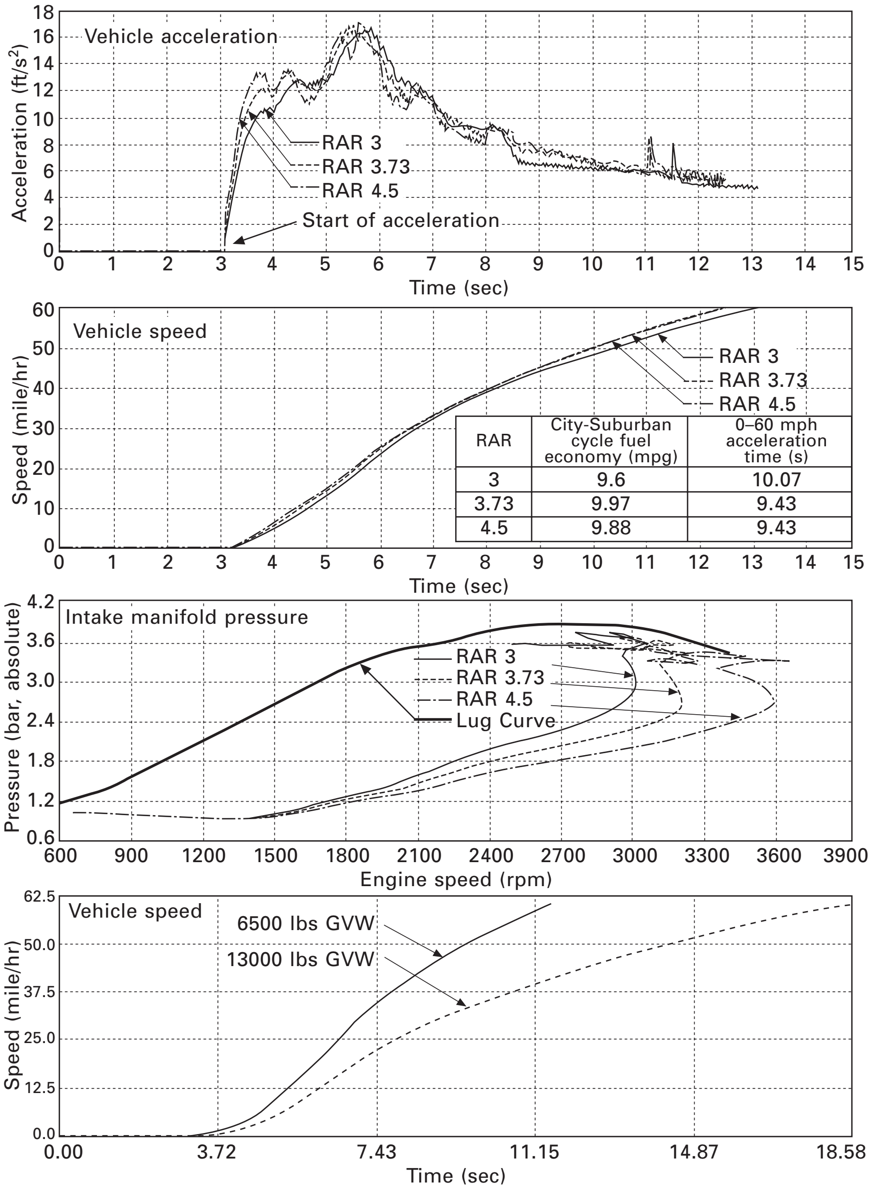
simulation coupled with the engine model is a very valuable tool to assess the engine derating requirement for coolant heat rejection control. Heavy-duty trucks usually encounter the most severe condition on the radiator inlet coolant temperature in high-altitude hot-climate operation due to the deterioration of the vehicle cooling system with the lower-density cooling





5.18 Light-duty truck driving cycle simulation with partially loaded weight (6500 lbs GVW).

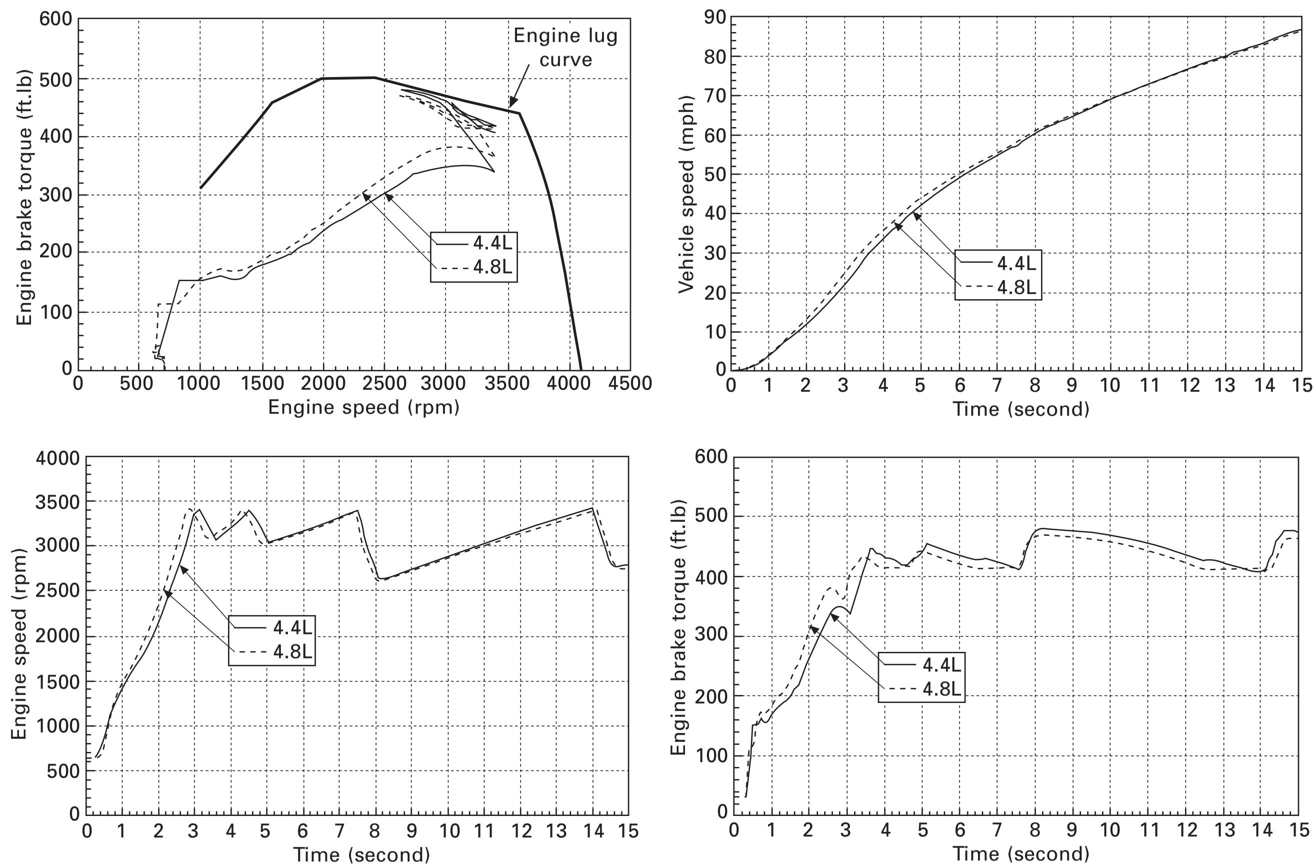
air flow occurring at high altitude. When the coolant temperature exceeds a certain limit, the engine fueling rate is reduced in order to protect the engine from being overheated. The derating obviously impairs the power capability of the engine, and this may become a serious competitive issue when the truck does not have enough power to maintain a high vehicle speed at high altitude in mountain climbing. When the coolant becomes sufficiently cool, the derating is over and more fuel can be injected again to recover the



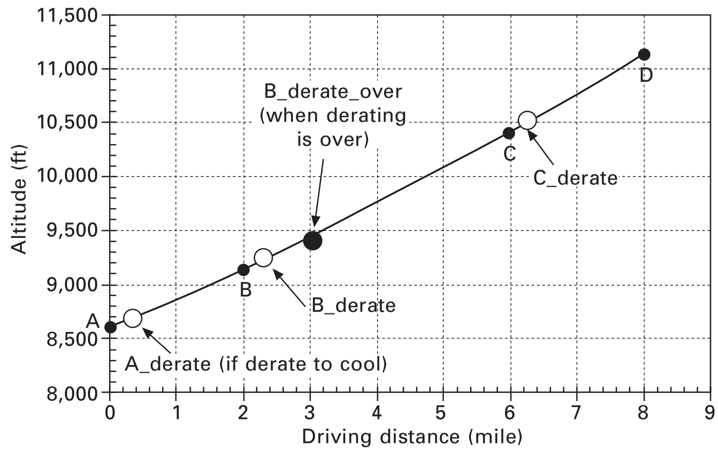
RAR	City-Suburban cycle fuel economy (mpg)	0–60 mph acceleration time (s)
3	9.6	10.07
3.73	9.97	9.43
4.5	9.88	9.43

5.19 Light-duty truck transient vehicle acceleration simulation on the effect of rear axle ratio and vehicle weight.

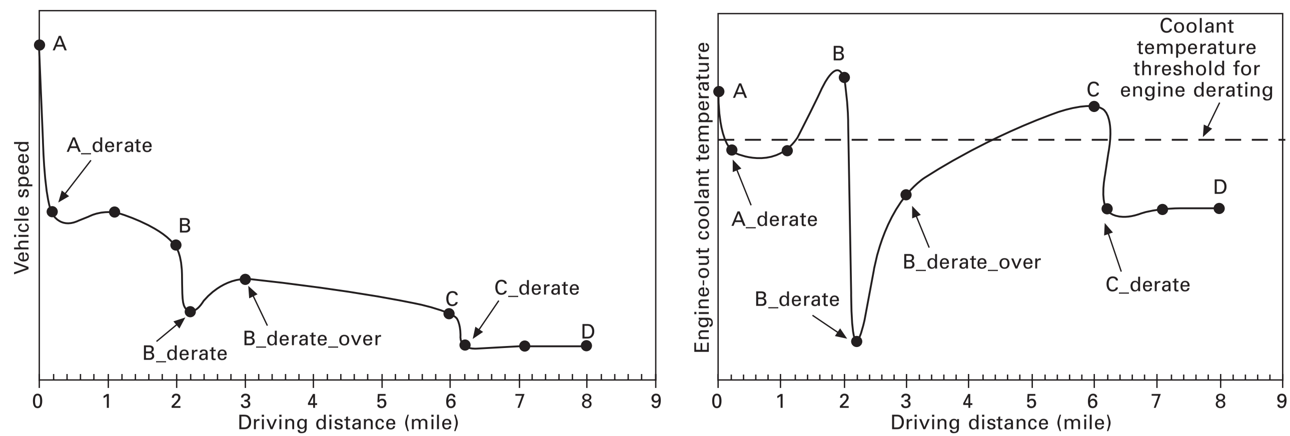
engine power. During high-altitude driving, such derating may occur several times, as shown in Fig. 5.21. This vehicle driving simulation helps identify the problems with heat rejection and derating control and can improve the



5.20 Light-duty truck transient acceleration simulation on the effect of engine displacement.



Driving station index	Driving station location	Driving distance (miles)	Altitude (ft)	Road grade
A	Highways 9 and 70 intersection near Silverthorn, CO	0	8600	5%
B	Laskey Gulch	2	9130	6%
C	Hamilton Gulch	6	10400	7%
C	Entry of Eisenhower Tunnel	8	11140	



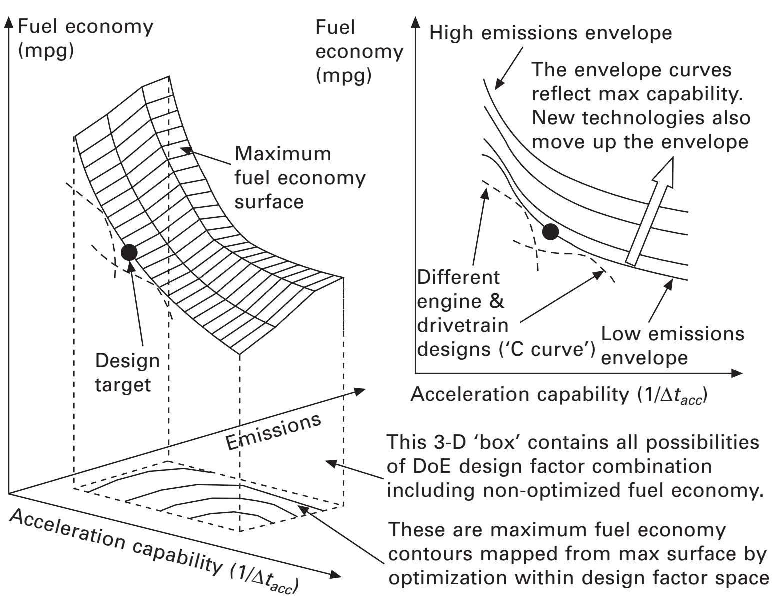
5.21 Vehicle driving simulation at high altitude and hot ambient on the effect of derating for coolant heat rejection control.

system design accordingly to eliminate, minimize, or optimize these derating effects based on all the trade-offs involved.

Vehicle powertrain dynamics performance simulation has also been conducted by Moskwa *et al*. (1997, 1999), Assanis *et al*. (1999, 2000), Bi *et al*. (2002), Vandenplas *et al*. (2003), Sahraeian *et al*. (2004), and Abel *et al*. (2006).

5.4 Optimization of engine–vehicle powertrain performance

In powertrain concept selection, with a given engine–vehicle–road configuration, the above-mentioned steady-state and transient analyses can be automated by computer simulation or programming, and the final analysis can be summarized as one high-level data point within a three-dimensional domain of 'fuel economy (e.g. mpg)' vs. 'vehicle acceleration capability (e.g. the reciprocal of 0–60 mph acceleration time)' and 'engine emissions (NO_x, PM, HC or their weighted average)', as illustrated in Fig. 5.22. Different configurations can be repeated in a DoE analysis, including various drive axle ratios, transmission gear ratios, gear numbers, torque converters, engine rated power and speeds, engine displacements, emission control technologies, or any other system-level parameters. Finally, all the summary data points



5.22 Method of powertrain performance optimization in engine system design.

(one from each configuration) are processed with DoE emulator optimization techniques (i.e., building mathematical equations to link the input data; to the output summary data, see Chapter 3 for details on DoE). Fuel economy can be compared on a fair basis at fixed acceleration capability and emissions. The calculated trade-off envelope curves are identified as the maximum limit of the system capability. The minimum fuel economy constant contours and its parametric dependency can be computed. Such an optimization design approach may greatly enhance the quality of powertrain concept selection and help cascade the best concept to the lower level of design hierarchy with a truly 'top-down' system approach.

5.5 Hybrid powertrain performance analysis

5.5.1 Importance of hybrid powertrain modeling in engine system design

As an alternative to conventional powertrain technologies, hybrid electric vehicles (HEV) and hybrid hydraulic vehicles (HHV) for both light-duty and heavy-duty applications are being developed to meet the current stringent regulations on fuel economy and emissions. The increasing interest in hybrid commercial vehicles requires a detailed understanding of the hybrid powertrain technologies and their system integration.

Diesel engine system design has been focusing largely on engine cycle simulation analysis for the air system (valvetrain, turbocharger, and EGR systems), heat rejection, and conventional vehicle performance for powertrain integration. As the energy conversion technologies have become more versatile, hybrid powertrains have become an integral part of diesel engine system design. The main focus of the hybrid powertrain analysis in diesel engine system design can be summarized as follows.

1. Use powertrain system simulation to evaluate the technical feasibility of various hybrid architectures and quantify the performance benefits of hybridization in fuel economy, emissions reduction and vehicle acceleration in different drive cycles and vehicle applications. Compare the performance of hybrid powertrains with that of conventional vehicles.
2. Develop high-level system design specifications for hybrid powertrains in principal subsystem hardware sizing (i.e., engine, motor, battery).
3. Help evaluate supervisory control strategies by engine–vehicle simulations.
4. Examine the sensitivity impact on performance due to the variations in the design parameters of the battery and the electric machines caused by design limitations, manufacturing tolerances, aging, etc.

5.5.2 Supervisory control strategies and hybrid-electric configurations

The objectives of good supervisory control logic include the following: realizing good driving performance, improving fuel economy, reducing emissions, and maintaining good battery energy management. One of the critical issues in the supervisory controller design is the battery life. It is imperative to keep the state of charge within certain limits in order to avoid undercharge or overcharge so that the battery life can be preserved. The operation of the controller is determined by energy management rules that are a function of the state of charge, the power demand and the braking power. Regenerative braking provides the ability to recover the kinetic energy that otherwise would be lost during the braking events by using the service brakes.

Typically, hybrid electric vehicles are classified into two basic configurations, series and parallel hybrids, although other configurations exist. In the series hybrid mode, there is no mechanical connection between the internal combustion engine and the wheels. The engine is coupled with an electric generator to produce electricity to charge the battery. The battery is connected to an electric traction motor that drives the wheels. Due to various operational possibilities of the traction motor (i.e., controlled as a generator or motor), regenerative braking energy may be captured to charge the battery.

In the parallel hybrid mode, both the engine and the electric motor are coupled to the wheels to propel the vehicle. The engine supplies its power mechanically to the wheels (like in a conventional vehicle), and the engine is assisted by an electric motor if necessary. Some of the advantages of the parallel mode operation include a reduced traction motor size, no need for a generator, and the elimination of efficiency losses in multiple power conversion processes between the electric machines and the battery.

In the electric-vehicle mode, an external energy source (e.g., battery, ultra-capacitor, fuel cell, etc.) and an electric motor are used for traction. The electric vehicle mode does not produce emissions and can be operated with high efficiency and low noise and vibration. The deterioration in vehicle acceleration and gradeability is probably the main disadvantage of the electric vehicle mode, compared with a conventional vehicle.

5.5.3 Hybrid powertrain modeling

There are several commercial software packages available for hybrid powertrain modeling (e.g., Gamma Technologies' GT-DRIVE, LMS's Imagine.Lab, and Ricardo's EASY-5 hybrid powertrain library). Vehicle models can be created from various built-in library components in the object-oriented programming environment in the software. The primary challenge in hybrid powertrain simulation is the modeling of supervisory control strategy to command the engine and the hybrid components. For example, the key elements in a hybrid electric powertrain performance model include the following (Fig. 5.23):

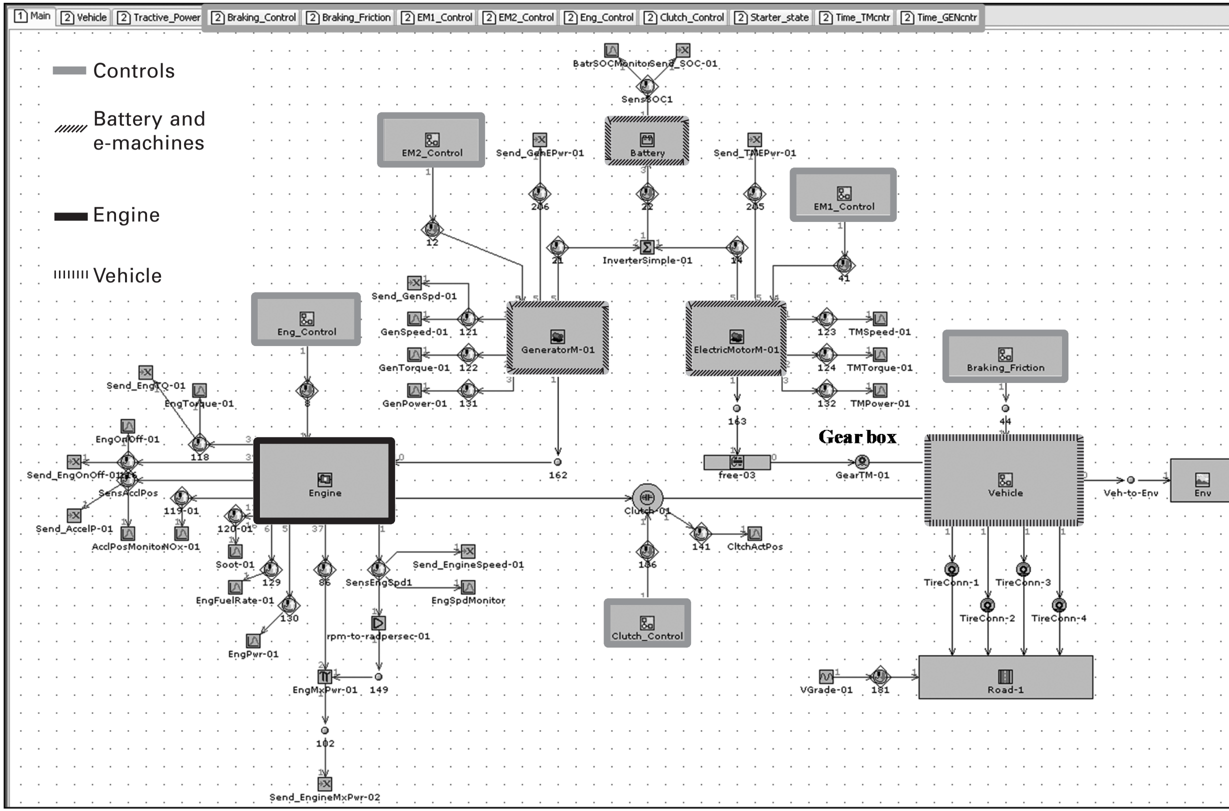
- a map-based engine model for BSFC, emissions, and gas temperatures
- driveline components (e.g., clutch, torque converter, transmission, differential, driveshaft, axle, tire)
- electro-mechanical components (i.e., motor, generator)
- energy storage component (i.e., battery)
- electronic control and signal processing elements.

Once the vehicle architecture is created in the model, a controller for the supervisory control strategy can be designed and implemented using detailed electrical and control components. Typical scenarios integrated within the supervisory control strategy include the following key factors:

- the transition process from electric launch to engine drive
- electric motor assist
- regenerative braking, braking load distribution (mechanical vs. electrical)
- battery 'state of charge' (SOC) maintenance
- interactions, inter-dependence and interlocks between the above factors.

5.5.4 Outlook of hybrid powertrain analysis in system design

The simulation techniques in hybrid powertrain have reached a mature level that can play a major role in hybrid technology development. The simulation conducted in a 'virtual world' is critical for hybrid powertrain development in order to ensure design optimization and cost reduction in testing and development. The modeling of supervisory control strategy and optimizing the complex subsystem interactions are two key challenges in hybrid powertrain simulation. The optimization needs to occur at a powertrain and engine system level in order to ensure the system specifications of performance, durability, packaging, and cost are considered simultaneously. Complete references on hybrid powertrains (mostly for diesel engines) are organized in Sections 5.6.2–5.6.4 below.



5.23 Hybrid-electric powertrain modeling (GT-DRIVE model).

5.6 References and bibliography

5.6.1 References and bibliography on conventional vehicle powertrain

Abel A, Schreiber U and Schindler J (2006), 'Engine and gearbox modeling and simulation for improving the shifting behavior of powertrains with manual or automated transmission', SAE paper 2006-01-1641.

Assanis D, Bryzik W, Chalhoub N, Filipi Z, Henein N, Jung D, Liu X, Louca L, Moskwa J, Munns S, Overholt J, Papalambros P, Riley S, Rubin Z, Sendur P, Stein J and Zhang G (1999), 'Integration and use of diesel engine, driveline and vehicle dynamics models for heavy duty truck simulation', SAE paper 1999-01-0970.

Assanis D, Filipi Z, Gravante S, Grohnke D, Gui X, Louca L, Rideout G, Stein J and Wang Y (2000), 'Validation and use of SIMULINK integrated, high fidelity, engine-in-vehicle simulation of the International class VI truck', SAE paper 2000-01-0288.

Balfour G, Dupraz P, Ramsbottom M and Scotson P (2000), 'Diesel fuel injection control for optimum driveability', SAE paper 2000-01-0265.

Bata R, Yacoub Y, Wang W, Lyons D, Gambino M and Rideout G (1994), 'Heavy duty testing cycles: survey and comparison', SAE paper 942263.

Bi X, Han S, Ma Z and Liu X (2002), 'Transient modeling for heavy tracked vehicle performance and fuel consumption under accelerating conditions', SAE paper 2002-01-0630.

Callahan B J, Kee R J, Mccartan C D, Fleck R, Kenny R G, Mackey D O (2003), 'Drivetrain effects on small engine performance', SAE paper 2003-32-0067.

Chana H E, Fedewa W L and Mahoney (1977), 'An analytical study of transmission modifications as related to vehicle performance and economy', SAE paper 770418.

Ciesla C R and Jennings M J (1995), 'A modular approach to powertrain modeling and shift quality analysis', SAE paper 950419.

Ciesla C, Keribar R and Morel T (2000), 'Engine/powertrain/vehicle modeling tool applicable to all stages of the design process', SAE paper 2000-01-0934.

Clark N N, Gautam M, Wayne W S, Nine R D, Thompson G J, Lyons D W, Maldonado H, Carlock M and Agrawal A (2003), 'Creation and evaluation of a medium heavy-duty truck test cycle', SAE paper 2003-01-3284.

D'Anna T, Govindswamy K, Wolter F and Janssen P (2005), 'Aspects of shift quality with emphasis on powertrain integration and vehicle sensitivity', SAE paper 2005-01-2303.

Dopson A, Taitt D and Sandford M H (1995), 'Powertrain systems definition process', SAE paper 950811.

Dorey R E and Holmes C B (1999), 'Vehicle driveability – its characterisation and measurement', SAE paper 1999-01-0949.

Dorey R E, McLaggan J D, Harris J M S, Clarke D P and Gondre B A C (2001), 'Transient calibration on the testbed for emissions and driveability', SAE paper 2001-01-0215.

Ehlbeck J M and Mayenburg M V (1991), 'Increasing heavy-duty truck fuel economy', SAE paper 912662.

Everett R L (1971), 'Measuring vehicle driveability', SAE paper 710137.

Ferraz L A S, Coelho E P D, Araujo N S S, Sarno I C and Silva A C (2001), 'Powertrain process to translate customer perceptions into technically driven database', SAE paper 2001-01-3838.

Fluga E C (1993), 'Modeling of the complete vehicle powertrain using ENTERPRISE', SAE paper 931179.

Frijlink P, Schaerlaeckens W, Tillaart E and Haas J (2001), 'Simulation of a vehicle with an ICE, CVT, and ISG powertrain – a pre-study for concept evaluation and dimensioning', SAE paper 2001-01-3453.

Giannelli R A, Nam E K, Helmer K, Younglove T, Scora G and Barth M (2005), 'Heavy-duty diesel vehicle fuel consumption modeling based on road load and powertrain parameters', SAE paper 2005-01-3549.

Giannelli R A, Fulper C, Hart C, Hawkins D, Hu J, Warila J, Kishan S, Sabisch M A, Clark P W, Darby C L, Ensfield C, Henry D and Yoder R (2010), 'In-use emissions from non-road equipment for EPA emissions inventory modeling (MOVES)', SAE paper 2010-01-1952.

Guzzella L and Sciarretta A (2005), *Vehicle Propulsion Systems – Introduction to Modeling and Optimization*, Springer-Verlag, Berlin.

Haggis S F, Hansen T A, Hicks K D, Richards R G and Marx R (2008), 'In-use evaluation of fuel economy and emissions from coal haul trucks using modified SAE J1321 procedures and PEMS', SAE paper 2008-01-1302.

Hayat O, Lebrun M and Domingues E (2003), 'Powertrain driveability evaluation: analysis and simplification of dynamic models', SAE paper 2003-01-1328.

Jawad B A, Trimboli B A and Ranspach P (1999), 'A steady state vehicle model to predict engine and transmission performance', SAE paper 1999-01-0742.

Jones C R (1991), 'Heavy duty drivetrains – the system and component application', SAE paper 910035.

Kluger M A and Fussner D R (1997), 'An overview of current CVT mechanisms, forces and efficiencies', SAE paper 970688.

Kluger M and Greenbaum J J (1993), 'Automatic transmission efficiency characteristics and gearbox torque loss data regression techniques', SAE paper 930907.

Kluger M A and Long D M (1999), 'An overview of current automatic, manual and continuously variable transmission efficiencies and their projected future improvements', SAE paper 1999-01-1259.

Lee H, Kim C, Kim T and Kim H (2004), 'CVT ratio control algorithm by considering powertrain response lag', SAE paper 2004-01-1636.

List H O and Schoeggl P (1998), 'Objective evaluation of vehicle driveability', SAE paper 980204.

McManus J G and Anderson S R (2001), 'Co-simulation: developing a virtual vehicle modeling capability', SAE paper 2001-01-0038.

Mercure R A (1979), 'Review of the automotive torque converter', SAE paper 790046.

Merrion D and Weber K E (1999), 'Trucks and buses', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 553–585.

Mikulec A and Li X (2000), 'Rapid evaluation of powertrain subsystems and components', SAE paper 2000-01-3545.

Miller J (2004), *Propulsion Systems for Hybrid Vehicles*, The Institution of Electrical Engineers, Stevenage.

Millo F, Ferraro C V and Pilo L (2000), 'A contribution to engine and vehicle performance prediction', SAE paper 2000-01-1266.

Min B, Matthews R, Duoba M, Ng H and Larsen B (2003), 'Direct measurement of powertrain component efficiencies for a light-duty vehicle with a CVT operating over a driving cycle', SAE paper 2003-01-3202.

Montazeri-Gh M, Varasteh H and Naghizadeh M (2005), 'Driving cycle simulation for heavy duty engine emission evaluation and testing', SAE paper 2005-01-3796.

Morel T, Keribar R, Silvestri J and Wahiduzzaman S (1999), 'Integrated engine/vehicle simulation and control', SAE paper 1999-01-0907.

Moskwa J J, Munns S A and Rubin Z J (1997), 'The development of vehicular powertrain system modeling methodologies: philosophy and implementation', SAE paper 971089.

Moskwa J J, Wang W, Chen S X and Munns S A (1999), 'Using dynamic modular diesel engine models to understand system interactions and performance', SAE paper 1999-01-0976.

Naruse T, Nakashima Y, Murayama Y, Akiyoshi Y and Kurachi T (1993), 'A study on evaluation method and improvement of shift quality of automatic transmission', SAE paper 930673.

Nunney M J (1998), *Automotive Technology*, 3rd edn, SAE International, Warrendale, PA.

Osamura K, Itoyama H and Iwano H (2001), 'Study of an integrated diesel engine-CVT control algorithm for improving drivability and exhaust emission performance', SAE paper 2001-01-3452.

Pagerit S, Sharer P and Rousseau A (2006), 'Fuel economy sensitivity to vehicle mass for advanced vehicle powertrains', SAE paper 2006-01-0665.

Patton K J, Sullivan A M, Rask R B and Theobald M A (2002), 'Aggregating technologies for reduced fuel consumption: a review of the technical content in the 2002 national research council report on CAFÉ', SAE paper 2002-01-0628.

Petrushov V A (1997), 'Coast down method in time-distance variables', SAE paper 970408.

Phillips A W, Assanis D N and Badgley P (1990), 'Development and use of a vehicle powertrain simulation for fuel economy and performance studies', SAE paper 900619.

Regner G, Loibner E, Krammer J, Walter L and Truemner R (2002), 'Analysis of transient drive cycles using CRUISE-BOOST co-simulation techniques', SAE paper 2002-01-0627.

SAE Surface Vehicle Information Report J2084 (1993), 'Aerodynamic testing of road vehicles – testing methods and procedures'.

SAE Surface Vehicle Information Report J2071 (1994), 'Aerodynamic testing of road vehicles – open throat wind tunnel adjustment'.

SAE Surface Vehicle Information Report J1087 (2000), 'One-way clutches – nomenclature and terminology'.

SAE Surface Vehicle Information Report J2408 (2002), 'Clutch requirements for high-torque truck and bus engines'.

SAE Surface Vehicle Information Report J1270 (2006), 'Measurement of passenger car, light truck, and highway truck and bus tire rolling resistance'.

SAE Surface Vehicle Information Report J1715 (2008), '(R) Hybrid electric vehicle (HEV) & electric vehicle (EV) terminology'.

SAE Surface Vehicle Recommended Practice J1252 (1981), 'SAE wind tunnel test procedure for trucks and buses'.

SAE Surface Vehicle Recommended Practice J1321 (1986), 'Joint TMC/SAE fuel consumption test procedure – type II'.

SAE Surface Vehicle Recommended Practice J1526 (1987), 'Joint TMC/SAE fuel consumption in-service test procedure type III'.

SAE Surface Vehicle Recommended Practice J678 (1988), 'Speedometers and tachometers – automotive'.
SAE Surface Vehicle Recommended Practice J2177 (1992), 'Chassis dynamometer test procedure – heavy-duty road vehicles'.
SAE Surface Vehicle Recommended Practice J1594 (1994), 'Vehicle aerodynamics terminology'.
SAE Surface Vehicle Recommended Practice J2264 (1995), 'Chassis dynamometer simulation of road load using coastdown techniques'.
SAE Surface Vehicle Recommended Practice J1667 (1996), 'Snap-acceleration smoke test procedure for heavy-duty diesel powered vehicles'.
SAE Surface Vehicle Recommended Practice J647 (1997), '(R) Transmissions – schematic diagrams'.
SAE Surface Vehicle Recommended Practice J1711 (1999), 'Recommended practice for measuring the exhaust emissions and fuel economy of hybrid-electric vehicles'.
SAE Surface Vehicle Recommended Practice J2452 (1999), 'Stepwise coastdown methodology for measuring tire rolling resistance'.
SAE Surface Vehicle Recommended Practice J2453 (1999), 'Manual transmission and transaxle efficiency and parasitic loss measurement'.
SAE Surface Vehicle Recommended Practice J649 (2000), 'Automatic transmission functions – terminology'.
SAE Surface Vehicle Recommended Practice J1540 (2000), 'Manual transmission efficiency and parasitic loss measurement'.
SAE Surface Vehicle Recommended Practice J1618 (2000), 'Continuously variable transmission test code for passenger cars'.
SAE Surface Vehicle Recommended Practice J923 (2001), 'Nomenclature and terminology for truck and bus drive axles'.
SAE Surface Vehicle Recommended Practice J1266 (2001), 'Axle efficiency test procedure'.
SAE Surface Vehicle Recommended Practice J1608 (2001), 'Manual transmission shift patterns'.
SAE Surface Vehicle Recommended Practice J2548 (2001), 'Test method to obtain lug-up torque data for heavy-duty truck and bus engines'.
SAE Surface Vehicle Recommended Practice J2711 (2002), 'Recommended practice for measuring fuel economy and emissions of hybrid-electric and conventional heavy-duty vehicles'.
SAE Surface Vehicle Recommended Practice J2188 (2003), 'Commercial truck and bus SAE recommended procedure for vehicle performance prediction and charting'.
SAE Surface Vehicle Recommended Practice J651 (2005), 'Passenger car and light truck automatic transmission and automatic transaxle test code'.
SAE Surface Vehicle Recommended Practice J1269 (2006), 'Rolling resistance measurement procedure for passenger car, light truck, and highway truck and bus tires'.
SAE Surface Vehicle Recommended Practice J1491 (2006), '(R) Vehicle acceleration measurement'.
SAE Surface Vehicle Recommended Practice J915 (2007), 'Automatic transmissions – manual control sequence'.
SAE Surface Vehicle Recommended Practice J670 (2008), '(R) Vehicle dynamics terminology'.
SAE Surface Vehicle Recommended Practice J2263 (2008), 'Road load measurement using onboard anemometry and coastdown techniques'.

SAE Surface Vehicle Recommended Practice J2807 (2008), 'Performance requirements for determining tow-vehicle gross combination weight rating and trailer weight rating'.

SAE Surface Vehicle Recommended Practice J645 (2009), '(R) Automotive transmission terminology'.

SAE Surface Vehicle Recommended Practice J1263 (2009), 'Road load measurement and dynamometer simulation using coastdown techniques'.

SAE Surface Vehicle Standard J1544 (1988), 'Revolutions per mile and static loaded radius for off-road tires'.

SAE Surface Vehicle Standard J2469 (2004), 'Clutch application powertrain startability rating requirements for truck and bus applications'.

SAE Surface Vehicle Standard J1082 (2008), 'Fuel economy measurement road test procedure'.

Sahraeian A, Shahbakhti M, Aslani A R, Jazayeri S A, Azadi S and Shamekhi A H (2004), 'Longitudinal vehicle dynamics modeling on the basis of engine modeling', SAE paper 2004-01-1620.

Schoeggl P, Ramschak E and Bogner E (2001), 'On-board optimization of driveability character depending on driver style by using a new closed loop approach', SAE paper 2001-01-0556.

Shannon J V (1991), 'The truck as a system: where are we, and where are we going?', SAE paper 912685.

Simner D (1996), 'The contribution of transmissions to vehicle fuel economy', IMechE paper C498/34/135/95.

Soltic P and Guzzella L (2001), 'Performance simulations of engine-gearbox combinations for lightweight passenger cars', *Proc. IMechE, Part D: Journal of Automobile Engineering*, Vol. 215, pp. 259–271.

Sovran G and Blaser D (2003), 'A contribution to understanding automotive fuel economy and its limits', SAE paper 2003-01-2070.

Steinberg P and Goblau D (2004), 'Fuel Consumption', in Basshuysen R and Schafer F (eds), *Internal Combustion Engine Handbook*, SAE International, Warrendale, PA, pp. 737–751.

Sutton T, Pope R, Brandewie K and Henry J (2001), 'International truck and engine high performance vehicle driveability – ride and handling', SAE paper 2001-01-2787.

Thring R H (1981), 'Engine transmission matching', SAE paper 810446.

Vandenplas B, Gotoh K and Dutre S (2003), 'Predictive analysis for engine/driveline torsional vibration in vehicle conditions using large scale multi body model', SAE paper 2003-01-1726.

Walker M and Ford R (2000), 'Hardware-in-the-loop, dynamometer based driver and vehicle simulator', SAE paper 2000-01-0289.

Watson H C (1995), 'Effects of a wide range of drive cycles on the emissions from vehicles of three levels of technology', SAE paper 950221.

Wehrwein D and Mourelatos Z P (2005), 'Modeling and optimization of vehicle drivetrain dynamic performance considering uncertainty', SAE paper 2005-01-2371.

Wong J Y (1993), *Theory of Ground Vehicles*, 2nd edn, John Wiley & Sons, New York.

Wong L T and Clemens W J (1979), 'Powertrain matching for better fuel economy', SAE paper 790045.

Yonekawa T, Okada A, Ehara M and Ogasawara T (1984), 'Analysis of driveability on diesel vehicles', SAE paper 840257.

5.6.2 References and bibliography on electric hybrid powertrain

An F and Santini D (2003), 'Assessing tank-to-wheel efficiencies of advanced technology vehicles', SAE paper 2003-01-0412.

Anderson C and Pettit E (1995), 'The effects of APU characteristics on the design of hybrid control strategies for hybrid electric vehicles', SAE paper 950493.

Anderson S R, Lamberson D M, Blohm T J and Turner W (2005), 'Hybrid route vehicle fuel economy', SAE paper 2005-01-1164.

Atkins M J and Koch C R (2003), 'A well-to-wheel comparison of several powertrain technologies', SAE paper 2003-01-0081.

Boesel J, Romeo S and Brotherton T (2003), 'Advanced fuels and technology forecasts for heavy-duty vehicles', SAE paper 2003-01-3404.

Burke A (2003), 'Saving petroleum with cost-effective hybrids', SAE paper 2003-01-3279.

Chen W, Chen H, Zhang Y and Qin G (2009), 'Modeling and optimization of vehicle acceleration and fuel economy performance with uncertainty based on Modelica', SAE paper 2009-01-0232.

Chu L and Wang Q (2001), 'Energy management strategy and parametric design for hybrid electric transit bus', SAE paper 2001-01-2748.

Cuddy M R and Wipke K B (1997), 'Analysis of the fuel economy benefit of drivetrain hybridization', SAE paper 970289.

Dembski N, Guezennec Y and Soliman A (2002), 'Analysis and experimental refinement of real-world driving cycles', SAE paper 2002-01-0069.

Ehsani M, Gao Y, Gay S E and Emadi A (2005), *Modern Electric, Hybrid Electric, and Fuel Cell Vehicles – Fundamentals, Theory and Design*, CRC Press LLC, Boca Raton, FL.

Gonder J and Markel T (2007), 'Energy management strategies for plug-in hybrid electric vehicles', SAE paper 2007-01-0290.

Inman S B, Haworth D C and El-Gindy M (2002), 'Integration of WAVE and ADVISOR simulations for optimization of a hybrid electric sport utility vehicle', SAE paper 2002-01-2856.

Juda Z (2001), 'Simulation of energy conversion in advanced automotive vehicles', SAE paper 2001-01-3341.

Kasseris E P and Heywood J B (2007), 'Comparative analysis of automotive powertrain choices for the next 25 years', SAE paper 2007-01-1605.

Kiuchi T, Ogata M, Iizawa M, Yanase T, Susuki Y, Takeda N and Kumagai N (2001), 'A report on advantage of a heavy-duty hybrid electric vehicle', SAE paper 2001-01-2777.

Kwon J, Kim J, Fallas E, Pagerit S and Rousseau A (2008), 'Impact of drive cycles on PHEV component requirements', SAE paper 2008-01-1337.

Lindenkamp N, Stöber-Schmidt C-P and Eilts P (2009), 'Strategies for reducing NO_x and particulate matter emissions in diesel hybrid electric vehicles', SAE paper 2009-01-1305.

Louckes T (2002), 'The opportunity for IC engines in a hybrid world', SAE paper 2002-01-1910.

Louis J J J (2001), 'Well-to-wheel energy use and greenhouse gas emissions for various vehicle technologies', SAE paper 2001-01-1343.

Miller J (2004), *Propulsion Systems for Hybrid Vehicles*, The Institution of Electrical Engineers, Stevenage.

Moon P, Burnham A and Wang M (2006), 'Vehicle-cycle energy and emission effects of conventional and advanced vehicles', SAE paper 2006-01-0375.
Morita K, Shimamura K, Sugiyama G, Hori M, Itai Y, Sekiyama S, Motooka A, Sasaki M and Suenaga K (2005), 'R&D and analysis of energy consumption improvement factor for advanced clean energy HEVs', SAE paper 2005-01-3828.
Olikara C, Steiber J and Shahed S M (1995), 'Analysis of a hybrid powertrain for heavy duty trucks', SAE paper 952585.
Petris C D, Giglio V and Police G (1993), 'Can hybrid vehicles reduce the pollutant emission in urban environments?', SAE paper 930046.
Plassat G (2004), 'Pollutants emissions, global warming potential effect, first comparison using external costs on urban buses', SAE paper 2004-01-2015.
Poeti L, Marco J and Vaughan N D (2009), 'Object oriented plant models for HEV controller development', SAE paper 2009-01-0148.
Pritchard E and Johnson R R (2005), 'Hybrid electric school bus technical feasibility', SAE paper 2005-01-3615.
Rizoulis D, Burl J and Beard J (2001), 'Control strategies for a series-parallel hybrid electric vehicle', SAE paper 2001-01-1354.
Rousseau A, Pagerit S, Monnet G and Feng A (2001), 'The new PNGV system analysis toolkit PSAT V4.1 – evolution and improvement', SAE paper 2001-01-2536.
SAE Surface Vehicle Recommended Practice J1711 (1999), 'Recommended practice for measuring the exhaust emissions and fuel economy of hybrid-electric vehicles'.
SAE Surface Vehicle Recommended Practice J2711 (2002), 'Recommended practice for measuring fuel economy and emissions of hybrid-electric and conventional heavy-duty vehicles'.
Takada Y, Ueki S and Saito A (2004), 'Study on fuel economy and NO_x emissions of medium duty hybrid truck in real traffic conditions', SAE paper 2004-01-1086.
Walters J, Husted H and Rajashekara K (2001), 'Comparative study of hybrid powertrain strategies', SAE paper 2001-01-2501.
Wang W, Wang Q and Zeng X (2009), 'Automated manual transmission shift strategy for parallel hybrid electric vehicle', SAE paper 2009-01-0144.
Williamson S S, Emadi A and Dewan A (2005), 'Effects of varying driving schedules on the drive train efficiency and performance characteristics of a parallel diesel-hybrid bus', SAE paper 2005-01-3477.
Zeng X, Wang Q, Wang W and Chu L (2005), 'Analysis and simulation of conventional transit bus energy loss and hybrid transit bus energy saving', SAE paper 2005-01-1173.

5.6.3 References and bibliography on hydraulic hybrid powertrain

Barr A and Veshagh A (2008), 'Fuel economy and performance comparison of alternative mechanical hybrid powertrain configurations', SAE paper 2008-01-0083.
Baseley S, Ehret C, Greif E and Kliffken M G (2007), 'Hydraulic hybrid systems for commercial vehicles', SAE paper 2007-01-4150.
Ivani Ž (2007), 'Data collection and development of New York city refuse truck duty cycle', SAE paper 2007-01-4118.
Jackey R A, Smith P and Bloxham S (2005), 'Physical system model of a hydraulic energy storage device for hybrid powertrain applications', SAE paper 2005-01-0810.

Johri R and Filipi Z (2009), 'Low-cost pathway to ultra efficient city car: series hydraulic hybrid system with optimized supervisory control', SAE paper 2009-24-0065.

Kim Y J and Filipi Z (2007a), 'Series hydraulic hybrid propulsion for a light truck – optimizing the thermostatic power management', SAE paper 2007-24-0080.

Kim Y J and Filipi Z (2007b), 'Simulation study of a series hydraulic hybrid propulsion system for a light truck', SAE paper 2007-01-4151.

Kokkolaras M, Mourelatos Z, Louca L, Filipi Z, Delagrammatikas G, Stefanopoulou A, Papalambros P and Assanis D (2005), 'Design under uncertainty and assessment of performance reliability of a dual-use medium truck with hydraulic-hybrid powertrain and fuel cell auxiliary power unit', SAE paper 2005-01-1396.

Nakazawa N, Kono Y, Takao E and Takeda N (1987), 'Development of a braking energy regeneration system for city buses', SAE paper 872265.

Surampudi B, Nedungadi A, Ostrowski G, Montemayor A and Gruenewald H (2009), 'Design and control considerations for a series heavy duty hybrid hydraulic vehicle', SAE paper 2009-01-2717.

Van Batavia B L (2009), 'Hydraulic hybrid vehicle energy management system', SAE paper 2009-01-2834.

Weber T R (1988), 'Analysis of unconventional powertrain systems', SAE paper 885023.

5.6.4 References on pneumatic hybrid powertrain

Trajkovic S, Tunestal P and Johansson B (2008), 'Investigation of different valve geometries and valve timing strategies and their effect on regenerative efficiency for a pneumatic hybrid with variable valve actuation', SAE paper 2008-01-1715.

6
Engine brake performance in diesel engine system design

Abstract: This chapter provides a comprehensive theory on engine brake performance. It first discusses vehicle braking requirement and the impact on engine–vehicle matching in engine brake operation, followed by a comparison between engine brakes and drivetrain retarders. It then introduces drivetrain retarders in detail including their torque and cooling characteristics. The performance characteristics of exhaust brakes and compression brakes are elaborated including their mechanisms and the interactions with valvetrain, variable valve actuation (VVA) and turbocharger. The principles of engine brake design are introduced through comprehensive simulation analysis on engine thermodynamic cycles in braking operation. A braking gas recirculation (BGR) theory is developed.

Key words: retarder, engine brake, exhaust brake, compression-release brake, retarder energy conversion ratio, retarding process efficiency, braking gas recirculation (BGR), engine brake noise.

6.1 Engine–vehicle powertrain matching in engine braking operation

Engine brake performance analysis is an integral part of diesel engine system design for two reasons. First, the topics of engine brake performance are focused mainly on power, gas pressure, temperature, and flow rate. Therefore, the theoretical foundation of engine braking is based directly on engine thermodynamic cycle analysis. Second, the design decision of engine brake affects air system sizing and controls. This chapter presents a comprehensive coverage of the vehicle braking requirements, drivetrain retarder and engine brake technologies, and analytical design for braking performance. Both fundamental principles and parametric sensitivity analysis of the engine brakes are presented. The parametric simulations include the effects of braking valve lift, valve timing, valve flow characteristics and turbocharging. The influence of engine braking on valvetrain dynamics is also included. Moreover, braking efficiency, heat rejection, braking noise, and high altitude operation are also discussed.

6.1.1 Vehicle braking requirement and the need for retarders

Vehicle force balance in braking

In order to analyze the braking force requirement for vehicle braking (retarding, no fuel), a lumped model of vehicle longitudinal dynamics is derived as follows based on vehicle force balance:

$$J_{E,r} + \frac{r_{tire}}{C_{tr} i_{gr} i_{ax}}[(F_{dr} + F_{br}) + (F_{rf} + F_a + F_{acc} + F_{df}) + F_{gl} - \xi m_V a_V] = 0 \qquad 6.1$$

where $J_{E,r}$ is the engine retarding torque acting on the crankshaft (including the engine motoring torque), F_{dr} is the resistance force of drivetrain retarders, and F_{br} is the force of the service brakes (also called foundation brakes or wheel brakes) applied on the vehicle wheels. Those three terms are the retarding forces beyond the natural retarding capability of the vehicle. The forces in the second parentheses in equation 6.1 are the natural retarding forces of the vehicle. They are: F_{rf}, the tire–road rolling friction force; F_a, the aerodynamic resistance force; F_{acc}, the vehicle accessory resistance force; and F_{df}, the drivetrain friction force. F_{gl} is the longitudinal gravity force on a grade, and $-\xi m_V a_V$ is the vehicle inertia force. The a_V is vehicle acceleration, m_V is the total vehicle mass, and ξ is the rotational mass coefficient. Other parameters in equation 6.1 are introduced in equation 5.1 in Section 5.1.2.

The need for retarding

When the vehicle needs to be propelled to overcome all resistance forces, it needs the engine to provide firing power. When the vehicle needs to be stopped, it needs the service brakes to slow it down and eventually bring it to a complete stop. When a heavy vehicle travels at a constant speed on a downhill grade, it needs a retarder to balance the gravity force to secure safe driving by reducing the use of the service brakes in order to prevent brake fade. Fast deceleration and high-speed downhill operation are equally important as fast acceleration and uphill driving performance in engine and vehicle design. In mountainous regions, when slowly crawling downhill is not a preferred option, the retarder is almost as equally important to a vehicle as the firing engine for motion control. The retarding device (or called ‘retarders’ in general here) can be a drivetrain retarder or an engine brake. In diesel engine system design, it is always desirable to design the engine as a two-way power conversion device (i.e., power producing and power absorbing or retarding) to possess the retarding function as much as possible. If the engine can be used to replace the driveline retarder, the overall cost of the vehicle can be reduced. It should be noted that although a retarding device can supplement or temporarily substitute for the service brakes, the device should be used for downhill driving or vehicle-slowing on level ground, rather than used as a vehicle stopping device.

The design efforts for continuous vehicle fuel economy improvement have resulted in large reductions in the natural retarding power of the vehicle over the past years. For example, radial tires are used to replace the conventional bias tires to reduce rolling resistance. The reduction of other drivetrain frictional losses (e.g., wheel bearing losses, drive axle gear churning) also reduces the natural retarding capability of the vehicle. Air deflectors are used to reduce aerodynamic drag. Fan clutches are used to reduce cooling fan power. Smaller displacement (engine downsizing) with enhanced turbocharging, reduced rated speed, and decreased parasitic losses all result in a decrease in engine friction and motoring power. As indicated by Meistrick (1992), as a result of the increase in vehicle legal gross weight and the reduction of vehicle natural retarding power since the 1970s, a US Class-8 highway truck in the 1990s required an additional 110 kW (148 hp) to maintain a downhill control speed of 62.5 mph on a 4% grade. On the other hand, engine firing power is mainly sized for the requirements of fast acceleration and high-speed uphill driving. Vehicle operation in mountainous terrains often requires the use of high-firing-power engines. The contrast between the ever increasing firing power and the gradually decreasing vehicle natural retarding power reflects an increasing mismatch between the uphill and the downhill driving, i.e., the ratio between the uphill and the downhill speeds becomes more and more unfavorable for heavy vehicles without additional retarding devices. In engine system design the engine retarding power should be increased in order to minimize the gap compared to the disproportionally increased firing power.

High-speed downhill driving is important for trucking economy. If there are no retarders the heavy vehicle has to move very slowly downhill to ensure the vehicle can be stopped as needed at any moment. Higher vehicle downhill speeds add more burdens on the service brakes, whose design is largely limited by the brake cooling capacity. There is an increasing need to use retarders to maintain a high control speed downhill. The control speed is the constant vehicle speed at which the gravity force pushing the vehicle forward on a downgrade is equal to the total of the resistance forces holding it back without using the service brakes. The ideal vehicle operation is to drive downhill and uphill with at least the same speed with a safety reserve available.

Vehicle retarding is mainly an issue for heavy vehicles. The vehicle potential energy and kinetic energy are proportional to the vehicle weight. To control the speed of the vehicle, the available braking force per unit of vehicle weight needs to be examined. The need for retarders results from insufficient natural retarding capability and service brake capability on a normalized weight basis. A passenger car may have 1 ton of weight, while a typical heavy vehicle may weigh 36 tons. The natural retarding power per vehicle weight of the heavy vehicle is in the order of 10% of the passenger car. The service brake area or braking force per weight of the heavy vehicle is in the order of one half of the passenger car (Habib, 1992). The insufficient service braking capability per vehicle weight in heavy vehicles is the primary reason for using retarders. Retarders can minimize the speed difference between cars and trucks on downgrades.

Theoretical analysis on required retarding power

The parametric dependency of the required vehicle retarding power is analyzed as follows. For level ground driving, the kinetic energy of a vehicle with the travelling speed N_V is given by

$$E_{k,V} = \frac{\xi m_V N_V^2}{2} \qquad 6.2$$

The total braking power required to slow down the vehicle is given by

$$\dot{W}_{r,level} = \frac{\mathrm{d}E_{k,V}}{\mathrm{d}t} = \xi m_V N_V a_V \qquad 6.3$$

Assuming a constant deceleration during the braking, i.e., $N_V = N_{V0} + a_V \cdot \Delta t$ or $a_V = (N_V - N_{V0})/\Delta t$, where N_{V0} is the initial vehicle speed, the total braking power can be derived as

$$\dot{W}_{r,level} = \xi m_V (N_{V0} + a_V \cdot \Delta t) a_V = \xi m_V \frac{N_V^2 - N_V N_{V0}}{\Delta t} \qquad 6.4$$

The total braking time is $\Delta t = - N_{V0}/a_V$. The total braking distance is given by $l_{r,V} = - N_{V0}^2/(2a_V)$. The required braking power reaches a maximum, $\xi m_V N_{V0} a_V$, at the beginning of the deceleration (i.e., at $t = 0$). It is observed that the maximum braking power is linearly proportional to the vehicle's total equivalent mass ξm_V, the initial vehicle speed N_{V0} and the deceleration a_V. The amount of braking power required to stop a vehicle can be enormous. For example, if accelerating a vehicle from standstill to 60 mph within one minute requires 100 hp engine power, to bring the same vehicle from 60 mph to a complete stop within six seconds would require 1,000 hp! Another observation is that while the deceleration may be constant during the braking event, the distance that the vehicle travels each second under braking varies greatly as the speed decreases. The distance being travelled each second is always greater at the beginning of the braking. Therefore, in order to minimize the stopping distance, it is important to make the brake fully effective immediately after it is engaged.

A more important application of the retarders is to maintain a constant descending vehicle speed along a downgrade. In this case, the deceleration is zero, and the potential energy of the vehicle is given by

$$E_{p,V} = m_V g l_{a,V} \qquad 6.5$$

where $l_{a,V}$ is the altitude level for the vehicle, which can be expressed as

$$l_{a,V} = N_V t \cdot \sin\theta \approx N_V t \cdot \tan\theta = N_V t G_r \qquad 6.6$$

where G_r is the road grade and θ is the road slope angle. The total braking power required to maintain the vehicle control speed is given by

$$\dot{W}_{r,grade} = \frac{\mathrm{d}E_{p,V}}{\mathrm{d}t} = m_V g \frac{\mathrm{d}l_{a,V}}{\mathrm{d}t} = m_V g N_V \cdot \sin\theta \approx m_V g N_V G_r \qquad 6.7$$

It is observed from equation 6.7 that the braking power is linearly proportional to the vehicle weight, descending control speed and the road grade. If the aerodynamic and tire–road rolling friction resistances are neglected, the amount of power required to maintain a constant-speed downgrade travel would be basically the same as the power required to propel the vehicle going uphill at the same speed. Since the engine rated power is usually reached for heavy vehicles travelling uphill, a similar power requirement is often specified for the engine brake retarding power used for travelling downhill. For a heavy vehicle running on a steep long downgrade at a high speed, a large amount of retarding power must be continuously supplied for a long time. Usually, only using the service brakes cannot meet such a requirement due to brake pad overheat or even fade. In mountainous regions the road grade can often be 5–8% for a length of several miles. For example, a typical European legal requirement is to maintain 30 km/hour (19 mph) on a 7% downhill grade for 6 km (3.73 miles). Long and steep mountainous grades with switchback curves are very common and therefore require downhill braking. Long descending ramps near bridges are also very common.

The required retarding powers for downhill cruising and level-ground deceleration have similar magnitude. Meistrick (1992) mentioned that to decelerate a Class-8 truck from 55 mph to 25 mph over 0.4 km (1/4 mile) distance, the retarding power required is equal to the retarding power that is required to maintain the vehicle speed at 62.5 mph on a 4% downgrade. The downhill retarding capability of the retarders can be expressed by a combination of vehicle weight, vehicle speed, and road grade (equation 6.7). The level-ground retarding capability can be expressed by a combination of vehicle weight, vehicle speed, and deceleration (equation 6.3).

Methods of retarding

Different levels of vehicle braking can be achieved by the following means.

- Disengage the engine from the drivetrain by placing the transmission in neutral gear and use the aerodynamic drag, tire–road rolling friction resistance, and uphill gravity resistance (if on a upgrade) to retard.
- Engage the engine with the drivetrain by selecting an appropriate transmission gear (and the resulting engine speed) and use the engine motoring power to retard.
- Activate an engine brake or a drivetrain retarder to retard.
- Apply the service brakes to retard.

Usually the aerodynamic and tire–road friction resistances are small compared to the retarding forces from the service brakes or retarders. Although downshifting the transmission can increase the engine motoring power via the increased engine speed to sufficiently maintain or slow down passenger cars, downshifting is usually not sufficient for heavy vehicles due to the substantial weight difference. The details of predicting retarder downhill performance and matching a retarder with the vehicle can be found in SAE J1489 (2000), which includes many typical values for the related parameters. SAE J1489 also illustrates the maps of retarding power and the maximum grades allowable vs. vehicle speed for each gear ratio without using the vehicle's service brakes.

6.1.2 The benefits and challenges of retarders

The service brakes are powerful for short application periods but they cannot be used continuously. Friction-type service brakes are ill-suited for prolonged braking in heavy vehicles due to the physical limitations of convective cooling for the brake pads. When the vehicle speed increases on a downhill grade, the frictional heat absorbed by the service brakes increases. This results in an increase in the brake temperature, regardless of using continuous braking or periodic repeated ('snub') braking. When the brake drum temperature is over 200°C, the brake lining average friction coefficient decreases rapidly to reach a point of fading. Avoiding brake overheating is important to maintain the nominal braking capacity of the service brakes. In downhill driving, brake overheating is related to vehicle weight, road grade, vehicle speed, and braking time (or travel distance). Less fade can only be achieved through lower values of braking energy such as less vehicle weight or speed, or through increased cooling capacity of the brake pad (Limpert, 1975). However, increased convective cooling coefficient or cooling area is difficult to achieve in current service brake designs without increased cost.

The service brake lining wear is nearly proportional to braking energy and increases rapidly with brake temperature. Reducing brake lining temperature is important for extended lining life. To stop or slow down an 80,000 lbs heavy vehicle or maintain a high vehicle speed on a long downgrade by only using the service brakes results in a large amount of wear on the brake pads. The pads have to be replaced frequently or may even fail due to overheat and fade. Vehicle retarders supplement and save the use of the service brakes, reduce service brake fade/failure and wear, and increase driving safety. The use of retarders has the following benefits:

- Control vehicle downhill speed for driving safety. Service brakes can remain cool and have sufficient reserve for safe stopping in an emergency. The primary use of vehicle retarders is on long downgrades where the service brakes would otherwise have to be frequently used to prevent the vehicle from over-speeding. O'Day and Bunch (1981) summarized the truck runaway events by severity, from low to high as follows: smoking brakes, loss of brakes – rode it out; loss of brakes – used runoff ramp (no damage); loss of brakes – used runoff ramp (damage); loss of brakes – crash; loss of brakes – injury; and loss of brakes – fatality. The probability of a runaway accident due to brake failure can be reduced by using the retarders. It should be noted that owing to their limited capabilities, some retarders cannot absolutely prevent runaways. In fact, retarders are so important that they are a necessary part of the continuous braking system for heavy vehicles subject to meeting the legislative requirements of effective and safe braking in many countries (e.g., in European Union countries and Switzerland).
- Reduce the thermal load and wear of the service brakes on brake linings and drums, and extend the brake life by reducing the use of the service brakes. Brake overhaul intervals of heavy vehicles can be increased. This is especially important for frequent downhill driving or in frequent start–stop driving cycles (e.g., transit buses).
- Extend tire casing life for more retreads by keeping the brake and tire casing cooler.
- Enhance the ease of vehicle control and reduce driver fatigue due to repeated use of the service brakes.
- Increase vehicle productivity and longer transportation mileages accumulated due to reduced trip time by maintaining a high vehicle control speed on long downgrades. Other road users will not be impeded when the vehicle speed is sufficiently high.
- Save overall cost for vehicle operation. The cost savings from the increased brake lining life, reduced long-term maintenance cost, less down-time, shorter travel time and the elimination of damage and injury due to a brake-failure accident usually exceed the cost of the retarder. The economic benefits have been confirmed by many studies including certain federally sponsored investigations. A detailed analysis of the economic benefits of the heavy truck retarders was conducted by O'Day and Bunch (1981). They used a return-on-investment model and a model

of the probability of a runaway incident. Their study provided a quantified example of cost–benefit analysis for retarding system designs. Other economic analysis on retarders can be found in Schreck *et al*. (1992). They concluded that the retarder has the advantage of a quick pay-off for itself and brings financial advantages to the transportation companies by increased mileages of retarder-equipped vehicles.

- Bring performance benefits to transient firing operation. Some retarders such as engine brakes can maintain engine component temperatures (e.g., in exhaust braking) and turbocharger speeds (e.g., in compression-release braking) at their operating levels during long braking, thus assuring rapid engine response and good transient performance in terms of emissions and noise when the engine is switched back from the no-fuel braking mode to the firing mode.

The challenges in retarder design include the following:

- Resolve the design complexity and compromise (e.g., turbocharger selection) from the powertrain system to the component when designing a highly capable retarder that has a high torque at low speeds and a high retarding power at high speeds.
- Achieve progressive or variable (continuously, if possible) levels of braking power for smooth and controllable braking. Progressive braking is a desirable feature to control the vehicle speed and reduce driver stress, just like the progressive fueling controlled by the accelerator pedal position in engine firing operation. Progressive braking power levels may also reduce or eliminate wheel lockup caused by over-braking in order to keep the vehicle stable.
- Resolve the dynamic braking coordination and system integration between the retarder and the service brakes under various operating conditions through powertrain controls on its interfaces with other systems (e.g., anti-lock brake, speed limiters).
- Accommodate additional vehicle cooling requirement resulting from the maximum retarder braking performance, especially for the drivetrain retarders, which operate by friction absorption and heat dissipation.
- Endure additional or alternating stress loads on the components.
- Accommodate additional weight and packaging space requirements of the retarder.
- Minimize the higher engine noise produced from certain compression-release engine brakes.
- Minimize the power dissipation (loss) during the shut-off conditions (e.g., for the hydrodynamic retarder).
- Achieve short response and cut-off times for fast engaging and disengaging of the retarder, respectively.
- Achieve a low-cost design of the retarder.

- Minimize additional wear in drivetrain. The potential higher wear on the tires and the braking axles to which the retarder is connected should not be neglected in retarder design and operation. Higher retarding torques at lower engine speeds can usually cause more wear on the driving tires. Moreover, non-smooth retardation may cause tire hop and lockup and hence increase tire wear. Although these effects can be offset by other non-braking axles, a balanced design and the minimum wear are always preferred.

6.1.3 Vehicle braking force distribution and retarder force target

Vehicle wheel lockup, braking efficiency, and vehicle instability

The braking effectiveness of the vehicle is not only related to the braking power but also to how the total available braking force is distributed among all the wheel axles and whether over-braking and under-braking occur. These topics are related to optimal braking, braking efficiency, and wheel lockup. In fact, they affect the realistic maximum design target of the retarder torque. The braking force of a retarder may alter the braking force distribution of the vehicle and hence affect the vehicle's directional stability. When the vehicle weight is low or the road adhesion coefficient is low, over-designing the retarding power capability may cause wheel lockup and vehicle instability.

According to the force and moment balances of the vehicle body and each axle, the load acting on the front axle decreases and the load on the rear axle increases during vehicle acceleration or hill climbing, when compared to their steady-state or level-ground driving. Similar force and moment balance analysis applied to vehicle braking operation reveals that the braking force distribution reaches an ideal situation when the tire–road adhesion limits of the front and rear axles are equal. Consequently, the minimum braking distance of the vehicle is achieved. However, when the braking force changes, the ideal distribution can be maintained only when the braking forces on the front and rear axles are properly adjusted. The concept of braking efficiency has been used to measure how close the braking is to the optimum. The braking efficiency refers to the ability of the braking system to utilize available tire–road friction for the vehicle to achieve a relatively high deceleration before wheel lockup and the corresponding loss of control occur (Radlinski, 1989). A 100% braking efficiency indicates a perfect braking force balance where all the wheels would lock up simultaneously. Braking force distribution directly affects braking efficiency. Braking force distribution also affects brake lining temperatures.

As in the vehicle tractive situation (engine firing), the maximum vehicle retarding (braking) force that the tire–ground contact can support is determined by the tire–road adhesion force. Above the adhesion limit, the tire will slide or lock up. Locking up rear tires or semitrailer tires in a trailer–semitrailer combination makes the vehicle completely lose directional stability in yaw motion. It creates the dangerous tractor 'jackknifing' and semitrailer swing, or makes a two-axle vehicle rotate by 180°. The lockup of front tires causes a loss of steering directional control but not yawing instability. For rear-wheel-driven vehicles on wet or icy roads with very low tire–road friction coefficient, over-braking on the rear wheels should be avoided in order to prevent the rear-wheel lockup and the associated vehicle instability.

Optimum braking force distribution between the front and rear tires ensures the maximum braking forces on the front and rear axles to be developed simultaneously to achieve the maximum deceleration rate and the minimum stopping distance. But the optimum braking force distribution varies with vehicle load, vehicle design, and road surface condition. With a non-optimum distribution, one of the axles will lock first. To prevent locking, less braking force must be applied, resulting in a reduced deceleration capability. The braking force distribution depends on the weight and the dimensions of the vehicle. Highway commercial motor vehicles have a wide range of vehicle weight, ranging from the single-unit trucks having three axles and 50,000–65,000 lbs to the turnpike double length combination vehicles having nine axles and 105,000–147,000 lbs. A detailed description of worldwide highway commercial motor vehicle weights, dimensions, and brake-related standards is provided by Freund (2007). SAE J2627 (2009) and J257 (1997) provide fundamental information on vehicle braking systems.

Service brake force distribution

Two comprehensive reviews on the braking force distribution on axles and heavy-vehicle braking performance have been provided by Radlinski (1987, 1989). He pointed out that it is difficult to optimize the braking performance for all operating conditions without adding complexity to the design. He also compared different braking design philosophies between the US and Europe as a result of their different brake regulations and design practices. For example, braking force distribution was regulated in European requirements, while the US regulations do not specify braking force distribution directly. As a result, European heavy vehicles must use relatively large brakes on the steering (front) axles to match the high front axle loads occurring during braking, and use load-sensing proportioning valves on the driven (rear) axles to reduce the braking force under unloaded or partially loaded conditions or on slippery road surfaces.

In the US design philosophy, balancing axle brake loads according to gross axle weight ratings (GAWR) is considered to be the general practice.

The dynamic braking force distributions among the various axles in the US and European designs are very different. Radlinski's studies may provide valuable insights for selecting the appropriate power target of vehicle retarders when they are integrated into the braking system. The retarding force will change the braking force distribution. Detailed descriptions of the vehicle force balance between axles are provided by Wong (1993). Vehicle service brake design and braking dynamics are elaborated by Limpert (1992).

Retarder force distribution

The required retarder power and force distribution should be analyzed by conducting a vehicle dynamic analysis and then cascaded to the engine system design team as a design target. The role of retarders is to supplement the service brakes to achieve sufficient retarding power and the optimum braking force distribution among all the axles under various vehicle operating conditions such as unloaded and loaded, on dry, wet, or icy road surfaces. The optimum distribution again refers to the condition that the tire–road adhesion limits of the axles are equal (i.e., fully utilizing the maximum braking potential of each axle). The retarding force acting on the driven axle should not exceed the road adhesion limit in order to prevent wheel lockup. This retarder characteristic needs to be considered in the service brake–retarder system design.

It should be noted that unlike the service brakes which are usually effective for both driven and non-driven wheels, the engine brake or vehicle retarder is only effective for the driven wheels. Another important issue when analyzing the wheel lockup with a retarder is the effect of the retarder's 'torque vs. speed' characteristics on longitudinal slip and wheel lockup. Because the retarder has different characteristics compared to the service brakes, the rotational and lockup characteristics of the wheels during retarder braking may be different from the characteristics of the braking using the service brakes. This topic is discussed in detail by Fancher and Radlinski (1983).

For lightly loaded vehicles or driving on slippery roads, the use of full powerful retarding may make the vehicle lose its directional stability. Such an improper use of the retarder was analyzed by Fancher and Radlinski (1983). The technique in their study can be used to analyze the maximum retarding power requirement under various vehicle operating conditions by considering the practical effect of wheel lockup. Vehicle directional stability with the operation of an engine brake and a hydrodynamic retarder was also analyzed by Göhring *et al*. (1992).

6.1.4 Classification of drivetrain retarders and engine brakes

In vehicle brake system design, all the retarders belong to the category of auxiliary brakes, which supplement the service brakes. By definition (Limpert, 1992), an auxiliary brake is a continuous brake in which the retarding torque is not generated by the friction between two sliding surfaces such as linings and a drum. Therefore, essentially they are wear-free brakes. All retarders can be classified into two categories: the drivetrain retarders (also called transmission or propeller shaft brakes, including the trailer axle retarders), and the engine brakes. In general, the drivetrain here may include the transmission and the driveline.

Some authors (e.g., Haiss, 1992) used the term 'primary or input retarders' to refer to the engine brakes or the drivetrain retarders that are installed in the powertrain (i.e., either in the engine or between the engine and the transmission). Their braking torque varies with the transmission gear selected. On the other hand, some authors used the term 'secondary or output retarders' to refer to the drivetrain retarders that are either bolted directly to the transmission output flange or located between the transmission and the driven axles. Their braking torque is irrelevant to which gear is engaged. Sometimes retarders are also installed on multi-axle towed vehicles. Such a classification is useful when evaluating the impact of transmission gear change on retarder performance. The primary retarders may increase the retarder speed by gear shifts (e.g., on a steep downgrade with a low vehicle speed), but they may lose the braking effect during the gear shift process in the case of manual transmissions. Among all the retarders/brakes, the compression-release engine brake is the most common type used in heavy vehicles equipped with diesel engines.

The drivetrain retarders include any retarding device that is located in the transmission or the driveline, essentially between the transmission and the rear axle, such as the friction retarder (the friction brake), the hydrodynamic and electric (electrodynamic and electromagnetic) retarders. Drivetrain retarders are usually used with mid-range-power diesel engines, for example in the city buses and the refuse trucks running at low speeds and with stop-and-start type of driving cycles. Drivetrain retarders are popular in Europe in a wide range of vehicles from the light-weight classes to 7.5 ton vehicles (Göhring *et al.*, 1992).

The engine brakes, strictly speaking, include all the retarding devices whose retarding power is generated from the engine or from the directly attached devices on the engine crankshaft before the transmission. They include the following five types:

1. The crankshaft/flywheel attachment brake. A typical example of this type of brake is Caterpillar's BrakeSaver (Darragh, 1974), which was

a flywheel hydraulic retarder. Other examples may include the power absorption devices through mechanical, hydraulic, pneumatic, or electrical means directly attached to the engine crankshaft or flywheel.
2. Engine motoring as a brake (i.e., by simply shutting off fueling).
3. The intake-throttle brake.
4. The exhaust brake.
5. The compression-release brake (or simply called compression brake or decompression brake).

Among the above five types, the exhaust brake and the compression brake are most popular and they are the most common 'engine brakes' people refer to. In the exhaust brake, there are three types: (1) the conventional exhaust brake by using a flap or butterfly valve installed at the turbine outlet; (2) the variable-geometry turbine (VGT) exhaust brake by modulating the flow area of the VGT; and (3) the variable-valve-actuation (VVA) exhaust brake.

There are generally four types of compression brakes, namely (1) the conventional compression brake; (2) the bleeder brake (i.e., the braking valve constantly held open during the engine cycle); (3) the exhaust-pulse-induced compression brake (also called the EPI or exhaust pulse brake; the braking valve actuated by exhaust manifold pressure pulses); and (4) the VVA compression brake. Compression brakes are usually used on Class-7 and Class-8 heavy vehicles (e.g., big semi-trailers, 18-wheel trucks, buses) and off-highway equipment. More than 90% of the Class-8 trucks in North America and a majority of the big bore diesel engines (larger than 10 liters in displacement) are equipped with compression brakes.

Very often the terms 'engine brake' or 'engine retarder' are mistakenly used to only refer to the compression-release brake, or even only to the 'Jake Brake'. For example, there are road traffic restriction signs of 'No Engine Braking' which result from the noise control ordinance in some towns. The 'Jake™', 'Jake Brake™' and 'Jacobs Engine Brake™' are registered trademarks of the Jacobs Vehicle Systems™ (JVS, formerly known as Jacobs Manufacturing Company and Jacobs Vehicle Equipment Company). The idea of the compression-release brake was conceived by Clessie Cummins, the founder of the Cummins Engine Company, in 1934 after a hair-raising and nearly fatal truck ride down a mountain pass with faded service brakes. Later, the Jacobs Manufacturing Company applied Cummins' idea (Cummins, 1959) and successfully implemented the compression brake in production, producing the first Jake Brake in 1961. JVS has been a leading manufacturer of engine brakes in the US and has a long history of implementing braking technologies. The compression brakes made by JVS were so famous that the term 'Jake brake' is sometimes incorrectly used to refer to compression-release engine brakes in general. In fact, Jake Brake refers to all of the JVS retarding products including the Jacobs Engine Brake, Jacobs Exhaust Brake and Jacobs Driveline Brakes. It should be noted that there are many different types of compression brakes such as the conventional compression brake, the bleeder brake, and the exhaust-pulse-induced compression brake. Both JVS and other companies produce many different types of engine brakes and driveline retarders. For example, in addition to the Jacobs Engine Brake, several compression brakes were made by other companies. These include Mack's Dynatard (Greathouse *et al.*, 1971), Cummins' C-brake, Mercedes-Benz's 'Konstantdrossel' brake, and Pacbrake's Engine Brake.

6.1.5 Braking mechanisms of retarders

The hydrodynamic retarder converts the mechanical energy of the drive shaft to the heat of the fluid inside the retarder via a rotor-stator design. Its retarding torque depends on the amount of fluid flow and the pressure of the viscous fluid inside the retarder. The retarding torque is controlled by adjusting the hydraulic pressure. The fluid under viscous damping is cooled by the engine coolant through the vehicle radiator or a separate cooler (in the case of a retarder-equipped trailer).

The electromagnetic retarder uses a disc rotating in a magnetic field to generate the eddy current which produces the braking torque. The braking toque is controlled by adjusting the exciting current.

Caterpillar's retarder 'BrakeSaver' (Darragh, 1974) used engine oil as a working medium churning between a rotor and a stator. It was mounted between the engine crankshaft and the flywheel in order to utilize the speed–torque multiplication by the transmission.

The retarding power of an engine brake consists of the contributions from mechanical rubbing friction power, engine accessory power (such as the fuel pump, the water pump and the oil pump), the pumping loss power during the intake and exhaust strokes, and the indicated power during the compression and expansion strokes.

The simplest form of engine braking is to use the engine's motoring power (without fueling). The motoring power is basically a second order function of the engine speed. Downshifting the transmission gear (e.g., from 5th gear to 4th) or turning off overdrive (in an automatic transmission) may increase the engine speed thus increase the retarding power. However, the engine speed should not exceed its maximum allowable limit. The engine mechanical friction here does not include the parasitic losses from certain vehicle accessories such as the air compressor, the power steering pump, the radiator fan, etc. The retarding power from these vehicle accessories is a part of the vehicle's natural retarding capability (see equation 6.1).

In the intake-throttle brake (also called 'vacuum engine brake'), high pumping loss or retarding power is achieved by the fact that the piston has to resist the high vacuum created in the cylinder when the piston moves down during the intake stroke. One example of the intake-throttle brake is that the gasoline engine develops retarding power by closing the intake throttle plate which restricts the amount of intake air. Gasoline engines commonly have an intake throttle to regulate the air flow to match the fuel flow in order to achieve stoichiometric combustion. Note that the non-EGR old diesel engines did not have an intake throttle. When the intake throttle is closed during engine braking, a gasoline engine may generate higher pumping loss and higher retarding power than a diesel engine of the same displacement without an intake throttle or an exhaust brake, if compared at the same engine speed. It should be noted that gasoline engines are designed to run at much higher engine speeds than diesel engines and hence they may develop higher retarding power due to their higher speed.

Similar to the intake-throttle brake, in the conventional exhaust brake with an exhaust throttle valve, high pumping loss and retarding power are achieved by the fact that the piston has to overcome the high exhaust manifold pressure during braking when the piston moves up in the exhaust stroke. The maximum potential of the intake-throttle brake comes from only 1 bar maximum in engine delta P (i.e., the difference between the 1 bar absolute pressure in the exhaust manifold and zero bar vacuum in the intake manifold), while the potential of the exhaust brake is much higher than 1 bar in terms of engine delta P. This is the reason why the exhaust brake is usually used instead of an intake-throttle brake in diesel engines. In fact, the exhaust brake is the most widely used retarder to date. It is usually used with mid-range-power diesel engines in a wide range of Class-2 through Class-7 vehicles, for example from the light pickup trucks that haul heavy loads (e.g., 10,000–15,000 lbs GCVW) in hilly terrain to heavy trucks and buses.

A compression-release engine brake is a device that can convert a diesel engine from a power-producing machine to a power-absorbing machine, for example to a reciprocating (piston) air compressor. During engine braking, with the fuel flow terminated, the power required to compress the air in the cylinder comes from the kinetic or potential energy of the moving vehicle and the turning engine crankshaft. In this case, the engine valves remain closed during the compression stroke. If the exhaust valve also remains closed during the expansion stroke (as in the normal valve event in firing operation), the compressed air will return the positive power to the vehicle via the crankshaft during the expansion stroke so that the earlier absorbed (negative) power is cancelled out. In this situation, the only retarding power offered by the engine is the pumping loss during the intake and exhaust strokes as well as the mechanical friction from the moving parts. If the exhaust valve is opened near the end of the compression stroke or in the very early portion of the expansion stroke, the compressed air is allowed to discharge to the exhaust manifold thus the energy stored in the compressed air is no longer retained in the cylinder during the expansion stroke to produce positive power. As a result, the net effect is that the engine becomes a power-absorbing air compressor to use the kinetic or potential energy of the moving vehicle to pump the ambient air through the engine cylinders and then discharge the hotter air to the atmosphere. The negative power produced during the compression and expansion strokes contributes greatly to the total retarding power of the engine. In this operation, such an exhaust valve is called an engine braking valve.

In fact, the braking valve can be the exhaust valve (like in the Jake Brake) or an additional decompression valve in the cylinder head (like in Mercedes-Benz's Konstantdrossel engine brake and Mitsubishi's Powertard engine brake). There are numerous designs and inventions to realize the optimum functions of the braking valve(s) through mechanical, hydraulic, and electromagnetic valvetrain mechanisms.

Although some different types of retarders can be combined to use on one vehicle, not all of them are compatible in retarding processes or mechanisms because the operating principle of one retarder may conflict with another. For example, an exhaust brake and a compression brake should not be used simultaneously at high engine speeds in a turbocharged engine because the exhaust brake operation tends to reduce the turbine pressure ratio and the turbine speed, while a compression brake relies on high turbocharger speed in order to deliver high boost pressure to enhance the compression-release effect. However, for naturally aspirated engines a combination of an exhaust brake and a compression brake has proven effective (Schmitz *et al.*, 1992, 1994). Another example is that if the cylinder charge is basically completely blown off during the expansion stroke in a compression brake, there will be little air to be compressed to raise the cylinder pressure during the exhaust stroke. Therefore, the use of an exhaust brake may not be effective in conjunction with such a compression brake. SAE J1621 (2005) and J2458 (1998) provide information on dynamometer testing and capability rating of engine brakes and exhaust brakes.

6.1.6 Comparison between engine brakes and drivetrain retarders

Retarder performance is measured by the following characteristics: retarder energy conversion ratios (related to thermodynamics), heat rejection (i.e., heat absorbed by the retarder and transferred to the vehicle cooling system), maximum power, retarding power ratio (the ratio of the maximum retarding power to the maximum firing power), low-speed torque (essentially the shape of 'torque vs. speed' curve), weight, power-to-weight ratio (indicating packing space and the stress level), power density (i.e., specific power or the ratio of power to engine displacement), transient response, and noise (related to operating mechanism). Other design attributes include durability, packaging (installability), and cost.

In general, a hydrodynamic retarder may be able to develop a retarding power as high as double that of the engine firing rated power. However, its heat rejection is also very high. The cooling system used to dissipate the heat is a great design challenge. A compression brake may be able to develop a similar amount of retarding power as the engine firing rated power. An exhaust brake usually produces significantly lower retarding power. With improved braking mechanisms and stronger engine structures, more powerful compression brakes are being developed.

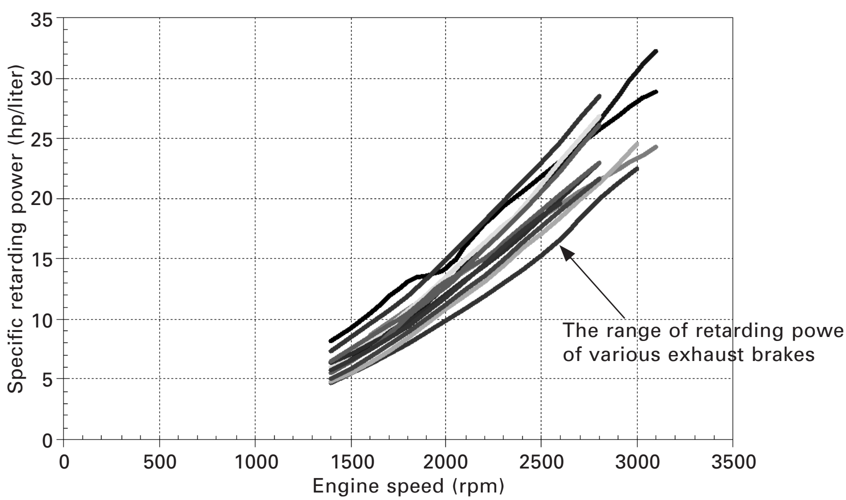
High retarding torque in the mid-speed range is very important. The torque capability of the hydraulic and electric retarders at medium to high speeds is usually limited by the temperature of the retarder's cooling medium. The drivetrain retarders often deliver poor performance (low power) at low vehicle speeds but good performance at higher vehicle speeds, while the engine brakes may deliver high power at low vehicle speeds. Engine brakes can achieve this by downshifting the transmission gear to run the engine at high speeds so that they can deliver high power on a much steeper downgrade compared to a drivetrain retarder. The low-speed torque of an engine brake can be increased by using an exhaust brake that throttles the turbine outlet or inlet.

The retarder power capability or retarder design limit is often translated into an allowable road grade for downgrade performance, or a vehicle deceleration in a fraction of the acceleration due to gravity for level-ground performance. For instance, a typical European legal requirement for downhill braking is that the retarder needs to maintain 30 km/hour (18.6 mph) vehicle speed on a 7% downgrade for 6 km (3.7 miles). Customers can be more demanding on these requirements, for example, requiring 80 km/hr speed on a 5% downgrade. Limpert (1975) reported that the mean vehicle deceleration when using the engine alone was approximately 0.015 *g*; and using the exhaust brake increased the mean vehicle deceleration to nearly 0.03 *g*. A 0.05 *g* (0.5 m/s^2) deceleration was reported in the literature for braking from 80 to 60 km/hr for a 40 ton truck (Schreck *et al.*, 1992). As a comparison, a 0.15 *g* deceleration (1.47 m/s^2) is typical in many normal vehicle stops using the service brakes (Spurlin and Trotter, 1982), as seen in typical transit buses from normal operating speeds to low speeds (Klemen *et al.*, 1989).

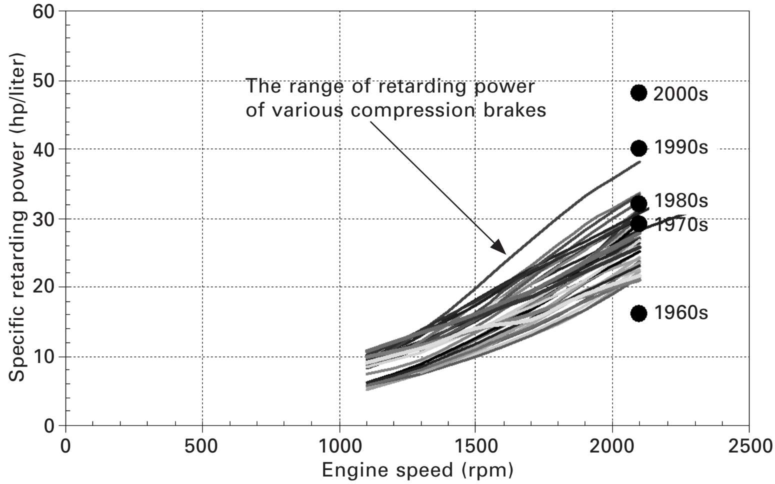
Hydrodynamic and electromagnetic retarders are usually heavy, large and expensive. A conventional compression brake must be structurally strong/heavy enough to support the high loading due to the in-cylinder pressure near the TDC (top dead center) transmitted via the braking valve, while a compression brake actuated by variable valve timing devices can be much lighter (e.g., reduced weight by half; Hu *et al.*, 1997b). The exhaust-pulse compression brake can be even lighter. The exhaust brake has very low weight (only several kilograms) and is usually less expensive than other retarders.

Typical engines have 5–9 hp/liter power density in motoring at 2000 rpm (SAE J1489, 2000). Exhaust brakes usually have around 10–15 hp/liter retarding power density at 2000 rpm (Fig. 6.1). Note that the retarding power data in this chapter are all presented as positive values for convenience. Compression brakes may have at least 30–50 hp/liter power density (Fig. 6.2). Compression brakes usually offer higher retarding power than exhaust brakes due to more effective retarding mechanisms of the compression-release process during the compression and expansion strokes. Figure 6.2 shows the retarding power of 9–16 L diesel engines equipped with conventional compression brakes. In 1994, the highest available retarding power was 343 kW (460 hp) at 2100 rpm for the 12–15 L engine market (Freiburg, 1994). There is a growing trend in retarding power density over the years through design improvement. Note that the firing power density is typically about 40 hp/liter at the rated speed (e.g., 1600–2300 rpm) for big bore engines. The retarding power density is definitely short around 1600 rpm but may become capable at 2100 rpm. As the engine rated speed is reduced, there is an increasingly large mismatch between the required firing speed and retarding speed. Engine system design needs to address such a mismatch. Finally, Fig. 6.3 shows there is a large variation in the engine retarding power at a fixed engine displacement due to the variation in brake design details.

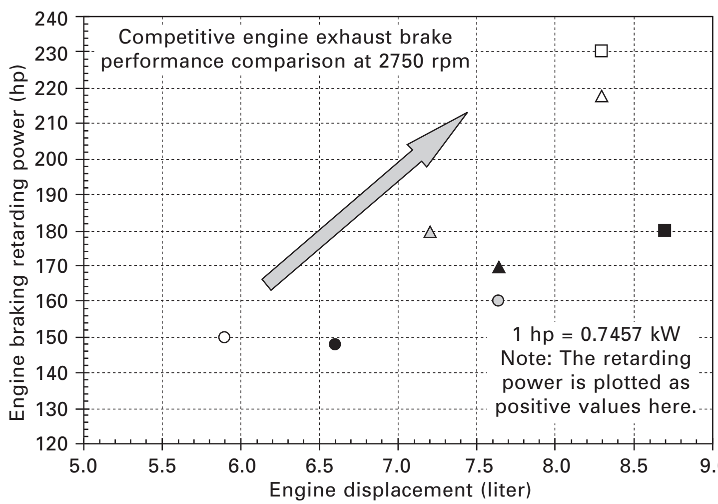
The power-to-weight ratio indicates the compactness of the retarder. The power density of the hydrodynamic retarder is in the order of 17 hp/kg



6.1 Conventional exhaust brake performance.



6.2 Conventional compression-release brake performance.



6.3 Competitive benchmarking analysis on exhaust brake retarding power.

(e.g., Voith Turbo's Aquatarder). The power density of the exhaust brake is around 30 hp/kg at 2000 rpm engine speed. The power-to-weight ratio of the conventional compression brake is around 10–11 hp/kg.

The friction brake usually has very fast transient response and high torque at low speed. The hydrodynamic retarder has a filling time and a delay in response, especially for a large-volume retarder. A small torus volume can improve the response time. The electromagnetic retarder usually has very fast response time. The transient response sensitivity of the hydraulic system in the conventional compression brake can be fairly rapid (Greathouse *et al.*, 1971). The compression brake can achieve full operating status in about 0.2 s and deactivate in less than 0.1 s, as indicated by Morse and Rife (1979). However, turbocharger lag may affect the transient response in terms of the intake manifold boost pressure for the compression brake in turbocharged engines.

Drivetrain retarders and the exhaust brake are very quiet (usually no louder than a plain engine). The conventional compression brake is noisy due to the high-frequency exhaust pressure pulse waves caused by the sudden compression-release process. The bleeder brake is less noisy. Engine brake noise is discussed in more details in Section 6.4.8.

6.1.7 Retarder energy conversion ratios

In drivetrain retarders, all the mechanical power is basically converted to frictional heat and there is a challenge to manage this heat transfer rate. There are two methods to manage the heat: using a waste heat recovery device, or using a cooling system to dissipate the heat to the ambient.

For engine brakes, their energy conversion needs to be analyzed based on the first law of thermodynamics for the engine system. It is well known that the fuel energy basically goes into three parts in engine firing operation based on energy balance: the brake (shaft) work, the exhaust enthalpy, and the coolant heat. Similarly, in engine brake operation, when the fuel energy is zero, according to the engine energy balance the mechanical brake power (i.e., retarding power) needs to be balanced by the exhaust enthalpy rate and the heat rejection to the coolant. In other words, because the engine is essentially an energy conversion device (e.g., as an air compressor in either the compression brake or the exhaust brake), the higher the retarding power absorbed by the engine, the higher the exhaust enthalpy rate and the heat rejection.

The energy conversion of the retarders needs to be evaluated and compared for their effects on engine performance. The two subjects of concern are: (1) the heat rejection and the associated difficulty for the cooling system design; (2) the exhaust gas enthalpy and its impact on aftertreatment operation (e.g., in regeneration) and turbocharger transient operation. It should be noted that the exergy of the coolant heat is so low that it is usually not worth the effort of waste heat recovery.

In order to compare different retarders, two retarder energy conversion ratios are defined as follows. The retarder heat dissipation ratio is defined as

$$\eta_{r,heat} \triangleq \frac{\dot{Q}_{hrj,retarding}}{\dot{Q}_{hrj,firing,max}} \quad 6.8$$

where $\dot{Q}_{hrj,retarding}$ is the unrecoverable heat dissipation to the cooling system during the retarder operation, and $\dot{Q}_{hrj,firing,max}$ is the maximum engine coolant heat rejection during the firing operation.

The retarder available exhaust energy ratio is defined as

$$\eta_{r,exh} \triangleq \frac{\dot{H}_{exh,retarding}}{\dot{Q}_{exh,firing,max}} \quad 6.9$$

where $\dot{H}_{exh,retarding}$ is the exhaust enthalpy flow rate during the retarder operation, and $\dot{H}_{exh,firing,max}$ is the maximum exhaust enthalpy flow rate during the firing operation.

Obviously, a lower retarder heat dissipation ratio less than 1 and a high retarder available exhaust energy ratio are desirable for any retarders. The $\eta_{r,heat}$ of drivetrain retarders is usually close to or much greater than 1, which imposes a significant design challenge for the vehicle cooling system. The $\eta_{r,heat}$ of engine brakes is usually much less than 1. Only when the engine brake's retarding power is several times higher than the engine rated power is its heat dissipation ratio close to 1.

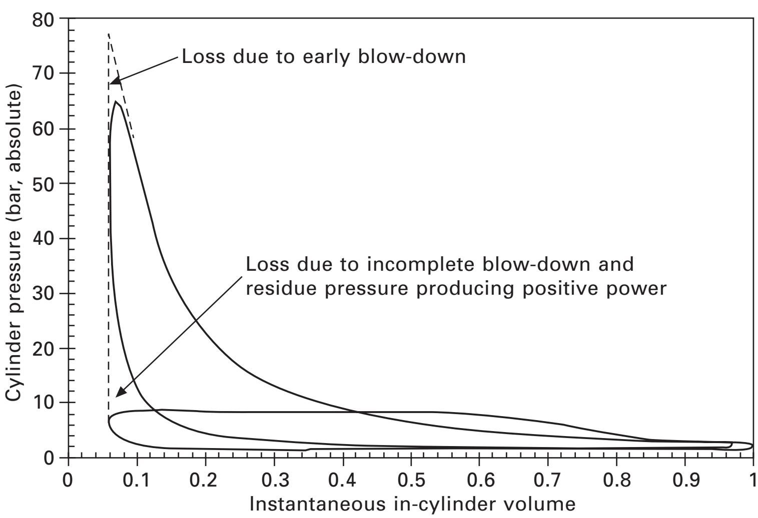
The $\eta_{r,exh}$ of drivetrain retarders is very low because only the exhaust heat from engine motoring contributes to the $\eta_{r,exh}$. This means during a long downgrade the exhaust temperature available for engine aftertreatment and turbocharger operations is very low. In contrast, the $\eta_{r,exh}$ of engine brakes is usually greater than or close to 1 at the full retarding power. The exhaust flow in engine braking contains a significant amount of energy to keep the aftertreatment and the turbocharger at good operating conditions.

Therefore, it is clear that engine brakes are far superior to drivetrain retarders from a thermodynamic perspective. This is especially true if the comparison between the retarders is examined from the second law of thermodynamics. Drivetrain retarders convert all the mechanical energy of the vehicle motion to frictional heat, which has a very low availability or exergy. Engine brakes convert the mechanical energy of the vehicle to useable exhaust energy, which has a high availability.

6.1.8 Retarding process efficiency of compression brakes

The research on the retarding process efficiency of compression brakes was started by Israel and Hu (1993) when they analyzed the performance sensitivity of the Jacobs compression brake to the braking system compliance at different ambient conditions. Later, Hu *et al.* (1997a) applied retarding process efficiency to the study of a VVA compression brake to try to quantify its performance benefit and compare it with the conventional compression brake. They pointed out that the maximum retarding power available comes from an isentropic compression followed by an instantaneous sudden release of the compressed cylinder charge at the TDC. In a real system, inevitable losses are caused by all the factors contributing to the insufficient compression and the non-instantaneous release of the cylinder charge (Fig. 6.4). Some examples are as follows:

- There are heat losses during the compression stroke.
- There is a leakage or insufficient compression near the end of the compression process caused by the opening of the braking valve. The braking valve needs to be opened before the TDC because it cannot open instantaneously with a sufficiently large flow area right at the TDC. Note that there is a restriction of valve-to-piston clearance at the TDC. Moreover, the braking valve opening timing is selected to try to limit the maximum cylinder pressure and the associated gas loading acting on the braking assembly from a structural durability point of view.
- The residue cylinder pressure during the expansion stroke caused by incomplete blow-down result in some positive power. This is caused mainly by the restriction of the supersonic flow across the braking valve.
- The hydraulic compliance of the engine brake system may lower the braking valve lift and increase the residue pressure in the expansion stroke.



6.4 Illustration of retarding process efficiency of compression brake.

Israel and Hu (1993) defined the retarding process efficiency of compression brakes as the ratio of the actual retarding power to the isentropic retarding power during the compression and expansion strokes, given by

$$\eta_{r,CR} \triangleq \frac{\dot{W}_{r,actual}}{\dot{W}_{r,isentropic}} \qquad 6.10$$

The pumping loss and mechanical friction are excluded in the efficiency definition of equation 6.10. In order to calculate the retarding process efficiency, the actual retarding power can be computed based on the cylinder p–V diagram or can be measured in the compression and expansion strokes. Israel and Hu (1993) proposed that the isentropic retarding power can be computed by the following equation during the compression stroke, assuming the power is zero during the expansion stroke:

$$\dot{W}_{r,isentropic} = \frac{c_{p,C}}{\kappa_c}(\Omega^{\kappa_c-1} - 1)T_1\left\{\frac{1}{\eta_C}\left[\left(\frac{p_2}{p_1}\right)^{\frac{\kappa_c-1}{\kappa_c}} - 1\right] + 1\right\} \qquad 6.11$$

where p_1 is the compressor inlet pressure, p_2 is the compressor outlet pressure, T_1 is the compressor inlet temperature, η_C is the compressor efficiency, and Ω is the engine compression ratio. It should be noted that equation 6.11 is only valid for a simplified closed thermodynamic system under the assumption that all the engine valves are closed from the BDC to TDC for an isentropic compression. In reality, the intake valve actually closes about 40° crank angle after the BDC. Moreover, equation 6.11 is not applicable for the case of having braking gas recirculation (BGR, to be detailed later) where the exhaust valve is open during the compression stroke and the initial state of the gas at the beginning of the compression is complicated by the BGR gas mixing.

The above retarding process efficiency for the compression brakes is a measure of the perfection of the thermodynamic process. The definition provides a fair basis to compare the performance of a given compression brake at different design or operating conditions (e.g., engine speeds, ambient conditions), or to compare the effectiveness of different types of compression brakes. In order to increase the actual retarding power, either the retarding process efficiency or the ideal isentropic maximum available power needs to be increased. In order to increase the isentropic power, intake manifold pressure or engine compression ratio needs to be increased. This will result in an increase in peak cylinder pressure, which in fact is limited by the maximum gas loading that can be sustained by the braking assembly.

It is difficult to define a similar efficiency for the retarding process in the pumping loss strokes. Recall from Chapter 4 that the pumping loss is characterized by the in-cylinder pressure differential $\Delta p_{cyl} = (p_{ExhaustManifold} - p_{IntakeManifold}) + (\Delta p_{in} + \Delta p_{ex})$ = (engine delta P) + $(\Delta p_{in} + \Delta p_{ex})$. The Δp_{in} is the pressure drop across the flow restrictions of the intake manifold, ports, and valves. The Δp_{ex} is the pressure drop across the flow restrictions of the exhaust valves, ports, and manifold. The pumping loss consists of two parts: engine delta P and those flow restrictions related to the volumetric efficiency. In order to maximize the pumping loss, it is desirable to increase the engine delta P and/or $\Delta p_{in} + \Delta p_{ex}$.

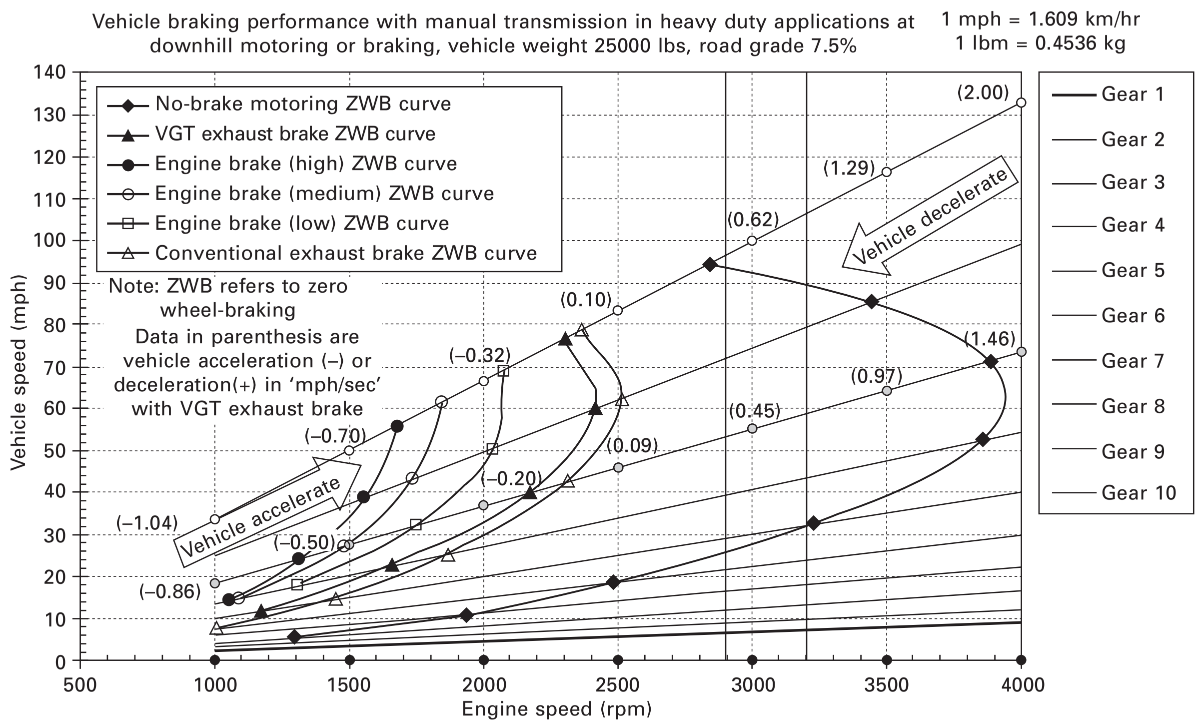
6.1.9 Analysis of engine brake retarding power requirement

In order to determine an appropriate design target for engine brake retarding power, its effect on vehicle braking performance can be analyzed using the method shown in Fig. 6.5 for an example of downgrade driving. The 'high, medium, and low' levels shown in the figure refer to three different design targets for the retarding power of a compression brake. The 'ZWB' refers to 'zero-wheel-braking', which means that the driver does not need to use the service brakes. The ZWB anchor points are calculated using the retarding power of different engine brakes with the vehicle torque balance equation 6.1. The ZWB curves are characteristic curves on which the vehicle travels at different control speeds as an ideal situation. Different vehicle weights, road grades, or tire–road rolling friction coefficients give different ZWB curves. Along each gear line, on the right side of the ZWB point, the vehicle decelerates. On the left side of the ZWB point, the vehicle accelerates. This approach can intuitively analyze the vehicle braking power requirement and determine an appropriate design target for the engine brake retarding power. The analysis is also useful for determining a suitable design target for the maximum allowable operating speed of the valvetrain (i.e., the valvetrain separation speed or no-follow limit) based on driving and braking conditions.

6.1.10 Integrated design of engine brake and powertrain system

Engine brake design integration with the engine

Engine brakes directly control the engine air path with various types of air control valves. Engine brake design is an integral part of the powertrain system design for vehicle safety. It is also an integrated part of engine air flow management for modern diesel engines, not only for braking but also related to firing. In the 1960s–1980s, compression brakes were primarily an aftermarket accessory added to the base engine to control exhaust valve



6.5 Effect of engine brake retarding power on vehicle braking performance.

actuation, installed by vehicle manufacturers or various types of service outlets. Later, the engine companies added compression brakes as a factory installed option (Freiburg, 1994). Exhaust brakes went through a similar history. The integration of a compression brake on an engine requires consideration of turbocharger matching, valvetrain flexibility, braking noise, etc. The integration of an exhaust brake requires consideration of using the exhaust brake as a flexible device for engine warm up, aftertreatment thermal management, and exhaust restriction control for the air/EGR system. Moreover, in the structural design related to the compression brake, the driveline load during retarding can be similar or even higher than that in engine firing operation. Therefore, the vehicle system design should consider retarding as the possible worst or the highest loading case. In summary, engine brake design cannot be isolated to a supplier or restricted to a subsystem. The design direction of engine brakes is moving toward increased integrated functionality to the engine to assist both braking and firing operations.

Retarder design optimization in the braking system

This subsection discusses the optimization criteria for all the retarders in general, including engine brakes. In the optimization of the integrated retarder and service brake system, optimal design can be conducted for either maximum deceleration or maximum brake lining life by an appropriate distribution of the retarding energy absorbed between the retarder and the service brakes (Limpert, 1975). To maximize the deceleration or minimize the stopping distance, the entire brake system needs to be designed such that all the axles are braked simultaneously near their tire–road adhesion limits with a minimum delay. On the other hand, to maximize the service brake lining life, a 'retarder brake-life extension factor' may be considered. O'Day and Bunch (1981) proposed such a factor defined as the ratio of the brake wear rate using the retarder to the brake wear rate without using the retarder. They reported that the range of this factor varied from slightly over 1.0 to a high value of 8.0–9.0, and tended to cluster as functions of retarder type, vehicle size, vehicle application and geographic region of operation. For instance, the factor in transit operations is around 8.0 with the use of electric retarders. Schreck *et al.* (1992) provided a brake lining life comparison between using and not using a ZF hydraulic retarder. Their investigation on 168 heavy vehicles in seven transportation companies showed that the brake lining life with the retarder was on average approximately four times that of the brake lining life without the retarder. Greathouse *et al.* (1971) reported that Mack's Dynatard compression brake reduced the need for service brake usage by as much as 87% in downhill driving, and the brake was activated about once per mile in normal over-the-road driving.

Coordinated engine and transmission controls for engine brakes

There are several system integration issues related to vehicle controls for engine brakes, transmissions, and service brakes. In the engine fueling control, operating the conventional compression brake during fueling would reduce the retarding power and increase the valvetrain loading unacceptably because the braking valve would experience a higher cylinder pressure. Operating the exhaust brake during fueling may result in excessively high soot or smoke due to the choked low air flow rate.

In order to activate the engine brake in retarder–transmission integration, the clutch needs to be engaged, and no fuel needs to be injected into the cylinder (i.e., foot off the pedal). Engine brakes may experience some losses of retarding power in the torque converter of automatic transmissions. Transmission lockup may ensure the minimum retarding power loss for the wheels. Many electronically controlled automatic transmissions can be programmed to downshift when the retarding operation is requested, allowing the optimum engine speed to be selected for the best engine brake performance. If a faster speed than the downhill control speed is desired, a higher gear in the transmission can be selected to reduce the running speed of the engine or the primary retarder, or alternatively a lower retarding power level can be selected if the retarder has progressive/variable levels of power.

The retarding torque of the engine brake transmitted to the driven wheels can be interrupted by disengaging the clutch or by placing the transmission in neutral. As a comparison, in the case of driveline retarders, once the brake is applied it can be disconnected from the driven wheels only through the release of the brake control. If the engine brake is operated with the clutch disengaged, the engine may inadvertently stall to a very low speed because no vehicle load is attached to the engine so that the engine speed will drop very quickly to a low speed and then stall. Stalling the engine is not acceptable for the later firing operation. The engine brake needs to be turned off below a certain engine speed in order to avoid stalling the engine.

Coordinated service brake controls for engine brakes

The retarder is not a substitute for the service brakes. The service brakes need to be used to bring the vehicle to a complete stop. The retarder can be used in conjunction with the service brakes to provide a combined retarding power. The engine brake design and control need to be coordinated with the existing wheel-antilock brake system to ensure a safe and optimal braking performance under all operating conditions. Ideally, the driven wheels need to be kept within the optimum slip range without over-braking and under-braking.

Because the vehicle retarders act only on the driven axle, there is a high risk of over-braking the driven axle to lock it prematurely due to the change in braking force distribution between the axles. The over-braking and the associated directional instability are especially prone to occur with empty vehicles travelling on road surfaces that have low adhesion coefficients. Overuse of a retarder on a slippery road or with empty load can lead to the dangerous 'jackknife' instability for trucks. If there is a tendency of wheel lockup during braking, a fast cut-out of the retarder is immediately needed in order to avoid over-braking. Haiss (1992) indicated that cut-out times of a maximum 0.2 s up to 10% of the residual braking torque pose hardly any problems on braking comfort and directional stability. The antilock brake system may compensate for the altered braking force distribution caused by the retarders. When the antilock system detects a slippage or a wheel lock, it may instantly shut off the retarder in addition to controlling/adjusting the service brakes. A central brake control management system needs to coordinate the operations between the service brakes, the engine brake, the drivetrain retarder, the antilock system, and the coupling force control system.

6.2 Drivetrain retarders

6.2.1 Types of drivetrain retarders

Drivetrain retarders include the following types: (1) the friction brake, i.e., the transmission brake or the clutch brake (Jahn, 1989) and the friction clutch (Spurlin and Trotter, 1982); (2) the hydrodynamic or hydraulic retarder; (3) the electric eddy-current, electromagnetic and permanent-magnet retarder (Kubomiya *et al.*, 1992); or (4) a combination of them.

The driveline retarder can be located before the transmission input (i.e., between the torque converter and the transmission gear range package) or after the transmission output. Different locations result in differences in design packaging, speed–torque capability and braking smoothness. A primary retarder can utilize the transmission gear shift to change its operating speed and braking torque. A secondary retarder operates at the driveline speed and there is no abrupt torque change or torque spike due to transmission gear shift as in the case of a primary retarder. Although the capability of regulating the retarder torque through changing its speed is lost, the secondary retarder may be beneficial in applications where braking smoothness and comfort are the major requirements.

The detailed introduction of drivetrain retarder design, construction, performance and the impact on vehicle performance is elaborated by Göhring *et al.* (1992). An important review of the design and system integration considerations for drivetrain retarders is provided by Haiss (1992). The basic principle of hydraulic retarders is explained by Rao (1968). Klemen *et al.*

(1989) introduce the design approach, control and durability testing for an Allison transmission output retarder.

6.2.2 Torque characteristics of drivetrain retarders

On very steep downgrades (e.g., 10–12%) the vehicle speed has to be maintained low enough to ensure a safe emergency stop. At low vehicle speeds, many drivetrain retarders installed after the transmission output become insufficient due to fast torque drop at the low shaft speed. Schreck *et al*. (1992) indicated that the application of the secondary retarders in 40 ton trucks is limited up to approximately 10% road grade, depending on the axle ratio. Such a capability may cover the majority of European downhill driving applications. It should be noted that the performance of the engine brakes and the primary retarders installed before the transmission are not limited by the low vehicle speed because a low gear in the transmission can be selected to increase the speed of the primary retarder to achieve high power.

The retarding power of the hydrodynamic retarder is a function of the relative speed of the retarder elements. It is important to increase the low-speed torque while keeping the high-speed power below the cooling design limit in retarder design. The hydrodynamic retarder usually has a relatively flat torque curve from the medium speed to the high speed. It is very capable of generating much higher power than the engine firing rated power. At low speeds, the blades stall and the torque is low. The hydrodynamic retarder also has non-negligible power loss at idle running (Jahn, 1989).

The electromagnetic or electric eddy-current retarder may develop higher torque at low speeds than the hydrodynamic retarder. It is suitable for stop-and-go driving applications (e.g., pickup and delivery trucks, refuse trucks, city buses). As pointed out by Göhring *et al.* (1992), the electrodynamic retarder is usually more suitable at low vehicle speeds, while the hydrodynamic retarder is usually more effective at high vehicle speeds. They are both free from wear and very quiet.

Spurlin and Trotter (1982) described a design of an Allison transmission output retarder which combined a friction clutch and a hydraulic retarder so that the low-speed torque was significantly increased by the friction clutch.

6.2.3 Cooling and thermal protection of drivetrain retarders

Usually, air cooling is not sufficient for the hydrodynamic retarder due to its high heat rejection. The cooling medium in the hydrodynamic retarder can be oil or transmission fluid (e.g., the ZF or Allison retarders) or water (e.g., the Voith Turbo's Aquatarder). The vehicle cooling capacity (e.g., the size of the radiator and the fan) often needs to be determined based on the heat rejection requirement in retarding instead of firing for the hydrodynamic retarder. The continuous braking power capability of the hydrodynamic retarder is dependent on the cooling capacity of the vehicle radiator. If the cooling capability is not sufficient, a retarding torque limit based on coolant temperature control is required in order to avoid overheating the coolant. An analysis of the continuous braking performance of a ZF hydraulic retarder was provided by Schreck *et al.* (1992). They addressed the retarder power derating based on coolant temperature control.

The heat generated in the electromagnetic retarder is dissipated to the surrounding air (i.e., ambient-air cooled) and does not need cooling water or oil (Habib, 1992). The electromagnetic retarder is challenging in frictional heat management. If the electromagnetic retarder runs hot in its rotor, its braking torque decreases due to the required thermal protection, for example by cutting off the current in some coils. Therefore, some electric retarders may exhibit certain time-dependent characteristics of retarding power decay due to overheating during a period of braking operation (e.g., a 20-minute braking).

6.3 Exhaust brake performance analysis

6.3.1 Conventional exhaust brake

The conventional exhaust brake uses a shut braking valve (usually a butterfly or sometimes a flat slide throttle valve) in the exhaust pipe mounted normally at the turbine outlet. Mounting the brake at the turbine inlet introduces design challenges such as tight packaging space, interference to the exhaust flow in the exhaust manifold and at the turbine entry, harsh environment due to the higher exhaust temperature, and so on. The braking valve is usually pneumatically actuated by a compressed-air type actuator (e.g., in large vehicles) or a vacuum actuator (e.g., in medium or small vehicles). Sometimes the braking valve is electrically activated by solenoid without the air. The exhaust brake has low weight, small size and low cost with high payback. It has been used as a standard device in many vehicles. Several brands of the exhaust brake exist in the aftermarket, including the Jacobs Exhaust Brake, Pacbrake's exhaust brake, US Gear's D-celerator, and BD's Engine Brake.

As pointed out by Kuwano *et al.* (1983), the popularity of diesel vehicles has been driving the wide usage of the exhaust brake. The exhaust brake is usually used with diesel engines because gasoline engines cannot handle the high exhaust manifold pressure created during braking. Another reason is that hydraulic lash adjusters have been widely used in gasoline engines, and they cannot tolerate the exhaust valve floating or bouncing caused by the exhaust brake operation. The use of the exhaust brake on an engine with hydraulic lash adjusters can cause engine damage due to valve-to-piston contact when the increased exhaust manifold pressure causes the exhaust valve to float off its seat and the associated lifter pumping-up occurs.

Usually, when the engine brake is activated fueling should be terminated. However, if the fuel injection amount is set to idle fueling, a small amount of fuel will be injected into the cylinder to produce some firing power. This is undesirable because it reduces the engine retarding power. Moreover, using the exhaust brake without the interruption of fuel injection will create a large increase in the carbonaceous residue in the combustion chamber and in the engine oil, as well as high smoke due to insufficient air, as reported by Sequeira and Faria (1984).

The exhaust brake is often used together with the glow plug to help speed up the engine warm-up. With the braking valve partially closed in fueling condition, the engine is loaded with higher back pressure and hence has to burn more fuel than in the normal idle mode after the engine start. This can promote the rise of coolant temperature during the warm-up.

Durability issues of the exhaust brake normally include valve sticking, thermal shock and thermal fatigue of the braking valve. The heat-resistant shaft seal of the brake valve and the carbon deposits on the movable parts are sometimes also problematic. The design and construction of the exhaust brake are elaborated by Kuwano *et al.* (1983). Exhaust brake design constraints related to engine gas pressures and temperatures are discussed in Section 6.3.4.

Although their retarding power at high engine speeds may be satisfactory, both the exhaust brake and the compression brake are poor in retarding torque at low engine speeds. The response of the exhaust brake is instantaneous to full retarding power, while a compression brake usually requires some time for the turbocharger to spool up to reach the maximum boost level. For that reason, the exhaust brake is superior in fast changing and repetitive driving cycles since it does not rely on the buildup of turbocharger boost pressure. On the other hand, the compression brake is superior in continuous steady driving conditions such as on long downgrades.

During exhaust braking, the in-cylinder cycle processes in the compression stroke and the expansion stroke are not adiabatic. The resulting indicated retarding power, although small, increases with the engine compression ratio. Pumping loss is the dominant source of the retarding power of the exhaust brake. The pumping loss power is proportional to engine delta P and is also related to the square of the air flow rate at the intake valve and the exhaust valve. The retarding power characteristic of the exhaust brake is essentially a function of engine speed, engine displacement, and exhaust manifold pressure. Unlike the compression brake, the turbocharger has little influence on the retarding power of the conventional exhaust brake because its retarding power comes from the pumping loss strokes instead of the indicated strokes. On the other hand, the use of the exhaust brake at high engine speeds may actually adversely affect the turbocharger performance by reducing the pressure ratio across the turbine. The interaction between the exhaust brake and the compression brake as well as their combined functions for naturally aspirated diesel engines are discussed by Schmitz *et al*. (1992, 1994) and Imai *et al*. (1996).

When a fixed opening of the exhaust braking valve is sized for high engine speeds, the opening size would be too large at low speeds, resulting in low retarding power. Advanced exhaust brake designs employ an electronically-controlled pressure-limiting bypass valve or essentially a variable-geometry braking valve so that the braking valve opening can be adjusted to much smaller at low speeds to build up very high exhaust manifold pressures to produce high power or torque. The valve area is opened more at high speeds to prevent the exhaust manifold pressure from being excessively high. Such a variable-geometry exhaust brake is the modern design direction to enhance the low-speed retarding performance of the exhaust brake.

6.3.2 Variable geometry turbine (VGT) exhaust brake

As use of the variable geometry turbine has gained popularity on diesel engines, using the VGT as a braking device has also become common. By closing the vane opening of the VGT, the turbine speed increases and a high exhaust manifold pressure can be developed. However, the intake manifold boost pressure also increases at the same time, although its increase is less than that of the exhaust manifold pressure. As a result, the engine delta P and the pumping loss increase as the VGT vane is closed. Compared with the conventional exhaust brake, the VGT exhaust brake usually has a lower engine delta P and retarding power, but a much higher turbocharger speed and engine air flow rate. The higher air flow helps reduce the exhaust manifold gas temperature and the thermal load on the components during braking.

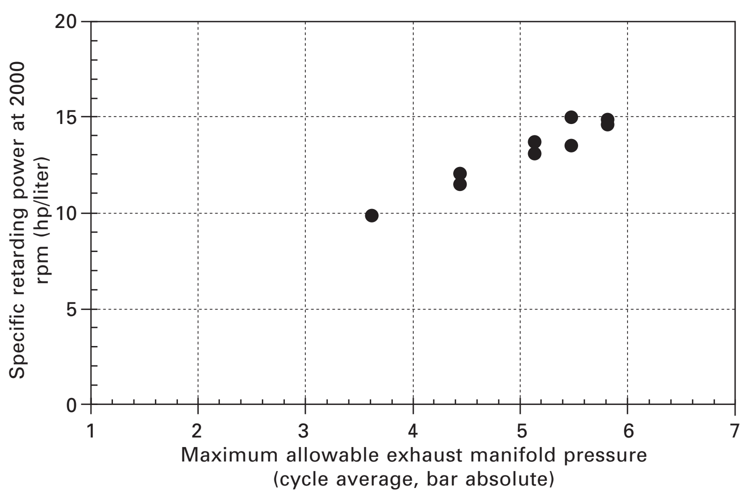
6.3.3 Variable valve actuation (VVA) exhaust brake

Variable valve actuation used as a braking technique provides another means to increase the retarding power. Generally, there are two methods to increase the pumping loss: increase engine delta P and reduce volumetric efficiency. The latter method can be achieved by increasing the pressure drops across the intake and exhaust valves as the air flows through the valves. The VVA brake uses the second method by throttling the engine intake or exhaust valve (or both) to decrease the cylinder pressure during the intake stroke or increase the cylinder pressure during the exhaust stroke. This technique is more effective for the exhaust valves due to the much greater potential. The benefit of an intake VVA brake is small just like the intake-throttle brake.

6.3.4 Effect of exhaust pressure pulses and interaction with valvetrain

Exhaust brake retarding performance

The design constraints of the exhaust brake include: (1) the exhaust gas temperature and the associated component temperatures such as the injector tip temperature at very choked engine air flow rate; and (2) the exhaust manifold pressure and the associated valvetrain dynamics issues. The exhaust brake retarding power is governed mainly by the following factors: engine displacement, engine speed, exhaust manifold configuration, engine firing order, and exhaust valve spring preload. The last three factors are related to the instantaneous exhaust manifold pressure pulses in braking. Figure 6.6 shows that the retarding power density of the conventional exhaust brake is basically linearly proportional to the cycle-average exhaust manifold pressure. The allowable exhaust manifold pressure is limited by excessive exhaust valve floating. If the limit is too low, the performance of the exhaust brake would not be substantial. The maximum allowable exhaust pressure in the non-EGR engines usually varies from 20 psi (1.4 bar) to 60 psi (4.1 bar) gauge pressure, depending on the exhaust valve spring preload. In order to achieve higher retarding power without excessive valve floating, it would require an upgraded (higher) spring preload to increase the allowable pressure limit, for example to 70 psi (4.8 bar). The upgrade needs to be conducted carefully to ensure the exhaust cam stress is still acceptable.



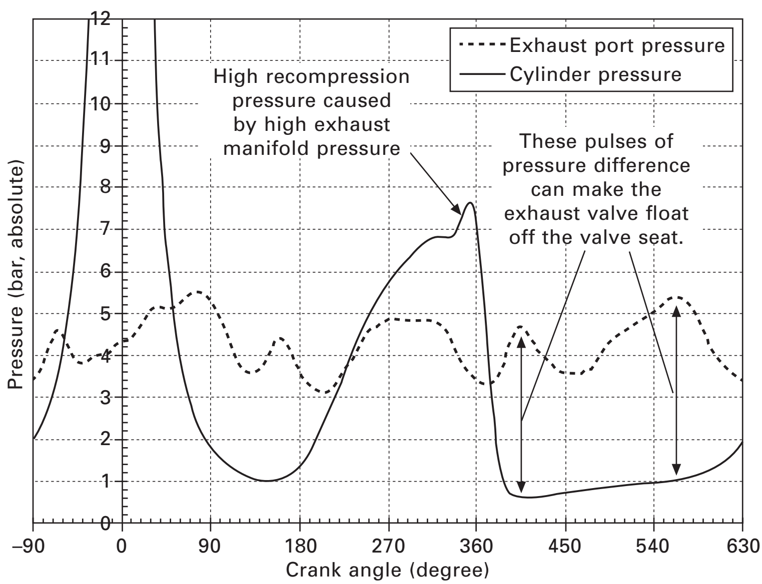
6.6 Exhaust brake performance as a function of maximum allowable exhaust manifold pressure.

Impact of exhaust brake gas loading on valvetrain

The potential impact of the exhaust brake operation on the valvetrain can be summarized as follows on the gas load acting on the valvetrain components:

1. Excessive exhaust valve floating in the intake or compression stroke.
2. Exhaust valvetrain separation near the BDC in the exhaust stroke.
3. High exhaust cam force and stress near the end of the exhaust stroke.
4. High intake cam force at the beginning of the intake stroke.
5. Intake valvetrain separation caused by high recompression pressure.

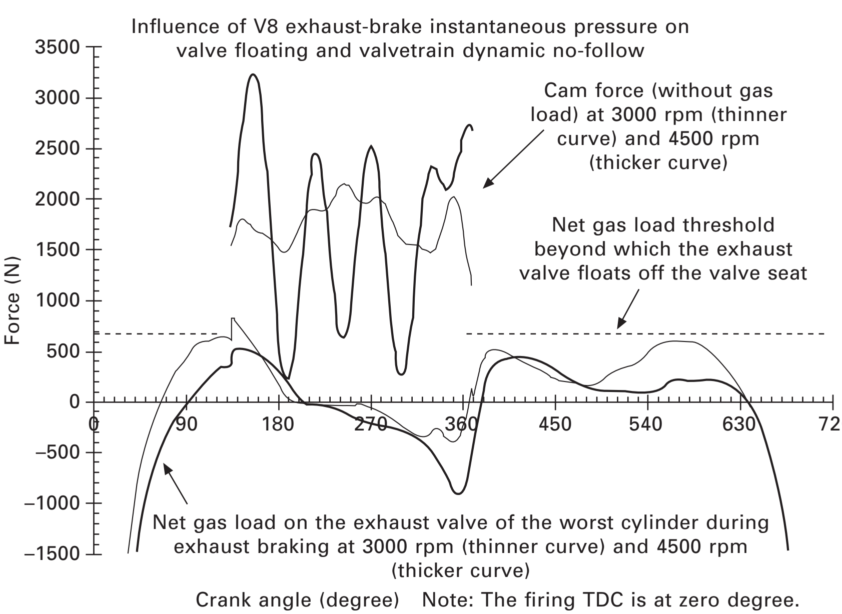
Figure 6.7 shows that there are two exhaust pressure pulses in the early portion of the intake stroke (360–540°) and the compression stroke (540–0°), respectively. These pulses may overcome the cylinder pressure and the exhaust valve spring preload so that the exhaust valve momentarily re-opens and closes abruptly with bouncing and uncontrolled high valve seating velocity, if the exhaust manifold pressure is too high. Moderate exhaust valve floating is acceptable (Schmitz *et al*., 1992), but not excessively large. The valve floating also causes other durability problems such as the fretting of the valve stem, the keeper and the retainer.



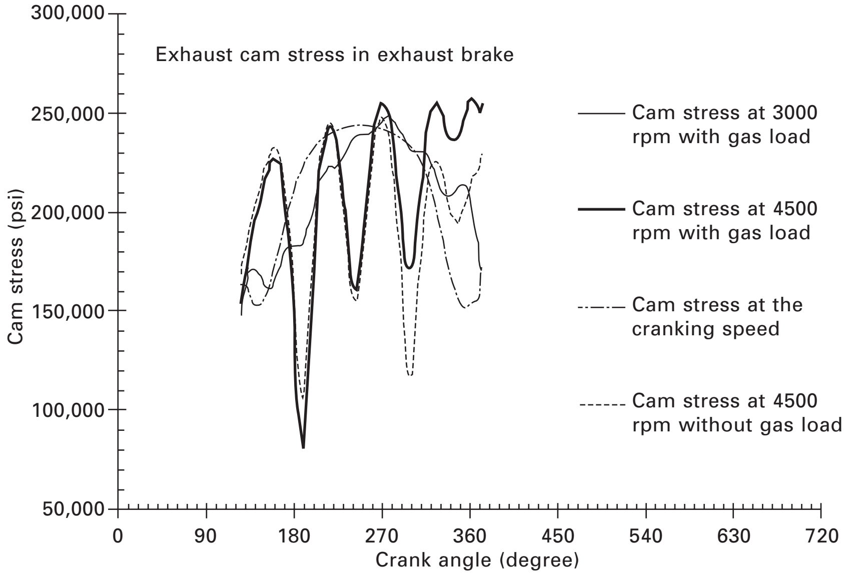
6.7 I6 engine exhaust port pressure pulse with conventional exhaust brake (2600 rpm).

Figure 6.7 also shows that there is a high recompression pressure around the 360° valve overlap TDC during exhaust braking at high engine speeds. The high recompression pressure is caused by high exhaust manifold pressure. The recompression pressure excites the violent vibration of the intake valvetrain so that the intake peak pushrod force or cam force can become excessively high just after the valve overlap TDC, and intake valvetrain separation may occur.

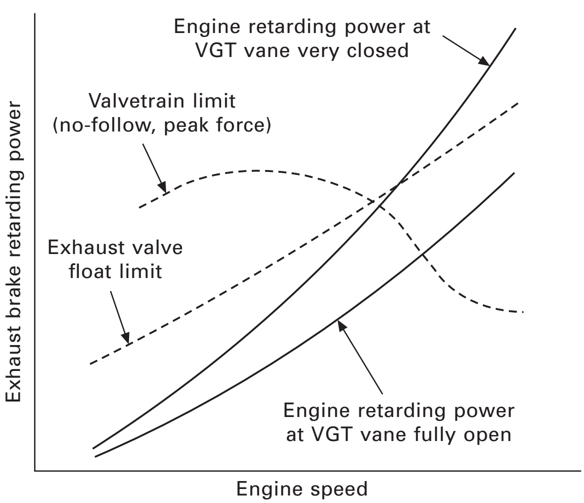
Figures 6.8 shows the net gas force acting on the exhaust valve during one engine cycle of the exhaust brake. The positive net gas force tends to push the exhaust valve open, while the negative net gas force tends to push the valve to close (or to increase the pushrod force). It is observed that during the exhaust cam event, the effect of the net gas loading in exhaust braking is to aggravate the exhaust valvetrain separation at high speeds near the BDC (e.g., at 190° crank angle), and to increase the exhaust cam force and stress at the end of the exhaust stroke. The exhaust cam stress increase in exhaust braking is shown in Fig. 6.9. During the intake and compression strokes, the effect of exhaust braking is to cause the exhaust valve floating, high intake pushrod force or cam force, and intake valvetrain separation induced by the high recompression pressure. The above valvetrain dynamics issues limit the retarding power of the exhaust brake. Figure 6.10 summarizes these constraints imposed by the valvetrain design.



6.8 Impact of V8 exhaust manifold pressure on valvetrain performance with exhaust brake.



6.9 Impact of exhaust braking on cylinder pressure gas loading and exhaust cam stress.



6.10 Engine exhaust brake design constrained by valvetrain limits.

Exhaust valve floating analysis

A detailed analysis of the exhaust valve floating is provided as follows. The exhaust valve floating dynamics can be analyzed based on a force balance in the following:

$$m_{VAL}\ddot{l}_{VAL} = F_{pre} + K_{s,SP}l_{VAL} + (p_{cyl}A_{VAL,cyl} - p_{port}A_{VAL,port}) - F_{VAL,f} \quad 6.12$$

where m_{VAL} is the exhaust valve mass, l_{VAL} is the lift of the floating motion of the valve, F_{pre} is the exhaust valve spring preload, $K_{s,SP}$ is the spring rate (stiffness), p_{cyl} is the in-cylinder pressure, $A_{VAL,cyl}$ is the exhaust valve head area exposed at the in-cylinder side, p_{port} is the exhaust port pressure, $A_{VAL,port}$ is the exhaust valve area exposed at the exhaust port side, and $F_{VAL,f}$ is the valve stem friction force. This force balance of the exhaust valve indicates that the valve will float off its seat once the net gas load overcomes the spring preload.

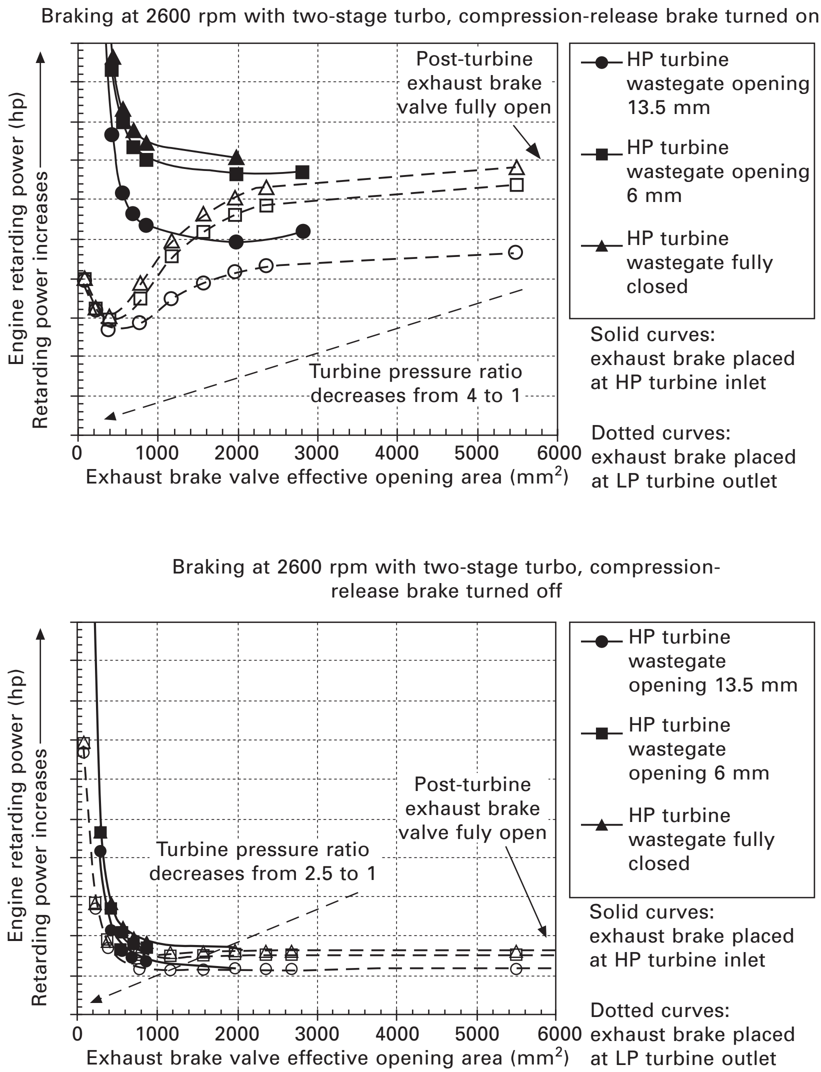
As pointed out by Akiba *et al*. (1981) in their simulation and experimental analysis of the dynamic motion of exhaust valve floating in the exhaust brake operation, the valve floating affects engine cycle simulation results and the predicted exhaust manifold pressure. Therefore, the simulation of the valve floating and bouncing motions should be interactive or coupled in nature between the valve dynamics and the engine cycle simulation.

Exhaust runner or manifold design can reduce the peaks in the exhaust pressure pulses that act on the exhaust valve. The reduced pulses may prevent the large unacceptable valve floating so that a high retarding power can be achieved. Akiba *et al*. (1981) showed that exhaust manifold grouping and the manifold cross-sectional area had a large impact on the retarding power. There was also an effect on the pumping loss due to large cylinder-to-cylinder variations of the exhaust pressure pulses during exhaust braking.

6.3.5 Interaction with turbocharger performance and compression brakes

The instantaneous exhaust pressure pulses are affected by the use of compression brakes. For turbocharged engines at low speeds where the turbine pressure ratio is close to one, or for naturally aspirated engines, the pulsation amplitude of the exhaust pressure pulses can be reduced if a compression brake is used in conjunction with an exhaust brake. In this case, the exhaust braking valve opening can be made smaller to achieve a higher cycle-average exhaust manifold pressure by targeting for the same limit of the instantaneous peak pulse value. As a result, a higher retarding power can be achieved without making the valve floating issue worse. Schmitz *et al*. (1992) reported that the retarding power could be 10–15% higher by using this technique.

The retarding performance of the conventional exhaust brake is also complicated by turbine performance, especially at high engine speeds. When the turbine pressure ratio is much greater than 1, if the exhaust brake is used together with a compression brake at high engine speeds, closing the exhaust braking valve at the turbine outlet actually results in a reduction in turbine pressure ratio and air flow rate. As a consequence, both the engine delta P and the retarding power decrease. When the turbine pressure ratio is closer to 1, closing the exhaust braking valve leads to a monotonic increase in engine delta P and retarding power. Figure 6.11 shows the simulation of such a complex behavior of the exhaust brake and its interaction with the compression brake.



6.11 Interaction between compression brake and conventional exhaust brake.

6.4 Compression-release engine brake performance analysis

6.4.1 Types of compression brakes

The types of compression brakes can be classified according to the following three criteria: (1) the method of how the braking valve motion is controlled (i.e., valve opening and closing timings precisely timed or not); (2) the measures used to activate the valve; and (3) whether the valve actuation is only dedicated to the retarding operation. In the conventional compression brake, the braking valve's opening and closing timings are precisely controlled by mechanical, hydraulic or electromagnetic means and dedicated to the retarding operation. In the bleeder brake, the braking valve continuously opens by a small lift amount during either the entire engine cycle or during the compression, expansion, and exhaust strokes without precisely timed controls for the valve opening and closing timings. In the exhaust-pulse-induced compression brake, the braking valve opening and closing timings are naturally controlled by the available exhaust pressure pulses in the exhaust port inherent in the engine (i.e., by floating the braking valve) rather than controlled by a precisely defined mechanical motion such as in the conventional compression brake. The exhaust-pulse brake can be further classified into turbine-inlet pulse control brake and turbine-outlet pulse control brake. In the VVA compression brake, the braking valve's opening and closing timings are controlled by a VVA device which is not dedicated only to the retarding operation (i.e., it is also used for firing). An example of the conventional brake is the Jake brake. Examples of the bleeder brake are Mercedes-Benz's decompression brake (Schmitz *et al.*, 1992) and Navistar's bleeder brake used on the I6 engines. Examples of the exhaust-pulse brake are Navistar's (including MWM International Motores at Brazil) and MAN's EVBec type of brakes (e.g., Barbieri *et al.*, 2010).

These four types of brakes have their own design characteristics and different constraints. For example, the conventional compression brake may have a fairly high cam force caused by the high peak cylinder pressure near the braking TDC acting on the braking valve. This cam force and the associated cam stress often are the limiting factor for the achievable retarding power. On the other hand, the exhaust-pulse brake does not have this type of loading due to its different hydraulic locking mechanism used to produce the braking valve lift near the TDC. Therefore, it can tolerate a much higher peak cylinder pressure thus producing a much higher retarding power. These different types of brakes also have different noise characteristics due to the difference in the frequency of the exhaust pressure pulses. The conventional brake is very noisy. The bleeder brake and the exhaust-pulse brake are much quieter (e.g., by more than 10–15 dBA).

6.4.2 Principles of compression-release brake design

Fundamental mechanisms of compression braking

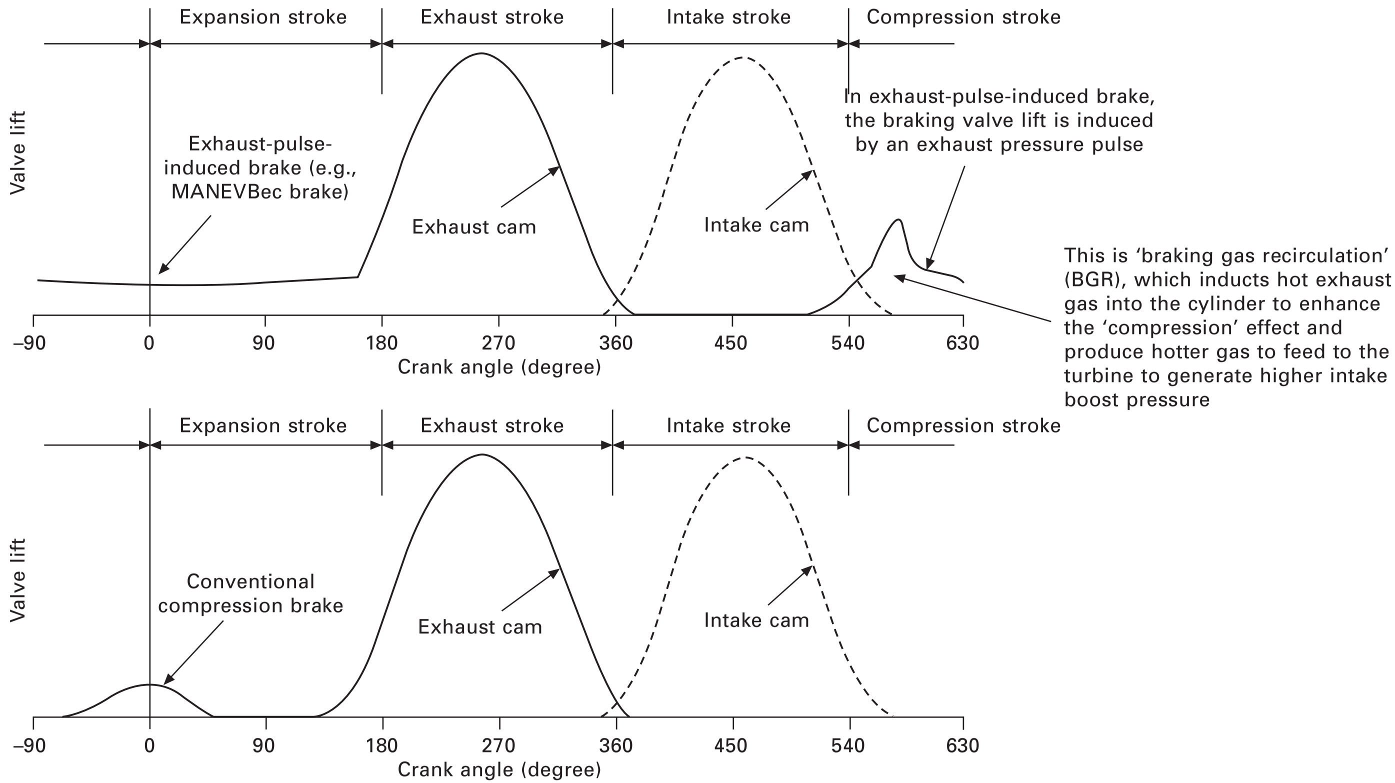
The design objectives for compression-release brakes include achieving high retarding power and low-speed torque, variable power levels, low noise, low weight, low cost, and high reliability while satisfying all the design constraints such as peak cylinder pressure, component loading, exhaust manifold gas temperature, and cylinder head component metal temperatures. Specifically, the major design challenges related to braking performance include the complexity of the following: (1) the valvetrain design for air flow control; (2) the exhaust system design for noise control; and sometimes (3) the mechanical loading control.

Compression brake design and performance have been sporadically reported during the past half a century. The Jake brake design and performance were introduced by Cummins (1966). The relationship between the turbocharger and the compression brake was discussed by Morse and Rife (1979). Braking valve timing/event control for improved retarding performance was addressed by Meistrick (1992). A development process of the compression brake to solve the design challenges at Jacobs Vehicle Equipment Company was described by Freiburg (1994).

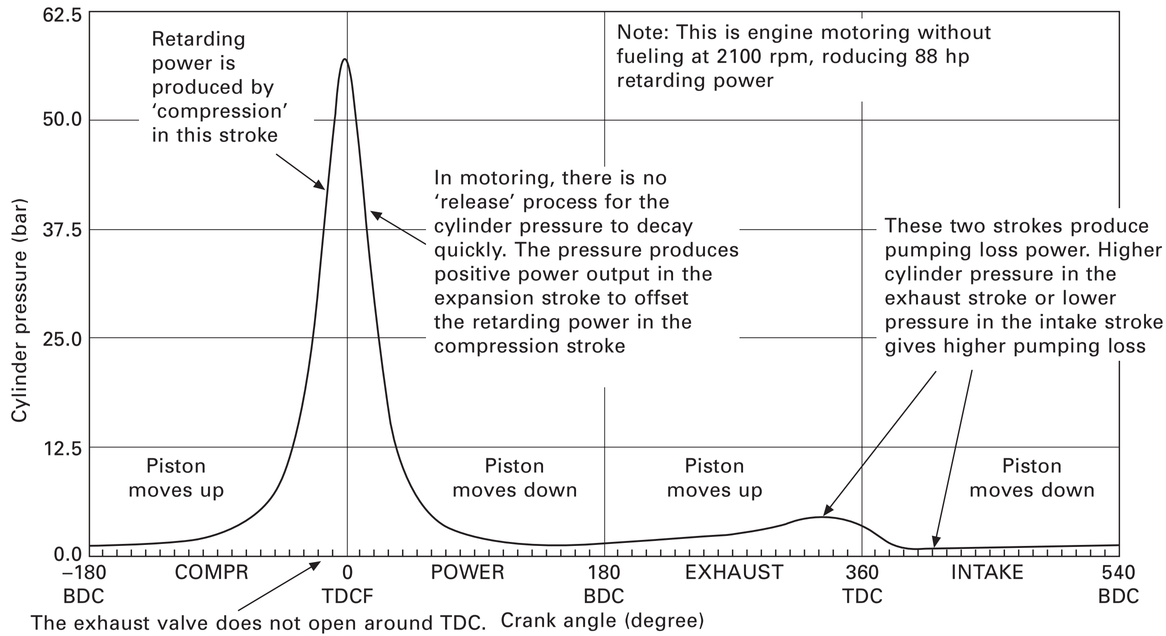
The design principles of compression brakes can best be understood by an in-depth engine cycle simulation since the core of the principles is based on thermodynamic cycles, volumetric efficiency, valve flows, and gas wave dynamics. As mentioned earlier, the retarding power of a compression brake consists of the contributions from the indicated strokes and the pumping loss strokes. In the indicated strokes, the concept of retarding efficiency introduced earlier clearly shows that an increase in the peak cylinder pressure followed by a fast blow-down is highly desirable in order to increase the retarding power. Two fundamental mechanisms should be involved in any high-performance compression brakes. They are the compression-release mechanism and the braking gas recirculation (BGR) mechanism (Fig. 6.12). The compression-release mechanism has been well known since the birth of the first compression-release brake. The efficiency of the compression-release process at a given intake manifold boost pressure level depends on the braking valve events (e.g., opening timing and lift profile). Figures 6.13–6.15 illustrate the compression-release process.

The braking gas recirculation (BGR) theory

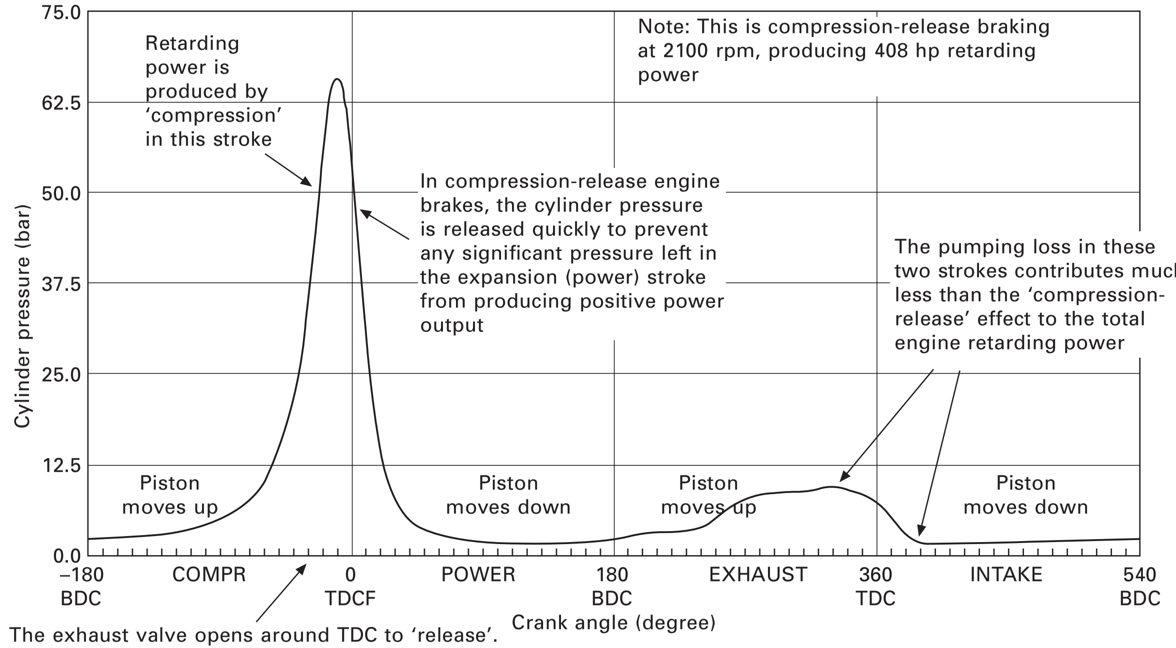
BGR is a relatively new mechanism related to turbocharged engines discovered in recent years by a group of researchers at Jacobs Vehicle Systems and the author at Navistar (Hu *et al*., 1997b; Yang, 2002; Meistrick *et al*., 2004; Xin, 2008). BGR is a technique to supplement the cylinder charge with the



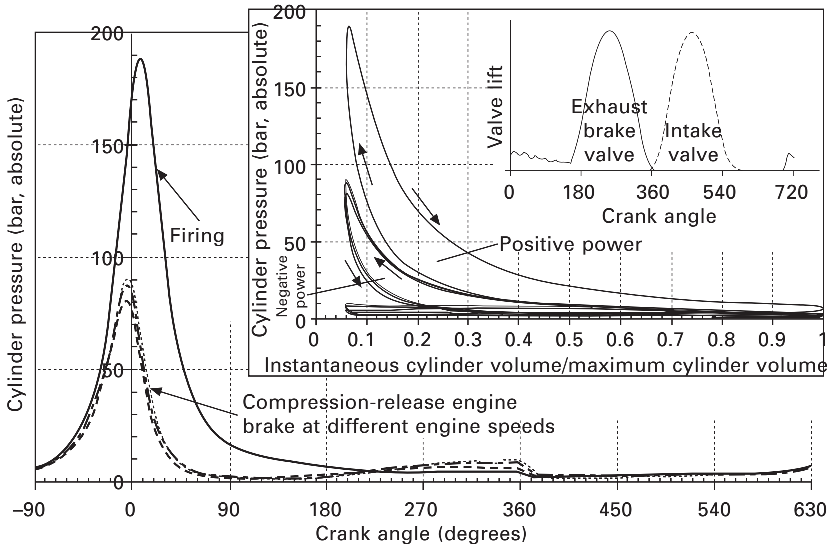
6.12 Mechanisms of compression release engine brakes.



6.13 Cylinder pressure trace and engine retarding power produced by pumping loss in motoring.



6.14 Cylinder pressure trace and engine retarding power produced by compression brake.



6.15 The engine cycle process of compression-release brake.

hot gas recirculated from the exhaust manifold during engine retarding in order to provide higher retarding power. It is essentially an 'internal EGR' technology used in the retarding operation. As shown earlier in Fig. 6.7 in the discussions on the exhaust brake, the engine inherently generates several exhaust pressure pulses in the exhaust port. These pulses try to open the exhaust valve during the intake or compression stroke to flow the exhaust gas into the cylinder. BGR utilizes this phenomenon. In BGR, the braking valve is opened in late intake stroke until the early part of the compression stroke to induct the hot exhaust gas into the cylinder via a pressure differential between the exhaust port pressure pulse and the in-cylinder pressure. The braking valve lift event for BGR can be generated by any of the following: a mechanical cam, a bleeder event, a VVA device, or by the exhaust-pressure-pulse-induced free motion of the braking valve such as in the case of the exhaust-pulse-induced brake. The BGR valve lift event is sometimes also called a 'secondary lift' in braking. BGR can help increase the turbocharger speed and achieve high peak cylinder pressure. In the open literature, the BGR concept has not been fully explored, and its mechanism has not been clearly explained. The BGR theory is developed as follows.

High peak cylinder pressure is the key for high retarding power. The following factors can increase the cylinder pressure: high engine compression ratio, high intake manifold boost pressure, and a large amount of gas mass trapped in the cylinder during the compression stroke. The boost pressure is related to turbocharger operation. The trapped gas mass is related to the valve events used in each stroke and the associated reverse flows (if any) as well as the engine volumetric efficiency. The intake boost pressure can be increased by many design measures (e.g., using higher turbocharger efficiency or a smaller turbine area such as a VGT, or reducing the turbine outlet pressure). One important and effective method of boosting the flow is to increase the turbine inlet temperature. Higher exhaust energy or exhaust manifold gas temperature received by the turbine can increase the turbine speed to deliver high intake boost pressure. The key issue is where to obtain this high exhaust temperature for the turbine. BGR conducts an 'exhaust gas recovery' to harvest the hot exhaust gas mass back into the cylinder, and compresses the hot charge by the piston to even hotter at the TDC, and release it to the turbine. Note that the hot exhaust mass is essentially originated from the moving vehicle's kinetic or potential energy during the braking process when the engine brake works like an air compressor. Transferring this energy from the vehicle to the turbine inlet to speed up the turbine is inherently a very efficient mechanism. With the blow-down process this hot exhaust gas is fed into the turbine as high pressure and high temperature pulse energy flow to increase the turbocharger speed. As a consequence, the boost pressure becomes higher, and the cylinder charge is compressed to an even higher temperature. Such a compounding effect continues until the entire charging process reaches an equilibrium condition (e.g., steady state). The key enabler for this highly effective cylinder-pressure building process is BGR. A higher BGR valve lift event with appropriate BGR valve timing will increase the retarding power significantly. Analysis and experimental work have shown that BGR can increase the retarding power of the compression brakes to extremely high (i.e., much higher than the firing rated power), provided the high cylinder pressure does not create structural design or stress problems on the braking components (depending on the type of the compression brakes). In summary, any efficient compression brake design should be directed toward the two important fundamental braking mechanisms: the compression-release process, and the BGR process.

For four-stroke engines, the right BGR valve lift location is the crank angle durations in the late intake stroke and the early compression stroke where the intake valve is almost closed and the exhaust port pressure is higher than the in-cylinder pressure (i.e., around 500–600° after the firing TDC). For different engines (I4, I6, with divided or undivided turbine entry or exhaust manifold) and at different speeds, the exhaust port pressure pulsation pattern can be different. The optimum BGR valve lift location can change accordingly.

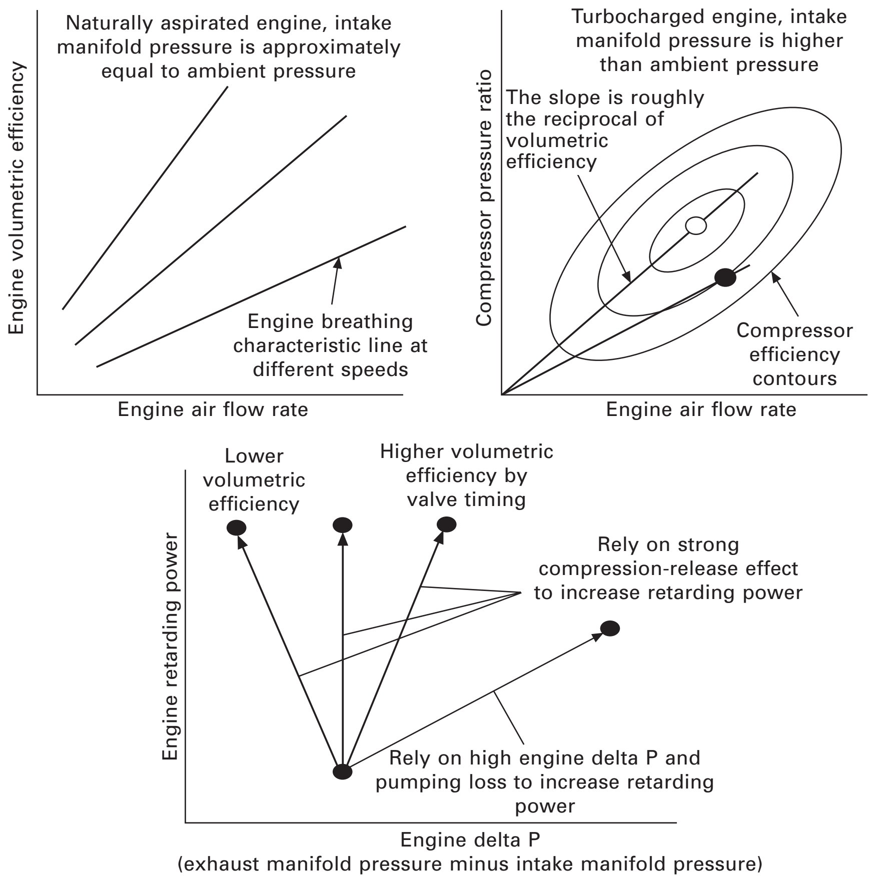
Engine air flow rate affects the in-cylinder gas temperature and the exhaust manifold gas temperature. In general, the lower the air flow rate, the higher the temperatures. A large air flow rate is desirable in engine braking because it cools the cylinder head components, the injector nozzle tip and the exhaust manifold for better durability. It should be noted that increasing the turbine outlet pressure causes a reduction in the turbine power and the air flow rate.

Also note that because the air flow rate is high in BGR operation, the increased exhaust manifold gas temperature is usually still below the allowable temperature limit. In fact, it can be lower than the high exhaust temperatures encountered in the exhaust brake operation where the air flow is choked. The air flow rate and the boost pressure are related via the engine volumetric efficiency. It should be noted the engine volumetric efficiency in compression braking with BGR becomes complex due to the change in the valve events, compared to the case in firing operation. Considering the requirements on both retarding power and component thermal durability, a good design objective for compression brake design can be stated as to achieve simultaneous high engine air flow rate and high exhaust manifold temperature within the design constraints. BGR is the mechanism to achieve this goal in a balanced manner. The ultimate retarding power limit is bounded by allowable peak cylinder pressure and exhaust manifold gas temperature.

Turbocharger matching for engine braking

Turbocharged diesel engines have different retarding power capability compared to the naturally aspirated diesel engines owing to their different levels of peak cylinder pressure and engine delta P. The engine delta P is governed mainly by turbine area and engine air flow rate. The volumetric efficiency is related to valve events. It should be noted that a fixed-geometry turbine or a wastegated turbine selected for good firing operation is usually inferior to a VGT for good retarding power capability. Therefore, both firing and retarding operations need to be considered in engine air system design and turbocharger selection/matching.

Figure 6.16 illustrates the engine breathing characteristics and turbocharger matching. It shows that if the engine volumetric efficiency can be altered by valve events, the turbocharging performance in engine braking can be adjusted for a specific brake design purpose. It also shows an illustration of the methods of improving engine retarding power relative to the critical parameter – the engine delta P. The engine delta P is not only related to the pumping loss and retarding power, but also to the design constraints such as exhaust valve floating, the exhaust valve spring preload, and the exhaust manifold pressure. Different design strategies, either low or high engine delta P, can be selected in engine brake designs. Figure 6.17 summarizes the root causes and the corresponding design solutions for the common problems related to compression brake performance.

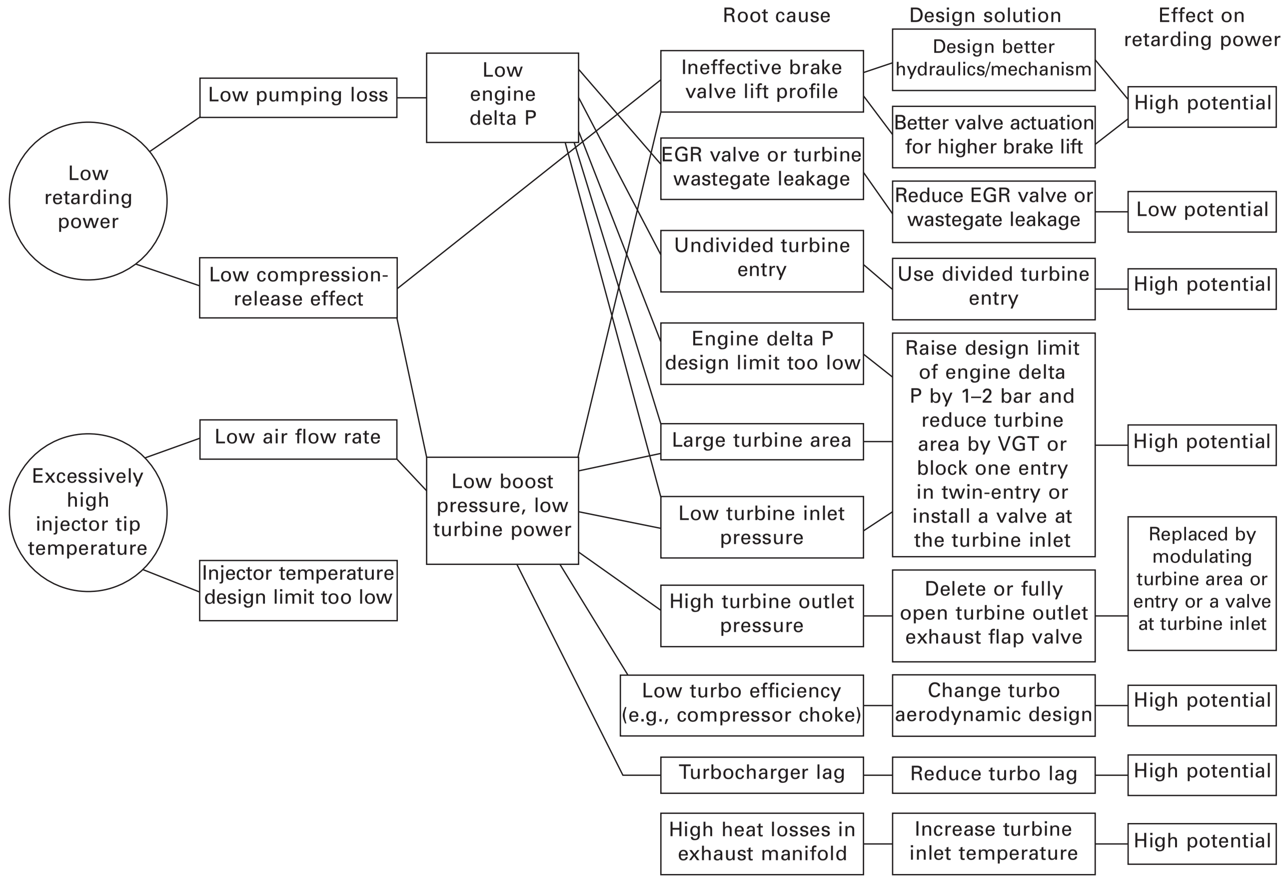


6.16 The principle of compression-release engine brake.

Summary of key design principles for compression brakes

The key elements in a powerful compression brake are summarized as below:

- Achieve high intake manifold boost pressure and peak cylinder pressure. This ensures a powerful compression-release process and a high air flow.
- Do not throttle turbine outlet at high engine speeds where the turbine pressure ratio should be much greater than 1. This ensures a high turbine speed with a large pressure ratio to deliver high intake boost pressure.
- Use BGR. The retarding power of the conventional compression brake is usually limited by the peak cylinder pressure that can be sustained by the maximum allowable braking cam stress, while the exhaust-pulse compression brake does not have such a cam stress limit because its



6.17 Logic flowchart of engine brake performance problems and design solutions.

braking valve lift is generated by the exhaust pressure pulses. Moreover, the exhaust-pulse brake naturally achieves the BGR effect. The maximum lift of the BGR event in the conventional brake is limited by the peak cylinder pressure. The exhaust-pulse brake does not have this limit and is more advanced. The key to make the exhaust-pulse brake successful is to produce all the braking valve lift events by throttling the turbine inlet rather than the turbine outlet. Because both the intake boost pressure and engine delta P can be high in the exhaust-pulse brake, the resulting required high exhaust manifold pressure should be managed carefully.

- Generate a braking valve lift profile around the braking TDC and a BGR valve lift with a cost-effective design without the excessive design constraints of braking cam stress due to the peak cylinder pressure.

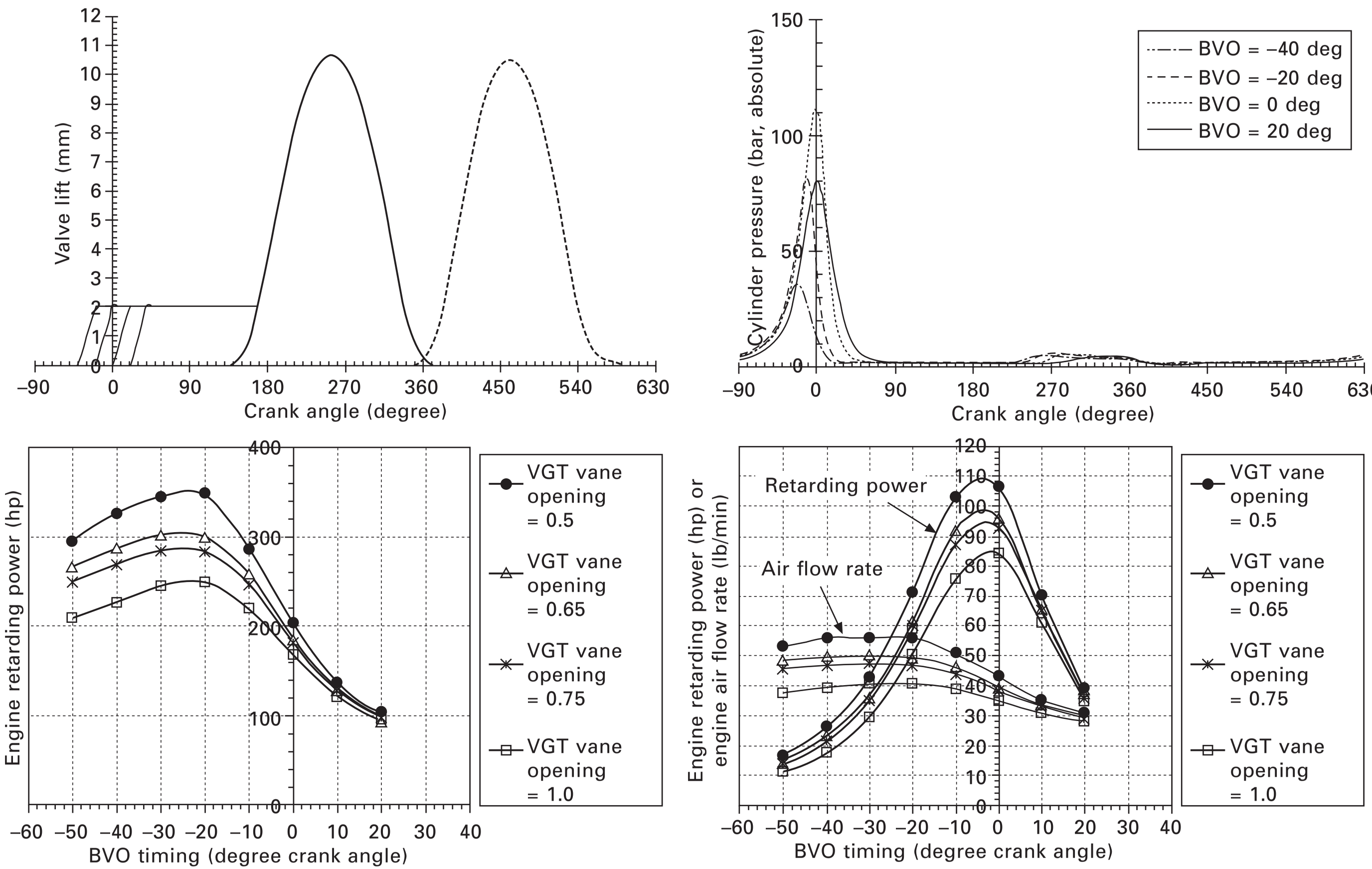
The above theory illustrates that, when turbocharger matching and valvetrain design are conducted, the requirements from engine braking performance also need to be considered. It is emphasized again that the engine performance design must be integrated as a whole system, not only because of the complex interactions between different subsystems, but also for the purpose of coordinating the needs of different applications.

6.4.3 Performance characteristics of compression brakes

The retarding power of compression brakes is sensitive to many design and operating factors including engine speed, total engine displacement in braking (or the number of cylinders), engine compression ratio, braking valve effective flow area, braking valve lift/timing/duration, and turbine area.

The optimum braking valve effective flow area for the maximum retarding power may vary with the engine speed. Usually, a high engine speed and a large air flow rate demand a large valve flow area. Using one valve in braking can reduce the gas load acting on the braking components (e.g., the injector, the exhaust pushrod and cam). In order to provide a sufficient braking valve flow area, one-valve braking needs to have basically twice the valve lift used in two-valve braking, thus demanding a larger valve-to-piston clearance. One-valve braking and two-valve braking designs are elaborated by Price and Meistrick (1983).

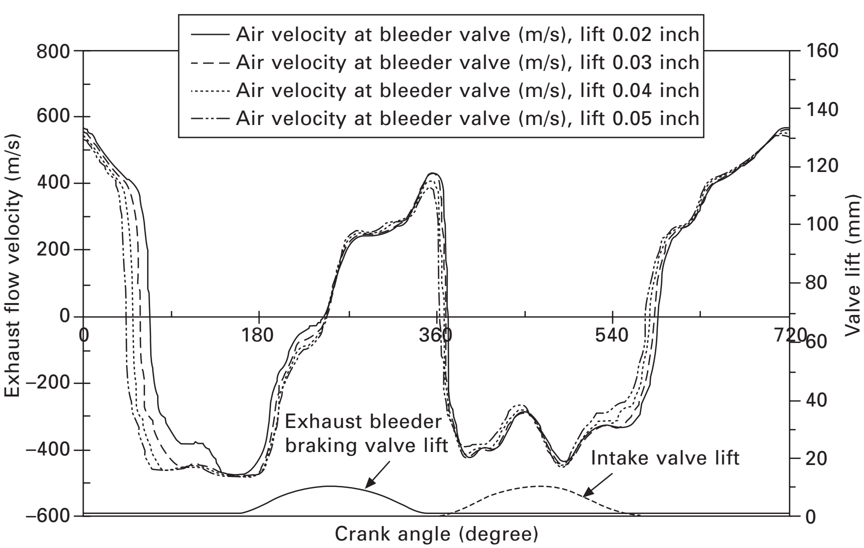
The optimum braking valve opening timing also depends on the engine speed. The optimum timing occurs earlier at higher speeds in order to assure adequate blow-down of the cylinder charge. Figure 6.18 presents the simulation results by using GT-POWER for a conventional compression brake with one-valve braking for an I6 heavy-duty diesel engine. The optimum closing timing (i.e., re-set timing) of the braking valve after the TDC following the



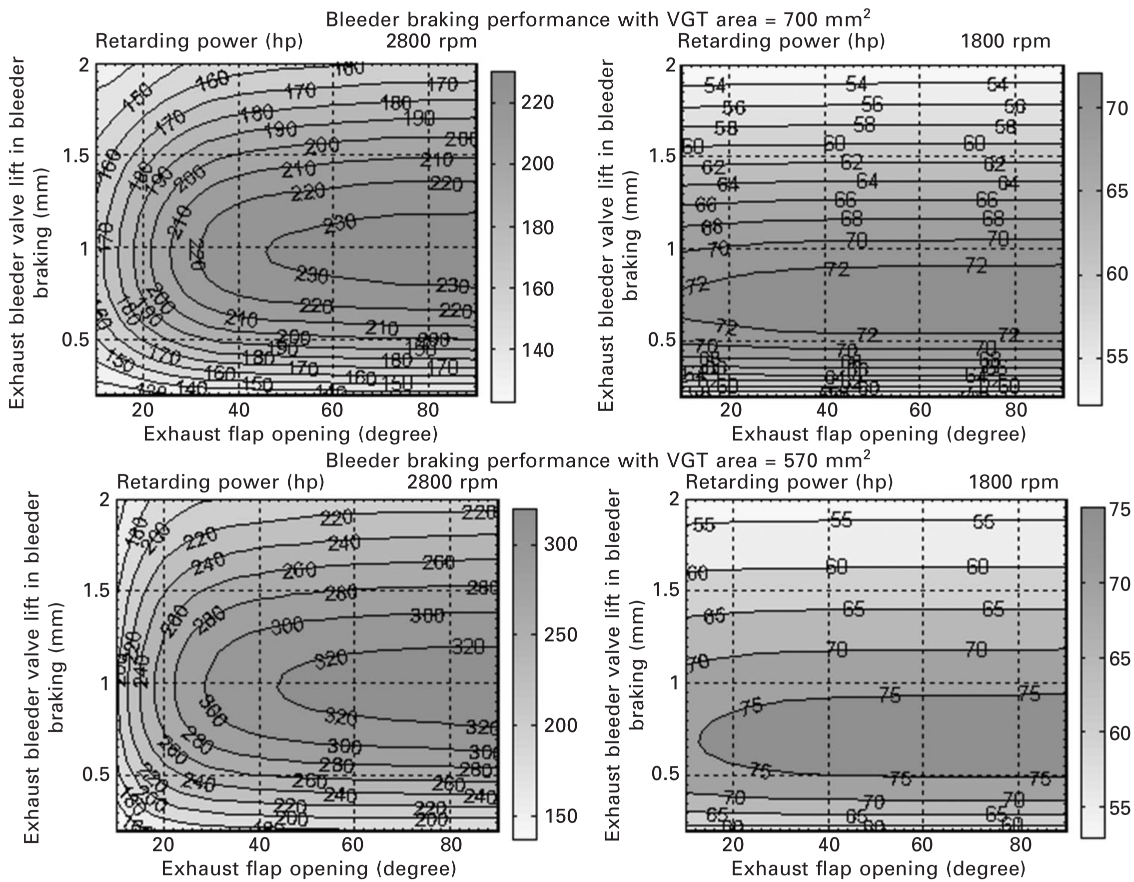
6.18 Conventional compression brake performance sensitivity at 2100 rpm with one-valve braking.

blow-down may yield an additional gain in retarding power because closing the valve as the piston moves down in the expansion stroke may minimize the cylinder pressure (see Fig. 6.19 for the illustration of the instantaneous valve flows in a bleeder brake simulation). Figure 6.20 presents a comprehensive simulation analysis on the effect of bleeder valve lift, turbine-out flap valve opening, VGT area, and engine speed for a bleeder brake. This analysis also reveals the interaction between the compression brake and the exhaust brake.

Superior low-speed retarding torque has been a design challenge. From a vehicle drivability point of view, running at 1500 rpm instead of 2100 rpm during braking is probably easier for the driver because down-shifting is not needed. Low-speed braking can also reduce compression brake noise. Using VGT, BGR or a turbine-inlet exhaust brake may help increase the low-speed torque. At very low engine speeds where the turbine pressure ratio is low or close to 1, a turbine-outlet exhaust brake can be considered. Note again that the throttling effects at the turbine inlet and outlet with an exhaust brake are fundamentally different. Throttling the turbine inlet does not affect the turbine pressure ratio or the turbine speed as drastically as throttling the turbine outlet where the effect of increased exhaust restriction is produced. A careful design balance between the brake, the engine, and the turbocharger is required in order to fulfill all the design requirements. The retarding performance of different compression brakes combined with a variable exhaust brake is discussed by Hu *et al.* (1997b). More information



6.19 Illustration of braking exhaust valve flow velocity in bleeder brake at 2100 rpm.



6.20 Engine bleeder brake performance sensitivity to the effects of bleeder lift, turbine-outlet exhaust flap valve, engine speed, and VGT opening area.

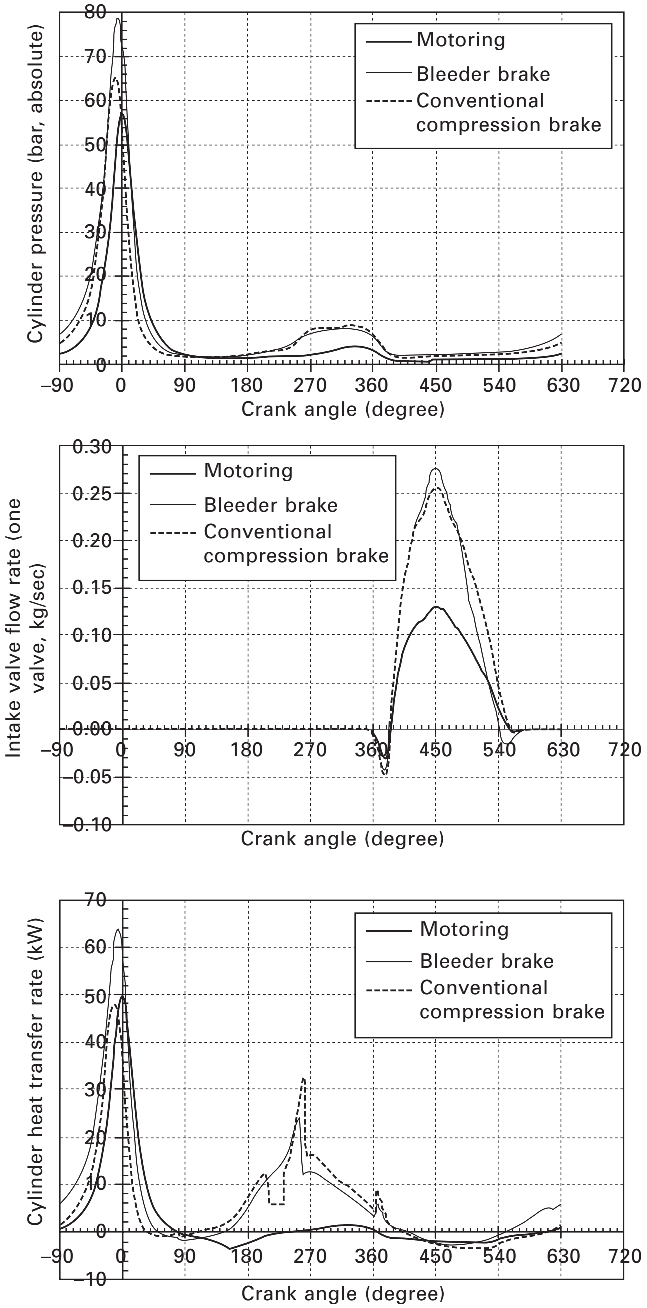
on the design effects on braking performance is provided by Schmitz *et al.* (1992) and Imai *et al.* (1996).

A detailed comparison between engine motoring, the conventional compression brake and the bleeder brake on their instantaneous working processes within an engine cycle is illustrated in Fig. 6.21. The corresponding steady-state performance data are shown in Table 6.1. This example shows that engine cycle simulation plays a very powerful role in understanding the principles of the engine brakes.

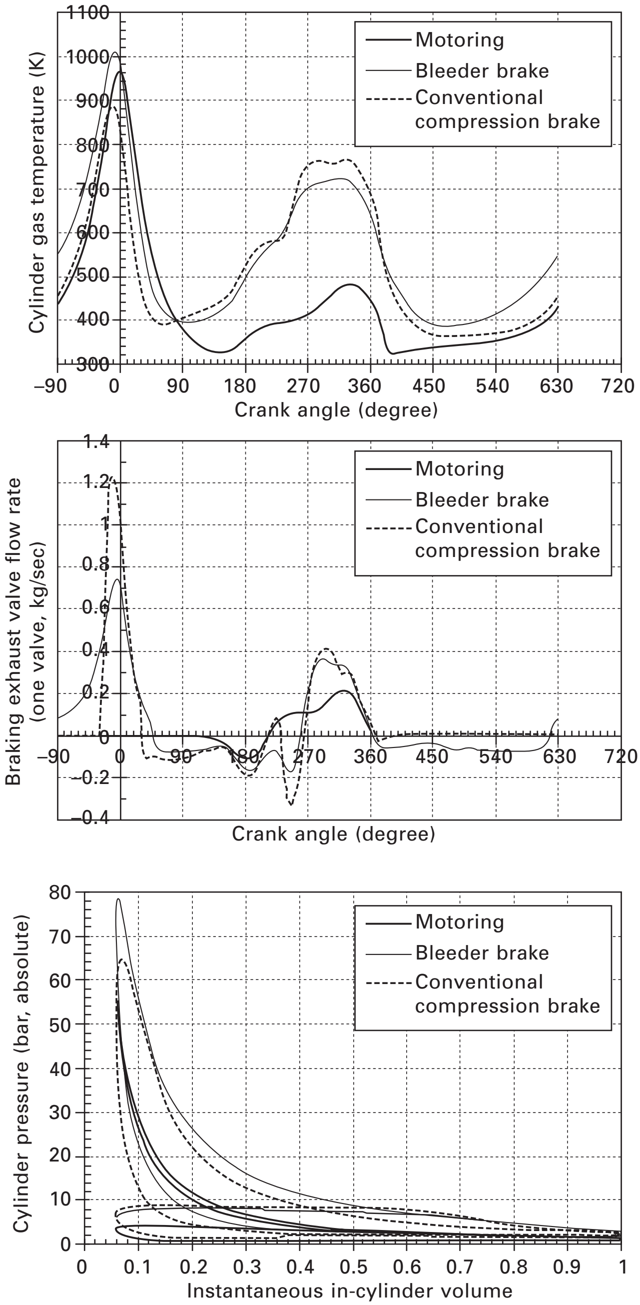
6.4.4 Design constraints of compression brakes

The key durability design constraints of the compression brake are summarized as follows. These design limits need to be examined carefully.

- Peak cylinder gas pressure, which affects engine structural durability.
- In-cylinder component metal temperature, for example the injector nozzle tip temperature. The temperature of the injector tip during engine braking is very different from that during the firing operation where a supply of fuel flow to the injector provides a cooling effect. The injector tip temperature depends directly on the in-cylinder heat flux, which is affected by the engine air flow rate and the compressor boost pressure.
- Exhaust manifold gas temperature, which is affected by BGR and engine air flow rate, which is in turn influenced by the turbocharger performance and the exhaust braking valve opening at the turbine outlet (if any).
- Valvetrain gas loading and component stress. These include the gas loading from the peak cylinder pressure near the braking TDC acting on the braking valve, the pushrod, and the cam. The braking load and the stress on the cam and the rocker arm may exceed the valvetrain loading capability in the firing operation. The loading on the valve system in the compression brake is discussed by Imai *et al.* (1996).
- Engine delta P. This limit is evaluated to prevent excessive exhaust valve floating. It is related to the exhaust spring preload and the associated exhaust cam stress.
- Braking house stress and structural/hydraulic compliance.
- Coolant heat rejection. Although the heat rejection of the compression brakes is far less than that of a hydrodynamic retarder, the coolant heat rejection may become significant at very high retarding power. It is necessary to check the heat rejection to ensure a sufficient cooling capacity.
- All other common design constraints that should be considered in the engine system design for firing operation, for example turbocharger speed, compressor outlet air temperature, etc.



6.21 Performance comparison between conventional compression brake and bleeder brake at 2000 rpm (instantaneous processes).



6.21 Continued

Table 6.1 Retarding performance comparison between conventional compression brake and bleeder brake

	Motoring	Conventional compression brake	Bleeder brake
Engine speed (rpm)	2000	2000	2000
Engine retarding power (hp)	79	385	387
LP-stage turbine outlet pressure (bar, absolute)	1.026	1.111	1.101
Exhaust manifold pressure (bar, absolute)	2.355	5.367	5.020
Intake manifold boost pressure (bar, absolute)	1.199	2.363	2.720
Engine delta P (bar)	1.157	3.004	2.300
Engine air flow rate (lb/min)	29.77	58.81	55.45
Exhaust manifold gas temperature (°F)	338	934	936
Peak cylinder temperature (°F)	1279	1138	1359
Cycle-average cylinder temperature (°F)	334	480	528
One cylinder cycle-average heat transfer rate (kW)	2.65	5.77	7.41

6.4.5 Ambient effect on compression brake performance

The air system theory of turbocharged engine performance is discussed in Chapter 4. Different ambient performance can be explained by the four 'core equations' in Chapter 4. For example, at high altitude, there are basically two effects that can increase the turbine pressure ratio: a lower turbine outlet pressure and a higher turbine inlet temperature. The higher temperature results from the lower air mass flow rate caused by a lower ambient air density. The increased turbine pressure ratio makes the turbine run at a higher shaft speed than at sea-level altitude. This partly compensates for the reduced compressor outlet boost pressure. At a given altitude, hotter ambient air temperature will reduce the engine air mass flow rate, turbine pressure ratio, compressor pressure ratio, and hence intake manifold boost pressure.

Engine retarding power depends largely on the intake manifold boost pressure in the compression-release process. The boost pressure is dependent on the particular turbocharger used. The effects of ambient temperature, altitude, and humidity were analyzed by Israel and Hu (1993). Their results showed that the retarding power decreased slightly as the relative humidity increased. For example, when the relative humidity changed from 0% to 100%, the retarding power only changed by 5 kW from 240 kW to 235 kW at 2100 rpm engine speed. This effect is minor compared to the effects caused by the changes in the ambient temperature and pressure. A formula was proposed by Israel and Hu (1993), ignoring the retarding system compliance resulting from the cylinder pressure, as follows:

$$\varpi_{IMEP,retarding} = -\ 2.9105T_{AMB} + 11.7884p_{AMB} + C \qquad 6.13$$

where $\varpi_{IMEP,retarding}$ is the retarding indicated mean effective pressure (kPa), T_{AMB} is the ambient air temperature (°C), p_{AMB} is the ambient absolute pressure (kPa), and C is a constant (kPa) defining the absolute level of the retarding power at a baseline ambient condition. The retarding power decreases with increasing ambient temperature and altitude. Strictly speaking, this formula is limited to the particular turbocharger that was used to generate their simulation results. Israel and Hu (1993) also concluded that the system compliance in the compression brake, which was affected by the cylinder pressure and ambient conditions, had a significant adverse impact on the braking valve timing and the retarding power, especially at high engine speeds.

6.4.6 Two-stroke braking with compression brakes

As the four-stroke braking power is limited by various design constraints, innovative two-stroke braking may be considered to increase the power without violating the design constraints such as the braking assembly loading. By removing the normal exhaust valve event, the exhaust stroke in the four-stroke engine cycle becomes a second compression stroke from which more retarding power can be obtained via the strong compression-release effect, compared to the low pumping loss during that stroke in the four-stroke braking. Yang (2002) reported that more than 40% braking power increase was achieved by using a two-stroke compression brake compared with the four-stroke brake. More analysis on two-stroke braking is presented in the next section on the VVA compression brake since VVA is a practical mechanism to generate two-stroke valve events.

6.4.7 VVA and camless compression brakes

Usually the performance of the conventional compression brake can only be optimized at a certain engine speed within a narrow range due to the limitation of the mechanical driving mechanisms involved in the braking valve motion control. The VVA compression brake with solenoid controls provides great flexibility to optimize the retarding performance at each engine speed by adjusting the braking valve motion and opening/closing timing so that the retarding process efficiency can be greatly increased across the entire engine speed range. For example, at higher speeds, an advance of the braking valve opening timing is required to achieve optimum blow-down of the cylinder charge and the maximum retarding power. As analyzed by Hu *et al.* (1997a), the fixed braking valve timing results in a sharp drop in retarding efficiency as the engine speed decreases, and an optimized valve timing achieved with a VVA at each speed results in a high and constant retarding efficiency of 84% across the entire engine speed range. The VVA brake can also easily achieve variable retarding power levels to satisfy the different braking power requirements of the vehicle during driving.

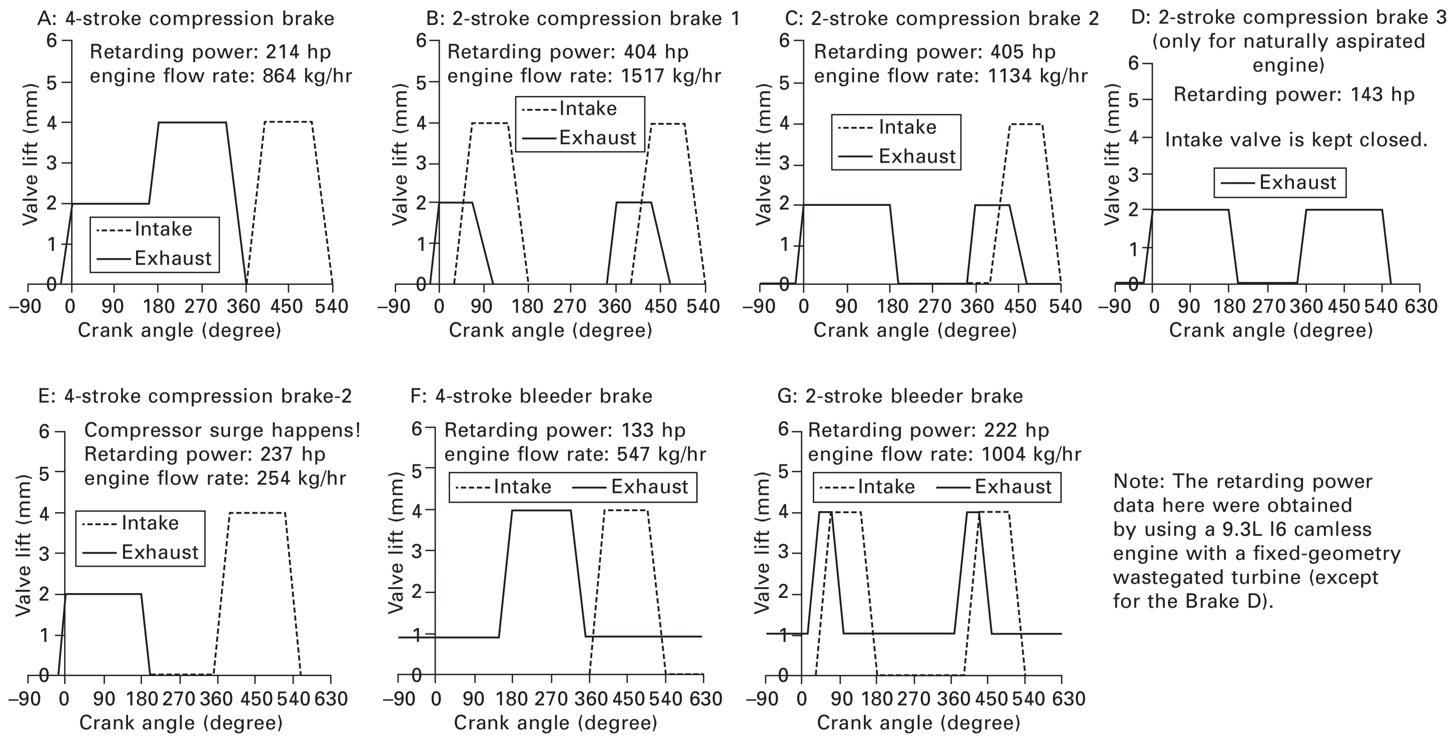
The conventional compression brake usually uses hydraulics added on top of the engine valvetrain and uses the motion of the injector cam or the exhaust cam to open one exhaust valve near the compression (braking) TDC. It introduces extra height and weight. The VVA brake can be much smaller, lighter (e.g., weight reduced by half) and quieter than the conventional brake. VVA brake can be less noisy because it can work like a bleeder brake to flexibly adjust the multiple pieces of the braking valve motion during the engine cycle in order to alter the gas flows through the exhaust valves and the exhaust pressure wave characteristics.

An integrated lost motion variable-valve-timing (VVT) diesel engine retarder was introduced by Hu *et al*. (1997b). A comprehensive simulation of VVA compression brake systems on performance, hydraulics, CFD, and finite element structural analysis was elaborated by Schwoerer *et al*. (2002). Diesel VVA compression brake performance was also investigated by Israel (1998) and Fessler and Genova (2004).

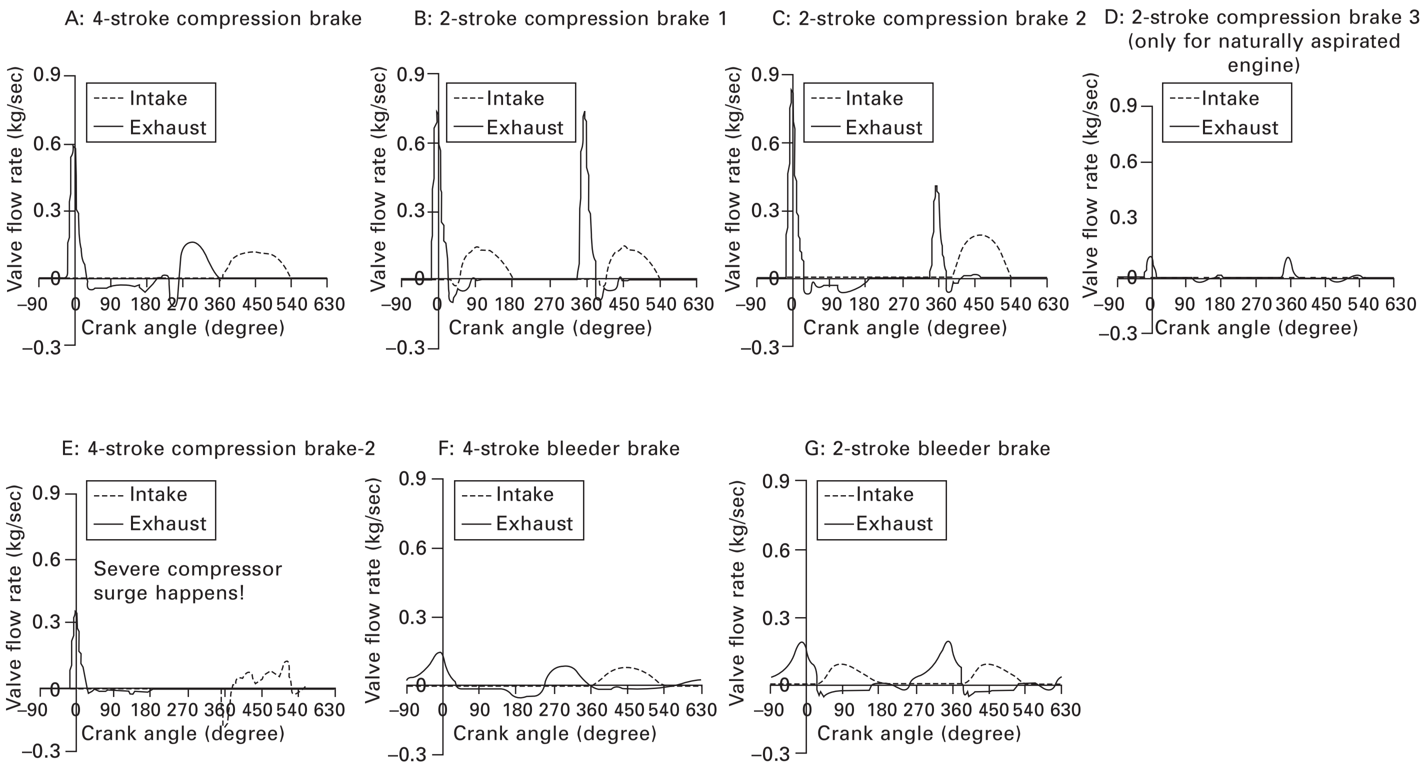
A comprehensive simulation analysis of various VVA or camless compression brakes is presented in Figs 6.22–6.25. Figure 6.22 shows the engine valve lift profiles used in seven different compression brakes, including both four-stroke and two-stroke brakes, as well as their retarding performance summary. Figure 6.23 shows the simulated instantaneous engine intake and exhaust valve flow rates of these brakes. Figure 6.24 shows the engine operating points located on the compressor map for these brakes. Figure 6.25 presents a comparison between the two-stroke and the four-stroke compression brakes. It is observed that turbine area and braking valve opening timing have a significant impact on the retarding power, and the two-stroke brake produces much higher retarding power than the four-stroke brake.

6.4.8 Engine brake noise

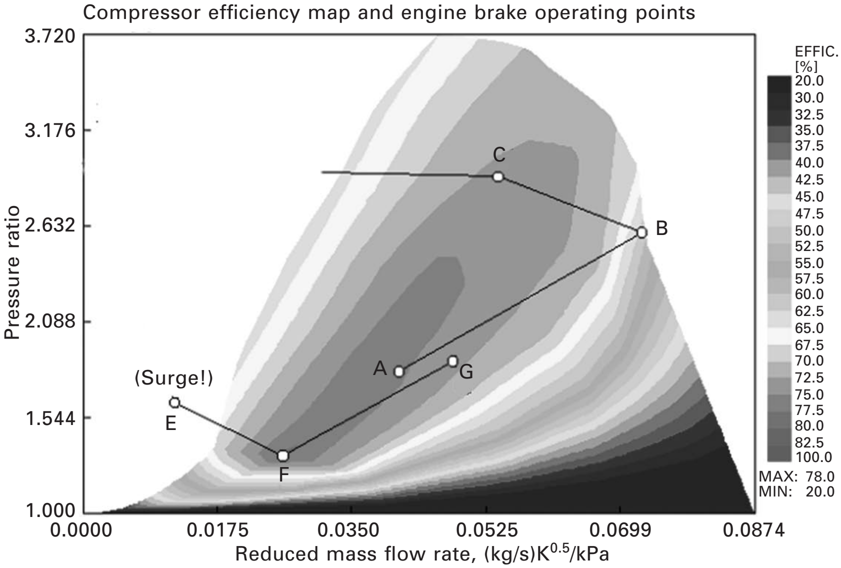
The US federal noise regulations effective since 1978 require that all vehicles meet noise requirements. According to US EPA's Noise Regulation 40 CFR Part 205, new trucks made in the US are required to emit less than 80 dBA of noise at 50-feet-away full-pedal drive-by acceleration. The noise must not exceed 83 dBA in any operating mode under 35 mph, and it cannot exceed 87 dBA over 35 mph. Some trucks with standard OEM exhaust systems may produce higher noise levels (e.g., 83 dBA) when certain types of compression brakes are turned on, compared to their 80 dBA noise during acceleration. A



6.22 HD diesel engine camless engine compression brake concept analysis – valve lift and summary.



6.23 HD diesel engine camless engine compression brake concept analysis – valve flow illustration.



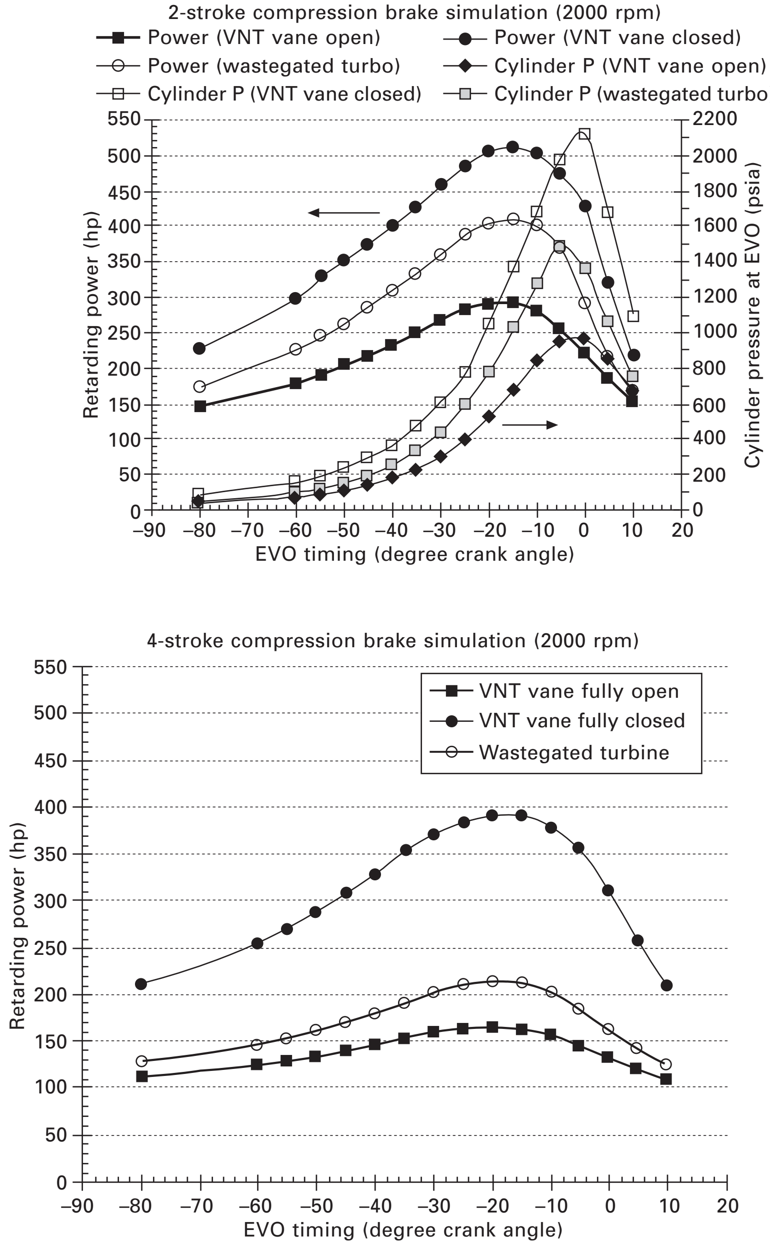
6.24 HD diesel camless engine compression brake concept analysis – compressor operating points.

truck without an exhaust muffler can produce a much higher noise level such as 100 dBA noise when the conventional compression brake is turned on.

In the conventional compression brake, the braking valve opens near the end of the compression stroke where the air is compressed to a very high pressure. The sudden, almost instantaneous, release of the air creates explosive pressure pulses in the exhaust manifold, causing a distinctive staccato (loud popping) braking noise. This effect can be seen from the simulation result in Fig. 6.26 on the exhaust port pressure pulse waves. It is observed from Fig. 6.26 that the sharp high-frequency pulse at 10° crank angle before the braking TDC (i.e., –10°) is the reason for the loud noise in the conventional compression brake. Note that the other sharp pulse wave at 480° crank angle comes from the braking event of another cylinder.

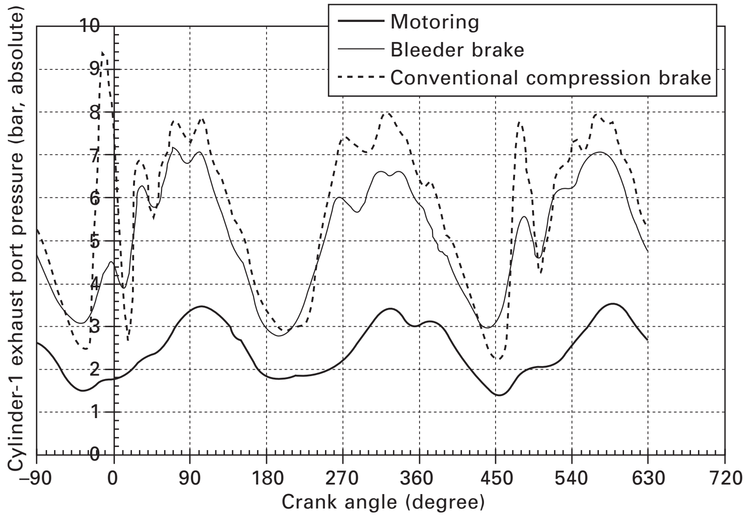
Many truck noise problems on the road are in fact due to defective mufflers or illegal exhaust systems (e.g., an unmuffled straight stack). The braking noise problem is most severe with illegally modified or defective exhaust systems. The braking noise causes many towns to ban the use of the conventional compression brake due to serious concerns for the quality of life in the community. Road signs such as 'Unmuffled engine brake use prohibited except emergency' or simply 'No Engine Brake' are seen, for example, near a downhill section of a fast road. In fact, the compression brake is a valuable legitimate safety device for the vehicle.

Widespread concerns about compression brake noise faced by many local



6.25 Comparison between two-stroke and four-stroke engine compression brake performance of an I6 engine.

legislation bodies for noise abatement indicate that a quiet compression-release engine brake design is highly desirable and will offer market competitiveness. The noise problem can be addressed by more advanced braking mechanism



6.26 Root cause of engine brake noise – simulation of exhaust pressure pulses of different engine brakes (2000 rpm).

and/or effective exhaust muffling. The advanced braking mechanism is preferred since it controls the noise problem at the source by utilizing better braking valve events that favorably affect the exhaust gas wave dynamics in the exhaust manifold and the tailpipe.

Different types of compression brakes produce different levels of noise. For example, a bleeder brake is quieter than a conventional compression brake without BGR (e.g., lower by 15 dBA at high engine speeds and 10 dBA at lower speeds). Bleeder or BGR may significantly reduce the braking noise due to its different exhaust flow and pressure pulse characteristics (Fig. 6.26). The noise of exhaust-pulse compression brake stays between the conventional brake and the bleeder brake, depending on its particular braking valve lift profiles. The noise of the conventional compression brake decreases as the engine speed decreases. The speed-dependency of the noise for the bleeder brake is much weaker (i.e., basically equally quiet at all speeds). Progressive braking using fewer cylinders also reduces the noise. It should be noted that the exhaust brake and the driveline brakes are nearly silent, and their noise levels are usually no louder than the plain engine noise. Schmitz *et al.* (1992) confirmed that their bleeder brake was more silent than the conventional brake. Schmitz *et al.* (1994) also reported that the use of a small decompression valve and throttling by a narrow overflow duct connected to the exhaust manifold reduced the compression brake noise, compared to using the larger exhaust valve as the braking valve in the compression brake.

In general, turbocharged diesel engines produce lower exhaust noise than non-turbocharged diesel engines. Compression brake noise is essentially a part of the exhaust noise and it can be controlled by a functioning muffler. Using proper mufflers with regular maintenance and inspections is important for controlling the compression brake noise. Advanced high-performance mufflers can be used with the compression brakes to reduce the noise without increasing the back pressure or losing fuel economy. Such a muffler can also reduce the acceleration noise. Donaldson's muffler, developed in cooperation with Jacobs Vehicle Systems, was reported to be able to reduce the sound pressure level of the exhaust noise by half during braking compared to a conventional muffler. More detailed studies on diesel engine brake noise are addressed by Reinhart (1991), Reinhart and Wahl (1997), and Wahl and Reinhart (1997).

6.4.9 Powertrain controls of compression brakes

Vehicle braking efficiency and stability control are very important for driving safety. The controls of a central braking management system were described by Göhring *et al.* (1992). Engine brake control with an automatic transmission and advanced estimates of vehicle weight and desirable deceleration were studied by Murahashi *et al.* (2006). Powertrain control of the compression brakes and controller design were extensively studied by a group of researchers at the University of Michigan. An adaptive controller design for a continuously variable compression brake to control the braking valve opening timing was proposed by Druzhinina *et al.* (2000). Their study was to ensure good vehicle speed tracking performance under large variations of vehicle weight and road grade. Engine brake control models and a PI controller for a variable compression brake actuator and the service brakes were developed by Moklegaard *et al.* (2001). A vehicle speed control study was conducted by Druzhinina and Stefanopoulou (2002) to address the coordinated controls on the demands between the service brakes and the compression brake by a PID closed-loop braking controller. Nonlinear controllers for vehicle speed control with a variable compression brake were developed by Druzhinina *et al.* (2002a) to coordinate the compression braking with gear selection and service brakes. Finally, Druzhinina *et al.* (2002b) summarized their study of adaptive controls for the variable compression brake to ensure good vehicle speed tracking performance under large variations of vehicle weight and road grade.

6.5 References and bibliography

Akiba K, Ohtani M and Yoshiki H (1981), 'The optimized design of the exhaust brake of the automotive diesel engine', SAE paper 810344.

Barbieri F A A, Andreatta E C, Argachoy C and Brandao H (2010), 'Decompression engine brake modeling and design for diesel engine application', SAE paper 2010-01-1531.
Cummins C L (1959), 'Diesel engine brake control', US Patent No. 2,876,876, March 10.
Cummins D D (1966), 'The Jacobs engine brake application and performance', SAE paper 660740.
Darragh C T (1974), 'BrakeSaver – Caterpillar's highway truck retarder', SAE paper 741129.
Druzhinina M and Stefanopoulou A G (2002), 'Speed control experiments for commercial heavy vehicles with coordinated friction and engine compression brakes', *Proceedings of the American Control Conference*, Anchorage, AK, May 8–10, pp. 2546–2551.
Druzhinina M, Moklegaard L and Stefanopoulou A G (2000), 'Compression braking control for heavy-duty vehicles', *Proceedings of the American Control Conference*, Chicago, IL, June, pp. 2543–2547.
Druzhinina M, Stefanopoulou A G and Moklegaard L (2002a), 'Speed gradient approach to longitudinal control of heavy-duty vehicles equipped with variable compression brake', *IEEE Transactions on Control Systems Technology*, 10(2), 209–220.
Druzhinina M, Stefanopoulou A and Moklegaard L (2002b), 'Adaptive continuously variable compression braking control for heavy-duty vehicles', *Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control*, 124, 406–414.
'Engine braking', *Truck Engineering*, November 1992, pp. 11–15.
Fancher P S and Radlinski R W (1983), 'Directional control of retarder-equipped heavy trucks operating on slippery surfaces', SAE paper 831788.
Fessler H and Genova M (2004), 'An electro-hydraulic "lost motion" VVA system for a 3.0 liter diesel engine', SAE paper 2004-01-3018.
Freiburg K E (1994), 'Design and development of the Jacobs model 340A engine retarder using concurrent development', SAE paper 940585.
Freund D M (2007), *Foundations of Commercial Vehicle Safety: Laws, Regulations and Standards*, the 52nd L. Ray Buckendale Lecture, SAE paper 2007-01-4298, SAE International, Warrendale, PA.
Göhring E, Glasner E C and Povel R (1992), 'Engine braking systems and retarders – an overview from a European standpoint', SAE paper 922451.
Greathouse J F, Pekar F J and Gibson R B (1971), 'The Mack Maxidyne, ENDT865 diesel with Dynatard engine brake', SAE paper 710557.
Habib G (1992), 'The present status of electro-magnetic retarders in commercial vehicles', SAE paper 922450.
Haiss J (1992), 'Demand criteria on retarders', SAE paper 922453.
Hu H, Israel M A and Vorih J M (1997a), 'Variable valve actuation and diesel engine retarding performance', SAE paper 970342.
Hu H, Vorih J M and Israel M A (1997b), 'The integrated lost motion VVT diesel engine retarder', SAE paper 973180.
Imai Y, Torii A, Kobayashi N, Miura Y and Maeda S (1996), 'Development of a new decompression brake for heavy duty diesel engine', *JSAE Review*, 17, 213–217, JSAE paper 9630381.
Israel M A (1998), 'Jacobs variable valve actuation system for diesel powering and retarding', *GT-Suite Users' Conference Proceedings*, November.
Israel M A and Hu H (1993), 'Impact of ambient parameters on operating efficiency of compression release engine retarder systems on heavy duty diesel engines', SAE paper 932972.

Jahn M (1989), 'A contribution on the application of downhill driving brakes', SAE paper 892501.

Kirchensteiner J (1927), 'Motor brake', US Patent No. 1,637,117, July 26.

Klemen D, Martin R K and Trotter J L (1989), 'A new Allison transmission output retarder', SAE paper 892530.

Kubomiya T, Kuwahara T and Araki K (1992), 'Permanent-magnet-type-retarder in commercial vehicles', SAE paper 922455.

Kuwano M, Endoh K and Suzuki I (1983), 'Exhaust brakes for Japanese trucks', SAE paper 831789.

Limpert R (1975), 'An investigation of integrated retarder/foundation brake systems for commercial vehicles', SAE paper 750126.

Limpert R (1992), *Brake Design and Safety*, SAE International, Warrendale, PA.

Meistrick Z (1992), 'Jacobs new engine brake technology', SAE paper 922448.

Meistrick Z, Usko J, Shoyama K, Kijima K, Okazaki T and Maeda Y (2004), 'Integrated internal EGR and compression braking system for Hino's E13C engine', SAE paper 2004-01-1313.

Moklegaard L, Druzhinina M and Stefanopoulou A G (2001), 'Brake valve timing and fuel injection: a unified engine torque actuator for heavy-duty vehicles', *Vehicle System Dynamics*, 36(2–3), 179–201.

Morse W H and Rife J M (1979), 'Compression engine brake performance with turbocharged diesel engines', SAE paper 790769.

Murahashi T, Sonoda S, Yamada N, Katoh H and Hayakawa K (2006), 'Development of engine brake control system for commercial vehicle with 6-speed automatic transmission', SAE paper 2006-01-1674.

O'Day J and Bunch H M (1981), 'Safety and economic benefits of heavy truck retarder use', SAE paper 811260.

Price R B and Meistrick Z S (1983), 'A new breed of engine brake for the Cummins L10 engine', SAE paper 831780.

Radlinski R (1987), 'Braking performance of highway US vehicles', SAE paper 870492.

Radlinski R (1989), 'Heavy vehicle braking – US versus Europe', SAE paper 892504.

Rao N N N (1968), 'The basic theory of hydraulic dynamometers and retarders', SAE paper 680178.

Reinhart T E (1991), 'US vehicle noise regulations and the effects of vehicle condition', SAE paper 912709.

Reinhart T E and Wahl T J (1997), 'Reducing compression brake noise', SAE paper 971870.

SAE Surface Vehicle Recommended Practice J257 (1997), '(R) Brake rating horsepower requirement – commercial vehicles'.

SAE Surface Vehicle Recommended Practice J2458 (1998), 'Exhaust brake dynamometer test and capability rating procedure'.

SAE Surface Vehicle Recommended Practice J1489 (2000), 'Heavy truck and bus retarder downhill performance mapping procedure'.

SAE Surface Vehicle Recommended Practice J1621 (2005), 'Engine retarder dynamometer test and capability rating procedure'.

SAE Surface Vehicle Recommended Practice J2627 (2009), 'Braking system definitions – truck and bus'.

Schmitz T N, Bergmann H and Daeuble H (1992), 'The new Mercedes-Benz engine brake with decompression valve', SAE paper 920086.

Schmitz T N, Holloh K D, Fleckenstein G and Juergens R (1994), 'The new Mercedes-Benz engine brake with pulsed decompression valve – decompression valve engine brake (DVB)', SAE paper 942266.

Schreck H, Kucher H and Reisch B (1992), 'ZF retarder in commercial vehicles', SAE paper 922452.

Schwoerer J, Huang S and Trzaska G (2002), 'Simulation tools for development of advanced engine braking and variable valve actuation systems', SAE paper 2002-01-0375.

Sequeira M M and Faria P P (1984), 'Adverse consequences for engines resulting from the frequent use of exhaust brakes', *Tribology International*, 17(6), 335–340.

Spurlin S F and Trotter J L (1982), 'The Allison transmission output retarder', SAE paper 821277.

Wahl T J and Reinhart T E (1997), 'Developing a test procedure for compression brake noise', SAE paper 972038.

Wong J Y (1993), *Theory of Ground Vehicles*, 2nd edn, John Wiley & Sons, New York.

Xin Q (2008), 'Engine brake concept analysis for improved engine brake performance', Navistar technical report 580.08.2515.

Yang Z (2002), 'A study of two-stroke engine braking for HD diesel engines', SAE paper 2002-01-0487.

7

Combustion, emissions, and calibration for diesel engine system design

Abstract: This chapter establishes the system design logic from emissions to the overall engine system, especially the air system. It links three sequential functions in an engine development cycle: combustion/emissions development, engine system design, and engine calibration. It first introduces combustion and emissions development and the corresponding requirement for air system design, followed by discussions on engine calibration optimization at both steady-state and transients. The emissions behavior is illustrated by using the minimum BSFC (brake specific fuel consumption) optimization contour map. The advanced emissions modeling approach suitable for engine system design is summarized. The concept of virtual calibration is emphasized at the system design stage.

Key words: combustion, emissions, engine system design, calibration, minimum BSFC (brake specific fuel consumption) optimization contour map, heuristic macro-parameter-dependent approach.

7.1 The process from power and emissions requirements to system design

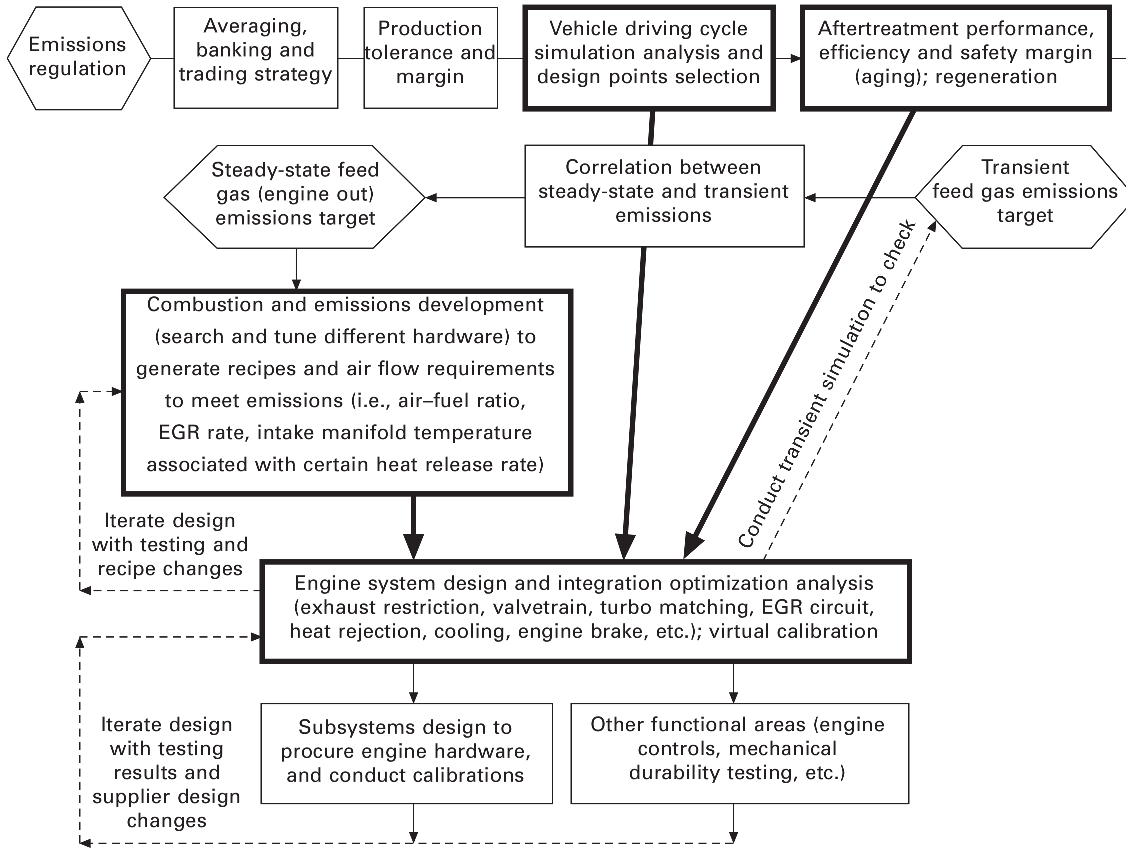
At the beginning of an engine program, performance simulation can be used extensively during the stage of choosing key technologies, for example hybrid powertrain, waste heat recovery, variable valve actuation (VVA), and cylinder deactivation. Once the technology path is defined, engine air system design can start with several possible configuration options, for example high-pressure-loop or low-pressure-loop exhaust gas recirculation (EGR), parallel or series EGR coolers, single-stage or two-stage turbocharger, air-cooled or coolant-cooled charge air cooler. The required input for system performance design is the established target of air flow requirements to meet the emission regulations. The air flow requirements generally include the gravimetric air–fuel ratio, the EGR rate and the intake manifold gas temperature at each critical engine speed and load mode. Those requirements are produced by conducting combustion hardware searching and tuning. The 'searching' refers to screening different sets of hardware (e.g., intake ports, combustion chambers, and fuel injection systems) to determine their ability to meet the emission requirements. 'Tuning' (referred to before the system design) or 'calibration' (referred to after the system design) means any optimization of the tunable parameters with a fixed set of hardware. Some examples of tuning and calibration include calibrating the fuel injection pressure, the injection timing, the VGT vane opening and the EGR valve opening to find the optimum air–fuel ratio and EGR rate for a minimum BSFC while meeting the emission requirements. The 'calibration' links the electronic software controls with the selected hardware for production engines. Engine system design resides in the center of this design process and bridges the upstream combustion development and the downstream production calibration (Fig. 7.1). When a particular set of combustion–fuel–air system hardware is selected, it is important to predict the outcome of engine calibration during the system design stage because any modifications to inadequately designed hardware during the calibration stage are extremely expensive. Engine system design not only produces a specification for each subsystem design, but also generates the simulated virtual performance and emissions regarding engine calibration prior to the actual calibration testing.

The same analysis methodology is used as the optimization techniques in the following three functional areas: combustion hardware screening, system simulation and design, and calibration testing. The analysis method is the design-of-experiments (DoE) combined with the response surface methodology (RSM), as introduced in Chapter 3. A thorough understanding of the hardware screening and calibration processes in the entire engine speed–load domain under various situations (e.g., cold and hot ambient, high altitude, aftertreatment regeneration) is critical for properly translating the emission requirements to the air flow functional target, which is needed by the engine system design as analysis input. This chapter introduces the sensitivity of emissions to the calibration parameters. The method of setting the engine system design target of airflow requirements to meet emission regulations is discussed. In addition, emissions modeling methodology is briefly reviewed.

7.2 Combustion and emissions development

7.2.1 Introduction to diesel emissions control

Compared with the gasoline engine, the diesel engine emits a similar amount of NO_x (engine-out NO_x), much more particulate matter (PM), and much less CO and HC. Improving the air–fuel mixing and combustion processes can reduce PM, but usually results in an increase in NO_x. There exists a fundamental trade-off between NO_x and PM in the diesel engine. The NO_x emissions from the gasoline engine may be reduced by a three-way catalyst aftertreatment device. But currently it is still very costly to use any aftertreatment devices to clean the NO_x from the diesel engine because the NO_x is formed with the air–fuel ratios much greater than those in the gasoline engine, and exists in an oxygen-rich exhaust flow environment. Reducing the engine-out NO_x



7.1 The process from vehicle, aftertreatment, combustion/emissions to system design.

and PM while still maintaining low fuel consumption poses a significant challenge to diesel engine combustion and air system design. Note that NO is largely formed during the premixed lean burn combustion phase with a local excess air–fuel ratio at 1.0–1.2, and soot is mainly formed in the diffusion combustion phase with insufficient air at a local excess air–fuel ratio lower than 0.5–0.6. The heterogeneous local mixing between the air and the fuel during the combustion process is the root cause of the soot emission in the diesel engine. Achieving better mixing and preventing high NO_x and soot is the design goal of the air–fuel–combustion system.

Emissions control is also closely related to mechanical design. A typical example is the lube oil consumption control to reduce the soluble organic fraction (SOF) in PM emissions by using high-tension piston rings or a better control on cylinder bore distortion with improved cooling jacket design. The high-tension piston rings result in increased friction losses and durability problems. Another example is using retarded fuel injection timing to reduce NO_x. This leads to higher soot emission. Faster soot accumulation in the lube oil would force more frequent oil changes for the customer. In fuel system design, the demand of using a higher injection pressure to control PM exerts a higher load on the injector mechanism. In that case, the lubricant oil experiences a higher shear rate, and the oil viscosity improver additive breaks down more quickly. More information on the effect of lubricant oil on diesel emissions can be found in Manni *et al*. (1997), Stunnenberg *et al*. (2001), Barr *et al*. (2003), and McGeehan *et al*. (2006). The effect of EGR on diesel engine oil is discussed by Sasaki *et al*. (1997).

The following topics are important for diesel engine system design:

- emissions certification and the impact on engine performance strategy
- the impact of the conventional and innovative combustion systems on engine system design
- the fuel system, ultra-low-sulfur fuels, and low-ash oils
- the intake manifold gas temperature and condensation controls
- the engine compression ratio
- cold start.

7.2.2 Combustion and emissions research

The following research findings in diesel combustion and emissions are important for engine system design. Automotive test drive cycles for emissions measurement is reviewed by Samuel *et al*. (2002) and Taylor *et al*. (2004). Early design guidelines for low-emission diesel engine design are summarized by Gill (1988). The technical trends of diesel engine designs to meet emission regulations are discussed by Mori (1997), Khair (1997), Browning (1997), Dexter and Kling (1999), Hountalas (2000), Walsh (1999, 2001), Moser *et al*. (2001), Wuensche *et al*. (2003), Pfeifer *et al*. (2003), Leet *et al*. (2004), Charlton (2005), and Dollmeyer *et al.* (2007). Vehicle integration of low emissions engines into commercial vehicles is summarized by a group of experts at Navistar (Baus *et al.*, 2006).

The heavy-duty diesel engine emission regulations are explained by Charmley (2004). The US EPA Not-to-Exceed (NTE) and in-use emissions are discussed by Krishnamurthy and Gautam (2005), Shade *et al.* (2008), Thompson *et al.* (2008), Darlington *et al.* (2008), and Johnson *et al.* (2008). The NTE regulation is a concept of heavy-duty emissions control that is applied to both laboratory and on-road in-use testing. The effect of the NTE regulation is to level the emission map and avoid or preclude the likelihood of defeat devices. The NTE regulation covers an 'NTE control zone' in the engine speed–load domain (i.e., speeds greater than the so-called 15% European Stationary Cycle (ESC) speed, and engine loads greater than or equal to the higher of 30% of the maximum power or 30% of the maximum torque). The NTE regulation also covers a range of ambient conditions (i.e., the temperatures up to 38°C or 100°F, a humidity range of 0–100%, altitude levels up to 1,676 meters or 5,500 feet). Engine manufacturers are required to test customer vehicles for compliance with the NTE regulation. The reader is referred to the current EPA website at www.epa.gov for more information on the NTE standards and exclusions (e.g., exclusions due to not capable of operation, 5% limited test, deficiencies and auxiliary emissions control device (AECD), cold temperature operation, and cold exhaust aftertreatment).

Diesel engine combustion is introduced in Bowman (1991), Borman and Ragland (1998), and Hsu (2002). The fundamental mechanisms of the emissions of modern diesel engines with EGR are reviewed by Reitz (1998). Fundamental diesel emissions research has been carried out by Flynn *et al.* (2000), Kamimoto and Bae (1988), Song *et al.* (2001), Akihama *et al.* (2001), and Bianchi *et al.* (2002). It should be noted that the combustion similarity research to compare different sizes of diesel engines was conducted by Chikahisa *et al.* (1992). Emissions control and combustion technologies in engine designs have been reviewed by Suzuki *et al.* (1997), Kanda *et al.* (2004), Thompson *et al.* (2004), Zhu *et al.* (2004), and Czerwinski *et al.* (2007).

The interaction between the fuel spray and the air motion in diesel engines is reviewed by Singal *et al.* (1993). The effect of air–fuel ratio on emissions is discussed by Page (1996). The effect of EGR on diesel combustion and emissions has been investigated by Shiozaki *et al.* (1996), Zelenka *et al.* (1998), Ladommatos *et al.* (1999), Desantes *et al.* (2000), Simescu *et al.* (2002), Tomazic and Pfeifer (2002), Langridge and Fessler (2002), and Jacobs *et al.* (2003). The effects of fuel injection parameters and other engine design/calibration parameters (e.g., swirl ratio, intake manifold gas temperature) on combustion and emissions have been researched by Mather and Reitz (1998), Wickman *et al.* (2000), Woods *et al.* (2000), Montgomery and Reitz (2001), Henein *et al.* (2001), Shin *et al.* (2001), Payri *et al.* (2003), Zhu *et al.* (2003), Dronniou *et al.* (2005), and Millo *et al.* (2007).

Advanced diesel engine combustion (e.g., low-temperature combustion, Yun and Reitz (2005)) is being actively developed in order to reduce the engine-out emissions. It will have a direct impact on the emissions recipes required for the system design. Understanding the advanced combustion modes is important for a system engineer. Summary information on the homogeneous charge compression ignition (HCCI) can be found in Zhao *et al.* (2003), Najt and Foster (1983), Thring (1989), Christensen *et al.* (1998), Stanglmaier and Roberts (1999), Baxter and Hiltner (2001), Epping *et al.* (2002), Ryan *et al.* (2004), Narayanaswamy *et al.* (2005), Schleyer (2006), Zhao (2007), and Cracknell *et al.* (2008). More discussions on HCCI are provided in Chapter 9 related to VVA.

7.2.3 Diesel fuel injection system

Fuel system design and matching with combustion and air systems is a highly important and specialized area for low-emission diesel engines. Fuel system design is very complex and includes a wide range of components such as low-pressure oil pump, high-pressure oil pump, fuel pump, fuel line, oil/fuel rail, injector-rail adaptor, and injector. The design issues include fatigue, distortion, vibration, noise, leakage, cavitation, lubrication, wear, scuffing, coking, flow restriction, flow distribution, and hydraulic pressure fluctuation, under a variety of hydraulic, thermo-mechanical, tribological, and mechanical impact loads. Actuator performance degradation over time, nozzle valve seat migration and nozzle flow loss due to lacquering are common durability problems. Different types of fuel system (e.g., the electronically controlled unit injector, the unit pump, or the common rail system) exhibit different performance, durability, and packaging characteristics.

Fuel injection dynamics and fuel system performance largely affect emissions, fuel economy, engine startability, load acceptance (acceleration), and combustion noise. Fuel injection pressure, nozzle number, nozzle hole size, injection rate, and many other design and calibration parameters need to be optimized. The air system (e.g., for air–fuel ratio and swirl ratio) is designed to match the capability of the fuel system for a given emissions target. Generally, at a fixed engine speed–load mode a simultaneous increase of fuel injection pressure (or initial injection rate) and EGR rate may reduce engine-out soot while maintaining a constant NO_x. Pilot injection may reduce combustion noise, especially at the idle mode or low speeds/loads. However, pilot injection sometimes may adversely affect the NO_x–soot trade-off. Moreover, injection stability becomes more difficult to control when the pilot quantities become very small. Flexible multiple injections are desirable to control the emissions, the combustion noise, the transient operation, and the aftertreatment regeneration. For example, post fuel injection can be used to assist DPF regeneration. However, there are challenges on the post injection timing. Too early post injection may cause difficulty in engine torque control, while too late post injection may cause oil dilution or bore-washing problems.

Engine system design is closely related to the following fuel system areas:

- the effect of the conventional diesel fuel properties on engine performance, emissions, and durability
- alternative-fuel and dual-fuel diesel engine performance and durability
- life cycle cost and benefit analysis for the alternative-fuel and dual-fuel engines
- the effect of fuel injection on engine performance, combustion, emissions, and noise; and the strategy of the required fuel injection rate profile
- fuel injection spray characteristics
- fuel injection system hydraulic dynamics for the predictive modeling of fuel injection rate shape and engine system optimization
- parameterization of fuel injection rate profiles for the engine steady-state and transient cycle simulations
- fuel system parasitic power losses, heat rejection, and the impact on engine BSFC and vehicle fuel economy
- injector tip temperature and injector coking
- real-time modeling of fuel system hydraulic dynamics for hardware-in-the-loop controls development (Woermann *et al.*, 1999; discussed in Chapter 14)
- fuel delivery unevenness detection and misfire detection, and their model-based controls based on system performance and dynamics parameters (Macián *et al.*, 2006; discussed in Chapter 14)
- model-based fuel path and governor controls (e.g., engine speed control for better stability and drivability; discussed in Chapter 14).

Engine cycle simulation is closely related to fuel system design and matching. The fuel injection rate profile used in the engine cycle simulation can be estimated empirically based on the fuel system bench test data or the hydraulic/dynamic simulation data. Fuel spray can be simulated to predict the air entrainment, the evaporation and the combustion by using the phenomenological or KIVA models. The effects of fuel system design and combustion bowl matching on fuel spray pattern, heat release and air utilization in the cylinder can be analyzed. The injector tip temperature is affected by the in-cylinder gas temperature, the heat flux, and the fuel flow rate. Injector coking is directly affected by the metal temperature and the additives in the fuel. The engine cycle simulation in system design may provide the simulated thermal boundary conditions for the cylinder head and the injector in the entire engine speed–load domain under different operating conditions. This can help develop engine control algorithms to alleviate the injector coking problem.

The ability of diesel fuels to lubricate the fuel injection components is referred to as its lubricity. Fuel lubricity (SAE J2265, 1995; Matzke *et al.*, 2009) and wear durability of the fuel system is mainly a component design issue. However, the impact of lubricity additives in the diesel fuel on engine emissions and aftertreatment performance should be considered at the system design level. Modern fuel systems offer very high injection pressures and hence the loaded tribological contact conditions in the fuel injection equipment will become more severe.

The following literature on diesel fuels and fuel systems can help an engine system design engineer to acquire the necessary knowledge of fuel system selection and matching. The chemistry of diesel fuels is introduced by Oven and Trevor (1995) and Song *et al.* (2000). Diesel fuel properties are reviewed by Batts and Zuhdan-Fathoni (1991), Majewski and Khair (2006), Ribeiro *et al.* (2007), and Matzke *et al.* (2009), and also explained in SAE J313 (2004) and J1498 (2005).

The effects of diesel fuels on emissions are reported by Den Ouden *et al.* (1994), Singal and Pundir (1996), Nylund *et al.* (1997), Boesel *et al.* (2003), Matthews *et al.* (2005), Kono *et al.* (2005), Hara *et al.* (2006), Zannis *et al.* (2008), Fanick (2008), Nanjundaswamy *et al.* (2009), and Hochhauser (2009). The sulfur impact on diesel emissions control is reviewed by Corro (2002). Diesel fuel system design and performance are summarized by Cuenca (1993), Gill and Herzog (1996), Bauer (1999), Stan (1999) and Zhao (2010).

The fundamentals of fuel injection system dynamics are introduced by Marcic (1993, 1995). Advanced simulation models of fuel injection system dynamics are developed by Kouremenos *et al.* (1999), Desantes *et al.* (1999), Yamanishi (2003), Gullaksen (2004), Mulemane *et al.* (2004, with the software AMESim), and Kolade *et al.* (2004, with GT-FUEL). The models were able to predict the fuel injection pressure, the needle lift, the injection flow rate shape, the hydraulic pressure fluctuation in the system, and the fuel spray condition at the nozzle exit. A Design-of-Experiments (DoE) simulation analysis applied to fuel injection system dynamics was presented by Amoia *et al.* (1997). Fuel injection system instability and cavitation were studied by Ficarella *et al.* (1999) with dynamic modeling. The effects of the geometry of the rail-to-injector connection pipe on the injection pressure oscillation and the injection rate in a common rail system were experimentally investigated by Beierer *et al.* (2007).

Simplified fluid dynamics models may produce substantial errors in fuel injection pressure and rate predictions if fuel cavitation in the high pressure system and the variations in bulk modulus with temperature and pressure are not considered (Lee *et al.*, 2002). Fuel spray behavior and injector nozzle flow cavitation are reviewed by Schmidt and Corradini (2001).

7.2.4 Combustion chamber design

In direct injection diesel combustion system design, the combustion chamber shape and the engine compression ratio are important for in-cylinder air motion and emissions. The compression ratio has a direct impact on system design because it affects unaided cold start, peak cylinder pressure at the full load, the allowable air–fuel ratio and EGR rate, friction losses and mechanical efficiency, emissions, engine cycle indicated thermodynamic efficiency, and BSFC. For low emissions and low fuel consumption the heat release rate during the premixed combustion phase needs to be reduced in order to lower the combustion temperature and reduce the NO_x emissions (for example, by using pilot injection). The heat release rate during the diffusion combustion phase needs to be raised in order to keep fast combustion rate and a proper in-cylinder temperature to reduce the HC, the CO, and the PM. The late stage of the diffusion combustion phase needs to be shortened in order to reduce the fuel consumption and the exhaust temperature.

Key issues in combustion design include the following:

1. The engine compression ratio needs to be selected to ensure acceptable peak cylinder pressure and reliable unaided cold start. During the cold start, self-ignition is achieved by a combination of sufficient time and the required temperature (Gardner and Henein, 1988). They are affected by a number of factors, including compression ratio, ambient temperature, fuel injection amount and timing, starting speed, the speed-related gas leakage past the piston rings and heat transfer losses, and other starting aids (e.g., excess fuel injection, glow plug, heaters, and ether). A low compression ratio causes a reduction in engine thermodynamic cycle efficiency and possibly undesirable increase in the ignition delay. Note that the indicated fuel conversion efficiency of the constant-volume cycle is

$$\eta_{th} = 1 - \Omega^{1-\kappa} \tag{7.1}$$

   where Ω is the compression ratio and κ is the ratio of the specific heat capacities, $\kappa = c_p/c_v > 1$. An excessively high compression ratio results in increased mechanical friction and higher NO_x emissions. When intake valve closing timing is changed, the 'effective' engine compression ratio may vary although the effect is not exactly the same as changing the geometric compression ratio. In general, a high compression ratio is better for low HC emission at the low-speed, low-load operation. At high speeds or high loads, a low compression ratio may reduce smoke and fuel consumption via calibration changes.
2. The combustion chamber design needs to match the high-pressure fuel injection system and the intake swirl for good mixing. The combustion chamber shape variation along the radial or the axial direction controls the swirl, the vortex, and the turbulence of the in-cylinder airflow and

the wall impingement of the fuel spray to achieve better air utilization and reduce the air–fuel ratio required.

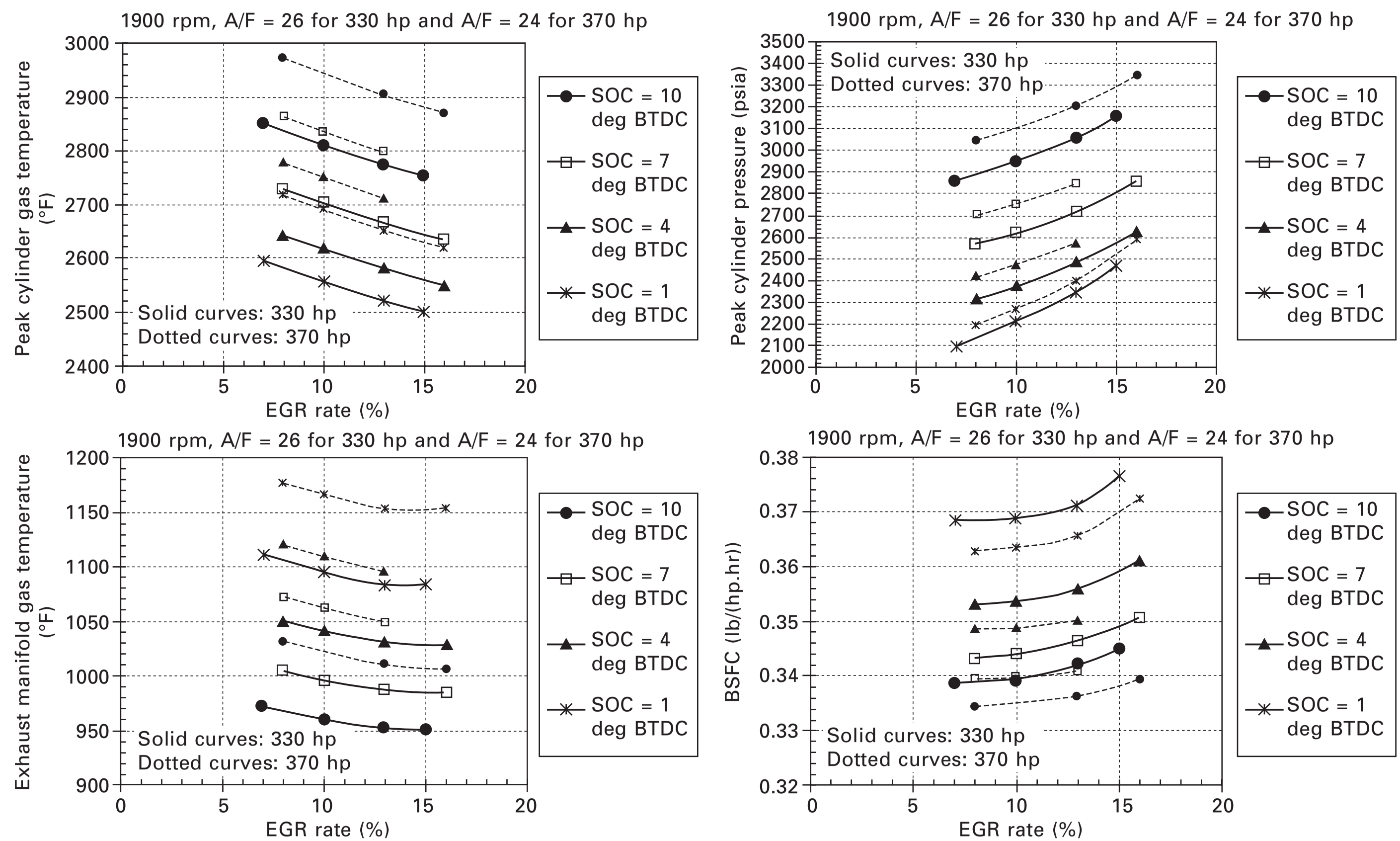
3. Optimum swirl intensity has a tendency to decrease so that the design of a more quiescent combustion bowl with a higher injection pressure is more favored for low to medium speed engines. Wall impingement of the fuel spray should be generally avoided in the design. Low swirl leads to low turbulence in the combustion chamber, resulting in low heat rejection through the wall of the combustion chamber.
4. A low surface-to-volume ratio is preferred in order to minimize the heat rejection losses (Siewert, 1978; Filipi and Assanis, 2000).
5. The metal temperature at the edge of the combustion bowl is related to the bowl shape. It needs to be optimized for the best trade-off between the airflow turbulence enhancement and the thermal fatigue life.
6. The combustion noise may be reduced by turbocharging and fuel injection strategies through a reduced rate of increase of the cylinder pressure. The combustion noise can also be reduced by the attenuation of structural damping. This topic is discussed in more detail in Chapter 11.

More information about diesel engine combustion chamber design can be found in Hikosaka (1997), Lu *et al*. (2000), Montajir *et al*. (2000), Mori *et al*. (2000), Bianchi *et al*. (2000), Regueiro (2001), Wickman *et al*. (2001), and Cursente *et al*. (2008). Diesel cold start performance was investigated by Gonzalez *et al*. (1991), Mitchell (1993), Liu *et al*. (2003), Zhong *et al*. (2007), Peng *et al*. (2008), MacMillan *et al*. (2008), and Pacaud *et al*. (2008). Variable compression ratio (VCR) engines have been researched by Roberts (2003), Rabhi *et al*. (2004), Hountalas *et al*. (2006), Tomita *et al*. (2007), Tsuchida *et al*. (2007), and Gérard *et al*. (2008).

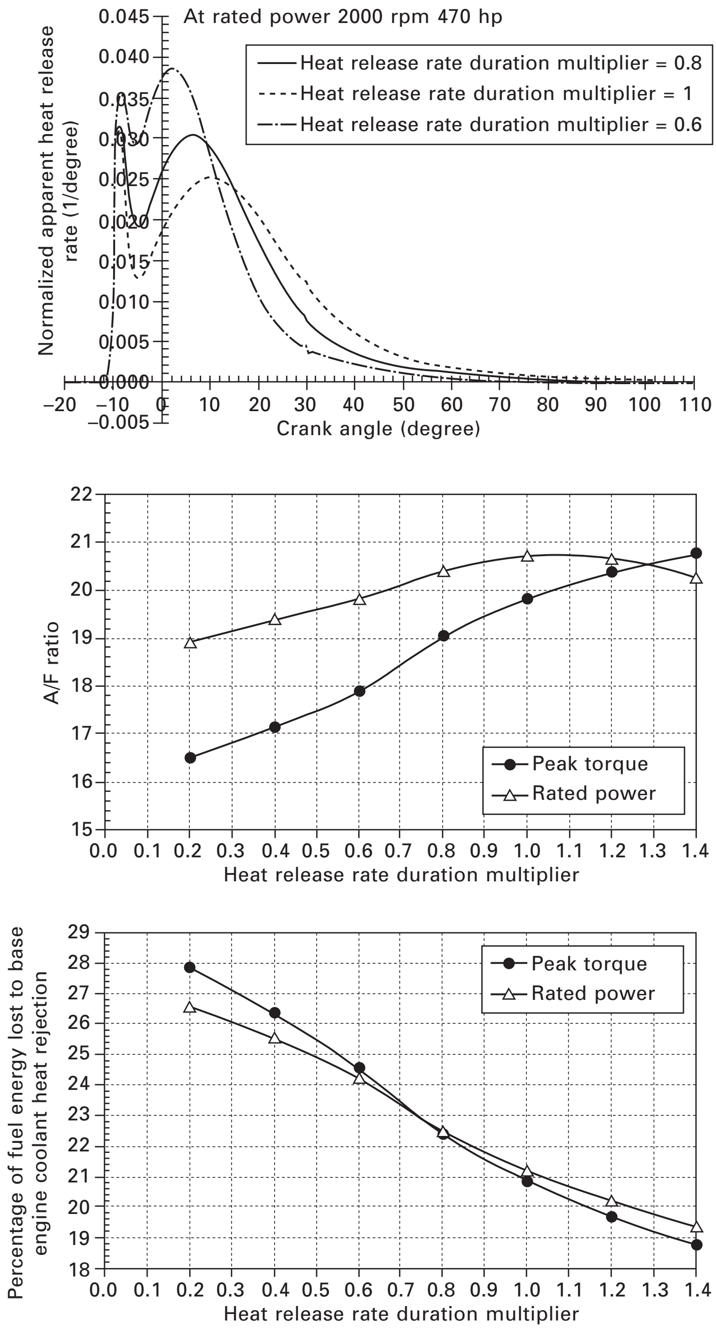
An engine cycle simulation is presented in Fig. 7.2 to illustrate the effect of start-of-combustion (SOC) timing and EGR rate on engine system performance. Figure 7.3 shows the simulation result of the effect of combustion duration on engine performance at peak torque and rated power. Figure 7.4 illustrates the effects of the rise rate and the duration of the heat release rate on exhaust manifold gas temperature and coolant heat rejection. It is found that the initial reduction of the coolant heat rejection is caused by a reduction in the in-cylinder heat transfer as the heat release rate duration is stretched longer. As the combustion duration increases further, the final sharp increase in the coolant heat rejection is due to the much higher fueling rate required to maintain the same rated power caused by the large increase in the BSFC.

7.2.5 Fundamental combustion/emissions test and air system requirement

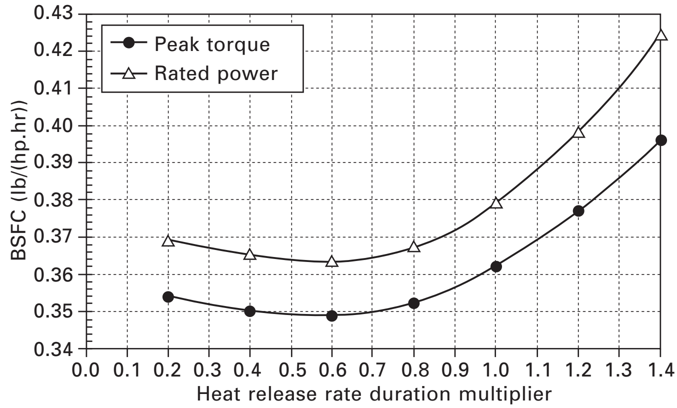
Either single-cylinder or multi-cylinder engine emissions experiments can be conducted in order to determine the combustion recipe and the air system

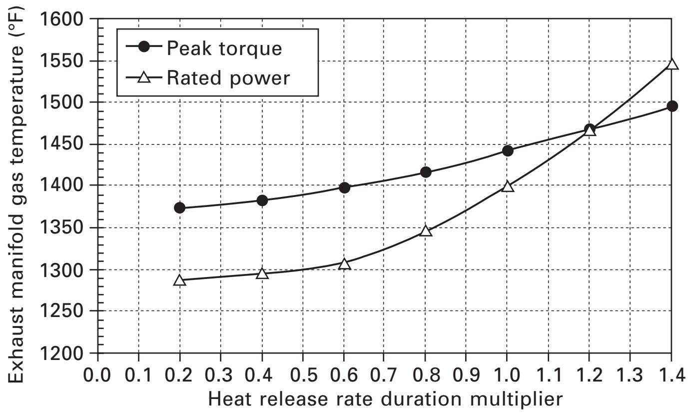


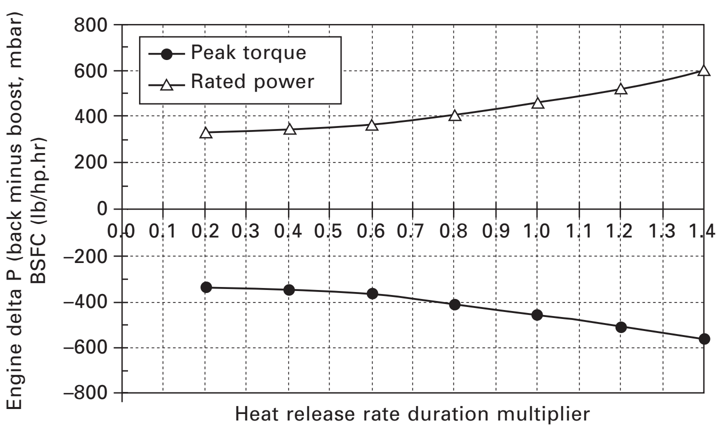
7.2 DoE simulation of the effect of start-of-combustion (SOC) timing and EGR rate.



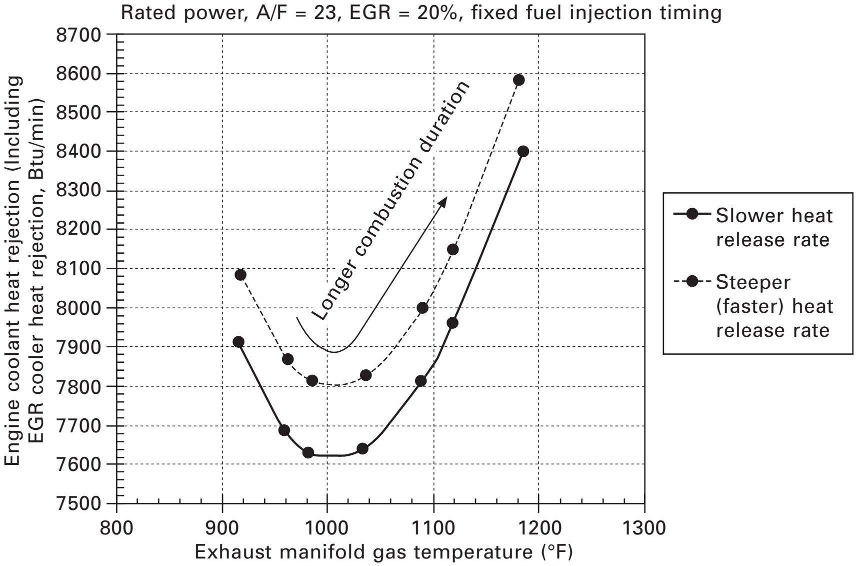
7.3 Simulation of the effect of heat release rate duration on engine performance (non-EGR high-power-density diesel engine).







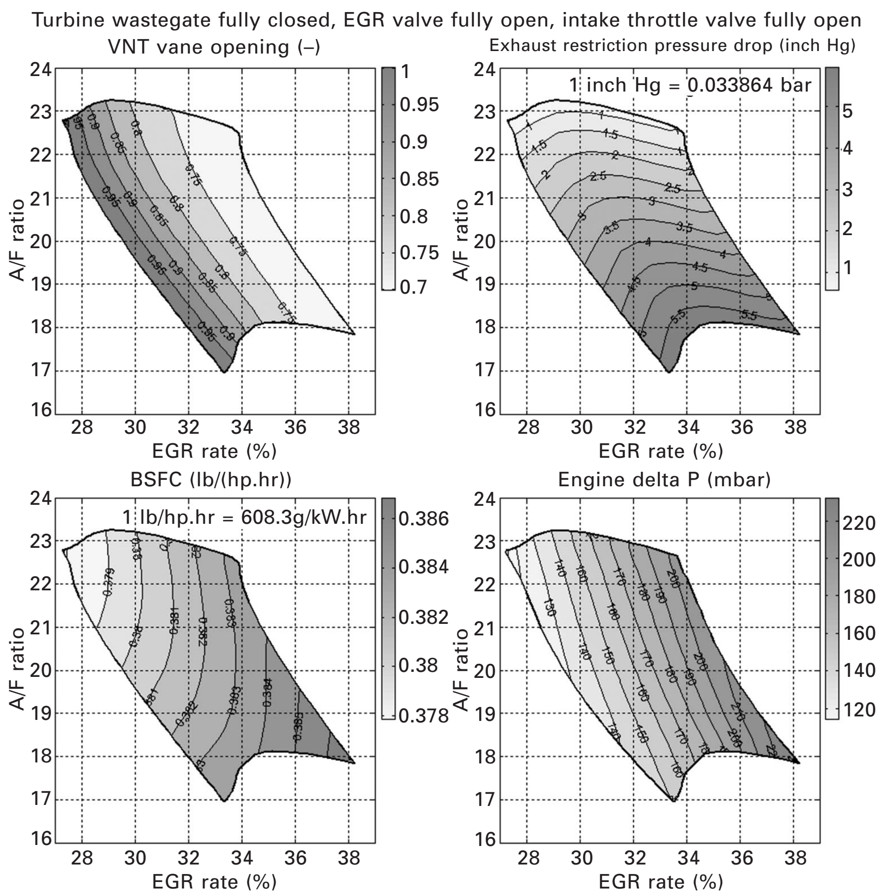
7.3 Continued



7.4 Effect of combustion heat release rate on exhaust temperature and coolant heat rejection.

requirement at the minimum engine delta P (pumping loss) and BSFC. At a given speed and torque, the air–fuel ratio and EGR rate required to meet the target emission requirements are contingent upon the BSFC produced in the test. Therefore, the engine delta P and the BSFC used in the emissions recipe test need to be as close to their respective optimums as possible. In the test, the EGR circuit hardware needs to be selected with the lowest possible flow restriction for the EGR cooler and the EGR valve. Hardware screening with the DoE method can be conducted by both the KIVA simulation and engine testing in order to select the optimum compression ratio, combustion bowl design, intake swirl ratio, and fuel injector nozzle design. Then, the DoE tuning test may be conducted at the critical speed–load modes used for the heavy-duty emissions certification. In the single-cylinder engine test, the DoE factors are intake manifold gas temperature (varied by EGR or charge air cooling), intake manifold pressure, exhaust manifold pressure, EGR valve opening, fuel injection timing and pressure. In the multi-cylinder engine test, the DoE factors are the same except for using the VGT vane opening or the turbine wastegate opening to replace the intake manifold pressure, and using the exhaust restriction setting to replace the exhaust manifold pressure. Intake throttle valve opening is an optional DoE factor that may be used near the peak torque region to drive the EGR flow or used in the high-speed, low-load region to largely reduce the air–fuel ratio in order to reduce the NO_x. The intake throttle may also be used to achieve the low minimum flow rates for smooth engine shutdown with a minimum NVH, or used in the fuel-rich regeneration of a NO_x adsorber. More DoE factors may be included for more flexible air systems (e.g., variable valve actuation). Those DoE factors are very similar to those used in the production engine calibration stage.

Before running the multi-cylinder engine test, in order to check whether the test can cover a desirable range of EGR rate and air–fuel ratio, the capability of the available turbocharger hardware needs to be checked by an engine cycle simulation through sweeping the VGT vane or the turbine wastegate opening, the exhaust restriction, or the intake throttle opening. The parametric sweeping simulation data in Fig. 7.5 illustrate how to build a sufficiently wide and controllable range of air–fuel ratio and EGR rate in such a fundamental emissions test. Note that each data point in Fig. 7.5 can



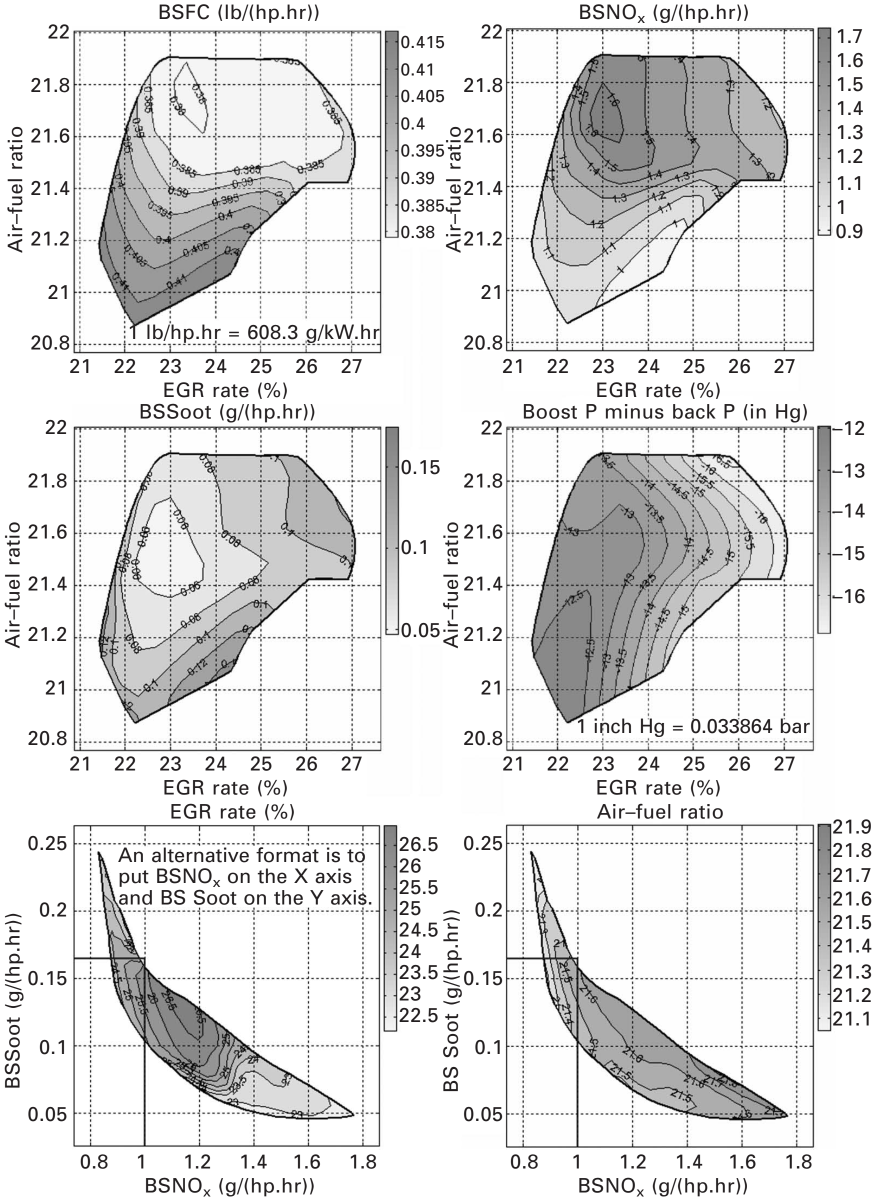
7.5 Effects of turbine area and exhaust restriction at rated power.

be solved by using the equations in Chapter 4. If the air–fuel ratio is too low, the exhaust restriction pressure drop can be artificially lowered. If the EGR rate is too high with the VGT (VNT) vane fully open, a turbine wastegate passage can be built in the test or cell to reduce the engine delta P in order to reduce the EGR rate.

After the test data are obtained, DoE emulators need to be built and an optimization needs to be conducted to compute the engine performance contour maps as shown in Fig. 7.6. The parameters on the contour maps usually include EGR rate, air–fuel ratio, intake manifold gas temperature, fuel injection timing and pressure, engine delta P, BSFC, and combustion noise. The required emissions recipe for a given emissions target of NO_x, PM, HC, and CO can be found in the 'PM or soot vs. NO_x', 'PM vs. NO_x+HC' or the 'air–fuel ratio vs. EGR rate' domains. In order to address the possible differences in turbocharger efficiency and EGR circuit restriction in a future design stage, two sets of contour plotting are useful: (1) an ideal situation with the EGR valve set at fully open, which produces the minimum engine delta P (pumping loss) and BSFC; (2) a more practical or deteriorated situation with the EGR valve set at partially open, which produces a higher pumping loss. The degree of EGR valve throttling depends on speed, torque, turbine area, and turbocharger efficiency.

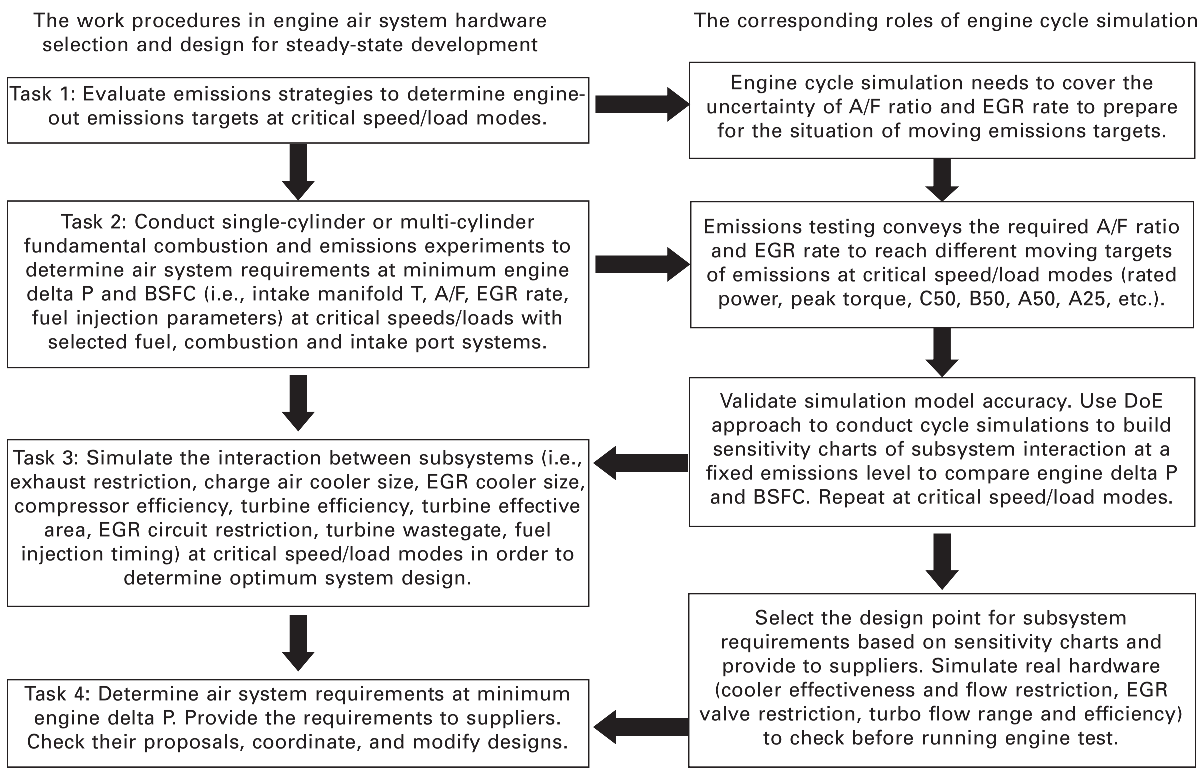
The above emissions test may also be conducted to mimic the high-altitude and hot-ambient emissions performance for the NTE emissions requirements. The DoE data of the extreme ambient conditions can be processed using the same approach to obtain the air system requirements for NTE. During the process of determining the combustion/emissions recipe, the correlation between the steady-state and the transient emissions with proper engine control strategies should be factored in. These air system functional requirements of air–fuel ratio, EGR rate, and intake manifold gas temperature are then sent to the system design engineer. Note that this combustion/emissions recipe is associated with a certain heat release rate and start-of-combustion timing and BSFC, as well as certain assumptions in the combustion–fuel system matching used by the combustion engineer. Engine system design specifications are then computed and optimized by the system engineer for hardware sizing of the turbocharger, the EGR circuit flow restriction, and the cooler size. The lower the air–fuel ratio and the EGR rate required for combustion, the easier the air system design. It should be noted that multiple turbochargers or cooling configurations may reach the same air system functional target. However, it is the system analysis engineer's responsibility to optimize the hardware selection. An example of combustion development can be found on Navistar's model year 2004 6.0 L V8 engine (Zhu *et al.*, 2004). A procedure of engine air system design based on combustion/emissions requirements is illustrated in Fig. 7.7.

It should be noted that engine system design often needs to translate the



7.6 Advanced DoE calibration optimization on emissions, BSFC, and air and EGR flows.

single-cylinder engine emissions recipe to the design specifications for the multi-cylinder engine. Moreover, engine system design needs to cover the following conditions: standard lab engine, different altitudes and ambient temperatures, vehicle in-use installations, and engine transients.



7.7 Analytical procedure of engine air system design for steady-state operation.

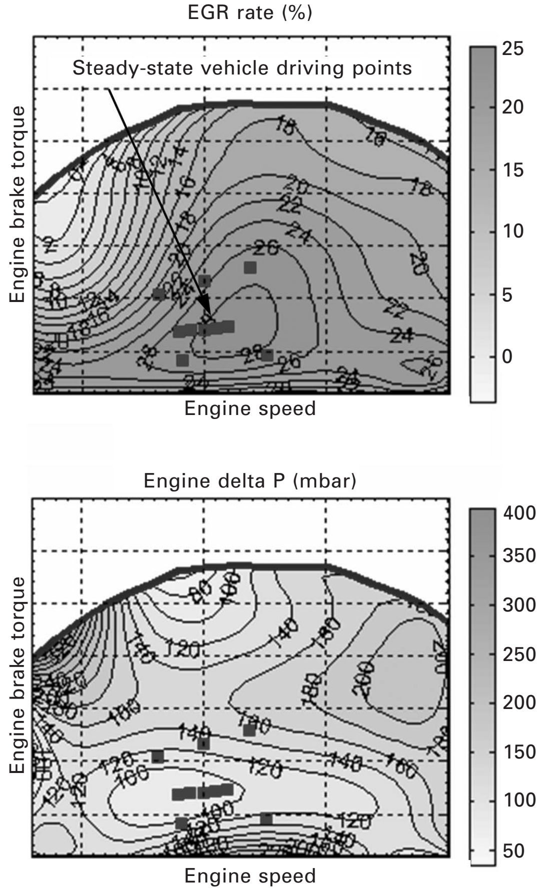
7.3 Engine calibration optimization

7.3.1 Engine emissions behavior and steady-state calibration optimization

Different combustion hardware provides different ranges of emissions capabilities. Within a fixed set of hardware, electronic controls of tunable actuators such as the EGR valve and the common rail fuel injection system are capable of altering the emission characteristics as a function of air system parameters. In the system design, because the air flow requirements may be an immature moving target, it is important to understand and document the emissions behavior as a result of air flow variations in order to design the system precisely and robustly. Emissions testing or simulation can reveal the complex parametric dependency between the air system parameters and the emissions. In contrast to the traditional trial-and-error 'knob-turning' rudimentary approach of tuning or calibration, modern automated engine calibration or hardware screening is based on DoE emulator models. Engine calibration can be so complex in a multi-dimensional factor space that even an experienced calibrator cannot always find the best settings without using optimization. Unlike the 'knob-turning' approach where the data quality is poor, the model-based calibration documents the engine's sensitivities completely, systematically, and concisely with optimization maps and mathematical models in order to facilitate any future reuse.

The model-based approach ensures the minimum BSFC is found by conducting optimization. Figure 7.6 shows the NO_x and soot behavior in the 'air–fuel ratio vs. EGR rate' domain at a fixed speed and load mode with a given set of hardware. Each data point on the maps has the minimum BSFC achievable within the factor range of the DoE. In this example, the maps are bounded by the ranges of the factors (i.e., VGT vane opening, EGR valve opening, fuel injection timing and injection pressure in this case). Different combustion hardware can 'shift' or 'rotate' the map data so that the emissions recipes may approach closer to or depart further away from the emission regulation 'box' with different levels of BSFC. These optimized contour maps in the sense of the minimum BSFC in the domain of emissions or air system parameters are a powerful and concise tool for the system engineer to judge the parametric sensitivities and to determine a precise and robust performance functional target.

The above local optimization needs to be conducted by combustion/emissions development at each critical speed–load mode. After the air system functional targets are derived, system simulation can be conducted, often with the DoE optimization method, to produce the hardware design specifications as well as the smooth steady-state performance simulation maps in the global speed–load domain. Such a 'virtual calibration', an example of which is shown in Fig. 7.8, may be used to identify the problems in the



7.8 Virtual engine calibration simulation in speed–load domain.

combustion/emissions recipe or the problems in the calibration optimization for best emissions, drivability, and fuel consumption.

Diesel engine performance and emissions calibration are summarized by the MathWorks-hosted SAE Panel Discussion (The MathWorks, 2007). Internal combustion engine calibration automation is introduced by Kampelmuhler *et al*. (1993), Schmitz *et al*. (1994), Ullmann *et al*. (2005), Jones and Muske (2007), and Foster (2008). Simulation-based calibration (related to virtual calibration) is explored by Onder and Geering (1995), Rask and Sellnau (2004), Kim and Guezennec (2005), Millich *et al*. (2005), and Neumeister *et al*. (2007). Model-based steady-state engine calibration is presented by Stuhler *et al*. (2002), Burk *et al*. (2003), Nozaki *et al*. (2005), Vora *et al*. (2005), Schlosser *et al*. (2006), and Diewald *et al*. (2009). Special application calibrations (e.g., dual-use) are described by Knafl *et al*. (2005).

Idle (including low idle) is an important mode in engine calibration for emissions, fuel economy, noise and vibration. The research on low idle performance was conducted by Khan *et al*. (2006), Pekula *et al*. (2003). Toback *et al*. (2004), MacMillan *et al*. (2009), Ghaffarpour *et al*. (1995, 2006), and Ghaffarpour and Noorpoor (2007).

7.3.2 Rapid transient calibration optimization

In order to conduct transient emissions calibration and modeling, the characteristics of the transient emissions need to be understood. Detailed information can be found in Chen and Yanakiev (2005), Urano *et al*. (2005), Kang and Farrell (2005), Hagena *et al*. (2006), and Alberer and Re (2009).

The above-mentioned DoE approach is well suited for steady-state engine emissions calibration, but is cumbersome for transient calibration of emission regulatory cycles due to its difficulty of surface-fitting the highly nonlinear, fast-changing, time-dependent response data. Transient calibration and control optimization has been a bottleneck in the engine development process and is very time-consuming and expensive. Neural network modeling provides a solution to this problem. A large amount of transient cycle engine test data with a systematic perturbation of all the calibratable parameters (including the transient gains) is used to train and build the neural network model. Then, the model can be used to predict the transient emissions and the BSFC. Optimization can be steered to minimize BSFC subject to the weighted constraints of emissions, drivability, and durability. Details of model-based transient calibration are given by Atkinson and Mott (2005). Transient performance analysis is elaborated in Chapter 14. Model-based transient calibration is discussed by Meyer and Greff (2002), Atkinson and Mott (2005), Knaak *et al*. (2005), and Atkinson *et al*. (2008). Neural network modeling of diesel engine emissions is investigated in Traver *et al*. (1999), Thompson *et al*. (2000), and Desantes *et al*. (2002).

7.4 Emissions modeling

7.4.1 Oxygen mass fraction calculation

NO_x emissions are related mainly to two fundamental parameters in the engine: the in-cylinder oxygen concentration and the gas temperature. Lower oxygen concentration and gas temperature lead to lower NO_x. The in-cylinder oxygen concentration is related to the air amount inducted and the trapped residue gas fraction. There is a direct correlation between the in-cylinder oxygen concentration and the intake manifold oxygen (O_2) mass fraction for EGR engines. The oxygen mass fraction and the in-cylinder gas temperature are affected by many engine design and operating parameters, for example ambient air humidity, EGR rate, air–fuel ratio, trapped residue gas fraction, fuel injection timing, heat release rate, intake manifold gas temperature, engine coolant temperature, and engine compression ratio. NO_x decreases as EGR rate increases since EGR reduces both the oxygen concentration and the in-cylinder flame temperature. In many cases NO_x decreases as air–fuel ratio decreases because of the reduction in the intake manifold oxygen concentration, although the calculated in-cylinder thermodynamic bulk gas temperature may appear to increase as the air–fuel ratio decreases (introduced in Chapter 4). The oxygen mass fraction is an important parameter in engine emissions analysis. Its calculation method is introduced below.

The stoichiometric air–fuel ratio can be calculated based on the chemical balance of the stoichiometric combustion of a diesel fuel with a given composition of carbon and hydrogen atoms (Heywood, 1988). The stoichiometric air–fuel ratio can be on the basis of either dry air or humid air. The engine air mass flow rate can be calculated as follows:

$$\dot{m}_{air} = f_{A/F}\,\dot{m}_{fuel} = f_{A/F,stoi}\,\dot{m}_{fuel} + (f_{A/F} - f_{A/F,stoi})\,\dot{m}_{fuel} \tag{7.2}$$

where the first term represents the reacting air flow in the combustion process and the second term represents the non-reacting air flow. Since there is no O_2 in the exhaust of the stoichiometric reacting flow, the oxygen mass fraction in the exhaust gas can be calculated by the following:

$$\begin{aligned} f_{O_2,exhaust} &= \frac{f_{O_2,air}[(f_{A/F} - f_{A/F,stoi})\,\dot{m}_{fuel}]}{(f_{A/F} - f_{A/F,stoi})\,\dot{m}_{fuel} + (f_{A/F,stoi} + 1)\,\dot{m}_{fuel}} \\ &= f_{O_2,air}\left(\frac{f_{A/F} - f_{A/F,stoi}}{f_{A/F} + 1}\right) \end{aligned} \tag{7.3}$$

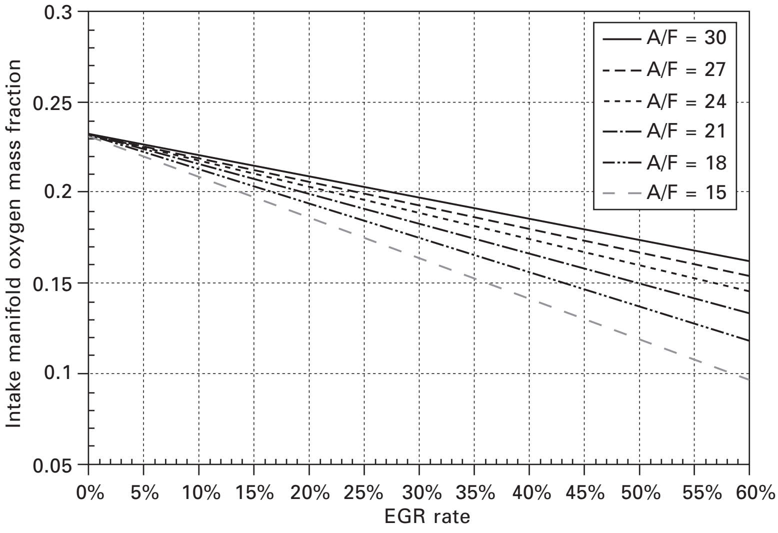
where $f_{O_2,air}$ is the mass fraction of O_2 in the ambient air, and $f_{O_2,DryAir} = 0.231$. For combustion of US No. 2 diesel fuel with dry air, $f_{A/F,stoi} = 14.5$.

The oxygen mass fraction of the intake manifold charge (i.e., the mixture of fresh air and EGR) is given by the following:

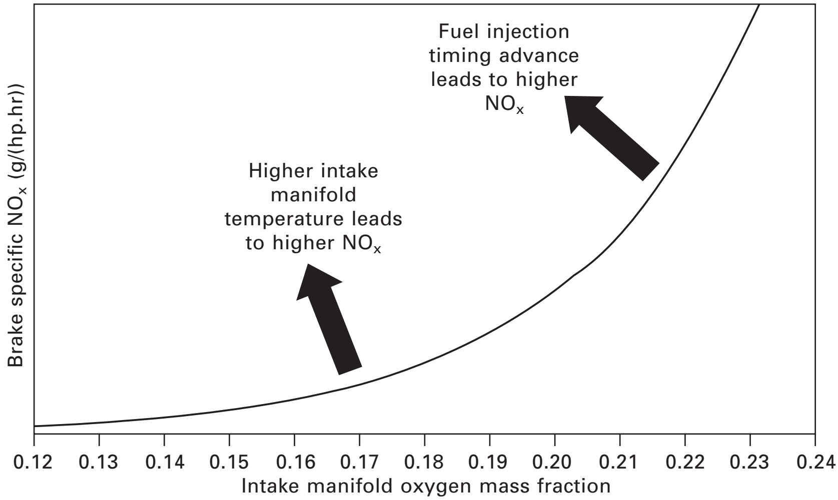
$$\begin{aligned} f_{O_2,IM} &= \frac{\dot{m}_{air} f_{O_2,air} + \dot{m}_{EGR} f_{O_2,exhaust}}{\dot{m}_{air} + \dot{m}_{EGR}} \\ &= f_{O_2,air}\left[\frac{1 + \left(\dfrac{f_{EGR}}{1 - f_{EGR}}\right)\dfrac{(f_{A/F} - f_{A/F,stoi})}{(f_{A/F} + 1)}}{\dfrac{1}{(1 - f_{EGR})}}\right] \\ &= f_{O_2,air}\left[1 - \frac{f_{EGR}(f_{A/F,stoi} + 1)}{f_{A/F} + 1}\right] \end{aligned} \tag{7.4}$$

For humid air, the terms $f_{O_2,air}$ and $f_{A/F,stoi}$ in equation 7.4 need to account for the effect of humidity. Figure 7.9 shows the oxygen mass fraction of the intake manifold charge as a function of EGR rate and air–fuel ratio for dry air.

Figure 7.10 illustrates a simple conceptual model for the diesel engine NO_x as a function of oxygen mass fraction. Emissions modeling is important



7.9 Intake manifold charge oxygen mass fraction as a function of EGR rate and air–fuel ratio (dry air).



7.10 Conceptual NO_x model as a function of intake charge oxygen fraction.

for engine system design. In addition to the experimental approach to predict the relationship between the emissions and the air system parameters, emissions simulation is also promising for establishing the air flow functional requirements in a more cost-effective way. Four types of emissions modeling approach are overviewed below.

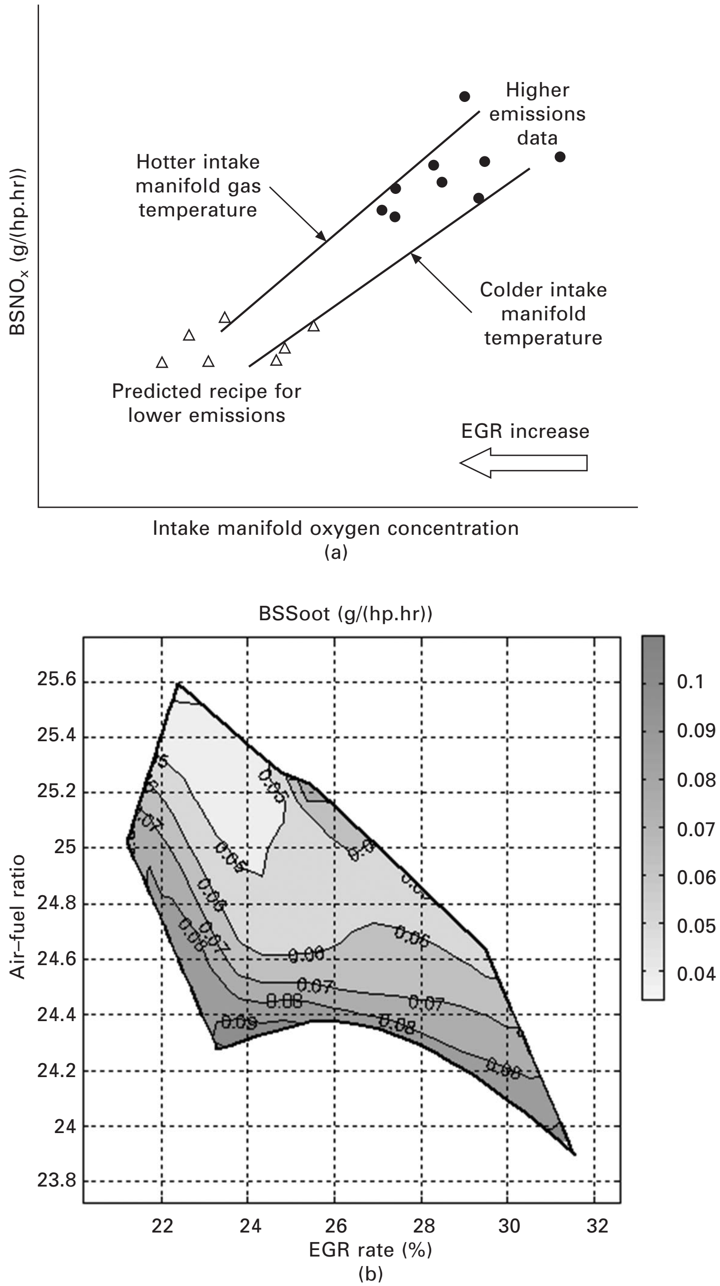
7.4.2 Empirical approach

A primitive but useful approach used in the engine industry to predict emissions is to build empirical parametric correlations among engine performance test data. There are two types of such models: (1) to directly plot without surface-fit, for example, plot with Microsoft Excel the curves of NO_x vs. oxygen concentration and intake manifold gas temperature, soot vs. air–fuel ratio or EGR rate; and (2) to surface-fit the DoE engine test data to build mathematical emulators of the emissions as functions of either the engine calibration parameters (e.g., the EGR valve opening and the fuel injection timing) or the performance parameters (e.g., the air–fuel ratio and the EGR rate). Two examples are shown in Fig. 7.11. The advantages of such an approach include the availability of a large amount of engine test data, ease of plotting, ability to obtain data trends through regression or surface-fit and a quick estimation of the emissions. The disadvantages include being purely empirical, lacking fundamental or governing physics in the deduced correlations or the emulator models, lacking real-time predictability and feasibility to be implemented in future advanced intelligent engine controls, and the fact that the predictability is limited to the engine hardware from which the test data are generated.

7.4.3 Zero-dimensional approach

In zero-dimensional models, the in-cylinder volume is treated as a single zone or divided into burned and unburned zones without modeling the details of the fuel spray and the turbulent air entrainment (Zhang, 1998; Lafossas *et al.*, 2007).

Useful information on the zero-dimensional emissions models can be found in Nightingale (1975), Lipkea and DeJoode (1994), Dodge *et al.* (1996), Zhang (1998), Easley *et al.* (2000), Yang *et al.* (2002), Kolade *et al.* (2004), Ponti *et al.* (2007), Lafossas *et al.* (2007), a group of researchers at the Lund University in Sweden (Egnell, 1998, 1999, 2000; Andersson *et al.*, 2006a, 2006b; Ericson *et al.*, 2006; Wilhelmsson *et al.*, 2009), and a group of researchers at the National Technical University of Athens in Greece (Kouremenos *et al.*, 1997; Rakopoulos and Hountalas, 1998; Rakopoulos *et al.*, 1999; Pariotis *et al.*, 2006).



7.11 Empirical emissions modeling: (a) direct plot; (b) DoE optimization emulator model.

7.4.4 Phenomenological approach

A typical traditional phenomenological combustion and emissions model is the diesel jet model developed from Cummins, Inc. The model simulates the mixing rate of fuel spray and air, and calculates the local gas temperature, the air–fuel ratio and the combustion in each zone of spray so that the computed in-cylinder gas properties vary with both space and time. The advantages of such a model include the fact that some combustion physics and chemistry are built into the model. In addition, with proper tuning the model has limited predictability on the heat release rate and the NO_x. The disadvantages include the following:

- many simplified assumptions concerning spray penetration, turbulence, swirl, heat transfer, vapor concentration distribution, wall impingement, flame propagation, combustion, and emissions formation
- the practical difficulty of obtaining in-cylinder experimental data to justify or calibrate each sub-model for industry use
- many model tuning parameters as a result of many detailed sub-models
- the difficulty of evaluating many design parameters related to the combustion chamber and the air–fuel–combustion system matching because those factors cannot be built into the model
- inability to predict particulate matter and soot accurately
- the fact that it is computationally intensive, and it is impossible to implement in real-time simulations or intelligent engine controls at current computer speeds.

It seems that the focus of the traditional phenomenological modeling may be directed toward assisting the development of the more practical heuristic macro-parameter-dependent modeling (to be introduced below), instead of trying to become a stand-alone design tool, at least from a diesel engine system design point of view.

More detailed information on the phenomenological emissions models can be found in Hiroyasu and Kadota (1976), Chiu *et al.* (1976), Kyriakides *et al.* (1986), Bazari (1992), Yoshizaki *et al.* (1993), Huang *et al.* (1996), Morel and Wahiduzzaman (1996), Wang *et al.* (1999), Hiroyasu and Long (2000), Torkzadeh *et al.* (2001), Kouremenos *et al.* (2001), Bayer and Foster (2003), Asay *et al.* (2004), Brahma *et al.* (2005), Pariotis *et al.* (2005), Arrègle *et al.* (2006), Mauviot *et al.* (2006), and Bagal *et al.* (2009).

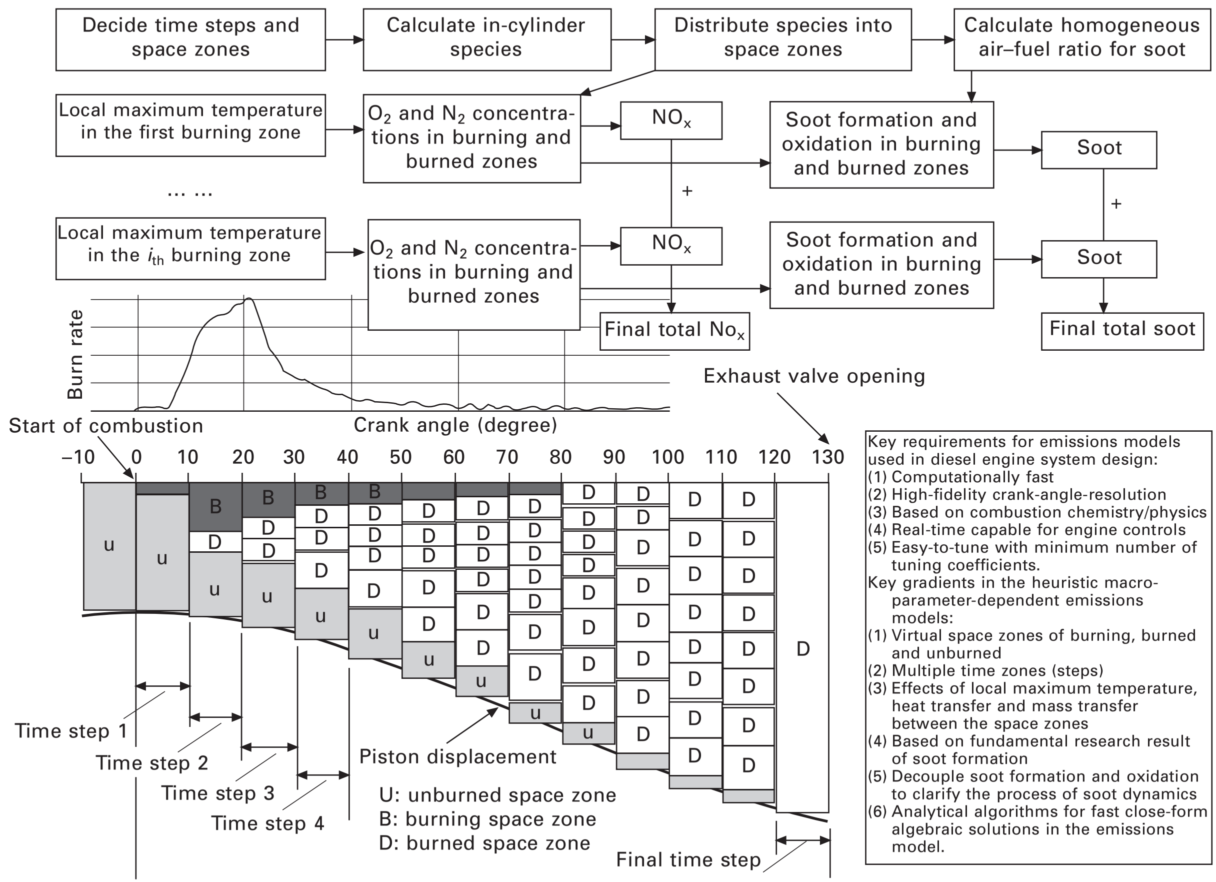
7.4.5 Computational fluid dynamics (CFD)–KIVA modeling

The three-dimensional KIVA is a very complex multi-zone combustion model based on the partial differential equations of viscous fluid dynamics, turbulence, chemical reaction, boundary layer, and heat transfer. It has demonstrated strong potential and practicality to be used as a design tool to predict diesel emissions and optimize combustion chamber design and air–fuel–combustion system matching. The challenges with the KIVA are the complexity of model tuning and the long computing time required.

There are numerous publications on the KIVA simulations. Those conducted by Han *et al*. (1996) and Bianchi *et al*. (2002) are good examples for a diesel engine system engineer to get familiar with this area. The KIVA modeling work is reviewed by Reitz and Rutland (1995) and Reitz and Sun (2010). The CFD simulation with the KIVA code and the commercial CFD codes (such as STAR-CD, VECTIS, FIRE, and FLUENT) on the characteristics of the diesel engine in-cylinder flow, turbulence, and spray is reviewed by Basha and Raja Gopal (2009).

7.4.6 Heuristic macro-parameter-dependent approach

Heuristics are the doctrine of obtaining new insights and universally valid design and calculation methods with the law of similarity or partially empirical simplification. It is a counterbalance to the theoretical KIVA method. The heuristics lead to more detailed insights if detailed models are established. The 'macro' means the global apparent engine performance parameters and the zero-dimensional in-cylinder bulk parameters. Diesel emissions are extremely complex to model in micro detail due to the various mechanisms involved in the ignition delay, the premixed and diffusion flame combustion processes. For example, the creation of soot is determined by the local temperature and the oxygen concentration during the events of the radical chain reaction in the core of the fuel jet, the adsorption of polycyclic aromatic hydrocarbons, the processes of polymerization, cyclization, coagulation and agglomeration in the soot formation and reoxidation phases. The idea of using a macro-parameter-dependent approach to build emissions models is to try to bypass the difficulties encountered by the traditional phenomenological model. The approach tries to link the emissions to more fundamental engine performance parameters, which are either the apparent macro cycle-average parameters such as the air–fuel ratio, or the macro instantaneous crank-angle based parameters such as the in-cylinder bulk temperature, in a relatively simplified combined theoretical form. It still conforms to the basic principles of emission chemistry. Its theoretical foundation is far beyond the empirical emissions data processing. It is indeed a practical semi-empirical approach, especially suitable for the needs of engine system design and analysis. The advantage of such a model is that it may eliminate many of the disadvantages in the empirical approach and the traditional phenomenological approach. Most importantly, it is the only feasible emission prediction model that can be used in future advanced intelligent real-time emissions-model-based



7.12 Summary of model structure of heuristic macro-parameter-dependent emissions model.

electronic engine controls to handle both the steady-state and the transient control needs. The technique of building such a model can also be applied to build the macro emissions models for aftertreatment devices, where the technical challenge is actually less because the variation of the gas properties within an engine cycle is not as complex as the in-cylinder combustion processes. This type of macro model is not intended to predict the heat release rate or evaluate the combustion chamber design. Heat release rate is used as input in the model. The predictability of the model is limited to the given hardware of combustion and fuel systems. If the effects of different combustion hardware design need to be simulated, the KIVA simulation must be used. The real-time heuristic macro models of NO_x and soot were proposed and detailed by Zheng and Xin (2009). Figure 7.12 illustrates the model theory and the model structure.

The best emissions model is not necessarily the most complicated theoretical model; rather it should be a simple, practical, and most efficient model for engineering design applications.

Transient emissions modeling is another great challenge in engine system design. Information in this area can be found in Hamburg and Throop (1984), Jiang and Gerpen (1992), Bazari (1994), Bi and Han (1995), Ramamurthy *et al.* (1998), Cui *et al.* (2001), Shirota *et al.* (2001), Ericson *et al.* (2005), Andersson *et al.* (2006a, 2006b), Zhang *et al.* (2007), Brahma *et al.* (2009), and Hirsch and Re (2009).

7.5 References and bibliography

Akihama K, Takatori Y, Inagaki K, Sasaki S and Dean A M (2001), 'Mechanism of the smokeless rich diesel combustion by reducing temperature', SAE paper 2001-01-0655.

Alberer D and Re L (2009), 'Fast oxygen based transient diesel engine operation', SAE paper 2009-01-0622.

Amoia V, Ficarella A, Laforgia D, De Matthaeis S and Genco C (1997), 'A theoretical code to simulate the behavior of an electro-injector for diesel engines and parametric analysis', SAE paper 970349.

Andersson M, Johansson B, Hultqvist A and Nöhre C (2006a), 'A real time NO_x model for conventional and partially premixed diesel combustion', SAE paper 2006-01-0195.

Andersson M, Johansson B, Hultqvist A and Nöhre C (2006b), 'A predictive real time NO_x model for conventional and partially premixed diesel combustion', SAE paper 2006-01-3329.

Arrègle J, López J J, Martín J and Mocholí E M (2006), 'Development of a mixing and combustion zero-dimensional model for diesel engines', SAE paper 2006-01-1382.

Asay R J, Svensson K I and Tree D R (2004), 'An empirical, mixing-limited, zero-dimensional model for diesel combustion', SAE paper 2004-01-0924.

Atkinson C and Mott G (2005), 'Dynamic model-based calibration optimization: an introduction and application to diesel engines', SAE paper 2005-01-0026.

Atkinson C, Allain M and Zhang H (2008), 'Using model-based rapid transient calibration to reduce fuel consumption and emissions in diesel engines', SAE paper 2008-01-1365.

Bagal N L, Rutland C J, Foster D E, Narayanaswamy K and He Y (2009), 'CO emission model for an integrated diesel engine, emissions, and exhaust aftertreatment system level model', SAE paper 2009-01-1511.

Barr D, Boyer M and McAtee R (2003), 'Formulating for lower emissions – issues and solutions', SAE paper 2003-01-1966.

Basha S A and Raja Gopal K (2009), 'In-cylinder fluid flow, turbulence and spray models – a review', *Renewable and Sustainable Energy Reviews*, 13(6–7), 1620–1627.

Batts B D and Zuhdan-Fathoni A (1991), 'A literature review on fuel stability studies with particular emphasis on diesel oil', *Energy & Fuels*, 5(1), 2–21.

Bauer H (ed.) (1999), *Diesel-Engine Management*, 2nd edn, Robert Bosch GmbH/SAE International, Warrendale, PA.

Baus M, Cook A and Schaller D (2006), *Integrating New Emissions Engines into Commercial Vehicles: Emissions, Performance & Affordability*, the 51st L. Ray Buckendale Lecture, SAE paper 2006-01-3545, SAE International, Warrendale, PA.

Baxter J M and Hiltner J (2001), 'Working toward homogeneous diesel combustion: a fresh look at the work of Max Fiedler', ASME paper 2001-ICE-418.

Bayer J and Foster D E (2003), 'Zero-dimensional soot modeling', SAE paper 2003-01-1070.

Bazari Z (1992), 'A DI diesel combustion and emission predictive capability for use in cycle simulation', SAE paper 920462.

Bazari Z (1994), 'Diesel exhaust emissions prediction under transient operating conditions', SAE paper 940666.

Beierer P, Huhtala K and Vilenius M (2007), 'Experimental study of the hydraulic circuit of a commercial common rail diesel fuel injection system', SAE paper 2007-01-0487.

Bi X and Han S (1995), 'Prediction of transient exhaust soot for a turbocharged diesel engine', SAE paper 952434.

Bianchi G M, Pelloni P, Corcione F E, Mattarelli E and Bertoni F L (2000), 'Numerical study of the combustion chamber shape for common rail H.S.D.I. diesel engines', SAE paper 2000-01-1179.

Bianchi G M, Cazzoli G, Pelloni P and Corcione F E (2002), 'Numerical study towards smoke-less and NO_x-less HSDI diesel engine combustion', SAE paper 2002-01-1115.

Boesel J, Romeo S and Brotherton T (2003), 'Advanced fuels and technology forecasts for heavy-duty vehicles', SAE paper 2003-01-3404.

Borman G L and Ragland K W (1998), 'Direct-injection engine combustion', *Combustion Engineering*, McGraw-Hill, Boston, MA.

Bowman C T (1991), 'Chemistry of gaseous pollutant formation and destruction', in Bartok W and Sarofim A F (eds) *Fossil Fuel Combustion – A Source Book*, John Wiley & Sons, New York, pp. 215–260.

Brahma I, Rutland C J, Foster D E and He Y (2005), 'A new approach to system level soot modeling', SAE paper 2005-01-1122.

Brahma I, Sharp M C and Frazier T R (2009), 'Empirical modeling of transient emissions and transient response for transient optimization', SAE paper 2009-01-1508.

Browning L H (1997), 'Technologies and costs for on-road heavy-duty engines meeting 2004 emissions standards', SAE paper 973256.

Burk R, Jacquelin F and Wakeman R (2003), 'A contribution to predictive engine calibration based on vehicle drive cycle performance', SAE paper 2003-01-0225.

Charlton S J (2005), *Developing Diesel Engines to Meet Ultra-low Emission Standards*, the 50th L. Ray Buckendale Lecture, SAE paper 2005-01-3628, SAE International, Warrendale, PA.

Charmley W J B (2004), *The Federal Government's Role in Reducing Heavy Duty Diesel Emissions*, the 49th L. Ray Buckendale Lecture, SAE paper 2004-01-2708, SAE International, Warrendale, PA.

Chen S K and Yanakiev O (2005), 'Transient NO_x emission reduction using exhaust oxygen concentration based control for a diesel engine', SAE paper 2005-01-0372.

Chikahisa T, Kikuta K and Murayama T (1992), 'Combustion similarity for different size diesel engines: theoretical prediction and experimental results', SAE paper 920465.

Chiu W S, Shahed S M and Lyn W T (1976), 'A transient spray mixing model for diesel combustion', SAE paper 760128.

Christenson M (1998), 'Supercharged homogeneous charge compression ignition', SAE paper 980787.

Christensen M, Johansson B, Amneus P and Mauss F (1998), 'Supercharged homogeneous charge compression ignition', SAE paper 980787.

Corro G (2002), 'Sulfur impact on diesel emission control – a review', *Reaction Kinetics and Catalysis Letters*, 75(1), 89–106.

Corti E and Solieri L (2005), 'Rapid control prototyping system for combustion control', SAE paper 2005-01-3754.

Cracknell R, Ariztegui J, Barnes K, Bessonette P, Cannella W, Douce F, Kelecom B, Kraft H, Lampreia I, Rickeard D J, Savarese M C, Williams J and Rose K D (2008), 'Advanced combustion for low emissions and high efficiency: a literature review of HCCI combustion concepts', CONCAWE Report No. 4/08, www.concawe.org.

Cuenca R M (1993), 'Evolution of diesel fuel injection equipment – the last 20 years', SAE paper 933015.

Cui Y, Deng K and Wu J (2001), 'A direct injection diesel combustion model for use in transient condition analysis', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 215, 995–1004.

Cursente V, Pacaud P and Gatellier B (2008), 'Reduction of the compression ratio on a HSDI diesel engine: combustion design evolution for compliance the future emission standards', SAE paper 2008-01-0839.

Czerwinski J, Péterman J-L, Comte P, Lemaire J and Mayer A (2007), 'Diesel $NO/NO_2/NO_x$ emissions – new experiences and challenges', SAE paper 2007-01-0321.

Darlington T L, Kahlbaum D and Thompson G (2008), 'On-road NO_x emission rates from 1994–2003 heavy-duty diesel trucks', SAE paper 2008-01-1299.

Den Ouden C J J, Clark R H, Cowley L T, Stardling R J, Lange W W and Maillard C (1994), 'Fuel quality effects on particulate matter emissions from light- and heavy-duty diesel engines', SAE paper 942022.

Desantes J M, Arregle J and Rodriguez P J (1999), 'Computational model for simulation of diesel injection systems', SAE paper 1999-01-0915.

Desantes J M, Arregle J, Molina S and Lejeune M (2000), 'Influence of the EGR rate, oxygen concentration and equivalent fuel/air ratio on the combustion behavior and pollutant emissions of a heavy-duty diesel engine', SAE paper 2000-01-1813.

Desantes J M, Lopez J J, Garcia J M and Hernandez L (2002), 'Applications of neural networks for prediction and optimization of exhaust emissions in a H.D. diesel engine', SAE paper 2002-01-1144.

Dexter S G and Kling W H (1999), 'Medium speed engine design changes caused by international emissions regulations', ASME paper 99-ICE-225.

Diewald R, Cartus T, Schuessler M and Bachler H (2009), 'Model based calibration methodology', SAE paper 2009-01-2837.

Dodge L G, Leone D M, Naegeli D W, Dickey D W and Swenson K R (1996), 'A PC-based model for predicting NO_x reductions in diesel engines', SAE paper 962060.

Dollmeyer T A, Vittorio D A, Grana T A, Katzenmeyer J R, Charlton S J, Clerc J, Morphet R G and Schwandt B W (2007), 'Meeting the US 2007 heavy-duty diesel emission standards – designing for the customer', SAE paper 2007-01-4170.

Dronniou N, Lejeune M, Balloul I and Higelin P (2005), 'Combination of high EGR rates and multiple injection strategies to reduce pollutant emissions', SAE paper 2005-01-3726.

Easley W L, Mellor A M and Plee S L (2000), 'NO formation and decomposition models for DI diesel engines', SAE paper 2000-01-0582.

Egnell R (1998), 'Combustion diagnostics by means of multizone heat release analysis and NO calculation', SAE paper 981424.

Egnell R (1999), 'A simple approach to studying the relation between fuel rate, heat release rate and NO formation in diesel engines', SAE paper 1999-01-3548.

Egnell R (2000), 'The influence of EGR on heat release rate and NO formation in a DI diesel engine', SAE paper 2000-01-1807.

Epping K, Aceves S, Bechtold R and Dec J (2002), 'The potential of HCCI combustion for high efficiency and low emissions', SAE paper 2002-01-1923.

Ericson C, Westerberg B and Egnell R (2005), 'Transient emission predictions with quasi stationary models', SAE paper 2005-01-3852.

Ericson C, Westerberg B, Andersson M and Egnell R (2006), 'Modelling diesel engine combustion and NO_x formation for model based control and simulation of engine and exhaust aftertreatment systems', SAE paper 2006-01-0687.

Fanick E R (2008), 'Diesel fuel keeping pace with diesel engine technology', SAE paper 2008-01-1808.

Ficarella A, Laforgia D and Landriscina V (1999), 'Evaluation of instability phenomena in a common rail injection system for high speed diesel engines', SAE paper 1999-01-0192.

Filipi Z and Assanis D N (2000), 'The effect of the stroke-to-bore ratio on combustion, heat transfer and efficiency of a homogeneous charge spark ignition engine of given displacement', *International Journal of Engine Research*, 1(2), 191–208.

Flynn P F (2001), 'How chemistry controls engine design', *Proceedings of the 2001 Fall Technical Conference of the ASME Internal Combustion Engine Division, Diesel Combustion and Emissions, Fuel injection and Sprays*, Volume 1, pp. 1–9.

Flynn P F, Hunter G L, Durrett R P, Farrell L A and Akinyemi W C (2000), 'Chemistry limits on minimum in-cylinder NO_x production for internal combustion engines', IMechE paper C588/027/2000, pp. 187–208.

Foster T (2008), 'Calibration scenario editor enables higher calibration efficiencies', SAE paper 2008-01-2714.

Gardner T P and Henein N A (1988), 'Diesel starting: a mathematical model', SAE paper 880426.

Gérard D, Besson M, Hardy J-P, Croguennec S, Thomine M, Aoyama S and Tomita M (2008), 'HCCI combustion on a diesel VCR engine', SAE paper 2008-01-1187.

Ghaffarpour M R, Baranescu R and Doan Thien (1995), 'Combustion noise and emissions reduction at idle in diesel engines through numerical analysis', ASME paper 95-ICE-12.

Ghaffarpour M, Azarfam M and Noorpoor A (2006), 'Emission reduction in diesel

engines using new fuel injection system', *JSME International Journal*, *Series B*, 49(4), 1298–1306.
Ghaffarpour M R and Noorpoor A R (2007), 'A numerical study of the use of pilot or split rate injection to reduce diesel engine noise', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 221, 457–464.
Gill A P (1988), 'Design choices for 1990's low emission diesel engines', SAE paper 880350.
Gill D W and Herzog P L (1996), 'Fuel injection technology for low emissions HSDI diesel engines', SAE paper 962369.
Gonzalez M A, Borman G L and Reitz R D (1991), 'A study of diesel cold starting using both cycle analysis and multidimensional calculations', SAE paper 910180.
Grondin O, Stobart R, Chafouk H and Maquet J (2004), 'Modelling the compression ignition engine for control: review and future trends', SAE paper 2004-01-0423.
Gullaksen J (2004), 'Simulation of diesel fuel injection dynamics using MATLAB', SAE paper 2004-01-2966.
Hagena J R, Filipi Z S and Assanis D N (2006), 'Transient diesel emissions: analysis of engine operation during a tip-in', SAE paper 2006-01-1151.
Hamburg D R and Throop M J (1984), 'A comparison between predicted and measured feedgas emissions for dynamic engine operation', SAE paper 841256.
Han Z, Uludogan A, Hampson G J and Reitz R D (1996), 'Mechanism of soot and NO_x emission reduction using multiple-injection in a diesel engine', SAE paper 960633.
Hara S, Kaneko T, Kakegawa T, Senbokuya S, Nakamura O, Shibuya M, Itoyama H, Okada M, Sugiyama G, Hasegawa T (2006), 'Effects of fuel properties on the performance of advanced diesel NO_x aftertreatment devices', SAE paper 2006-01-3443.
Henein N A, Lai M-C, Singh I, Wang D and Liu L (2001), 'Emissions trade-off and combustion characteristics of a high-speed direct injection diesel engine', SAE paper 2001-01-0197.
Heywood J B (1988), *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.
Hikosaka N (1997), 'A view of the future of automotive diesel engines', SAE paper 972682.
Hiroyasu H and Kadota T (1976), 'Models for combustion and formation of nitric oxide and soot in direct injection diesel engines', SAE paper 760129.
Hiroyasu H and Long Y (2000), 'The simulation of the distribution of temperature and mass of liquid and vapor fuels, and the wall impinging spray pattern in a diesel combustion chamber', SAE paper 2000-01-1887.
Hirsch M and Re L (2009), 'Adapted D-Optimal experimental design for transient emission models of diesel engines', SAE paper 2009-01-0621.
Hochhauser A M (2009), 'Review of prior studies of fuel effects on vehicle emissions', SAE paper 2009-01-1181.
Hountalas D T (2000), 'Available strategies for improving the efficiency of DI diesel engines – a theoretical investigation', SAE paper 2000-01-1176.
Hountalas D T, Zannis T C and Mavropoulos G C (2006), 'Potential benefits in heavy duty diesel engine performance and emissions from the use of variable compression ratio', SAE paper 2006-01-0081.
Hsu B D (2002), *Practical Diesel-Engine Combustion Analysis*, SAE International, Warrendale, PA.
Huang Y, Wang Z and Hu Y (1996), 'The application of phenomenological combustion model for experimental study of DI diesel engines', SAE paper 961051.

Husted H, Kruger D, Fattic G, Ripley G and Kelly E (2007), 'Cylinder pressure-based control of pre-mixed diesel combustion', SAE paper 2007-01-0773.

Jacobs T, Assanis D and Filipi Z (2003), 'The impact of exhaust gas recirculation on performance and emissions of a heavy-duty diesel engine', SAE paper 2003-01-1068.

Jiang Q and Gerpen J H V (1992), 'Prediction of diesel engine particulate emission during transient cycles', SAE paper 920466.

Johnson K C, Durbin T D, Cocker III D R, Miller J W, Agama R J, Moynahan N and Nayak G (2008), 'On-road evaluation of a PEMS for measuring gaseous in-use emissions from a heavy-duty diesel vehicle', SAE paper 2008-01-1300.

Jones J C P and Muske K R (2007), 'Automatic calibration of 1 and 2-D look-up tables using recursive least-squares identification techniques', SAE paper 2007-01-1343.

Kamimoto T and Bae M (1988), 'High combustion temperature for the reduction of particulate in diesel engines', SAE paper 880423.

Kampelmuhler F T, Paulitsch R and Gschweitl K (1993), 'Automatic ECU-calibration – an alternative to conventional methods', SAE paper 930395.

Kanda T, Kobayashi S, Matsui R and Sono H (2004), 'Study on Euro IV combustion technologies for direct injection diesel engine', SAE paper 2004-01-0113.

Kang H and Farrell P V (2005), 'Experimental investigation of transient emissions (HC and NO_x) in a high speed direct injection (HSDI) diesel engine', SAE paper 2005-01-3883.

Khair M K (1997), 'Technical and synergistic approaches towards the twenty-first century diesel engine', SAE paper 972687.

Khan A S, Clark N N, Thompson G J, Wayne W S, Gautam M, Lyons D W and Hawelti D (2006), 'Idle emissions from heavy-duty diesel vehicles: review and recent data', *Journal of the Air & Waste Management Association*, 56(10), 1404–1419.

Kim C-G and Guezennec Y G (2005), 'A methodology for virtual engine mapping test of CIDI engine with arbitrary fuel injection schedule for control purpose', SAE paper 2005-01-0230.

Klein K, Grüter R and Loffeld O (2007), 'Real-time estimation of the exhaust gas recirculation ratio based on cylinder pressure signals', SAE paper 2007-01-0493.

Klingbeil A E, Juneja H, Ra Y and Reitz R D (2003), 'Premixed diesel combustion analysis in a heavy-duty diesel engine', SAE paper 2003-01-0341.

Knaak M, Schoop U and Roepke K (2005), 'Comparison of different transient air charge models', SAE paper 2005-01-0051.

Knafl A, Hagena J R, Filipi Z S and Assanis D N (2005), 'Dual-use engine calibration: leveraging modern technologies to improve performance – emissions tradeoff', SAE paper 2005-01-1549.

Kolade B, Morel T and Kong S-C (2004), 'Coupled 1-D/3-D analysis of fuel injection and diesel engine combustion', SAE paper 2004-01-0928.

Kono N, Suzuki Y and Takeda H (2005), 'Effects of driving conditions and fuel properties on diesel emissions', SAE paper 2005-01-3835.

Kouremenos D A, Rakopoulos C D and Hountalas D T (1997), 'Multi-zone combustion modelling for the prediction of pollutants emissions and performance of DI diesel engines', SAE paper 970635.

Kouremenos D A, Hountalas D T and Kouremenos A D (1999), 'Development and validation of a detailed fuel injection system simulation model for diesel engines', SAE paper 1999-01-0527.

Kouremenos D A, Hountalas D T and Binder K B (2001), 'The effect of EGR on the performance and pollutant emissions of heavy duty diesel engines using constant and variable AFR', SAE paper 2001-01-0198.

Krishnamurthy M and Gautam M (2005), 'Comparison of averaging techniques employed in calculating not-to-exceed emissions for heavy-duty vehicles', SAE paper 2005-01-3787.

Kyriakides S C, Dent J C and Mehta P S (1986), 'Phenomenological diesel combustion model including smoke and NO emission', SAE paper 860330.

Ladommatos N, Abdelhalim S and Zhao H (1999), 'The effects of exhaust gas recirculation on diesel combustion and emissions', *International Journal of Engine Research*, 1(1), 107–126.

Lafossas F-A, Marbaix M and Menegazzi P (2007), 'Development and application of a 0D DI diesel combustion model for emissions prediction', SAE 2007-01-1841.

Langridge S and Fessler H (2002), 'Strategies for high EGR rates in a diesel engine', SAE paper 2002-01-0961.

Lee H-K, Russell M F, Bae C S and Shin H D (2002), 'Development of cavitation and enhanced injector models for diesel fuel injection system simulation', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 216, 607–618.

Leet J A, Simescu S, Froelund K, Dodge L G and Roberts C E (2004), 'Emissions solutions for 2007 and 2010 heavy-duty diesel engines', SAE paper 2004-01-0124.

Lipkea W H and DeJoode A D (1994), 'Direct injection diesel engine soot modeling: formulation and results', SAE paper 940670.

Liu H, Henein N A and Bryzik W (2003), 'Simulation of diesel engines cold-start', SAE paper 2003-01-0080.

Liu Y, Tao F, Foster D E and Reitz R D (2005), 'Application of a multiple-step phenomenological soot model to HSDI diesel multiple injection modeling', SAE paper 2005-01-0924.

Lu L, Duan S, Xiao J, Wu J and Gao X (2000), 'Effects of combustion chamber geometry on in-cylinder air motion and performance in DI diesel engine', SAE paper 2000-01-0510.

MacMillan D, Rocca A L, Shayler P J, Murphy M and Pegg I G (2008), 'The effect of reducing compression ratio on the work output and heat release characteristics of a DI diesel under cold start conditions', SAE paper 2008-01-1306.

MacMillan D, Rocca A L, Shayler P J, Morris T, Murphy M and Pegg I (2009), 'Investigating the effects of multiple pilot injections on stability at cold idle for a DI diesel engine', SAE paper 2009-01-0612.

Macián V, Luján J M, Guardiola C and Perles A (2006), 'A comparison of different methods for fuel delivery unevenness detection in diesel engines', *Mechanical Systems and Signal Processing*, 20(8), 2219–2231.

Majewski W A and Khair M K (2006), *Diesel Emissions and Their Control*, SAE International, Warrendale, PA.

Manni M, Florio S and Gommellini C (1997), 'An investigation on the reduction of lubricating oil impact on diesel exhaust emissions', SAE paper 972956.

Marcic M (1993), 'Computer simulation of the diesel fuel injection nozzle', SAE paper 930925.

Marcic M (1995), 'Calculation of the diesel fuel injection parameters', SAE paper 952071.

Mather D K and Reitz R D (1998), 'Modelling the influence of fuel injection parameters on diesel engine emissions', SAE paper 980789.

Matthews R, Hall M, Prozzi J, Machemehl R, Anthony J, Ullman T, Baker R and Lewis D (2005), 'The Taxas diesel fuels project, part 4: fuel consumption, emissions, and cost-effectiveness of an ultra-low-sulfur diesel fuel compared to conventional diesel fuels', SAE paper 2005-01-1724.

Matzke M, Litzow U, Jess A, Caprotti R and Balfour G (2009), 'Diesel lubricity requirements of future fuel injection equipment', SAE paper 2009-01-0848.

Mauviot G, Albrecht A and Poinsot T J (2006), 'A new 0D approach for diesel combustion modeling coupling probability density function with complex chemistry', SAE 2006-01-3332.

McGeehan J A, Moritz J, Shank G, Kennedy S, Totten W, Urbank M, Belay M, Goodier S, Cassim A, Runkle B, DeBaun H, Harold S, Chao K, Deere J, Herzog S, Stockwell R, Passut C, Fetterman P, Taber D, Williams L, Kleiser W M and Zalar J (2006), 'API CJ-4: New oil category for 2007 low emission diesel engines using particulate filters', *15th International Colloquium Tribology – Automotive and Industrial Lubrication*, January 17–19, Ostfildern, Germany.

Meyer S and Greff A (2002), 'New calibration methods and control systems with artificial neural networks', SAE paper 2002-01-1147.

Millich E, Bohn C, Braun H and Schultalbers M (2005), 'A highly efficient simulation-based calibration method exemplified by the charge control', SAE paper 2005-01-0052.

Millo F, Pautasso E, Delneri D and Troberg M (2007), 'A DoE analysis on the effects of compression ratio, injection timing, injector nozzle hole size and number on performance and emissions in a diesel marine engine', SAE paper 2007-01-0670.

Mitchell K (1993), 'The cold performance of diesel engines', SAE paper 932768.

Montajir R M, Tsunemoto H, Ishitani H and Minami T (2000), 'Fuel spray behavior in a small DI diesel engine: effect of combustion chamber geometry', SAE paper 2000-01-0946.

Montgomery D T and Reitz R D (2001), 'Effects of multiple injections and flexible control of boost and EGR on emissions and fuel consumption of a heavy-duty diesel engine', SAE paper 2001-01-0195.

Morel T and Wahiduzzaman S (1996), 'Modeling of diesel combustion and emissions', *XXVI FISITA Congress*, Prague, Czech Republic, June 16–23.

Mori K (1997), 'Worldwide trends in heavy-duty diesel engine exhaust emission legislation and compliance technologies', SAE paper 970753.

Mori K, Jyoutaki H, Kawai K and Sakai K (2000), 'New quiescent combustion system for heavy-duty diesel engines to overcome exhaust emissions and fuel consumption trade-off', SAE paper 2000-01-1811.

Moser F X, Sams T and Cartellieri W (2001), 'Impact of future exhaust gas emission legislation on the heavy-duty truck engine', SAE paper 2001-01-0186.

Mulemane A, Han J-S, Lu P-H, Yoon S-J and Lai M-C (2004), 'Modeling dynamic behavior of diesel fuel injection systems', SAE paper 2004-01-0536.

Najt P M and Foster D E (1983), 'Compression-ignited homogeneous charge combustion', SAE paper 830264.

Nanjundaswamy H, Tatur M, Tomazic D, Koerfer T, Lamping M and Kolbeck A (2009), 'Fuel property effects on emissions and performance of a light-duty diesel engine', SAE paper 2009-01-0488.

Narayanaswamy K, Hessel R P and Rutland C J (2005), 'A new approach to model DI-diesel HCCI combustion for use in cycle simulation studies', SAE paper 2005-01-3743.

Neumeister J, Taylor J and Gurney D (2007), 'Virtual air path calibration of a multi cylinder high performance GDI engine using 1D cycle simulation', SAE paper 2007-01-0490.

Nightingale D R (1975), 'A fundamental investigation into the problem of NO formation in diesel engines', SAE paper 750848.

Nozaki Y, Fukuma T and Tanaka K (2005), 'Development of a rule-based calibration method for diesel engines', SAE paper 2005-01-0044.

Nylund N-O, Aakko P, Mikkonen S and Niemi A (1997), 'Effects of physical and chemical properties of diesel fuel on NO_x emissions of heavy-duty diesel engines', SAE paper 972997.

Ohyama Y (2001), 'Engine control using a real-time combustion model', SAE paper 2001-01-0256.

Onder C H and Geering H P (1995), 'Model-based engine calibration for best fuel efficiency', SAE paper 950983.

Oven K and Trevor C (1995), *Automotive Fuels Reference Book*, 2nd edn, SAE International, Warrendale, PA.

Pacaud P, Perrin H and Laget O (2008), 'Cold start on diesel engine: is low compression ratio compatible with cold start requirements?', SAE paper 2008-01-1310.

Pacitti G C, Amphlett S, Miller P, Norris R and Truscott A (2008), 'Real-time crank-resolved engine simulation for testing new engine management systems', SAE paper 2008-01-1006.

Page D L (1996), 'Optimization of the air/fuel ratio for improved engine performance and reduced emissions', SAE paper 961714.

Pariotis E G, Hountalas D T and Rakopoulos C D (2005), 'Modeling the effects of EGR on a heavy duty DI diesel engine using a new quasi-dimensional combustion model', SAE paper 2005-01-1125.

Pariotis E G, Hountalas D T and Rakopoulos C D (2006), 'Sensitivity analysis of multi-zone modeling for combustion and emissions formation in diesel engines', SAE paper 2006-01-1383.

Payri F, Benajes J, Molina S and Riesco J M (2003), 'Reduction of pollutant emissions in a HD diesel engine by adjustment of injection parameters, boost pressure and EGR', SAE paper 2003-01-0343.

Pekula N, Kuritz B, Hearne J, Marchese A J and Hesketh R P (2003), 'The effect of ambient temperature, humidity, and engine speed on idling emissions from heavy-duty diesel trucks', SAE paper 2003-01-0290.

Peng H, Cui Y, Shi L and Deng K (2008), 'Improve combustion during cold start of DI diesel engine by EGR under normal ambient temperature', SAE paper 2008-01-1084.

Pfeifer A, Krueger M, Gruetering U and Tomazic D (2003), 'US 2007 – which way to go? Possible technical solutions', SAE paper 2003-01-0770.

Ponti F, Corti E, Serra G and De Cesare M (2007), 'Common rail multi-jet diesel engine combustion model development for control purposes', SAE paper 2007-01-0383.

Rabhi V, Beroff J and Dionnet F (2004), 'Study of a gear-based variable compression ratio engine', SAE paper 2004-01-2931.

Rakopoulos C D and Giakoumis E G (2009), *Diesel Engine Transient Operation – Principles of Operation and Simulation Analysis*, Springer-Verlag London, London.

Rakopoulos C D and Hountalas D T (1998), 'Development and validation of a 3-D multi-zone combustion model for the prediction of DI diesel engines performance and pollutants emissions', SAE paper 981021.

Rakopoulos C D, Hountalas D T and Agaliotis N (1999), 'Application of a multi-zone combustion model for the prediction of large scale marine diesel engines performance and pollutants emissions', SAE paper 1999-01-0227.

Ramamurthy R, Clark N N, Atkinson C M and Lyons D W (1998), 'Models for predicting transient heavy duty vehicle emissions', SAE paper 982652.
Rask E and Sellnau M (2004), 'Simulation-based engine calibration: tools, techniques, and applications', SAE paper 2004-01-1264.
Regueiro J F (2001), 'The case for new divided-chamber diesel combustion systems, part three: problems with current DI engines and divided-chamber solutions for future diesel engines', SAE paper 2001-01-0278.
Reitz R D (1998), 'Controlling DI diesel engine emissions using multiple injections and EGR', *Combustion Science and Technology*, 138(1–6), 257–278.
Reitz R D and Rutland C J (1995), 'Development and testing of diesel engine CFD models', *Progress in Energy and Combustion Science*, 21(2), 173–196.
Reitz R D and Sun Y (2010), 'Advanced computational fluid dynamics modeling of direct injection engines', in Zhao H (ed.), *Advanced Direct Injection Combustion Engine Technologies and Development, Volume 2: Diesel Engines*, Woodhead Publishing, Cambridge.
Ribeiro N M, Pinto A C, Quintella C M, da Rocha G O, Teixeira L S G, Guarieiro L L N, Rangel M C, Veloso M C C, Rezende M J C, da Cruz R S, de Oliveira A M, Torres E A and de Andrade J B (2007), 'The role of additives for diesel and diesel blended (ethanol or biodiesel) fuels: a review', *Energy & Fuels*, 21, 2433–2445.
Roberts M (2003), 'Benefits and challenges of variable compression ratio (VCR)', SAE paper 2003-01-0398.
Ryan T W, Challahan T J and Mehta D (2004), 'HCCI in a variable compression ratio engine – effects of engine variables', SAE paper 2004-01-1971.
SAE Surface Vehicle Information Report J1498 (2005), 'Heating value of fuels'.
SAE Surface Vehicle Standard J900 (1995), 'Crankcase emission control test code'.
SAE Surface Vehicle Standard J2265 (1995), 'Diesel engines – diesel fuel – performance requirement and test method for assessing fuel lubricity'.
SAE Surface Vehicle Standard J313 (2004), 'Diesel fuels'.
Samuel S, Austin L and Morrey D (2002), 'Automotive test drive cycles for emission measurement and real-world emission levels – a review', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 216, 555–564.
Sasaki M, Kishi Y, Hyuga T, Okazaki K, Tanaka M and Kurihara I (1997), 'The effect of EGR on diesel engine oil, and its countermeasures', SAE paper 971695.
Schleyer C H (2006), 'HCCI for heavy duty vehicles: new development and future prospects', *Hart World Refining & Fuels Conference*, May 30–June 1, Brussels.
Schlosser A, Kinoo B, Salber W, Werner S and Ademes N (2006), 'Accelerated powertrain development through model based calibration', SAE paper 2006-01-0858.
Schmidt D P and Corradini M L (2001), 'The internal flow of diesel fuel injector nozzles: a review', *International Journal of Engine Research*, 2(1), 1–22.
Schmitz G, Oligschlager U, Eifler G and Lechner H (1994), 'Automated system for optimized calibration of engine management systems', SAE paper 940151.
Schulze T, Wiedemeier M and Schuette H (2007), 'Crank angle-based diesel engine modeling for hardware-in-the-loop applications with in-cylinder pressure sensors', SAE paper 2007-01-1303.
Shade B C, Carder D K, Thompson G J and Gautam M (2008), 'A work-based window method for calculating in-use brake-specific NO_x emissions of heavy-duty diesel engines', SAE paper 2008-01-1301.
Shin B-S, Lyu M-S, Kim C-H and Choi K-H (2001), 'Combustion system development in a small bore HSDI diesel engine for low fuel consuming car', SAE paper 2001-01-1257.

Shiozaki T, Nakajima H, Kudo Y, Miyashita A and Aoyagi Y (1996), 'The analysis of combustion flame under EGR conditions in a DI diesel engine', SAE paper 960323.

Shirota K, Baba F, Horikoshi M and Sato K (2001), 'Development of virtual and real simulator or engine', SAE paper 2001-01-1355.

Siewert R M (1978), 'Engine combustion at large bore-to-stroke ratios', SAE paper 780968.

Simescu S, Ryan III T W, Neely G D, Matheaus A C and Surampudi B (2002), 'Partial pre-mixed combustion with cooled and uncooled EGR in a heavy-duty diesel engine', SAE paper 2002-01-0963.

Singal S K and Pundir B P (1996), 'Diesel fuel quality and particulate emissions: an overview', SAE paper 961185.

Singal S K, Pundir B P and Mehta P S (1993), 'Fuel spray-air motion interaction in DI diesel engines: a review', SAE paper 930604.

Song C, Hsu C S and Mochida I (2000), *Chemistry of Diesel Fuels*, Taylor & Francis, New York.

Song H, Ladommatos N and Zhao H (2001), 'Diesel soot oxidation under controlled conditions', SAE paper 2001-01-3673.

Stan C (ed.) (1999), *Direct Injection Systems for Spark-Ignition and Compression-Ignition Engines*, SAE International, Warrendale, PA.

Stanglmaier R H and Roberts C E (1999), 'Homogeneous charge compression ignition (HCCI): benefits, compromises, and future engine applications', SAE paper 1999-01-3682.

Stuhler H, Kruse T, Stuber A, Gschweitl K, Piock W, Pfluegl H and Lick P (2002), 'Automated model-based GDI engine calibration adaptive online DoE approach', SAE paper 2002-01-0708.

Stunnenberg F, Kleijwegt P and Feyens A W L V (2001), 'Future heavy duty diesel lubricants for low emission engines', SAE paper 2001-01-3768.

Suzuki T, Kakegawa T, Hikino K and Obata A (1997), 'Development of diesel combustion for commercial vehicles', SAE paper 972685.

Taylor S, Clark N, Gautam M and Wayne W S (2004), 'Diesel emissions prediction from dissimilar cycle scaling', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 218, 341–352.

The Mathworks (2007), 'SAE 2007 SAE Commercial Vehicle Engineering Congress – MathWorks hosted panel on analytical calibration', SAE Panel Discussion. Available at: www.mathworks.com/mason/tag/proxy.html?dataid=9962&fileid...

Thompson G J, Atkinson C M, Clark N N, Long T W and Hanzevack E (2000), 'Neural network modelling of the emissions and performance of a heavy-duty diesel engine', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 214, 111–126.

Thompson G J, Carder D K, Clark N N and Gautam M (2008), 'Summary of in-use NO_x emissions from heavy-duty diesel engines', SAE paper 2008-01-1298.

Thompson N, Wass U, Ntziachristos L, Samaras Z, Hausberger S, Aakko P and Sams T (2004), 'Overview of the European "particulates" project on the characterization of exhaust particulate emissions from road vehicles: results for heavy duty engines', SAE paper 2004-01-1986.

Thring R M (1989), 'Homogeneous charge compression ignition (HCCI) engines', SAE paper 892068.

Toback A T, Hearne J S, Kuritz B, Marchese A J and Hesketh R P (2004), 'The effect of ambient temperature and humidity on measured idling emissions from diesel school buses', SAE paper 2004-01-1087.

Tomazic D and Pfeifer A (2002), 'Cooled EGR – a must or an option for 2002/04', SAE paper 2002-01-0962.
Tomita M, Aoyama S, Ushijima K, Tanaka Y, Takahashi N, Moteki K and Takaba T (2007), 'Compact and long-stroke multiple-link VCR engine mechanism', SAE paper 2007-01-3991.
Torkzadeh D D, Längst W and Kiencke U (2001), 'Combustion and exhaust gas modeling of a common rail diesel engine – an approach', SAE paper 2001-01-1243.
Traver M L, Atkinson R J and Atkinson C M (1999), 'Neural network-based diesel engine emissions prediction using in-cylinder combustion pressure', SAE paper 1999-01-1532.
Tsuchida H, Hiraya K, Tanaka D, Shigemoto S, Aoyama S, Tomita M, Sugiyama T and Hiyoshi R (2007), 'The effect of a longer stroke on improving fuel economy of a multiple-link VCR engine', SAE paper 2007-01-4004.
Ullmann S, Reuss H-C and Zell A (2005), 'Calibration system prototype for increasing the level of automation in stationary engine testing and calibration', SAE paper 2005-01-1828.
Urano Y, Nakano Y, Takada H and Sugita M (2005), 'Optimization technique for transient emission reduction of heavy duty diesel engine', SAE paper 2005-01-1099.
Vora K A, Clark N N, Gautam M and Wayne W S (2005), 'Multidimensional correlation study using linear regression of PM and NO_x for heavy duty diesel vehicles', SAE paper 2005-01-1618.
Walsh M P (1999), 'Global trends in diesel emissions control – a 1999 update', SAE 1999-01-0107.
Walsh M P (2001), 'Global trends in diesel emissions regulation – a 2001 update', SAE 2001-01-0183.
Wang G, Li G, Liu Y, Chen L, Zhang X and Lu J (1999), 'A developed model for emissions prediction of a DI diesel engine', SAE paper 1999-01-0233.
Wickman D D, Tanin K V, Senecal P K, Reitz R D, Gebert K, Barkhimer R L and Beck N J (2000), 'Methods and results from the development of a 2600 bar diesel fuel injection system', SAE paper 2000-01-0947.
Wickman D D, Senecal P K and Reitz R D (2001), 'Diesel engine combustion chamber geometry optimization using genetic algorithms and multi-dimensional spray and combustion modeling', SAE paper 2001-01-0547.
Wilhelmsson C, Tunestål P, Johansson B, Widd A and Johansson R (2009), 'A physical two-zone NO_x model intended for embedded implementation', SAE paper 2009-01-1509.
Woermann R J, Theuerkauf H J and Heinrich A (1999), 'A real-time model of a common rail diesel engine', SAE paper 1999-01-0862.
Woods M, Kamo R and Bryzik W (2000), 'High pressure fuel injection for high power density diesel engines', SAE paper 2000-01-1186.
Wuensche P, Moser F X, Dreisbach R and Sams T (2003), 'Can the technology for heavy duty diesel engines be common for future emission regulations in USA, Japan and Europe?', SAE paper 2003-01-0344.
Yamanishi M (2003), 'Application of a computer model to various specifications of fuel injection system for DI diesel engines', *JSME International Journal*, *Series B*, 46(2), 326–331.
Yang B, Mellor A M and Chen S K (2002), 'Multiple injections with EGR effects on NO_x emissions for DI diesel engines analyzed using an engineering model', SAE paper 2002-01-2774.

Yoshizaki T, Nishida K and Hiroyasu H (1993), 'Approach to low NO_x and smoke emission engines by using phenomenological simulation', SAE paper 930612.

Yun H and Reitz R D (2005), 'Combustion optimization in the low-temperature diesel combustion engine', *International Journal of Engine Research*, 6, 513–524.

Zannis T C, Hountalas D T, Papagiannakis R G and Levendis Y A (2008), 'Effect of fuel chemical structure and properties on diesel engine performance and pollutant emissions: review of the results of four European research programs', SAE paper 2008-01-0838.

Zelenka P, Aufinger H, Reczek W and Cartellieri W (1998), 'Cooled EGR – a key technology for future efficient HD diesels', SAE paper 980190.

Zhang H (1998), 'A predictive tool for engine performance and NO_x emission', SAE paper 982462.

Zhang J, Gao S and Jiang F (2007), 'A diesel engine real time NO_x emission simulation system based on RTW and VxWorks', SAE paper 2007-01-0025.

Zhao F, Asmus T W, Assanis D N, Dec J E, Eng J A and Najt P M (eds) (2003), *Homogeneous Charge Compression Ignition (HCCI) Engines – Key Research and Development Issues*, SAE International, Warrendale, PA.

Zhao H (ed.) (2007), *HCCI and CAI Engines for the Automotive Industry*, Woodhead Publishing, Cambridge.

Zhao H (2010), *Advanced Direct Injection Combustion Engine Technologies and Development, Volume 2: Diesel Engines*, Woodhead Publishing Limited, Cambridge, UK.

Zheng J and Xin Q (2009), 'Theoretical analysis of diesel engine NO_x and soot with heuristic macro-parameter-dependent approach and virtual multi-zone real-time models', SAE paper 2009-01-2836.

Zheng J and Xin Q (2009a), 'Theoretical analysis of diesel engine NO_x and soot with heuristic macro-parameter-dependent approach and virtual multi-zone real-time models', SAE paper 2009-01-2836, also in *SAE International Journal of Commercial Vehicles*, March 2010, 2 (2), 45–60.

Zhong L, Gruenewald S, Henein N A and Bryzik W (2007), 'Lower temperature limits for cold starting of diesel engine with a common rail fuel injection system', SAE paper 2007-01-0934.

Zhu Y, Zhao H and Ladommatos N (2003), 'Computational study of the effects of injection timing, EGR and swirl ratio on a HSDI multi-injection diesel engine emission and performance', SAE paper 2003-01-0346.

Zhu Y, Ricart-Ugaz L, Wu S, Cigler J, El-Beshbeeshy M, Bulicz T and Yan J (2004), 'Combustion development of the new international 6.0 L V8 diesel engine', SAE paper 2004-01-1404.

8
Diesel aftertreatment integration and matching

Abstract: This chapter addresses one of the central tasks in diesel engine system design, engine–aftertreatment matching. It begins with an overview of aftertreatment requirements on engine system design, including the performance of diesel oxidation catalyst (DOC), urea-based selective catalytic reduction (SCR), solid ammonia storage and release (SASR), lean NO_x trap (LNT) and diesel particulate filter (DPF). The chapter then discusses exhaust thermal management and aftertreatment calibration, followed by a description of DPF regeneration requirements for engine system design. The chapter is concluded by an analytical approach of engine–aftertreatment integration.

Key words: engine–aftertreatment matching, aftertreatment calibration, diesel oxidation catalyst (DOC), selective catalytic reduction (SCR), solid ammonia storage and release (SASR), lean NO_x trap (LNT), diesel particulate filter (DPF), DPF regeneration.

8.1 Overview of aftertreatment requirements on engine system design

Current design and calibration strategies for the reduction of diesel engine-out emissions are approaching their practical limits. The diesel particulate filter (DPF) is required for the US 2007 model year, and NO_x aftertreatment may be considered in the future. Diesel aftertreatment devices exhibit several characteristics which influence the engine design and operation: (1) performance variation (e.g., the change of DPF pressure drop as soot builds up); (2) the need for regeneration and flexible engine controls; (3) performance being closely related to vehicle duty cycle; and (4) the interactive chemistry nature between the different aftertreatment devices. There is a large body of literature on aftertreatment performance itself. However, there is very little literature on either the approach of engine–aftertreatment matching or the analysis of aftertreatment integration in engine system design and calibration.

Diesel exhaust gases exhibit several unique characteristics compared to the gasoline engine: lower HC and CO; higher particulate matter (PM); lower exhaust temperature due to higher air–fuel ratio; and oxidizing exhaust gas chemistry suppressing the chemical reduction of NO_x. Modern diesel exhaust aftertreatment devices consist of a diesel oxidation catalyst (DOC), a diesel particulate filter (DPF) and possibly a NO_x aftertreatment device such as selective catalytic reduction (SCR) or lean NO_x trap (LNT, also known as NO_x adsorber). The muffler can usually be replaced by a wall-flow DPF in commercial vehicles, or it may be used to further reduce the exhaust noise. The emissions formation mechanisms in the diesel engine are provided in detail by Eastwood (2000, 2008 and 2010) and Majewski and Khair (2006). Exhaust aftertreatment control technologies are reviewed by Khair and McKinnon (1999), Johnson (2000a, 2000b, 2001, 2002, 2003, 2004, 2006, 2007a, 2007b, 2008a, 2008b, 2009a, 2009b and 2010), Graves *et al*. (2001), Corro (2002), Blakeman *et al*. (2003), Edgar *et al*. (2003), Eastwood (2000, 2008), and Majewski and Khair (2006).

8.1.1 Diesel oxidation catalyst (DOC) performance

The DOC reduces HC (including PAH), CO, SOF in PM, and diesel exhaust odor by converting them to H_2O and CO_2. The de-NO_x capability of the DOC in the presence of HC is insignificant so that the NO_x level is almost unaffected. The level of nanoparticles is not affected either. The combustion of a high-sulfur diesel fuel produces SO_2, which is converted to SO_3 and a large amount of sulfate particulate in DOC at high engine loads under high exhaust temperatures. Therefore, the SOF removal ability of the DOC may be compromised by the addition of the sulfate particulate matter. Using a low-sulfur diesel fuel can reduce the sulfate particulate matter and avoid catalyst poisoning. Because NO_2 may assist the DPF regeneration, the oxidation of NO to NO_2 by O_2 and the reduction of NO_2 by HC in the DOC need to be optimized. DOC performance was investigated by Phillips *et al*. (1999), Khair and McKinnon (1999), Nieuwstadt *et al*. (2005), and Kozlov *et al*. (2010) and reviewed by Majewski and Khair (2006). A real-time-capable DOC model was developed by Wahiduzzaman *et al*. (2008) and Wenzel *et al*. (2009). An engine CO emission model was developed by Bagal *et al*. (2009).

The efficiency of the catalytic converters is related to air–fuel ratio, light-off temperature and light-off time, space velocity, flow restriction characteristics and the durability features. The light-off temperature is defined as the temperature at which 50% conversion efficiency is achieved. The space velocity is the ratio of the normalized exhaust volume flow rate to the aftertreatment component volume. The flow restriction can be estimated by using the curve of 'pressure drop vs. flow rate'. The light-off temperatures of HC and CO become lower when the air–fuel ratio is higher. The three-way catalytic converter used in the gasoline engine has a high conversion efficiency above 80% simultaneously for CO, HC, and NO_x only when the excess air–fuel ratio stays around 1 (i.e., at the stoichiometric ratio) with a narrow fluctuation window about 0.01. Diesel engines operate under the 'lean burn' conditions with a much higher air–fuel ratio due to the heterogeneous nature in diesel combustion and the high PM emissions produced. In the oxidizing exhaust environment, the reductants such as HC and CO are more reactive to the excess oxygen rather than to the NO_x. Therefore, the three-way catalyst used in the gasoline engine usually cannot be used in the diesel engine although some research was conducted on using a three-way catalyst on diesel engines to explore the potential of emissions control (e.g., Sung *et al*., 2009; Simescu *et al*., 2010). Urea-based SCR, solid ammonia storage and release, and LNT are three promising technologies for diesel NO_x aftertreatment with up to 80–90% NO_x reduction. They all require reductants such as urea ($CO(NH_2)_2$), ammonia (NH_3), or hydrocarbons.

8.1.2 Urea-based selective catalytic reduction (SCR) performance

The urea-based SCR is believed by many people in the industry to be more suitable than the LNT for heavy-duty applications. Urea is used to generate ammonia by thermal decomposition. The aqueous urea is sprayed into the exhaust line, where under ideal conditions it evaporates and decomposes completely to ammonia and CO_2. In the SCR, ammonia is less reactive to oxygen than the hydrocarbons, and can be mainly reduced by NO_x rather than the other oxidant, O_2. Therefore, the NO_x conversion efficiency of the SCR is greater than the active de-NO_x catalytic reduction using hydrocarbons. In the SCR, NO_x is reduced to nitrogen and water. The dosing rate of the urea must match the temperature, the exhaust composition, and the space velocity of the exhaust flow. Precise dynamic dosage control is required for the SCR in order to prevent excessive slip of the toxic ammonia during engine transients in automotive applications. A second DOC may be added downstream of the SCR to convert the slipped ammonia to nitrogen. The freezing point of the urea in aqueous solution at –11°C could be a problem. The ammonium sulfate accumulated in the SCR needs to be purged at high temperature periodically. Moreover, the SCR is effective in removing HC and some SOF. Urea-based SCR performance is elaborated by Müller *et al*. (2003), Scarnegie *et al*. (2003), Klingstedt *et al*. (2006), Hosoya *et al*. (2007), Iitsuka *et al*. (2007), and Hirata *et al*. (2009). Urea-based SCR control has been investigated by Song and Zhu (2002), Willems *et al*. (2007), Devarakonda *et al*. (2008), Wang *et al*. (2008), and Herman *et al*. (2009). Urea-based SCR modeling has been conducted by Chi and DaCosta (2005) and Kim *et al*. (2007).

8.1.3 Solid ammonia storage and release for NO_x reduction

The urea-based SCR is an aqueous solution that has a number of issues for automotive applications, for example, large storage volume, freezing at cold climate, injector dosing reliability, risk of deposits with urea hydrolysis at low exhaust temperatures, and inconvenience for the end user. Another novel and promising technology for reducing the NO_x from the exhaust of diesel vehicles is to use solid ammonia storage and release (SASR). It does not have the above-mentioned shortcomings of the urea-based SCR. It uses compact solid metal ammine complexes (e.g., $M_g(NH_3)_6Cl_2$, Elmoe *et al.*, 2006) as a source to safely store ammonia and provides a controlled release in gas phase on demand for NO_x reduction. It can store a large amount of ammonia in the solid metal ammine complex with a high volumetric density that is very similar to pure liquid ammonia (NH_3), and hence provide long lasting ammonia storage on the vehicle. It requires much smaller storage volume than the urea-based SCR. Upon heating, the solid complex uses dynamic, accurate, and controlled dosing release of pure ammonia gas directly to the exhaust line of the engine, and does not require an injector for the dosing. This technology can perform reasonably well at all ambient temperatures and exhaust temperatures. It does not have the freezing problem and the issue of low urea-conversion rate at low exhaust temperatures as occurring with the urea-based SCR. More information on solid ammonia storage and release can be found in Elmoe *et al.* (2006), Johannessen *et al.* (2009), and Chakraborty *et al.* (2009).

8.1.4 Lean NO_x trap (LNT) performance

LNT converts NO to NO_2 through an oxidation catalyst when it is exposed to the oxidizing exhaust. Then, it adsorbs NO_2 as nitrates with alkaline earth compounds such as $Ba(NO_3)_2$ through acid–base neutralization reactions. The LNT is periodically and briefly regenerated to restore its NO_x reduction capability by using a reductant or electric heating. The working temperature window of the LNT is bounded by NO light-off on the lower side and by the nitrate instability on the upper side. The LNT regeneration in the diesel engine is more challenging than that in the lean-burn GDI gasoline counterpart because a rich air–fuel ratio cannot be utilized in the diesel engine due to concerns about smoke production. The LNT performance is very sensitive to sulfur content in the fuel, even as low as 3 ppm. Therefore, desulfation or other sulfur management is needed. The desulfation usually lasts several minutes with a rich air–fuel ratio at high temperatures. Other sulfur management includes using an ultra-low sulfur fuel or a sulfur-free fuel or sulfur traps. The LNT performance has been researched by Sluder and West (2001), Parks *et al.* (2002), Takahashi *et al.* (2004), Hinz *et al.* (2005), Theis *et al.* (2005), Hu *et al.* (2006), Nam *et al.* (2007), McCarthy Jr and Holtgreven (2008), Ottinger *et al.* (2009), and McCarthy *et al.* (2009). LNT modeling has been conducted by Kim *et al.* (2003) and He (2006).

8.1.5 Diesel particulate filter (DPF) performance

Achieving high filtration efficiency and low flow restriction with small dimensions and low costs is a challenging task in DPF design. The exhaust restriction at the turbine outlet refers to the pressure drop through all the aftertreatment components plus the exhaust pipes. Exhaust restriction depends on hardware design and the flow restriction characteristics of the components. The pressure drop of the DPF increases as the soot and the ash accumulate in the filter. It also varies with the amount of soot deposits in the filter before and after the DPF regeneration. Exhaust restriction also can be regulated by using an exhaust back-pressure valve, such as a flap valve installed at the turbine outlet. Closing the back-pressure valve reduces the exhaust flow rate or may assist engine braking. Exhaust restriction has a large impact on turbocharged engine performance.

DPF technologies are reviewed and summarized in three comprehensive books by Majewski and Khair (2006), Johnson (2007b), and Eastwood (2008). There is no need to elaborate on them here. The literature on DPF research areas is organized as follows:

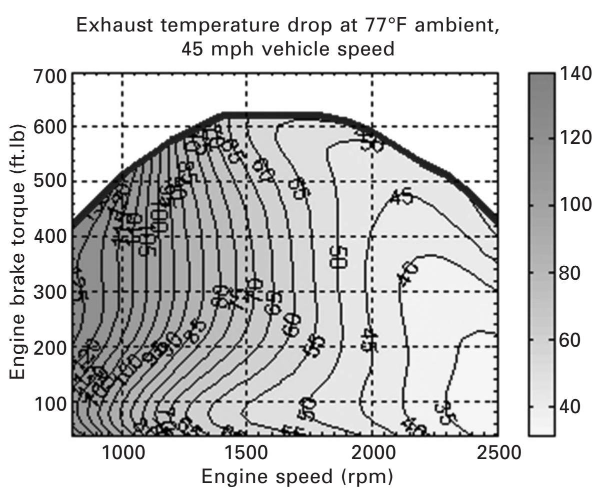
- The DPF selection for retrofitting vehicles is discussed by Mayer *et al.* (2001). DPF design effects were studied by Konstandopoulos *et al.* (1999, 2005b), Nikitidis *et al.* (2001), Merkel *et al.* (2001), Konstandopoulos and Kladopoulou (2004), Soeger *et al.* (2005), Yamaguchi *et al.* (2005), and Ido *et al.* (2005).
- DPF general performance was studied by Khair (2003), Khair and McKinnon (1999), Toorisaka *et al.* (2004), Herrmuth *et al.* (2004), Cutler (2004), and Kapetanovic *et al.* (2009).
- DPF pressure drop was researched by Taoka *et al.* (2001), Konstandopoulos *et al.* (2001b), Stratakis *et al.* (2002), Konstandopoulos (2003), Cunningham *et al.* (2007), and Ohyama *et al.* (2008).
- DPF regeneration performance was experimentally investigated by Tan *et al.* (1996), Gantawar *et al.* (1997), Park *et al.* (1998), Bouchez and Dementhon (2000), Salvat *et al.* (2000), Gieshoff *et al.* (2001), Locker *et al.* (2002), Hiranuma *et al.* (2003), Flörchinger *et al.* (2004), Mayer *et al.* (2005), Kong *et al.* (2005), Ogyu *et al.* (2007), and Ootake *et al.* (2007).
- DPF regeneration control algorithms were developed by Brewbaker and Nieuwstadt (2002), Nieuwstadt and Trudell (2004), Birkby *et al.* (2006), and Bencherif *et al.* (2009).
- DPF modeling has been carried out mainly by the following groups:
    - ➢ A group of researchers primarily at Aerosal & Particle Technology Laboratory in Greece (review papers by Konstandopoulos *et al.*, 2000, 2005a; Konstandopoulos and Kostoglou, 1999, 2004; Masoudi *et al.*, 2001; Konstandopoulos *et al.*, 2001a, 2002, 2003a, 2003b, 2004; Kladopoulou *et al.*, 2003).

- A group of researchers at Michigan Technological University (Huynh *et al.*, 2003; Kladopoulou *et al.*, 2003; Singh *et al.*, 2005; Mohammed *et al.*, 2006a, 2006b; Premchand *et al.*, 2007).
- A group of researchers at General Motors and the University of Wisconsin–Madison (Kapparos *et al.*, 2005; Strzelec *et al.*, 2006; England *et al.*, 2006; He, 2007; Rutland *et al.*, 2007; Gurupatham and He, 2008; Singh *et al.*, 2009).
- A group of researchers at Gamma Technologies for the software GT-POWER (Tang *et al.*, 2007, 2008; Wahiduzzaman *et al.*, 2007).
- Other sources (Rumminger *et al.*, 2001; Millet *et al.*, 2002; Kandylas and Koltsakis, 2002; Liu and Miller, 2002; Guo and Zhang, 2005a, 2005b; York *et al.*, 2005, 2009; Reader *et al.*, 2006; Yi, 2006; Bouteiller *et al.*, 2007; Cunningham and Meckl, 2007; Chiatti *et al.*, 2008; Subramaniam *et al.*, 2009).

8.1.6 Exhaust thermal management

The pressure drop of the engine exhaust system is important for engine performance. The thermal loss of the exhaust system also has a large impact on aftertreatment performance. Modeling the exhaust system with one-dimensional methods is discussed by Massey *et al.* (2002). The thermal losses of the diesel engine exhaust system are investigated by Kapparos *et al.* (2004) and Fortunato *et al.* (2007).

Figure 8.1 shows the exhaust gas temperature drop from the turbine outlet to the DOC inlet for a heavy-duty diesel engine. It is observed that

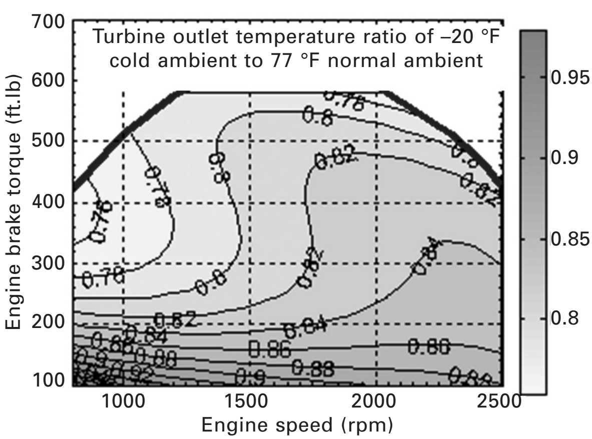


8.1 Simulation of exhaust temperature drop from turbine outlet to DOC inlet.

the temperature drop is primarily a function of the engine speed, which is related to the time scale in the exhaust pipe heat transfer. The temperature drop becomes greater as the vehicle speed decreases. The temperature drop is affected by the heat transfer loss of the exhaust pipe, which is in turn affected by engine exhaust flow rate, ambient temperature, vehicle speed and the orientation of the aftertreatment configuration (horizontal or vertical). The exhaust gas temperature decreases when the engine operates in cold climate (Fig. 8.2).

Flexibility of controlling the exhaust gas temperature and the flow rate is another important topic in air system design for engine thermal management and optimum aftertreatment performance. For example, the functions of NO_x adsorber during the adsorption phase and reduction phase require different exhaust temperatures and oxygen concentrations for the optimum performance. In order for the converter to reach the light-off temperature more easily or to carry out the aftertreatment regeneration more quickly, the following measures can be used to raise the exhaust temperature:

- minimizing the heat losses by either insulating the exhaust pipe or installing the aftertreatment device close to the exhaust manifold or runner
- raising the intake manifold gas temperature by flexible cooling or bypassing the coolers on the gas side
- raising the exhaust temperature by bypassing the turbine
- retarding the fuel injection timing
- using the in-cylinder post fuel injection (with engine torque balance) to reduce the air–fuel ratio and provide the reductant needed for the NO_x adsorber regeneration or the DPF regeneration



8.2 Simulation of exhaust temperature ratio between cold ambient and normal ambient at turbine outlet.

- using external fuel dosing in the exhaust stream
- reducing the engine air flow rate by intake throttling, exhaust throttling, or variable valve actuation to raise the exhaust temperature.

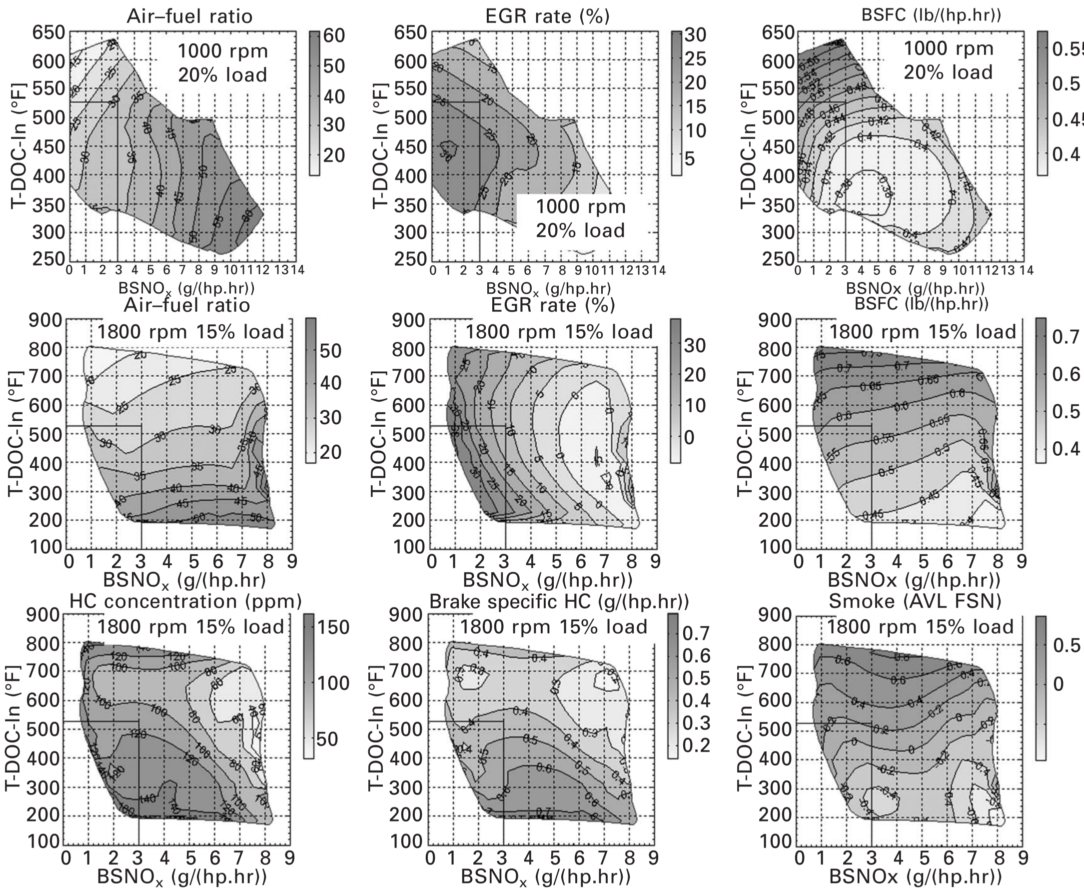
In some regeneration strategies, the EGR is turned off or reduced during the regeneration in order to make the exhaust gas hotter. Increasing the NO_x:C ratio by raising the concentration of feed-gas NO_x can assist soot burning. Hotter exhaust gas at the DOC outlet due to its exothermic reaction and faster light-off also helps to initiate the DPF regeneration.

8.1.7 Aftertreatment calibration

The DoE method can be used in aftertreatment calibration for the DPF active regeneration in order to optimize the effect of aftertreatment calibration parameters on engine performance and emissions. The main fuel injection timing and the air system control parameters such as intake throttle opening, VGT vane opening, and EGR valve opening are adjusted in order to increase the turbine outlet exhaust gas temperature to obtain a high light-off temperature of the DOC. Once the DOC is heated up, it can produce high-efficiency oxidization of the post-injected fuel for the DPF regeneration. Based on the DOC light-off temperature achieved, the in-cylinder post fuel injection along with further modulation of the air system parameters can be used to reach a sufficiently high inlet gas temperature for the DPF regeneration. The objective of the aftertreatment calibration optimization is to minimize the BSFC under the constraints of acceptable NO_x, soot, and HC emissions at the DOC inlet, a sufficiently high turbine outlet temperature, and an undetectably small variation of the engine torque for driving smoothness. Figure 8.3 shows an example of the advanced DoE optimization data at part load conditions for the DPF regeneration of a heavy-duty diesel engine. The control factors used in the DoE test include intake throttle opening, VGT vane opening, EGR valve opening, and main fuel injection timing. The 'T-DOC-In' shown on the vertical axis of the figures represents the turbine outlet exhaust gas temperature before the DOC. Engine BSFC is minimized in the entire domain of 'NO_x vs. T-DOC-In' by using the emulators built from the DoE test data. These maps illustrate the parametric dependency of the emission parameters and the exhaust temperature upon the air–fuel ratio and the EGR rate.

8.1.8 Cold-start emissions control

The engine-out and aftertreatment-out emissions during transient cold-start have received much greater attention since the 1990s after the type approval procedures in emission regulations became more stringent, especially for



8.3 Advanced aftertreatment calibration optimization for DPF regeneration.

gasoline engines. The diesel engine is less sensitive than the gasoline engine in cold-start emissions, although the diesel engine may experience problems of ‘white smoke’ formed from evaporated fuel droplets. Close-coupled catalysts for cold-start emissions control may affect the HC:NO_x ratio in the exhaust flow and hence the aftertreatment recipe. Overall, the emissions control at cold-start in a cold climate is more of a local design issue related to catalyst warm-up or cold-start aid rather than a system design focus.

8.2 Diesel particulate filter (DPF) regeneration requirements for engine system design

The DPF removes PM including the nanoparticles if a high filtration efficiency is achieved. Engine–DPF matching requires careful integration. The DPF is more complex than the DOC because the DOC is entirely passive, reacting continuously, and does not require regeneration. The DPF is usually the opposite.

A soot load factor is defined as the ratio of the DPF pressure drop at the fully loaded condition to the pressure drop at the clean condition (i.e., no soot accumulated). The regeneration efficiency is defined as the ratio of the pressure drop reduction after the regeneration to the pressure drop reduction after an ideal complete removal of the soot. The soot load factor and the DPF size are determined by the following: (1) the exhaust restriction level allowed for acceptable engine performance; (2) the regeneration frequency; (3) the peak burning temperature; (4) the temperature gradient during the exothermic regeneration process related to DPF thermo-mechanical durability; and (5) the regeneration efficiency.

The regeneration efficiency is defined by

$$\eta_{regen} = \frac{\Delta p_{loaded} - \Delta p_{regen}}{\Delta p_{loaded} - \Delta p_{clean}} \tag{8.1}$$

where Δp_{loaded} is the pressure drop of a soot-loaded DPF, Δp_{regen} is the pressure drop after the regeneration, and Δp_{clean} is the pressure drop of a clean DPF associated with certain incombustible ash loading. The filters that are only partially loaded are more difficult to regenerate and to maintain self-sustaining soot combustion. When the soot load accumulated in the DPF exceeds a certain level, it is desirable to have a regeneration, either passive or active, in order to burn off the soot to reduce the exhaust restriction. Uncontrolled regeneration occurs with too much accumulation of soot and deposited hydrocarbons. It generates excessively high burning temperatures or temperature gradients to melt or crack the filter. The related factors in catalytic soot regeneration include mainly the following: exhaust temperature; the concentrations of O_2 and NO_2 (for C+O_2 and C+NO_2 reactions of the carbon burning); the NO_2:NO (1) ratio; and the NO_x:C mass ratio. The following means can assist soot burning: higher temperature and higher O_2 concentration in the exhaust flow; (2) more deposited hydrocarbons (SOF) and its exothermic burning reaction to provide energy to ignite the soot; (3) lower space velocity or volumetric flow rate; and (4) lower soot loading.

DPF durability is related to the filter material and the design, the regeneration burning temperature, the temperature spatial and temporal gradients, and the regeneration frequency (related to the fatigue life). The regeneration frequency is affected by the installation position and the soot loading. It is also related to the filter size, the soot load-up rate and the driving cycle. Small filters located closer to the engine under high-load driving need to regenerate more frequently. DPF durability is discussed in more detail in Chapter 2.

In the non-catalyzed DPF regeneration, the soot burning lasts a few minutes and requires approximately 550–650°C exhaust temperature to start, which cannot be reached in most part-load operating conditions. It was reported that a complete burn of soot requires an exhaust temperature of above 600°C and an oxygen concentration of above 7% (Basshuysen and Schafer, 2004). In the catalyzed DPF (CDPF) regeneration, additives in the fuel or the catalytic coating inside the filter may be used in order to reduce the thermal energy required to initiate the regeneration. The soot ignition temperature can be significantly reduced to the range of 275–450°C. The catalyst, either for the SOF or for the carbon soot, is usually more effective at a higher oxygen concentration in the exhaust flow. The ignition and burning performance of the catalytic regeneration also depends on the amount of NO_x, the fuel sulfur level, and the compositions of SOF and PM.

Passive regeneration, also known as self regeneration, is usually assisted by catalytic means to lower the soot oxidation temperature. It offers certain advantages such as not using sensing/control systems and hence eliminating their durability troubles. In the passive catalytic regeneration, there is no control over the soot load, the pressure drop of the DPF, and the regeneration efficiency before and after the regeneration because the driving condition can initiate or halt the regeneration. Once started, the combustion of soot may become self-sustaining, and thermal runaway could happen. Therefore, low soot loading is preferred in the passive regeneration in order to avoid catastrophic thermal failures.

In the active DPF regeneration, when the Δp_{loaded} exceeds a certain preset threshold, regeneration is triggered to occur. In addition to the above-mentioned measures of raising the exhaust temperature, other methods may also be adopted, some with additional costs, in order to precisely control the course of the regeneration and the regeneration efficiency. The methods include: (1) using a fuel-fed burner; (2) electrical heating; (3) microwave; (4) compressed air; (5) using controlled pre-DOC or post-DOC dosing of hydrocarbons as a catalyst to generate sufficient heat energy to regenerate the DPF; (6) using a transmission to force the engine to operate at a high load; and (7) producing oxidants such as NO_2 turnover from the NO in the DOC or the CDPF in order to oxidize the carbon soot (with reactions: $NO + 0.5O_2 = NO_2$, $2NO = NO_2 + 0.5N_2$, $C + NO_2 = NO + CO$). It is noted that NO_2 is much more toxic than NO; but after leaving the tailpipe, NO will be converted to NO_2 in the atmosphere anyway. For low engine-out NO_x emissions, there is an impact on the NO_2 oxidant available for the DPF regeneration.

A dynamic and precise control of the DPF regeneration process is needed for several reasons. A dangerous phenomenon called 'hill-cresting' refers to the thermal runaway (meltdown) with uncontrolled, excessively high soot-burning temperatures inside the DPF when the regeneration occurs immediately before the end of a high-load condition. In other words, thermal runaway occurs when the engine speed/load decreases during the course of the DPF regeneration. After the engine load or the EGR rate decreases or the speed changes, the air–fuel ratio may increase. The exhaust flow rate may become lower, resulting in a reduction in the convective cooling rate inside the DPF. Consequently, the temperature may rise to unacceptably high levels to cause durability issues. Moreover, it is important to slow down the regeneration burn-off rate and to control the temperature spatial gradient because they affect the thermal stress and the fatigue life of the DPF. The burn-off rate can be reduced or optimized by changing the exhaust flow rate. It can also be decreased by reducing the oxygen concentration via a reduction in air–fuel ratio or an increase in EGR rate.

Secondary emissions may occur during the regeneration, such as the following species: (1) the CO from incomplete soot burning; (2) the evaporation of the adsorbed HC and SOF during a slow exhaust heating phase without being oxidized; (3) the unburned hydrocarbons of post-injection during the DPF heating; (4) the NO_x due to shutting off EGR or due to the soot oxidation reaction ($C + NO_2 = CO + CO_2 + NO$); (5) the sulfuric acid aerosol nanoparticles created by the catalyst; and (6) the organic compounds nucleated downstream of the DPF. Using an ultra-low sulfur fuel, a low-sulfur lube oil (McGeehan *et al.*, 2002, 2006, 2007), and a sulfur trap may eliminate the sulfur-containing aerosol nanoparticles. The effectiveness of DPF regeneration also depends on the sulfur content in the diesel fuel.

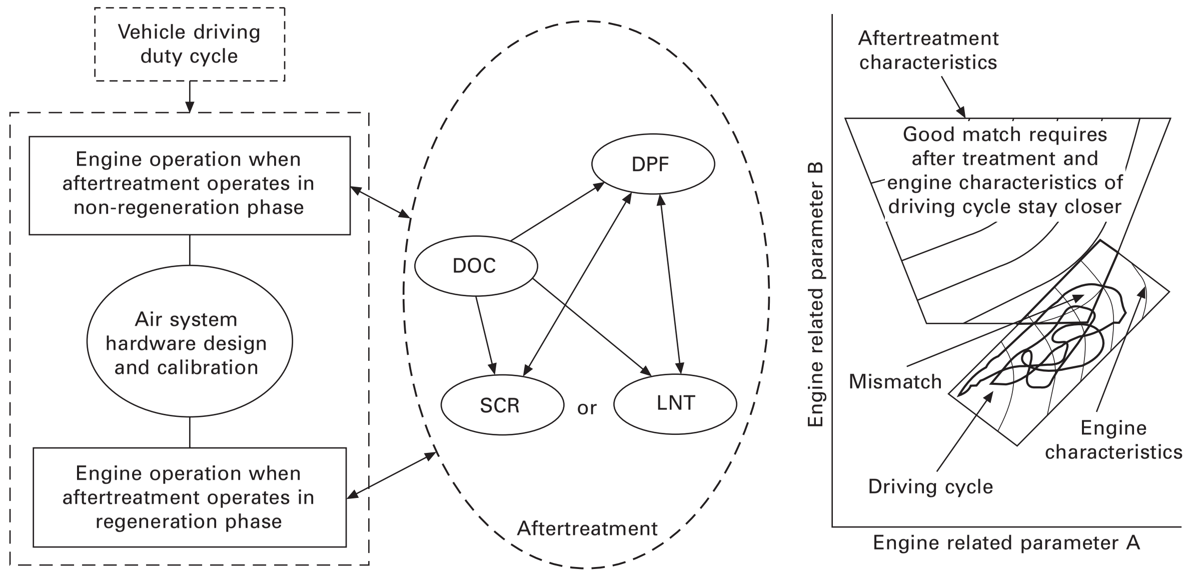
An important goal of engine air system design is to robustly adapt to the varying level of exhaust restriction caused by DPF soot loading changes, and to achieve the capability required for optimum DPF regeneration at all ambient conditions and operating modes. Vehicle driving cycle simulation plays a critical role in analyzing the exhaust temperature windows as functions of time and occurrence frequency for optimized DPF matching and regeneration.

8.3 Analytical approach of engine–aftertreatment integration

The analysis tasks of engine–aftertreatment integration or matching need to start with analyzing the requirements of exhaust temperature, flow rate, and constituent concentrations (e.g., O_2, HC, CO, NO) at the inlet of each aftertreatment device (e.g., DOC, SCR or LNT, DPF) for their normal operation and regeneration in the entire engine speed–load domain. The common requirements need to be consolidated, and the differences need to be identified. Finally, the optimum air handling system needs to be designed to minimize the BSFC and maintain good durability for the whole engine–aftertreatment system. A similar approach of engine–vehicle matching and engine–turbocharger matching may be used in engine–aftertreatment matching, i.e., plotting the engine operating data on aftertreatment characteristic maps for the non-regeneration and regeneration phases.

The process of engine–aftertreatment performance matching is summarized in the following five steps (Fig. 8.4):

1. Understand the aftertreatment design and the operating characteristics for both normal operation and regeneration, and create the characteristic maps for each device (e.g., DOC, DPF, SCR, or LNT) where the characteristics need to be presented as functions of engine performance-related parameters (such as flow rate, temperature and gas mass, and certain fundamental combinations of these parameters) and chemical parameters (such as the NO_2:NO_x ratio) in order to facilitate the matching in the next steps. The characteristic responses may be presented in contours, for example the chemical reactions involved, the light-off temperature, the DPF 'balance point temperature', temperature windows, species conversion efficiency, other efficiencies, component size, storage capacity, flow restriction, loading performance (e.g., DPF or LNT load-up rate), heat losses, regeneration energy and temperature, thermal durability, etc. Identifying the fundamental influential parameters and constructing such characteristic maps are the important tasks for aftertreatment suppliers when they cooperate with the engine manufacturer.
2. Analyze the engine–aftertreatment calibration DoE test data to understand the aftertreatment performance on the engine. The engine manufacturer is responsible for this step.
3. Match the engine and each aftertreatment device by plotting the engine performance characteristic data (e.g., exhaust gas temperature, flow rate and engine-out emissions in the engine speed–load domain, at various ambient and altitude conditions) on the aftertreatment characteristic maps; or vice versa, whichever is more convenient. The engine manufacturer holds a major responsibility for this step.
4. Discover any mismatch in both normal operation and regeneration modes



8.4 Engine–aftertreatment matching.

from the maps. Revise the engine air system controls and reconstruct the matching maps. Alternatively, reselect the aftertreatment. Finally, optimize all the links between different aftertreatment devices and the engine–aftertreatment interface. The suppliers and the engine manufacturer need to work together to accomplish this step. During the matching, engine air system performance simulation and vehicle transient driving cycle simulation need to be conducted. In addition to the packaging and cost attributes of the aftertreatment design, consideration should be given to the following important performance and durability topics during the matching:

- Select the optimum location and sequence of each aftertreatment device based on the steady-state and transient performance criteria of the engine and the aftertreament (e.g., during turbocharger lag).
- Minimize the mismatch in the temperature windows and the space velocity windows.
- Arrange exhaust thermal management to minimize the thermal energy required (e.g., bypass the gas flow or conducting in-line regeneration). Note that turbocharger efficiency, air–fuel ratio and EGR rate all have a direct impact on the turbine outlet exhaust temperature and aftertreatment performance.
- Select the optimum amount of the precious metal used in catalytic reaction based on the balance of product cost between the base engine and the aftertreatment devices.
- Control back-pressure variations and the pressure drop of the aftertreatment devices for optimum engine performance.
- Optimize the DPF regeneration methods (either active or passive) and control the regeneration process.
- Minimize the fuel economy penalty due to regeneration by balancing the engine–aftertreatment as a whole.
- Minimize the risk of temperature-induced durability problems for the aftertreatment devices.
- Minimize the engine power surge during aftertreatment regeneration for good vehicle drivability.

5. Finalize the engine-out emissions target and the air system requirements from an aftertreatment perspective for the designs of the engine air system and electronic controls.

Integrated system-level modeling to combine the engine and the aftertreatment as a whole is the direction for future advanced simulations. Analytical tools play a vital role in the engine–aftertreatment matching and integration. Simulation models of system-level aftertreatment integration are presented in Stamatelos *et al.* (1999), Peters *et al.* (2004), Kapparos *et al.* (2005), England *et al.* (2006), Strzelec *et al.* (2006), Wahiduzzaman *et al.* (2007), Tang *et al*. (2007, 2008), He (2007), Rutland *et al*. (2007), Gurupatham and He (2008), and Singh *et al*. (2009).

8.4 References and bibliography

Bagal N L, Rutland C J, Foster D E, Narayanaswamy K and He Y (2009), 'CO emission model for an integrated diesel engine, emissions, and exhaust aftertreatment system level model', SAE paper 2009-01-1511.

Basshuysen R and Schafer F (2004), *Internal Combustion Engine Handbook*, SAE International, Warrendale, PA.

Bencherif K, Benaicha F, Sadaï S and Sorine M (2009), 'Diesel particulate filter thermal management using model-based design', SAE paper 2009-01-1082.

Birkby N, Beesley S, Hatton A and El-Khatib H (2006), 'The development and testing of an active particulate regeneration system using model based control', SAE paper 2006-01-3258.

Blakeman P G, Chiffey A F, Phillips P R, Twigg M V and Walker A P (2003), 'Development in diesel emissions aftertreatment technology', SAE paper 2003-01-3753.

Bouchez M and Dementhon J B (2000), 'Strategies for the control of particulate trap regeneration', SAE paper 2000-01-0472.

Bouteiller B, Bardon S, Briot A, Girot P, Gleize V and Higelin P (2007), 'One dimensional backpressure model for asymmetrical cells DPF', SAE paper 2007-01-0045.

Brewbaker T and Nieuwstadt M (2002), 'Control of oxygen for thermal management of diesel particulate filters', SAE paper 2002-01-0427.

Brown K F, Rideout G, Rostkowski J and Meyer N (2004), 'Durability studies of a base metal catalyzed particulate filter in a severe non-road application', SAE paper 2004-01-0077.

Chakraborty D, Petersen H N, Elkjaer C, Cagulada A and Johannessen T (2009), 'Solid ammonia as energy carrier: current status and future prospects', *Fuel Cells Bulletin*, 10, 12–15.

Chi J N and DaCosta H F M (2005), 'Modeling and control of a urea-SCR aftertreatment system', SAE paper 2005-01-0966.

Chiatti G, Chiavola O and Falcucci G (2008), 'DPF soot loading and regeneration: a lumped parameter approach', SAE paper 2008-01-0441.

Corro G (2002), 'Sulfur impact on diesel emission control – a review', *Reaction Kinetics and Catalysis Letters*, 75(1), 89–106.

Cunningham P and Meckl P (2007), '1-D dynamic diesel particulate filter model for unsteady pulsating flow', SAE paper 2007-01-1140.

Cunningham P, Meckl P and Shah C (2007), 'Correlating dynamic pressure signal features to diesel particulate filter load', SAE paper 2007-01-0333.

Cutler W A (2004), 'Overview of ceramic materials for diesel particulate filter applications', *Ceramic Engineering and Science Proceedings*, 25(3), 421–430.

Devarakonda M, Parker G, Johnson J H, Strots V and Santhanam S (2008), 'Adequacy of reduced order models for model-based control in a urea-SCR aftertreatment system', SAE 2008-01-0617.

Eastwood P (2000), *Critical Topics in Exhaust Gas Aftertreatment*, Research Studies Press, Baldock, Hertfordshire.

Eastwood P (2008), *Particulate Emissions from Vehicles*, John Wiley & Sons, New York.

Eastwood P (2010), 'Exhaust gas aftertreatment for light-duty diesel engines', in Zhao H (ed.), *Advanced Direct Injection Combustion Engine Technologies and Development, Volume 2: Diesel Engines*, Woodhead Publishing Limited, Cambridge, UK.

Edgar B, Rumminger M and Streichsbier M (2003), 'A framework for evaluating aftertreatment PM control strategies', SAE paper 2003-01-2306.

Elmoe T D, Sorensen R Z, Quaade U, Christensen C H, Norskov J K and Johannessen T (2006), 'A high-density ammonia storage/delivery system based on $M_g(NH_3)_6Cl_2$ for SCR-DeNO$_x$ in vehicles', *Chemical Engineering Science*, 61, 2618–2625.

England S B, Rutland C J, Foster D E and He Y (2006), 'Investigation of the effect of DPF loading and passive regeneration on engine performance and emissions using an integrated system simulation', SAE paper 2006-01-0263.

Flörchinger P, Zink U, Cutler W and Tomazic D (2004), 'DPF regeneration-concept to avoid uncontrolled regeneration during idle', SAE paper 2004-01-2657.

Fortunato F, Caprio M, Oliva P, D'Aniello G, Pantaleone P, Andreozzi A and Manca O (2007), 'Numerical and experimental investigation of the thermal behavior of a complete exhaust system', SAE paper 2007-01-1094.

Gantawar A K, Opris C N and Johnson J H (1997), 'A study of the regeneration characteristics of silicon carbide and cordierite diesel particulate filters using a copper fuel additive', SAE paper 970187.

Gieshoff J, Pfeifer M, Schäfer-Sindlinger A, Hackbarth U, Teysset O, Colignon C, Rigaudeau C, Salvat O, Krieg H and Wenclawiak B W (2001), 'Regeneration of catalytic diesel particulate filters', SAE paper 2001-01-0907.

Graves R L, Kass M D, Lewis S A, Partridge W P, Thomas J F, Habeger C F, Aardahl C L, Rappe K G, Tran D N, Delgado M A and Singh G (2001), 'Exhaust aftertreatment research for heavy vehicles', SAE paper 2001-01-2064.

Guo Z and Zhang Z (2005a), 'A one-dimensional numerical model for diesel particulate trap performance study during loading and regeneration', SAE paper 2005-01-0961.

Guo Z and Zhang Z (2005b), 'A one-dimensional numerical model for diesel particulate trap performance study during loading and regeneration', *Proc. IMechE*, *International Journal of Engine Research*, 6, 247–262.

Gurupatham A and He Y (2008), 'Architecture design and analysis of diesel engine exhaust aftertreatment system and comparative study with close-coupled DOC-DPF system', SAE paper 2008-01-1756.

He Y (2006), 'Development and application of a lean NO$_x$ trap model', SAE paper 2006-01-0686.

He Y (2007), 'Development of an integrated diesel exhaust aftertreatment simulation tool with applications in aftertreatment system architecture design', SAE paper 2007-01-1138.

Herman A, Wu M-C, Cabush D and Shost M (2009), 'Model based control of SCR dosing and OBD strategies with feedback from NH_3 sensors', SAE paper 2009-01-0911.

Herrmuth H, Cartus T, Ducellari R and Derschmidt O (2004), 'Combined NO$_x$ and PM exhaust gas aftertreatment approaches for HSDI diesel engines', SAE paper 2004-01-1425.

Hinz A, Andersson L, Edvardsson J, Salomonsson P, Karlsson C-J, Antolini F, Blakeman P G, Lavenius M, Magnusson B, Walker A P and Chen H-Y (2005), 'The application of a NO$_x$ absorber catalyst system on a heavy-duty diesel engine', SAE paper 2005-01-1084.

Hiranuma S, Takeda Y, Kawatani T, Doumeki R, Nagasaki K and Ikeda T (2003), 'Development of DPF system for commercial vehicle – basic characteristic and active regenerating performance', SAE paper 2003-01-3182.

Hirata K, Masaki N, Yano M, Akagawa H, Takada K, Kusaka J and Mori T (2009), 'Development of an improved urea-selective catalytic reduction–diesel particulate filter system for heavy-duty commercial vehicles', *International Journal of Engine Research*, 10, 337–348.

Hosoya M, Kawada Y, Sato S and Shimoda M (2007), 'The study of NO_x and PM reduction using urea selective catalytic reduction system for heavy duty diesel engine', SAE paper 2007-01-1576.

Hu H, Reuter J, Yan J and McCarthy J (2006), 'Advanced NO_x aftertreatment system and controls for on-highway heavy duty diesel engines', SAE paper 2006-01-3552.

Huynh C T, Johnson J H, Yang S L, Bagley S T and Warner J R (2003), 'A one-dimensional computational model for studying the filtration and regeneration characteristics of a catalyzed wall-flow diesel particulate filter', SAE paper 2003-01-0841.

Ido T, Ogyu K, Ohira A, Hayashi M, Ohno K, Konstandopoulos A G (2005), 'Study on the filter structure of SiC-DPF with gas permeability for emission control', SAE paper 2005-01-0578.

Iitsuka Y, Yamauchi H, Prieto G, Takashima K and Mizuno A (2007), 'Ammonia production from solid urea using non-thermal plasma', *Conference Record – IAS (IEEE Industry Applications Society) Annual Meeting*, article number 4347979, pp. 1489–1493.

Johannessen T, Schmidt H, Frey A M and Christensen C H (2009), 'Improved automotive NO_x aftertreatment system: metal ammine complexes as NH_3 source for SCR using Fe-containing zeolite catalysts', *Catalysis Letters*, 128(1–2), 94–100.

Johnson T V (2000a), 'Diesel emission control in review', SAE paper 2000-01-0184.

Johnson T V (2000b), 'Diesel emission control – last 12 months in review', SAE paper 2000-01-2817.

Johnson T V (2001), 'Diesel emission control in review', SAE paper 2001-01-0184.

Johnson T V (2002), 'Diesel emission control: 2001 in review', SAE paper 2002-01-0285.

Johnson T V (2003), 'Diesel emission control in review – the last 12 months', SAE paper 2003-01-0039.

Johnson T V (2004), 'Diesel emission control technology – 2003 in review', SAE paper 2004-01-0070.

Johnson T V (2006), 'Diesel emission control in review', SAE paper 2006-01-0030.

Johnson T V (2007a), 'Diesel emission control in review', SAE paper 2007-01-0233.

Johnson T V (2007b), *Diesel Particulate Filter Technology*, SAE International, Warrendale, PA.

Johnson T V (2008a), 'Diesel emission control in review', SAE paper 2008-01-0069.

Johnson T (2008b), 'Diesel engine emissions and their control: an overview', *Platinum Metals Review*, 52(1), 23–37.

Johnson T V (2009a), 'Diesel emission control in review', SAE paper 2009-01-0121.

Johnson T V (2009b), 'Review of diesel emissions and control', *International Journal of Engine Research*, 10, 275–285.

Johnson T (2010), 'Overview of diesel emissions and control for heavy-duty diesel engines', in Zhao H (ed.), *Advanced Direct Injection Combustion Engine Technologies and Development, Volume 2: Diesel Engines*, Woodhead Publishing Limited, Cambridge, UK.

Kandylas I P and Koltsakis G C (2002), 'Simulation of continuously regenerating diesel particulate filters in transient driving cycles', *Proc. IMechE*, *Part D: Journal of Automobile Engineering*, 216, 591–606.

Kapetanovic S, Wallace J S and Evans G (2009), 'Investigation of sulfate nanoparticulate formation from a catalyzed diesel particulate filter on an engine fueled with ULSD and a biodiesel blend', SAE paper 2009-01-0691.

Kapparos D J, Foster D E and Rutland C J (2004), 'Sensitivity analysis of a diesel exhaust system thermal model', SAE paper 2004-01-1131.

Kapparos D J, Brahma I, Strzelec A, Rutland C J, Foster D E and He Y (2005), 'Integration of diesel engine, exhaust system, engine emissions and aftertreatment device models', SAE paper 2005-01-0947.

Khair M (2003), 'A review of diesel particulate filter technologies', SAE paper 2003-01-2303.

Khair M and McKinnon D L (1999), 'Performance evaluation of advanced emission control technologies for diesel heavy-duty engines', SAE paper 1999-01-3564.

Kim J Y, Cavataio G, Patterson J E, Laing P M and Lambert C K (2007), 'Laboratory studies and mathematical modeling of urea SCR catalyst performance', SAE paper 2007-01-1573.

Kim Y-W, Sun J, Kolmanovsky I and Koncsol J (2003), 'A phenomenological control oriented lean NO_x trap model', SAE paper 2003-01-1164.

Kladopoulou E A, Yang S L, Johnson J H, Parker G G and Konstandopoulos A G (2003), 'A study describing the performance of diesel particulate filters during loading and regeneration: a lumped parameter model for control applications', SAE paper 2003-01-0842.

Klingstedt F, Arve K, Eranen K and Murzin D Y (2006), 'Toward improved catalytic low-temperature NO_x removal in diesel-powered vehicles', *Accounts of Chemical Research*, 39(4), 273–282.

Kong Y, Kozakiewicz T, Johnson R, Huffmeyer C, Huckaby J, Abel J, Baurley J and Duffield K (2005), 'Active DPF regeneration for 2007 diesel engines', SAE paper 2005-01-3509.

Konstandopoulos A G (2003), 'Flow resistance descriptors for diesel particulate filters: definitions, measurements and testing', SAE paper 2003-01-0846.

Konstandopoulos A G and Kladopoulou E (2004), 'The optimum cell density for wall-flow monolithic filters: effects of filter permeability, soot cake structure and ash loading', SAE paper 2004-01-1133.

Konstandopoulos A G and Kostoglou M (1999), 'Periodically reversed flow regeneration of diesel particulate traps', SAE paper 1999-01-0469.

Konstandopoulos A G and Kostoglou M (2004), 'Microstructural aspects of soot oxidation in diesel particulate filters', SAE paper 2004-01-0693.

Konstandopoulos A G, Skaperdas E, Warren J and Allansson R (1999), 'Optimized filter design and selection criteria for continuously regenerating diesel particulate traps', SAE paper 1999-01-0468.

Konstandopoulos A G, Kostoglou M, Skaperdas E, Papaioannou E, Zarvalis D and Kladopoulou E (2000), 'Fundamental studies of diesel particulate filters: transient loading, regeneration and aging', SAE paper 2000-01-1016.

Konstandopoulos A G, Kostoglou M and Housiada P (2001a), 'Spatial non-uniformities in diesel particulate trap regeneration', SAE paper 2001-01-0908.

Konstandopoulos A G, Skaperdas E and Masoudi M (2001b), 'Inertial contributions to the pressure drop of diesel particulate filters', SAE paper 2001-01-0909.

Konstandopoulos A G, Skaperdas E, Masoudi M (2002), 'Microstructural properties of soot deposits in diesel particulate traps', SAE paper 2002-01-1015.

Konstandopoulos A G, Kostoglou M, Housiada P, Vlachos N and Zarvalis D (2003a), ‘Multichannel simulation of soot oxidation in diesel particulate filters’, SAE paper 2003-01-0839.

Konstandopoulos A G, Vlachos N, Housiada P and Kostoglou M (2003b), ‘Simulation of triangular-cell-shaped, fibrous wall-flow filters’, SAE paper 2003-01-0844.

Konstandopoulos A G, *et al*. (2004), ‘The diesel exhaust aftertreatment (DEXA) cluster: a systematic approach to diesel particulate emission control in Europe’, SAE paper 2004-01-0694.

Konstandopoulos A G, Papaioannou E, Zarvalis D, Skopa S, Baltzopoulou P, Kladopoulou E, Kostoglou M and Lorentzou S (2005a), ‘Catalytic filter systems with direct and indirect soot oxidation activity’, SAE paper 2005-01-0670.

Konstandopoulos A G, Kostoglou M, Vlachos N and Kladopoulou E (2005b), ‘Progress in diesel particulate filter simulation’, SAE paper 2005-01-0946.

Kozlov A, Harris T and Salyards C (2010), ‘Engine test for DOC quenching in DOC-DPF system for non-road applications’, SAE paper 2010-01-0815.

Kramer J, Pfahl U, Bruestle C, Diewald R and Hunter G (2009), ‘The PM-Metalit: a PM control technology for Tier 4 off-highway applications’, SAE paper 2009-01-2838.

Liu Z G and Miller R K (2002), ‘Flow distributions and pressure drops of wall-flow diesel particulate filters’, SAE paper 2002-01-1311.

Locker R J, Gunasekaran N and Sawyer C (2002), ‘Diesel particulate filter test methods’, SAE paper 2002-01-1009.

Majewski W A and Khair M K (2006), *Diesel Emissions and Their Control*, SAE International, Warrendale, PA.

Masoudi M, Konstandopoulos A G, Nikitidis M S, Skaperdas E, Zarvalis D, Kladopoulou E and Altiparmakis C (2001), ‘Validation of a model and development of a simulator for predicting the pressure drop of diesel particulate filters’, SAE paper 2001-01-0911.

Massey S, Williamson P S and Chuter R J (2002), ‘Modelling exhaust systems using one-dimensional methods’, SAE paper 2002-01-0005.

Mayer A, Nöthiger P, Zbinden R and Evéquoz R (2001), ‘Particulate trap selection for retrofitting vehicle fleets based on representative exhaust temperature profiles’, SAE paper 2001-01-0187.

Mayer A, Kany S, Nöthiger P, Richards P, Andreassen L, Andreoni T and Sem T (2005), ‘Retrofitting TRU-diesel engines with DPF-systems using FBC and intake throttling for active regeneration’, SAE paper 2005-01-0662.

McCarthy Jr J and Holtgreven J (2008), ‘Advanced NO_x aftertreatment system performance following 150 LNT desulfation events’, SAE paper 2008-01-1541.

McGeehan J A, *et al*. (2002), ‘API CI-4: The first oil category for diesel engines using cooled exhaust gas recirculation’, SAE paper 2002-01-1673.

McGeehan J A, *et al*. (2006), ‘API CJ-4: Diesel oil category for both legacy engines and low emission engines using diesel particulate filters’, SAE paper 2006-01-3439.

McGeehan J A, Moritz J, Shank G, Kennedy S, Stehouwer D, Urbank M, Belay M, Goodier S, Cassim A, Runkle B, DeBaun H, Harold S, Chao K, Herzog S, Stockwell R, Passut C, Fetterman P, Taber D, Williams L, Kleiser W M, Zalar J, Scinto P, Santos E and Rutherford J A (2007), ‘API CJ-4: Diesel oil category for pre-2007 engines and new low emission engines using cooled exhaust gas recirculation and diesel particulate filters’, SAE paper 2007-01-1966.

McCarthy J, Korhumel T and Marougy A (2009), ‘Performance of a fuel reformer, LNT and SCR aftertreatment system following 500 LNT desulfation events’, SAE 2009-01-2835.

Merkel G A, Beall D M, Hickman D L and Vernacotola M J (2001), 'Effects of microstructure and cell geometry on performance of cordierite diesel particulate filters', SAE paper 2001-01-0193.

Millet C-N, Menegazzi P, Martin B, Colas H and Bourgeois C (2002), 'Modeling of diesel particulate filter regeneration: effect of fuel-borne catalyst', SAE paper 2002-01-2786.

Mohammed H, Lakkireddy V R, Johnson J H and Bagley S T (2006a), 'An experimental and modeling study of a diesel oxidation catalyst and a catalyzed diesel particulate filter using a 1-D 2-layer model', SAE paper 2006-01-0466.

Mohammed H, Triana A P, Yang S-L and Johnson J H (2006b), 'An advanced 1D 2-layer catalyzed diesel particulate filter model to simulate: filtration by the wall and particulate cake, oxidation in the wall and particulate cake by NO_2 and O_2, and regeneration by heat addition', SAE paper 2006-01-0467.

Müller W, Ölschlegel H, Schäfer A, Hakim N and Binder K (2003), 'Selective catalytic reduction – Europe's NO_x reduction technology', SAE paper 2003-01-2304.

Nam G-W, Park J-W, Lee J-H and Yeo G-K (2007), 'The effect of an external fuel injection on the control of LNT system; the diesel NO_x reduction system', SAE 2007-01-1242.

Niemi S A, Paanu T P J and Laurén M J (2004), 'Effect of injection timing, EGR and EGR cooling on the exhaust particle number and size distribution of an off-road diesel engine', SAE paper 2004-01-1988.

Niemi S, Lundin K, Karhu T, Laurén M, Ekman K, Nousiainen P and Paanu T (2009), 'Exhaust particle number in off-road engines of different generations', SAE paper 2009-01-1869.

Nieuwstadt M J and Trudell D F (2004), 'Diagnostics for diesel particulate filters', SAE paper 2004-01-1422.

Nieuwstadt M, Upadhyay D and Yuan F (2005), 'Diagnostics for diesel oxidation catalysts', SAE paper 2005-01-3602.

Nikitidis M S, Manikas T, Zarvalis D, Divinis N, Altiparmakis C and Konstandopoulos A G (2001), 'Collection efficiency of various filter media in diesel exhaust', SAE paper 2001-01-065.

Ogyu K, Oya T, Ohno K, Dolios I, Kladopoulou E, Lorentzou S and Konstandopoulos A G (2007), 'Study on catalyzed-DPF for improving the continuous regeneration performance and fuel economy', SAE paper 2007-01-0919.

Ohyama N, Nakanishi T and Daido S (2008), 'New concept catalyzed DPF for estimating soot loadings from pressure drop', SAE paper 2008-01-0620.

Ootake M, Kondou T, Ikeda M, Daigo M, Nakano M, Yokoyama J and Miura M (2007), 'Development of diesel engine system with DPF for the European market', SAE paper 2007-01-1061.

Ottinger N A, Nguyen K, Bunting B G, Toops T J and Howe J (2009), 'Effects of rapid high temperature cyclic aging on a fully-formulated lean NO_x trap catalyst', SAE paper 2009-01-0634.

Park D S, Kim J U, Cho H and Kim E S (1998), 'Considerations on the temperature distribution and gradient in the filter during regeneration in burner type diesel particulate trap system (II)', SAE paper 980188.

Parks J, Watson A, Campbell G and Epling B (2002), 'Durability of NO_x absorbers: effects of repetitive sulfur loading and desulfation', SAE paper 2002-01-2880.

Peters B J, Wanker R J, Münzer A and Wurzenberger J C (2004), 'Integrated 1D to 3D simulation workflow of exhaust aftertreatment devices', SAE paper 2004-01-1132.

Phillips P R, Chandler G R, Jollie D M, Wilkins A J J and Twigg M V (1999), 'Development of advanced diesel oxidation catalysts', SAE paper 1999-01-3075.

Premchand K C, Johnson J H, Yang S-L, Triana A P and Baumgard K J (2007), 'A study of the filtration and oxidation characteristics of a diesel oxidation catalyst and a catalyzed particulate filter', SAE paper 2007-01-1123.

Reader G T, Banerjee S, Wang M and Zheng M (2006), 'Energy efficiency analysis of active-flow operations in diesel engine aftertreatment', SAE paper 2006-01-3286.

Rumminger M D, Zhou X, Balakrishnan K, Edgar B L and Ezekoye O A (2001), 'Regeneration behavior and transient thermal response of diesel particulate filters', SAE paper 2001-01-1342.

Rutland C J, England S B, Foster D E and He Y (2007), 'Integrated engine, emissions, and exhaust aftertreatment system level models to simulate DPF regeneration', SAE paper 2007-01-3970.

Salvat O, Marez P and Belot G (2000), 'Passenger car serial application of a particulate filter system on a common rail direct injection diesel engine', SAE paper 2000-01-0473.

Scarnegie B, Miller W R, Ballmert B, Doelling W and Fischer S (2003), 'Recent DPF/SCR results targeting US2007 and Euro 4/5 HD emissions', SAE paper 2003-01-0774.

Simescu S, Ulmet V, Neely G and Khair M (2010), 'A novel approach for diesel NO_x/PM reduction', SAE paper 2010-01-0308.

Singh N, Johnson J H, Parker G G and Yang S-L (2005), 'Vehicle engine aftertreatment system simulation (VEASS) model: application to a controls design strategy for active regeneration of a catalyzed particulate filter', SAE paper 2005-01-0970.

Singh N, Rutland C J, Foster D E, Narayanaswamy K and He Y (2009), 'Investigation into different DPF regeneration strategies based on fuel economy using integrated system simulation', SAE paper 2009-01-1275.

Sluder C S and West B H (2001), 'Performance of a NO_x adsorber and catalyzed particle filter system on a light-duty diesel vehicle', SAE paper 2001-01-1933.

Soeger N, Mussmann L, Sesselmann R, Leippe G, Gietzelt C, Bailey O and Hori M (2005), 'Impact of aging and NO_x/soot ratio on the performance of a catalyzed particulate filter for heavy duty diesel applications', SAE paper 2005-01-0663.

Song Q and Zhu G (2002), 'Model-based closed-loop control of urea SCR exhaust aftertreatment system for diesel engine', SAE paper 2002-01-0287.

Stamatelos A M, Koltsakis G C, Kandylas I P and Pontikakis G N (1999), 'Computer aided engineering in the development of diesel exhaust aftertreatment systems', SAE paper 1999-01-0458.

Stratakis G A, Psarianos and Stamatelos A M (2002), 'Experimental investigation of the pressure drop in porous ceramic diesel particulate filters', *Proc. IMechE*, *Part D: Journal of Automobile Engineering*, 216, 773–784.

Strzelec A, Rutland C J, Foster D E and He Y (2006), 'Development of a system level soot–NO_x trap aftertreatment device model', SAE paper 2006-01-3287.

Subramaniam M N, Joergl V, Keller P, Weber O, Toyoshima T and Vogt C D (2009), 'Feasibility assessment of a pre-turbo after-treatment system with a 1D modeling approach', SAE paper 2009-01-1276.

Sung K, Kim J and Ritz R D (2009), 'Experimental study of pollutant emission reduction for near-stoichiometric diesel combustion in a three-way catalyst', *International Journal of Engine Research*, 10, 349–357.

Takahashi Y, Takeda Y, Kondo N and Murata M (2004), 'Development of NO_x trap system for commercial vehicle – basic characteristics and effects of sulfur poisoning', SAE paper 2004-01-0580.

Tan J C, Opris C N, Baumgard K J and Johnson J H (1996), 'A study of the regeneration process in diesel particulate traps using a copper fuel additive', SAE paper 960136.

Tang W, Wahiduzzaman S and Leonard A (2007), 'A lumped/1-D combined approach for modeling wall-flow diesel particulate filters – applicable to integrated engine/aftertreatment simulations', SAE paper 2007-01-3971.

Tang W, Wahiduzzaman S, Wenzel S, Leonard A and Morel T (2008), 'Development of a quasi-steady approach based simulation tool for system level exhaust aftertreatment modeling', SAE paper 2008-01-0866.

Taoka N, Ohno K, Hong S, Sato H, Yoshida Y and Komori T (2001), 'Effect of SiC-DPF with high cell density for pressure loss and regeneration', SAE paper 2001-01-0191.

Theis J R, Ura J A, Goralski C T, Caine J, Davies M, Kay D, Todd A and Dinsdale S (2005), 'The effects of aging temperature and PGM loading on the NO_x storage capacity of a lean NO_x trap', SAE paper 2005-01-1117.

Toorisaka H, Minamikawa J, Narita H, Muramatsu T, Kominami T and Sone T (2004), 'DPR developed for extremely low PM emissions in production commercial vehicles', SAE paper 2004-01-0824.

Wahiduzzaman S, Tang W, Wenzel S and Leonard A (2007), 'Modeling of integrated aftertreatment systems: a highly configurable system level approach', SAE paper 2007-01-4127.

Wahiduzzaman S, Wenzel S, Despujols B and Tang W (2008), 'Development of diesel engine and catalyst model derived from detailed model for engine and powertrain control design', SIA, 28–29 May 2008, INSA de Rouen, France.

Wang D Y, Yao S, Shost M, Yoo J-H, Cabush D, Racine D, Cloudt R and Willems F (2008), 'Ammonia sensor for closed-loop SCR control', SAE paper 2008-01-0919.

Wenzel S, Despujols B, Wahiduzzaman S and Papadimitriou I (2009), 'Development of real time catalyst model for engine and powertrain control design', SAE paper 2009-01-1273.

Willems F, Cloudt R, Eijnden E, Genderen M, Verbeek R, De Jager B, Boomsma W and Heuvel I (2007), 'Is close-loop SCR control required to meet future emission targets?', SAE paper 2007-01-1574.

Yamaguchi S, Fujii S, Kai R, Miyazaki M, Miyairi Y, Miwa S and Busch P (2005), 'Design optimization of wall flow type catalyzed cordierite particulate filter for heavy duty diesel', SAE paper 2005-01-0666.

Yeh J T, Pennline H W and Hoffmann J S (1996), 'Life cycle tests of the moving-bed copper oxide process for SO_2 and NO_x control', 89th AWMA Annual Meeting, Nashville, TN, June 23–28.

Yi Y (2006), 'Simulating the soot loading in wall-flow DPF using a three-dimensional macroscopic model', SAE paper 2006-01-0264.

York A P E, Cox J P, Watling T C, Walker A P, Bergeal D, Allansson R and Lavenius M (2005), 'Development and validation of a one-dimensional computational model of the continuously regenerating diesel particulate filter (CR-DPF) system', SAE paper 2005-01-0954.

York A P E, Watling T C, Ahmadinejad M, Bergeal D, Phillips P R and Swallow D (2009), 'Modeling the emissions control performance of a catalyzed diesel particulate filter (CDPF) system for light duty diesel applications', SAE paper 2009-01-1266.

9
Advanced diesel valvetrain system design

Abstract: As the first chapter in Part III, this chapter starts to bridge the thermo-fluid topics and the mechanical considerations such as dynamic vibration, friction, and noise in diesel engine system design. As a unique system that has dual characteristics of both air system and mechanical system, the valvetrain controls engine gas flows, volumetric efficiency and engine delta P, subject to mechanical design constraints. This chapter first summarizes the design guidelines of the conventional valvetrain with an integrated analytical approach by simultaneously considering engine breathing performance and valvetrain dynamics in valvetrain design that includes the cam and the valve spring. It then investigates the performance of variable valve actuation (VVA) and its interactions with other air system components. A theory of using VVA to control pumping loss is developed. The benefits of cylinder deactivation for the diesel engine are also analyzed.

Key words: valvetrain, valve timing, volumetric efficiency, engine delta P, cam design, valve spring, valvetrain dynamics, recompression pressure, no-follow, variable valve actuation (VVA), Miller cycle, cylinder deactivation.

9.1 Guidelines for valvetrain design

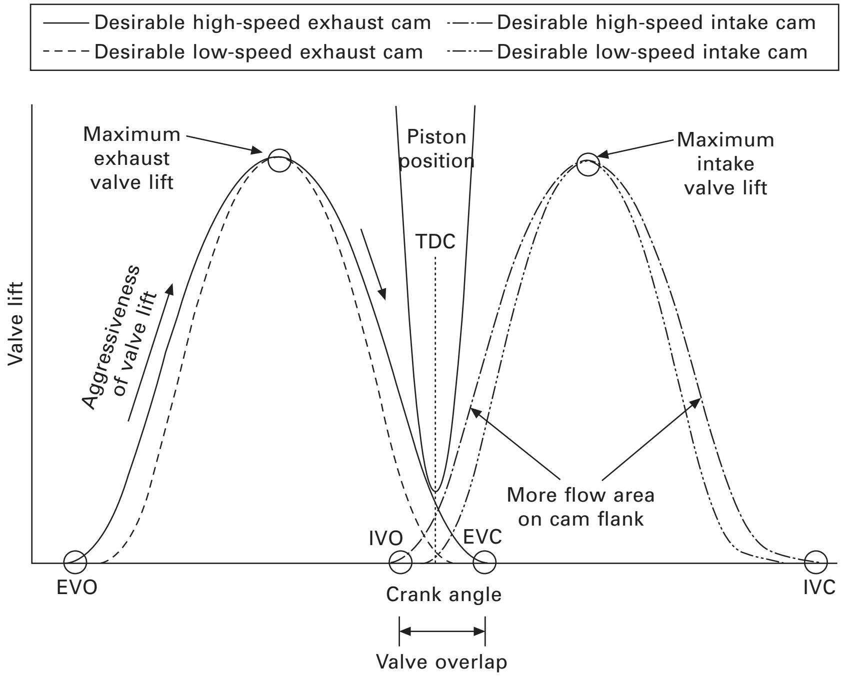
Engine valvetrain consists of valves, springs, valve bridges, rocker arms, pushrods, followers, and rocker shafts. In the pushrod and overhead cam valvetrains, the camshaft is either chain-driven or belt-driven. The drive ratio is 1:2 in four-stroke engines so that one revolution of the camshaft corresponds to two revolutions of the crankshaft (i.e., 1° cam angle is equivalent to 2° crank angle). Valvetrain design is important for engine system performance as it affects the following aspects:

- volumetric efficiency (i.e., air flow or air–fuel ratio capability)
- pumping loss and engine delta P control
- mechanical friction
- engine noise
- regulating engine air/gas flow to achieve advanced features.

The air system theory in Chapter 4 illustrates that pumping loss is affected by volumetric efficiency (equation 4.40) and pressure losses across the intake/exhaust valves. The definition of volumetric efficiency shows that in order to reach a given engine flow rate a higher volumetric efficiency would require a lower intake manifold boost pressure so that a lower engine delta P and better brake specific fuel consumption (BSFC) can be achieved. Different valvetrain concepts have different component weight, stiffness, configuration height, complexity and friction loss. In order to understand the interactions among these characteristics, it is necessary to comprehend the functions of valve lift profile and cam lift profile, their impact on engine performance and valvetrain dynamics, and finally the implications on design. This section explains these topics. The research on valvetrains has a long history in the development of the internal combustion engine. In order to understand the advanced features of the valvetrain (including both the conventional fixed cam system and variable valve actuation or VVA systems) in modern diesel engines, it is necessary to first understand the fundamentals of the basic functions and design issues related to the valvetrain. For more information on valvetrain design fundamentals, the reader is referred to Newton and Allen (1954), Tauschek (1958), and Wang (2007).

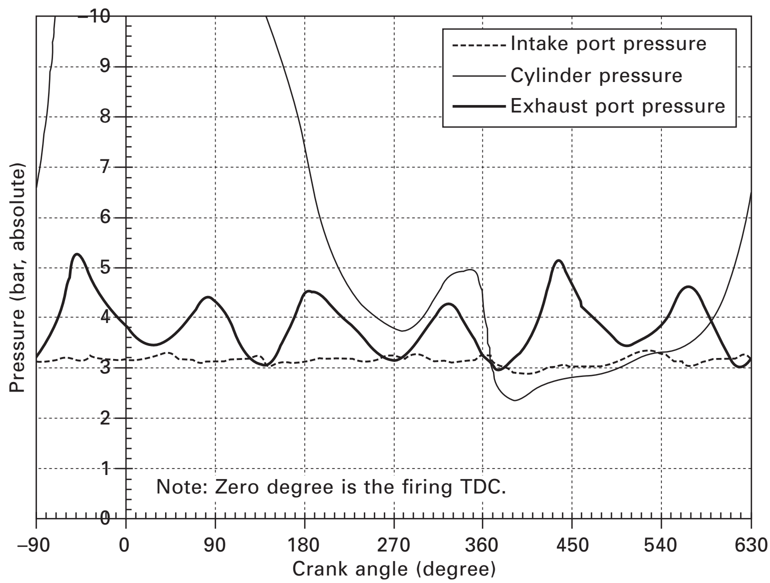
9.1.1 Valve lift profile and valve timing

Typical engine valve lift profiles of the intake and exhaust valves in the conventional cam-driven valvetrain are shown in Fig. 9.1. The intake and exhaust port pressures and the in-cylinder pressure are illustrated in Fig.

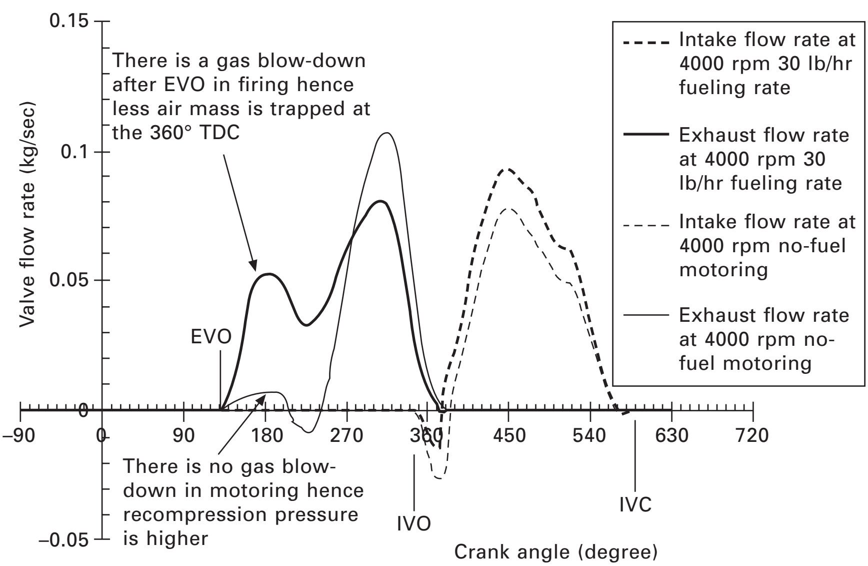


9.1 Control points on engine valve lift profiles.

9.2. Typical engine valve flow characteristics are shown in Fig. 9.3. Good volumetric efficiency is achieved via appropriate valve lift and valve timing designs. It is observed that there are several control points on the valve lift



9.2 Intake and exhaust port pressure pulses at full load firing at 2000 rpm of an I6 engine.



9.3 Engine valve flow characteristics in firing and motoring.

profiles that govern the shape and the timing of the valve lift curve, and they are (in terms of valve timing locations in crank angle): exhaust valve opening (EVO), exhaust valve closing (EVC), intake valve opening (IVO), and intake valve closing (IVC). The valve overlap is usually referred to by the duration from IVO to EVC around the TDC where both the intake valve and the exhaust valve are open. The profile control points also include the maximum valve lift (i.e., the 'nose lift') for the intake and exhaust valves. Moreover, the lift profile is controlled by the rising rate of the lift at the opening and closing flanks. The rising rate characterizes the steepness or aggressiveness of the lift profile.

The optimum EVO timing is usually determined by an appropriate balance in engine performance between the useable work during the expansion stroke and the pumping loss during the exhaust stroke. Changing EVO alters the effective engine expansion ratio and hence has a large impact on the thermodynamic cycle efficiency of the engine. An excessively advanced (or early) EVO timing (i.e., toward the firing TDC) results in excessive loss in indicated power and causes a penalty in BSFC. An excessively retarded (or late) EVO timing results in high pumping loss, especially at high engine speeds when the engine air flow rate is high, thus also causing an increase in BSFC.

When the cylinder pressure during the blow-down process reduces to a level that is lower than the intake manifold pressure due to rapid decompression, the resulting 'under-pressure' during the exhaust stroke may produce a positive pumping work and hence reduces pumping loss. Such a phenomenon of the under-pressure in the cylinder at the end of the blow-down process is known as the 'Kadenacy' effect (Stas, 1999; Parvate-Patil *et al.*, 2004). This effect may happen when the exhaust valve is opened very suddenly and rapidly. The rapid opening results in a rarefaction wave so that a lower pressure level may occur during the exhaust stroke to reduce the exhaust pumping loss. The intensity of the Kadenacy effect and the cylinder pressure immediately after EVO are affected by the rate of valve opening, the valve flow area, and the pressure ratio across the exhaust valve. Usually, the cylinder under-pressure after the blow-down becomes smaller as the engine speed increases. Stas (1999) estimated that using the Kadenacy effect may provide a possible gain in the engine thermal efficiency up to 3%.

The optimum EVC timing is a balance among the following three parameters: exhaust valve recession or piston cutout, pumping loss, and the residue gas trapped. The valve recession or piston cutout affects the 'K-factor' (detailed in Chapter 7) and combustion quality. Residue fraction is essentially hot internal EGR. It affects the breathing quality of the engine, and reduces the amount of fresh air and external cooled EGR that the engine can induct. A too-early EVC timing may result in excessively high pumping loss and too much residue gas trapped. A too-late EVC timing requires a large valve recession in order to avoid the valve-to-piston contact near the TDC. It may also induce undesirable backflow (reverse flow) of the burned exhaust gas from the exhaust manifold into the cylinder. It should be noted that a later EVC timing may reduce pumping loss due to the larger exhaust valve flow area during the later part of the exhaust stroke.

The optimum IVO timing is determined by the balance among the following three parameters: intake valve recession or piston cutout, pumping loss, and the residue gas trapped, similar to the effect given by EVC. For example, a too-early IVO timing may require a large valve recession in order to avoid the valve-to-piston contact and may also induce backflow of the exhaust gas flowing from the cylinder or the exhaust manifold into the intake manifold. The backflow reduces the breathing capacity of the engine for inducting the fresh air. Also note that a too-early IVO timing may reduce pumping loss due to the larger intake valve flow area during the early part of the intake stroke. On the other hand, EVC and IVO cause different valvetrain dynamics issues such as the recompression pressure level accumulated near the valve overlap TDC, the intake valvetrain vibration induced by the recompression pressure, and the exhaust valve seating velocity.

The optimum IVC timing in the conventional cam-driven fixed valvetrain is usually determined by minimizing the backflow in order to maximize volumetric efficiency. Moreover, changing IVC timing alters the effective engine compression ratio and may largely impact the thermodynamic cycle performance and emissions of the engine. Changing IVC timing is also the most effective way to reduce the amount of engine air flow without incurring throttle loss as would occur using an intake throttle valve. Many VVA mechanisms regulate IVC timing (e.g., the throttleless operation in gasoline engines for load control, the Miller cycle or the Atkinson cycle for lower BSFC and NO_x, and the compression ratio and charge mass controls in HCCI).

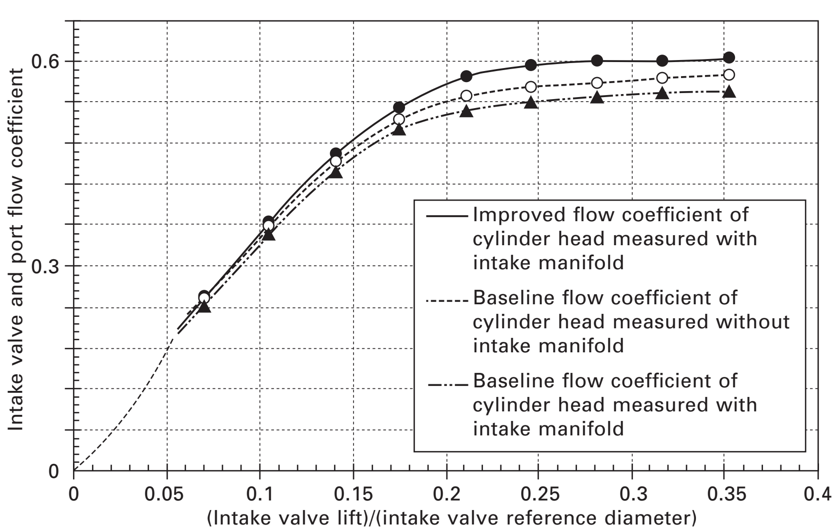
The primary design objective related to EVC, IVO, and IVC timings in the conventional cam-driven fixed valvetrain is both volumetric efficiency and BSFC, while the primary concern on EVO timing is BSFC. A higher volumetric efficiency gives a higher air–fuel ratio at the same level of intake manifold boost pressure without the need to raise the turbine inlet pressure in order to achieve the high air–fuel ratio. Consequently, soot can be reduced, and this emissions benefit can also be indirectly converted to a BSFC benefit through other design/calibration means.

The valve event duration (i.e., from EVO to EVC for the exhaust valve, or from IVO to IVC for the intake valve) is also a consideration in valvetrain dynamics and cam mechanical design, but these dynamic considerations are usually secondary compared to the requirements of breathing performance, especially when the engine speed is not very high (e.g., less than 4000 rpm). Moreover, note that engine valve flow behavior changes as the engine speed changes due to the effects of gas wave dynamics. The optimum valve timing is a trade-off between low speeds and high speeds as illustrated in Fig. 9.1. This is particularly true when the engine has a very wide range of operating speeds.

The maximum valve lift and the aggressiveness of the lift rising rate are determined by the balance among cam stress, valvetrain dynamics, and the possible diminishing return of the effective valve flow area at very high lift. An example of valve and port flow coefficient is shown in Fig. 9.4. The valve and port flow coefficient is highly dependent on the valve lift but nearly independent of the pressure ratio across the valve. Note that there is a loss in the flow coefficient when the manifold is tested alone with the cylinder head. The loss is due to more flow restrictions introduced by the manifold. The effective valve flow area is usually calculated by the following:

$$A_{VAL,eff} = C_f A_{ref} = C_f \frac{\pi d_{VAL,ref}^2}{4} \qquad 9.1$$

where A_{ref} can be any reference area, but usually the valve disc diameter is used as the reference diameter. Figure 9.3 shows that the valve flow coefficient profile becomes flat at high valve lift, and this means the port gradually becomes a choking bottleneck for the flow capacity of the cylinder head. When this happens, any higher maximum valve lift will only provide a small and diminishing gain on volumetric efficiency but cause a large difficulty in cam design and controlling the valvetrain dynamics. The interaction between valve timing and the port flow coefficient occurs mainly in the low lift region



9.4 Illustration of valve flow coefficient.

because the effective valve flow area at the low lift affects the backflow behavior around EVC, IVO, and IVC.

The valve lift profile also has an interaction with the turbocharger. When a high volumetric efficiency is achieved via appropriate valve timing or valve lift profile, the same air flow rate required to meet the emissions target can be achieved with a lower intake manifold boost pressure, and this alleviates the burden on the turbocharger to deliver the air flow. This is especially important for the air–fuel ratio at low engine speeds such as at peak torque. The turbine area is basically determined based on the air and EGR flow requirements at peak torque (if the turbine is not a flexible VGT). If the valve timing is excessively biased toward high speeds so that the volumetric efficiency is very low at peak torque, an excessively small turbine area would have to be used to compensate for the air/EGR flow requirements. The fixed small turbine area will result in a large penalty in pumping loss at high engine speeds.

Finally, it should be noted that valve timing and volumetric efficiency also interact with turbocharger matching via the change in engine delta P and gas scavenging at valve overlap. Gas scavenging refers to a cooling effect by flowing the intake air from the intake manifold to the exhaust manifold during valve overlap in order to reduce the exhaust gas temperature and the thermal load on the cylinder head, the exhaust manifold, and the turbine. As shown in the theoretical analysis of the engine air system in Chapter 4, volumetric efficiency is one of the key parameters that affect engine delta P and EGR driving capability. For non-EGR engines, a negative engine delta P can be created by selecting a large turbine area in turbocharger matching to enable both a pumping work 'gain' and gas scavenging. In this case, a large valve overlap needs to be designed accordingly in order to facilitate the gas scavenging. However, for EGR engines the engine delta P basically needs to be positive in order to drive EGR. In this case, gas scavenging from the intake manifold to the exhaust manifold is impossible. Instead, undesirable backflow occurs during the valve overlap. Therefore, for EGR engines, the valve overlap needs to be designed very small in order to ensure high volumetric efficiency. A more detailed theoretical elaboration about the effect of valve timing on volumetric efficiency and pumping loss is provided in Section 9.7.

9.1.2 Cam lift profile

The 'cam lift' is actually an abbreviation of the 'follower lift at the cam side'. The term is usually only meaningful for the pushrod valvetrain or direct-acting overhead cam (OHC) valvetrain where the follower lift has a simple one-dimensional translational motion driven by the cam. In some other OHC valvetrains (e.g., with a end-pivot finger follower), the follower often has a two-dimensional motion so that the kinematic relationship between the valve lift and the follower lift becomes more complicated. Note that the cam lift is not equivalent to the contour shape of the cam lobe for the case of a roller follower. The cam contour shape refers to the radial distance beyond the cam base circle from the cam center to any point on the cam lobe profile. The cam lift is not equal to the radial distance because the radius of the roller follower is not equal to zero. For the pushrod and the direct-acting OHC valvetrains, cam lift or essentially the follower lift is usually used to describe the intended cam profile design, and the actual contour shape of the cam lobe can be calculated once the cam lift profile and the roller size are specified. On the other hand, for some OHC valvetrains (e.g., with a finger follower), it is often more convenient to directly use the actual contour shape of the cam lobe to define the cam profile design.

The major difference between the valve lift profile and the cam lift profile is that the cam profile has ramps at the opening and closing ends. The ramp height must be sufficient to accommodate the cold lash, the valvetrain elastic compression caused by spring preload, the thermal growth of the valvetrain under operating conditions, the dynamic deflections in the camshaft caused by the operation of adjacent valves, the wear allowance of the valvetrain components, the normal setting tolerances, the leak-down of the hydraulic lash adjuster (if any, for the closing ramp), and so on. The ramp height needs to ensure the valve opens and closes on the ramp with controlled velocities at any operating conditions to achieve low impact noise and good durability. Using hydraulic lash adjuster may obtain precise valve timing control at any hot or cold conditions because there is no lash. In contrast, a mechanical lash adjuster results in variations in valve timing in cold and hot conditions.

The aggressiveness of the cam lift profile is determined by the height and the width of the cam's positive acceleration peaks (humps). A higher peak results in a narrower width of the acceleration hump and usually causes larger vibrations of the valvetrain.

In pushrod cam design, a target valve lift profile first needs to be determined by the requirement of engine breathing performance. Once the target maximum valve lift and valve event duration are determined, a cam profile can be designed, for example with the popular 'Polydyne' cam design approach. In OHC design, due to the complexity of the two-dimensional motion of the finger follower, the cam profile cannot be directly computed from a prescribed valve lift motion. Usually, iterations are required by using a kinematic or dynamic model to calculate the valve lift profile based on a designed cam contour shape, which is then compared with the desired valve lift profile.

In engine performance analysis, usually a 'kinematic' or static valve lift is used for simplicity in order to conduct a large amount of engine cycle simulations to optimize the required valve timing and lift profile. The evaluation of the more complex dynamic or vibratory valve lift profile only occurs at the cam design stage by using valvetrain dynamics modeling. The definition of the kinematic valve lift profile is given below. If the valvetrain elastic vibration effect is ignored, the hot non-vibratory valve lift (i.e., hot kinematic valve lift) for a pushrod valvetrain or a direct-acting OHC valvetrain can be calculated as follows:

$$l_{VAL,hot}(\phi, N_E \dot{W}_E) = l_{CAM}(\phi) \cdot f_{RA} - c_{VT,hot}(N_E, \dot{W}_E) - \frac{F_{pre}}{K_{s,VT}} - \frac{F_{gas,VAL}(\phi, N_E, \dot{W}_E)}{\dfrac{K_{s,VT} \cdot K_{s,VAL}}{K_{s,VT} + K_{s,VAL}}} \qquad 9.2$$

where ϕ is the crank angle, N_E is the engine speed, $\dot{W}_E$ is the engine power, l_{CAM} is the cam lift as a function of crank angle, f_{RA} is the rocker arm ratio (either simplified as a constant or treated as a more complicated function of cam angle or valve lift), and $c_{VT,hot}$ is the valvetrain hot lash (for mechanical lash adjuster only) as a function of engine speed and load. F_{pre} is the valve spring preload. $F_{gas,VAL}$ is the net gas load acting on the valve, exerted by the cylinder gas and the port gas, as a function of crank angle, engine speed and load. $K_{s,VT}$ is the valvetrain stiffness defined at the valve side. $K_{s,VAL}$ is the valve or valve stem stiffness. The $l_{CAM}(\phi) \cdot f_{RA}$ is usually called gross valve lift. The kinematic lift is a static net lift when the valvetrain lash (gaps between the components) and all the elastic compressions are subtracted. The elastic compressions of the valvetrain refer to the total amount of compressions caused by various forces before the valve opens. It is obvious that the definition of valve event duration depends on how complex the valve lift profile is defined. For example, some engine catalogs use 0.004 inch above the cam base circle as 'advertised valve event duration'. Another method often used in the aftermarket catalogs measures valve duration at 0.050 inch above the cam base circle. When specifying or comparing valve event durations, it is important to clarify how the duration is defined.

In equation 9.2, it is usually difficult to estimate the valvetrain hot lash for the mechanical lash adjusters when the temperatures of the valvetrain components are uncertain. Moreover, the valvetrain hot lash changes at different engine speeds and loads. In addition, the cylinder gas load also changes at different crank angle locations. Therefore, ideally the hot kinematic valve lift needs to be calculated at each engine speed and load. It is obvious that such a treatment is neither convenient nor practical. For simplicity, the cold kinematic valve lift is usually used in engine system design and performance simulations. The cold kinematic valve lift is defined as follows by excluding the speed- or load-dependent factors in 9.2:

$$l_{VAL,cold}(\phi) = l_{CAM}(\phi) \cdot f_{RA} - c_{VT,cold} - \frac{F_{pre}}{K_{s,VT}} \qquad 9.3$$

where $c_{VT,cold}$ is the valvetrain cold lash set by a feeler gauge at the valve side. The cold kinematic valve lift is regarded as a theoretical and representative lift profile which can be used to compare the valve timings of different engines.

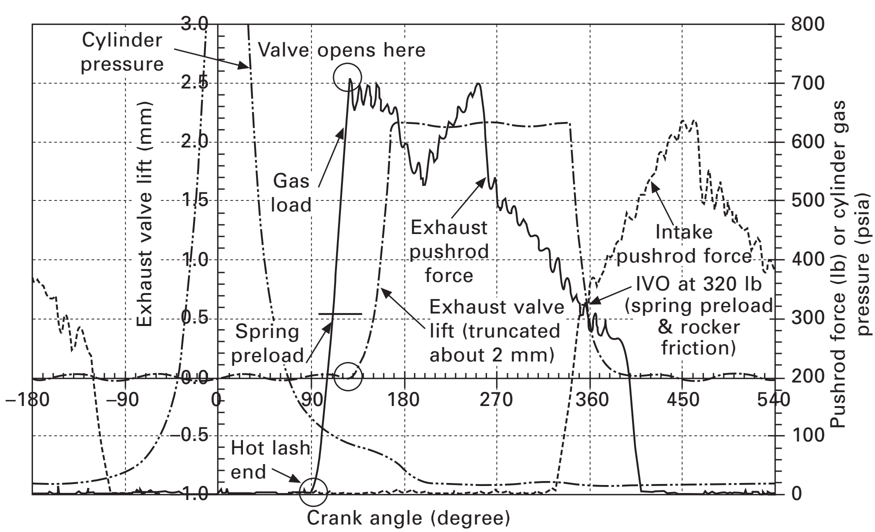
At hot operating conditions, the intake valvetrain is usually still relatively cold, and the cylinder gas loading at IVO is very low. At full load, the exhaust valve train is hot, and the thermal expansion of the valvetrain components is not negligible. But the thermal expansion and the elastic compression caused by gas loading may offset each other so that the cold kinematic valve lift may not be too bad an assumption to represent the real lift, especially around EVO. However, there could be large errors at EVC because the elastic compression effect due to gas loading is much smaller there. As pointed out by Morel and LaPointe (1994), the comparison between the simulated and measured port pressure traces can be used to check the dynamic valve timing. Moreover, note that usually the valvetrain elastic compression caused by the valve spring preload is much smaller than the cold lash for mechanical lash adjusters.

9.1.3 Valvetrain dynamics

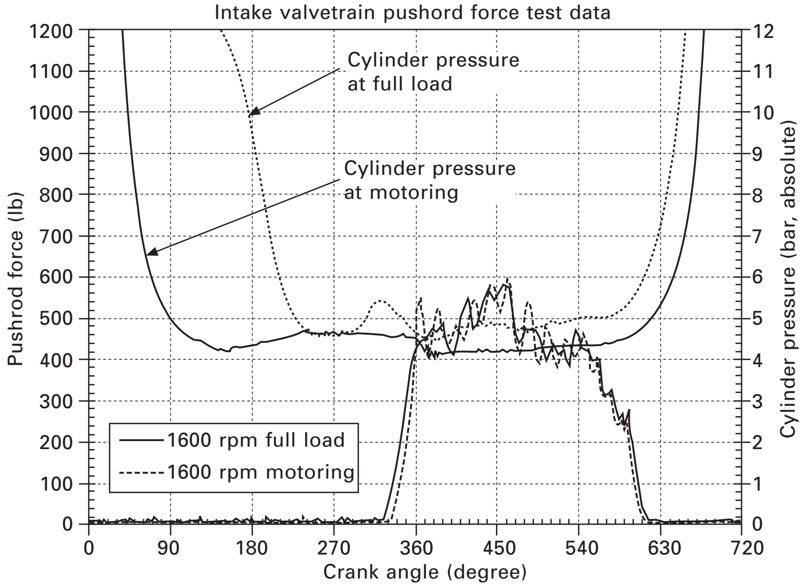
Cam design is a highly integrated system design topic with trade-offs between engine breathing performance and valvetrain dynamics. Valve timing, valve event duration, and the maximum valve lift are largely determined by the requirements of engine breathing performance. Combustion and emissions requirements affect cam profile design through the K-factor, valve recession (or piston cutout), and valve overlap height (related to valve-to-piston contact). Cam lift profile affects valvetrain dynamics via two factors: the gas loading acting on the valves and the inertia force excitation induced by cam acceleration. For example, the aggressiveness of the exhaust cam closing profile and EVC timing affect the effective exhaust valve flow area during the second half of the exhaust stroke and hence impact the recompression pressure gas loading at the valve overlap TDC that act on the intake valvetrain.

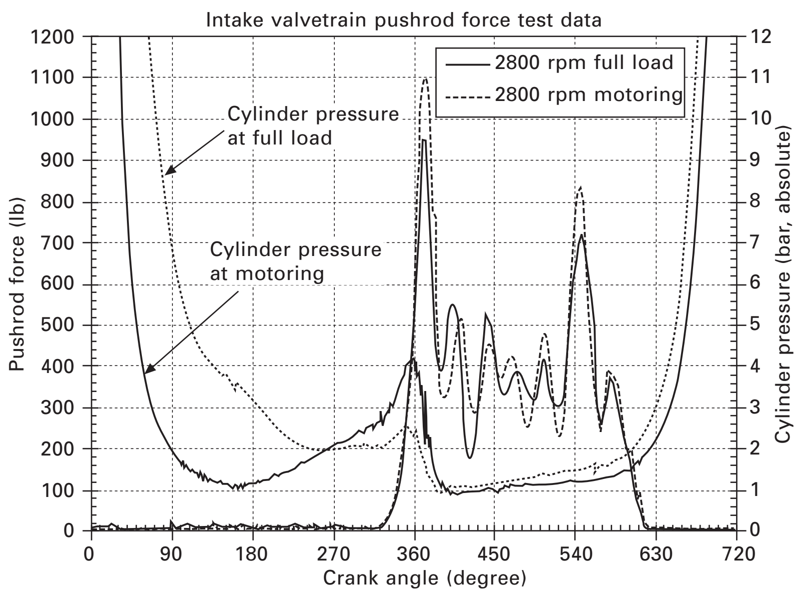
The critical design criteria in valvetrain dynamics usually include the following: no-follow (also called separation), vibration forces acting on the valvetrain joints (e.g., the fulcrum plate or the rocker pivot), cam stress, cam minimum oil film thickness, cam flash temperature, valve seating velocity, and valve spring surge. The valvetrain is an elastic dynamic system with certain stiffness and damping characteristics. Its natural frequency is affected by the mass and the stiffness of the components. Its damping is affected by friction and lubrication. Valvetrain vibration is excited by both cam lift profile and gas load. The firing gas load increases the exhaust pushrod force at full load (Fig. 9.5), and the recompression pressure gas load increases the vibration amplitude of the intake pushrod force at high-speed motoring (Fig. 9.6). The pushrod vibration force is a superposition of the static valve spring force, the dynamic forces due to valvetrain inertia and gas loading, and the valvetrain friction force. Note that the sudden drop of the pushrod force at the cam nose shown in Fig. 9.5 is caused by the rocker shaft friction force. The vibratory valve lift motion may cause variations in the opening and closing timing of the valves at different engine speeds and loads.

Valvetrain vibration may also cause no-follow between the valvetrain components, for example at the interface between the cam and the follower, or at the interface between the rocker arm and the valve bridge. The no-follow is characterized by a zero force at the cam–follower contact (or called zero cam force) during the cam deceleration phase. The cam force is the summation of the pushrod force and the follower's inertia force. Therefore, the cam force usually reaches zero at a lower engine speed than the speed at which the pushrod force reaches zero. When no-follow happens, impact and bounce between the components occur, and the valve motion is not well controlled. For valvetrains equipped with hydraulic lash adjusters, no-follow may cause the lifter to pump up to fill the dynamic clearances. If the no-follow lasts long enough, the lifter will eventually pump up to its full extent, preventing the valve from returning to its seat and causing a catastrophic failure of valve-to-piston contact. Valvetrain no-follow can be controlled by a combination of appropriate cam acceleration shape, low recompression pressure gas loading, strong spring preload and spring rate.



9.5 Typical valvetrain pushrod force trace at low speed full load.





9.6 Effect of recompression pressure gas loading on intake valvetrain vibration.

These need to be designed under the constraints of the maximum cam stress limit and the minimum penalty of engine breathing in terms of pumping loss and volumetric efficiency.

The performance of hydraulic lash adjusters is important for engine breathing and valvetrain dynamics. The design and modeling information about hydraulic lash adjusters is provided by Abell (1969), Herrin (1982), Kreuter and Maas (1987), Phlips and Schamel (1991), Porot and Trapy (1993), Zou and McCormick (1996), Dammers (1997), Zhao *et al*. (1999), and Okarmus *et al*. (2008). This topic is not covered in detail in this book.

Valvetrain design is highly complex because all the parameters involved are inter-related. They exhibit different dynamic behavior at different engine operating conditions such as low speeds, high speeds, full-load firing, no-fuel motoring, exhaust braking, etc. The design parameters need to be determined simultaneously through optimization subject to the constraints of engine performance and valvetrain dynamics. The effects of the design parameters on valvetrain dynamics are summarized in Table 9.1.

9.1.4 Valvetrain no-follow

When a vehicle is driven downhill, it is possible for the transmission speed to exceed the shaft speed desired for the present gear of the transmission. Such a condition is called an over-speed condition because the transmission over-speeds its gear. If the over-speed lasts long enough (e.g., two seconds) the transmission gear might be damaged. Similarly, there is also a condition of engine over-speed. Many sub-systems or components in an internal combustion engine have a speed limit subject to mechanical failure. The valvetrain is one of them. Above a certain engine speed, the following can happen to damage the valvetrain or the engine: valvetrain no-follow, excessively high valve seating velocity, high pushrod force exceeding the strength of valvetrain joints, etc. The limiting factor for most engines is valvetrain no-follow. Severe no-follow occurs when the pushrod force reaches zero and the duration of the zero force lasts for more than a few degrees of crank angle.

An engine can over-speed under three conditions. First, if the engine is provided with a sufficient amount of fuel, appropriate air, and a relatively low load, it will accelerate to the over-speed limit on its own. Engine governors are used to control this type of non-motoring condition by limiting fueling. Secondly, the engine can over-speed when the vehicle is in a gear and the vehicle load suddenly decreases to create a sufficient momentum to force the engine speed to go above the governed speed through back-loading the transmission. An analogy to this situation is that a marine engine over-speeds when the ship propeller is out of the water in a rough sea. Thirdly, the engine can be motored without fueling to the over-speed limit when an outside force increases the engine speed. An example is that the vehicle travels downhill with the transmission gear engaged and the operator's foot off the accelerator pedal. If the gravity force pulling the vehicle down the grade is higher than the retarding force provided from the engine and the vehicle, the vehicle

Table 9.1 General trend of valvetrain design effects on valvetrain dynamics

	Recompression pressure at high-speed motoring	Valvetrain no-follow speed	Peak cam force	Maximum cam stress and wear	Cam radius of curvature	Valve seating velocity	Valvetrain friction	Valvetrain noise	Valve spring surge
Reduce valve (or cam) event duration		– (worse)	+ (worse)	+ (worse)	– (worse)	+ (worse)			+ (worse)
Advance EVC timing	+ (worse)	– (intake worse)	+ (intake worse)	+ (intake worse)					
Retard IVO timing		+ (intake better)	– (intake better)	– (intake better)					
Reduce valve overlap size	+ (worse)	– (intake worse)	+ (intake worse)	+ (intake worse)					
Increase exhaust manifold pressure	+ (worse)	– (intake worse)	+ (intake worse)	+ (intake worse)					
Increase maximum valve or cam lift		– (worse)	+ (worse)	+ (worse)	– (worse)	+ (worse)	+ (worse)		+ (worse)
Reduce rocker arm ratio		– (worse)	– (better)	– (better)	– (worse)	+ (worse)			+ (worse)
Reduce valvetrain weight		+ (better)	– (better)	– (better)		– (better)		– (better)	– (better)
Increase valvetrain stiffness		+ (better)	– (better)	– (better)		– (better)		– (better)	– (better)

Increase the aggressiveness of valve lift slope with higher cam acceleration	– (better)	– (worse)	+ (worse)	+ (exhaust worse at full load)	– (worse)	+ (worse)		+ (worse)	+ (worse)
Reduce valve spring preload or spring rate		– (worse)	– (better)	– (better)			– (better)		
Reduce cam base circle diameter		– (worse)	+ (worse)	+ (worse)	– (worse)	+ (worse)			+ (worse)
Reduce roller follower's roller diameter				+ (worse)	– (worse)				

Note: '+' sign indicates increase. '–' sign indicates reduction. 'Intake' indicates intake valvetrain only. 'Exhaust' indicates exhaust valvetrain only.

speed and the engine speed will increase. Over-speed can occur especially when the driver downshifts to a too low gear (e.g., shift from fourth gear at 45 mph vehicle speed and 3000 rpm engine speed to third gear at 4500 rpm) or makes a skipped downshift in a manual transmission (e.g., shift from fourth gear to first gear directly).

Some vehicles and automatic transmissions are equipped with over-speed warning or prevention systems to prevent engine over-speed by disallowing downshifting the transmission to certain gears according to the driving condition. In addition, some systems with manual transmissions can alert the driver to the danger of motoring the engine beyond the over-speed limit during gear selection so that the operator is offered an opportunity to take corrective actions to avoid damaging the engine due to excessive motoring. The engine needs to be designed to possess a good mechanical capability up to the over-speed limit.

A target valvetrain no-follow speed is established in engine design as one of the over-speed limits of the engine. It is often used as the maximum allowable operating speed of the engine. An appropriate no-follow speed target is a compromise between the requirement of engine breathing capability and the requirement of protecting engine over-speed in manual transmission vehicles. The no-follow speed target should be selected above the high-idle speed with a certain margin. An excessively low no-follow speed target would produce the valvetrain separation problem within the normal operating speed range of the engine. On the other hand, an excessively high (or over-designed) no-follow speed target would force either the cam acceleration hump to be too low or the valve spring force to be too high. A low cam acceleration results in a low valve flow area and high pumping loss. A high spring load leads to high friction force and poor BSFC. Moreover, it should be noted that although the selection of the valvetrain no-follow speed target is based largely on the normal motoring condition, the more severe but less frequent operating conditions such as engine braking or transients should also be considered.

9.1.5 Valvetrain inertia force

Valvetrain inertia force is affected by component mass and valve acceleration. In the simplified single-degree-freedom valvetrain dynamics model (detailed later in Fig. 9.15), the inertia force at the valve side can be calculated as the product of the valvetrain equivalent mass and valve acceleration. The valvetrain equivalent mass usually can be calculated as follows:

$$\begin{aligned} m_{VT} \approx m_{valve} + m_{retainer} + m_{bridge} + \frac{m_{spring}}{3} \\ + m_{RA} \cdot \frac{l_{RA,0}^2}{l_{RA,ValveSide}^2} + \frac{m_{pushrod}}{3\, f_{RA}^2} \end{aligned} \qquad 9.4$$

where m_{valve} is the valve mass, $m_{retainer}$ is the retainer mass, m_{bridge} is the valve bridge mass, m_{spring} is the valve spring mass, $m_{pushrod}$ is the pushrod mass, m_{RA} is the rocker arm mass, $l_{RA,ValveSide}$ is the rocker arm length at the valve side, f_{RA} is the rocker arm ratio, and $l_{RA,0}$ is the radius of gyration. The effective mass of the rocker arm can be considered as the mass of the rocker arm concentrated at one point which is at a distance $l_{RA,0}$ from the center of rotation, i.e., $I_{RA} = m_{RA} l_{RA,0}^2$ where I_{RA} is the moment of inertia of the rocker arm with respect to the center of rotation. Lower cam acceleration, higher stiffness, lower weight, and higher natural frequency of the valvetrain result in lower vibration amplitude of the valve acceleration. The design guideline for valvetrain inertia force control is that the design needs to achieve a precise target of valvetrain no-follow speed with an optimized gas load. Neither over-design nor under-design is acceptable for a good balance between breathing performance and valvetrain dynamics.

9.1.6 Valvetrain gas loading and recompression pressure

The gas loading acting on the intake and exhaust valves has a great impact on valvetrain dynamics and cannot be ignored. The gas loading acting on the valve face is the in-cylinder pressure. The gas loading acting on the valve back is the pressure in the port. The mechanisms and effects of gas loading are different for different valves and at different engine operating conditions. The gas loading is affected by the valve flow characteristics of the engine (Fig. 9.3). In general, there are three issues in valvetrain dynamics related to gas loading: (1) the peak cam (or pushrod) force for both intake and exhaust valvetrains; (2) intake valvetrain no-follow; and (3) exhaust valve floating.

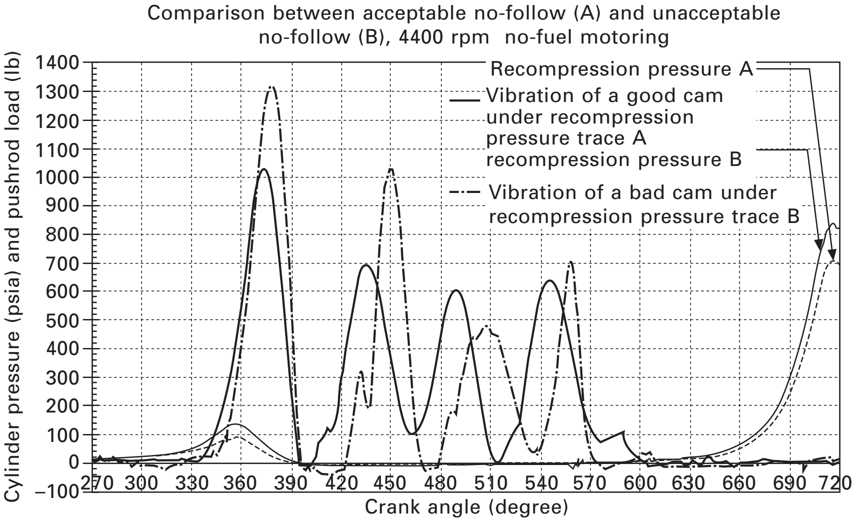
In engine firing and especially at full load, the exhaust valve needs to overcome a large in-cylinder pressure before it opens. The gas loading at EVO at full load has a dominant impact on the exhaust cam force. The cam stress at EVO at full load may sometimes become even higher than the cam stress at the cam nose during low-speed cranking. Once the exhaust valve opens, the cylinder pressure gradually decays with a strong blow-down process to expel the gas mass out of the cylinder. As a result, there is only very little gas mass trapped at the end of the exhaust stroke so that the recompression pressure at the valve overlap TDC is usually very low at full load. Therefore, the net gas loading acting on the intake valvetrain at full load is usually very low.

On the back of the exhaust valve, there are usually several exhaust port pressure pulses (in the case of pulse turbocharging) within an engine cycle to try to push the valve to open. The exhaust valve spring preload and the in-cylinder pressure counteract the exhaust pressure pulses. At full load, the intake manifold pressure is high. Therefore, the pressure differential between the exhaust port pressure pulse and the in-cylinder pressure during the intake stroke is usually not large enough to float the exhaust valve off the valve seat. Figures 9.5 and 9.6 illustrate the effects of gas loading on pushrod vibration.

In engine motoring, the in-cylinder pressure at EVO is low and the net gas loading acting on the exhaust valvetrain is very low. However, due to the lack of blow-down, a large amount of in-cylinder residue gas can be trapped in the cylinder and compressed to a very high recompression pressure at the valve overlap TDC (Fig. 9.6). A high recompression pressure is caused directly by a high exhaust manifold pressure, which results from the following: (1) high engine speed; (2) small turbine area; (3) small effective exhaust valve flow area at the cam closing side due to a small valve, a low port flow coefficient, an early EVC timing, or a less aggressive exhaust cam lift profile; and (4) exhaust braking. Note that an early EVC timing, may result in a small valve overlap, which usually needs to be used in modern high EGR engines. The recompression pressure has a sharp decaying characteristic just past the TDC when the piston moves away from the TDC. The high pressure level together with its inherent 'shock loading' characteristics make the recompression pressure the most dominant factor for intake valvetrain vibration in modern diesel engines. The gas loading and cam acceleration induce violent intake valvetrain vibration at high-speed motoring. The vibration can be so severe that the first peak of the cam (or pushrod) force can be even higher than the maximum cam force of the exhaust valvetrain. The peak force is often followed by a severe no-follow condition (i.e., a long duration of zero pushrod force as shown in Fig. 9.7). Due to the recompression pressure, valvetrain no-follow usually occurs first in the intake valvetrain rather than the exhaust valvetrain.

In exhaust braking, the exhaust manifold pressure is raised to a very high level in order to increase pumping loss and retarding power. The intake manifold pressure usually remains low. The recompression pressure in exhaust braking can be much higher than that in motoring. Severe intake valvetrain no-follow may occur if not designed properly. Moreover, the large pressure differential across the exhaust valve between the exhaust port pressure pulse and the in-cylinder pressure during the intake stroke may float the exhaust valve off the valve seat.

During transient acceleration, the EGR valve is usually closed in order to flow more fresh air into the cylinder to match the suddenly increased fueling amount. The increased turbine flow rate due to EGR valve shut-off causes an immediate large increase in the exhaust manifold pressure. However, the intake manifold pressure cannot respond instantly to quickly build up due to turbocharger lag. As a result, there is a large engine delta P which may create two problems for the valvetrain: (1) intake no-follow due to high



9.7 Valvetrain no-follow and its control with reduced recompression pressure.

recompression pressure; and (2) exhaust valve floating during the intake stroke.

The problem of excessively high peak exhaust cam force can be solved using a smaller rocker arm ratio or designing a stronger structural strength. The problem with exhaust valve floating can be solved or alleviated, if not totally avoided during transients or engine braking, through a proper selection of valve spring preload. Recompression pressure and intake valvetrain no-follow controls are the most sophisticated issues in valvetrain system design since they require a system-level approach to optimize the solutions in the early stage of the engine design. There are several methods to reduce the recompression pressure at high-speed motoring by either reducing the exhaust manifold pressure or bleeding off the in-cylinder gas mass: (1) choosing a large turbine area; (2) opening the VGT vane or the turbine wastegate; (3) using large exhaust valves; (4) using less restrictive exhaust ports; (5) retarding the EVC timing (i.e., using larger valve overlap in cam design); or (6) using a more aggressive exhaust cam closing lift profile (i.e., larger valve flow area at the closing side). Moreover, there are several methods to sustain a high recompression pressure in order to control the intake valvetrain no-follow: (1) using a less aggressive intake cam lift profile (i.e., lower cam acceleration); (2) using a longer intake cam duration by retarding IVC; (3) retarding IVO to reduce intake vibration; or (4) using a stronger intake valve spring at the expense of higher cam stress and valvetrain friction.

Turbocharger selection is largely determined by the boost requirement in the firing operation. Modern low-emissions EGR engines tend to use two-stage turbocharging with a very small turbine at the high-pressure stage. Such a configuration may generate higher exhaust manifold pressure and recompression pressure than a single-stage turbocharger. Consequently, the intake valvetrain no-follow control becomes a more challenging task. Typically, opening the VGT vane at high-speed motoring (e.g., 4000 rpm) may reduce the recompression pressure by approximately 25 psi (1.72 bar).

Using late EVC timing and large valve overlap has been the classical method to reduce the recompression pressure and pumping loss in the past for non-EGR engines. Typically, every 1° crank angle retardation of the exhaust cam closing lift profile leads to a 5 psi (0.34 bar) reduction in the recompression pressure. The disadvantages of this method in EGR engines include the higher residual fraction trapped in the cylinder (especially at low engine speeds), lower volumetric efficiency, and the larger valve recession required in order to avoid valve-to-piston contact. A lager valve recession results in a larger dead volume in the combustion chamber and higher soot emissions. More 'squared' exhaust cam closing profile combined with an early EVC timing can be considered as an alternative solution. If carefully designed, this method may avoid the penalties associated with the simple EVC retardation. However, the aggressiveness of the cam closing lift profile and the level of recompression pressure reduction are limited by the minimum allowable negative radius of curvature of the cam and the maximum allowable exhaust valve seating velocity. Compared to a normal exhaust cam, the specially designed aggressive exhaust cam may reduce the recompression pressure by 10–20 psi (0.69–1.38 bar) at an increased risk of high valve seating velocity.

The intake valve opening timing has a negligible influence on the recompression pressure. However, the intake valve opening timing affects the intake valvetrain vibration significantly due to the change in the location where the peak recompression pressure acts on the cam acceleration curve. A well-optimized smooth acceleration profile of the intake cam may sustain 30 psi (2.1 bar) recompression pressure at the same intake no-follow condition without a significant degradation in engine breathing performance. However, there is a slight penalty of higher cam stress at the cam nose. Typically, every 2° crank angle increase in the intake cam duration (e.g., by retarding IVC timing) results in the capability of sustaining an additional 12 psi (0.83 bar) recompression pressure at the same no-follow condition. The disadvantages with this IVC-retarding approach include negative impact on cold start capability and the degradation of volumetric efficiency at low engine speeds.

In summary, recompression pressure and intake valvetrain no-follow controls require an integrated solution and a balanced trade-off among cylinder head design, turbocharger selection, combustion chamber design, exhaust cam design, and intake cam design. The appropriate design sequence is as follows:

1. Reasonably establish a target no-follow speed based on engine application needs (e.g., in engine–transmission matching) without over-design.
2. Optimize the cylinder head design to achieve the maximum exhaust valve size and exhaust port flow coefficient.
3. Optimize turbocharger configuration, turbine area and wastegate valve opening and their control strategies at high engine speeds.
4. Determine the allowable exhaust valve recession and the corresponding dead volume in the combustion chamber.
5. Design a good exhaust cam profile to maximize the closing-side flow area.
6. Design a good intake cam profile to achieve smooth cam acceleration.
7. If all the above measures cannot solve the no-follow problem, the intake valve spring preload or spring rate has to be increased with increased cam stress or reduced maximum intake valve lift.

Among these design measures, the exhaust and intake cam profile designs have a vital importance. Figure 9.7 illustrates an example of intake valvetrain no-follow control by reducing the recompression pressure and designing a superior and smooth acceleration profile of the intake cam.

9.1.7 Valvetrain design criteria

The valvetrain basic competitive benchmarking parameters may include the following: valvetrain natural frequency; aggressiveness of cam acceleration; cam base circle radius; valve spring load; valve overlap duration; effective valve flow area; and the ratio of valve diameter to bore diameter.

The most important design criteria for the valvetrain in turbocharged EGR diesel engines are summarized as follows:

- low valvetrain weight and high stiffness (i.e., high natural frequency)
- coordinated design of cam profile and turbocharging to achieve high volumetric efficiency in a wide speed range with appropriate maximum valve lift, large effective valve flow area, and proper valve timing
- small valve overlap (i.e., short overlap duration and low overlap lift height) for low internal residue fraction and good transient acceleration
- no valve-to-piston contact but designed for tight clearance at the TDC
- low recompression pressure at the valve overlap TDC achieved by appropriate turbine area, exhaust cam lift profile, and EVC timing
- smooth shapes of cam acceleration and jerk, appropriate cam ramp height
- coordinated design of cam lift profile, valvetrain dynamics and engine

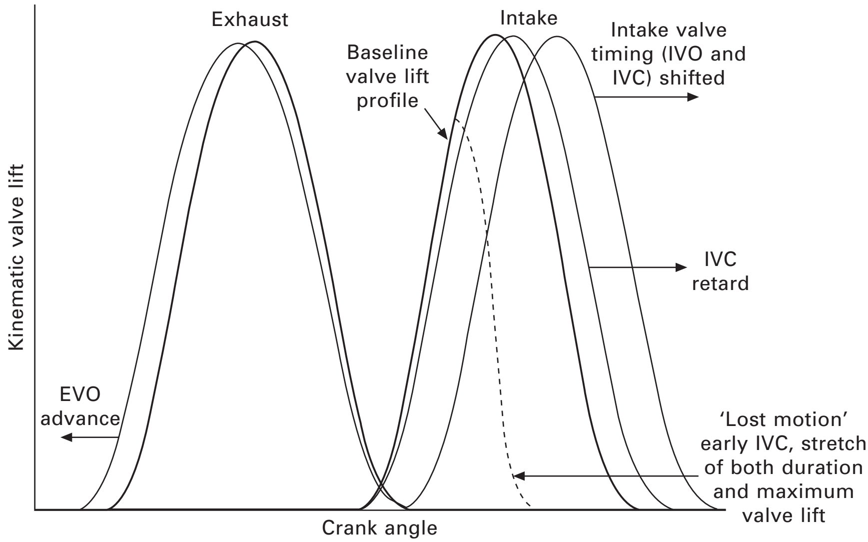
thermodynamic cycle performance to achieve an on-target valvetrain vibration level matched with the no-follow speed target with the maximum cam acceleration
- acceptable valve seating velocity and noise through cam profile design
- optimized cam base circle diameter, roller diameter, rocker arm ratio, and maximum valve lift
- acceptable minimum cam radius of curvature for cam manufacturing
- acceptable maximum pushrod force without pushrod buckling and exceeding the limit of valvetrain joint strength
- acceptable cam stress and oil film thickness
- minimum valve tip scrub achieved by optimum kinematic layout
- low friction and wear achieved by good kinematics, dynamics and appropriate valve spring force
- no valve spring surge or coil clash/bind by designing low cam harmonic amplitudes and appropriate valve spring design.

The heart of the valvetrain design is the cam profile, which is critical for both engine breathing performance and valvetrain dynamics. It should be noted that the design of the OHC valvetrain with the end-pivot finger follower is more complex than that of the pushrod valvetrain due to its two-dimensional oscillating follower motion rather than the simpler one-dimensional translational follower motion. The calculation methods of cam radius of curvature for the pushrod cam and the finger follower OHC are also different.

9.2 Effect of valve timing on engine performance

The effect of valve timing on engine performance is illustrated in the simulation results of Figs 9.8–9.10 for a light-duty EGR diesel engine. The optimum selection of the IVO, IVC, and EVC timing is usually based on both volumetric efficiency and BSFC. The optimum selection of the EVO is usually based on BSFC. Note that the optimum valve timing is a trade-off between high speeds and low speeds. Another typical method to analyze the optimum valve timing is shown in Fig. 9.11 by plotting the volumetric efficiency at a high speed vs. that at a low speed. In this way, the trade-off behavior of engine valve timing between high speeds and low speeds can be illustrated clearly to facilitate a design decision on the air flow capability.

Figure 9.12 illustrates the effects of valve overlap size and IVC timing for a heavy-duty non-EGR diesel engine that uses a large turbine area to achieve negative engine delta P at peak torque. The simulation data show that when the intake manifold pressure is higher than the exhaust manifold pressure, the larger valve overlap reduces the exhaust manifold gas temperature at peak

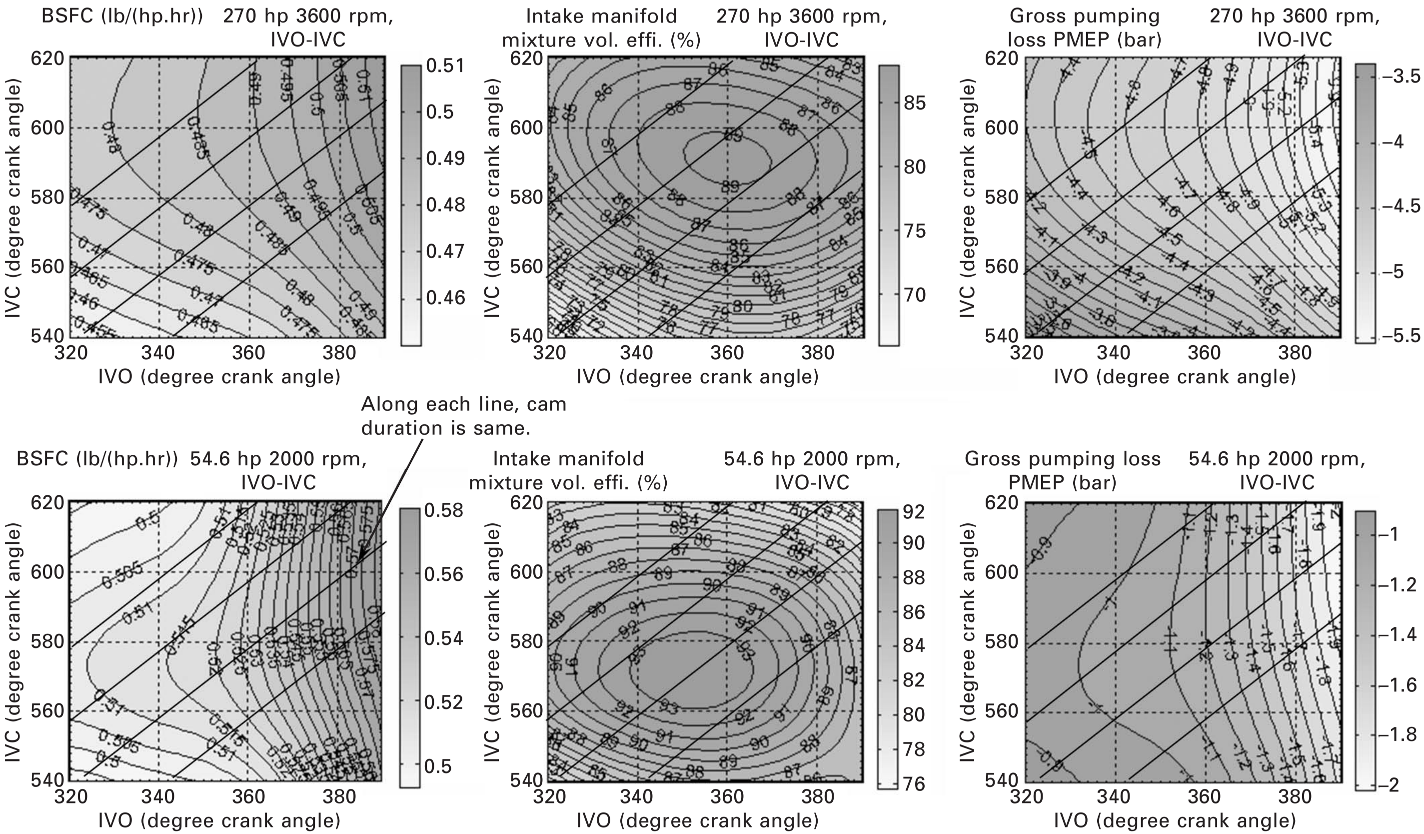


9.8 Illustration of valve timing changes in parametric simulation of engine breathing.

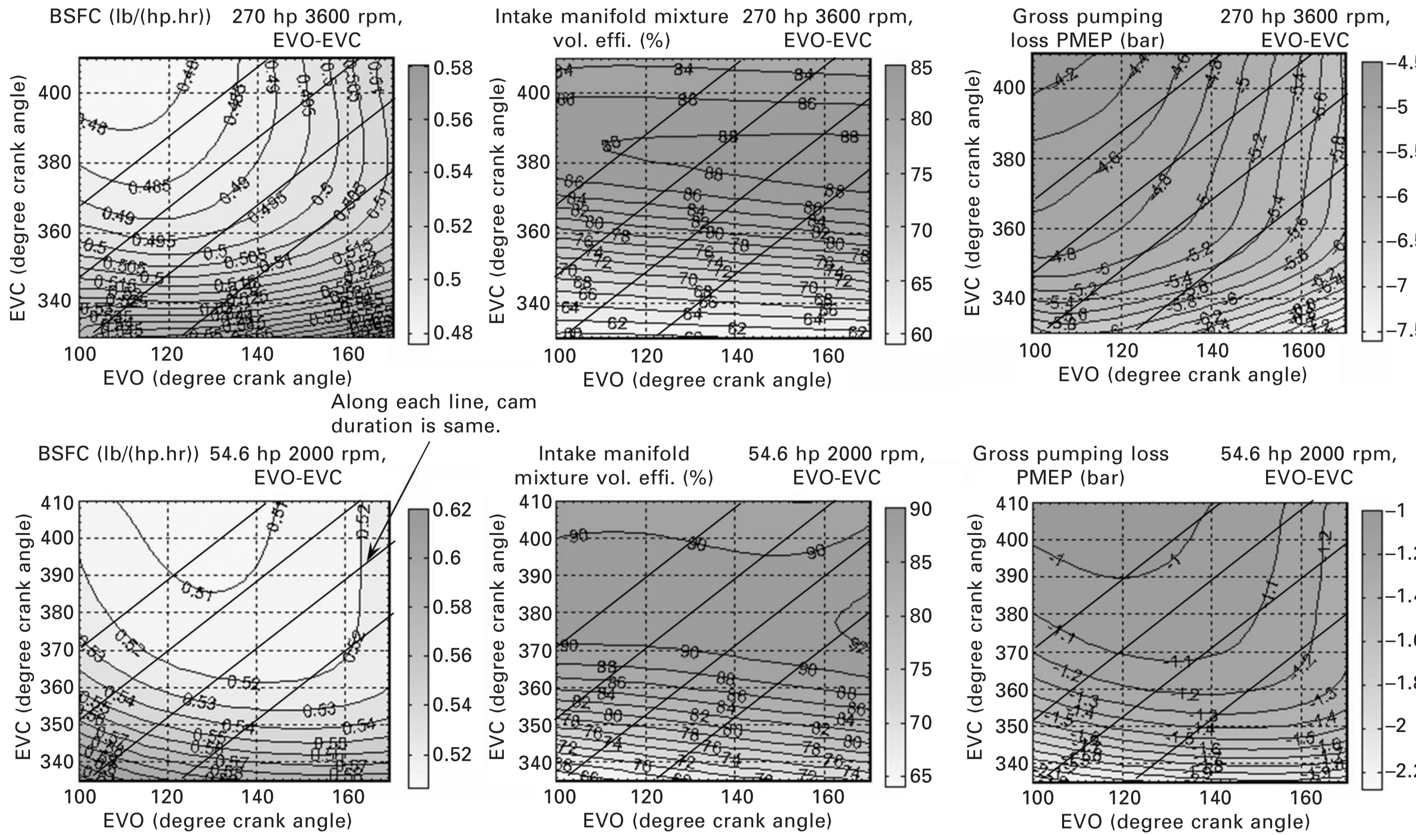
torque. In contrast, at rated power where the exhaust manifold pressure is higher than the intake manifold pressure, larger valve overlap results in an increase in the exhaust reverse flow (residue) and a reduction in air–fuel ratio and hence an increase in the exhaust manifold gas temperature. Note that the valve overlap also has an impact on engine delta P due to the variation in volumetric efficiency. Such an interaction can be explained by the core theoretical formula of the engine air system shown in Chapter 4. Figure 9.12(c) analyzes the interaction between IVC timing and turbocharging, and is explained in detail in Section 9.7.

The IVC timing affects the effective compression ratio and the cold-start capability of the engine. Figure 9.13 illustrates the effects of the IVC timing and engine geometric compression ratio on cold-start capability.

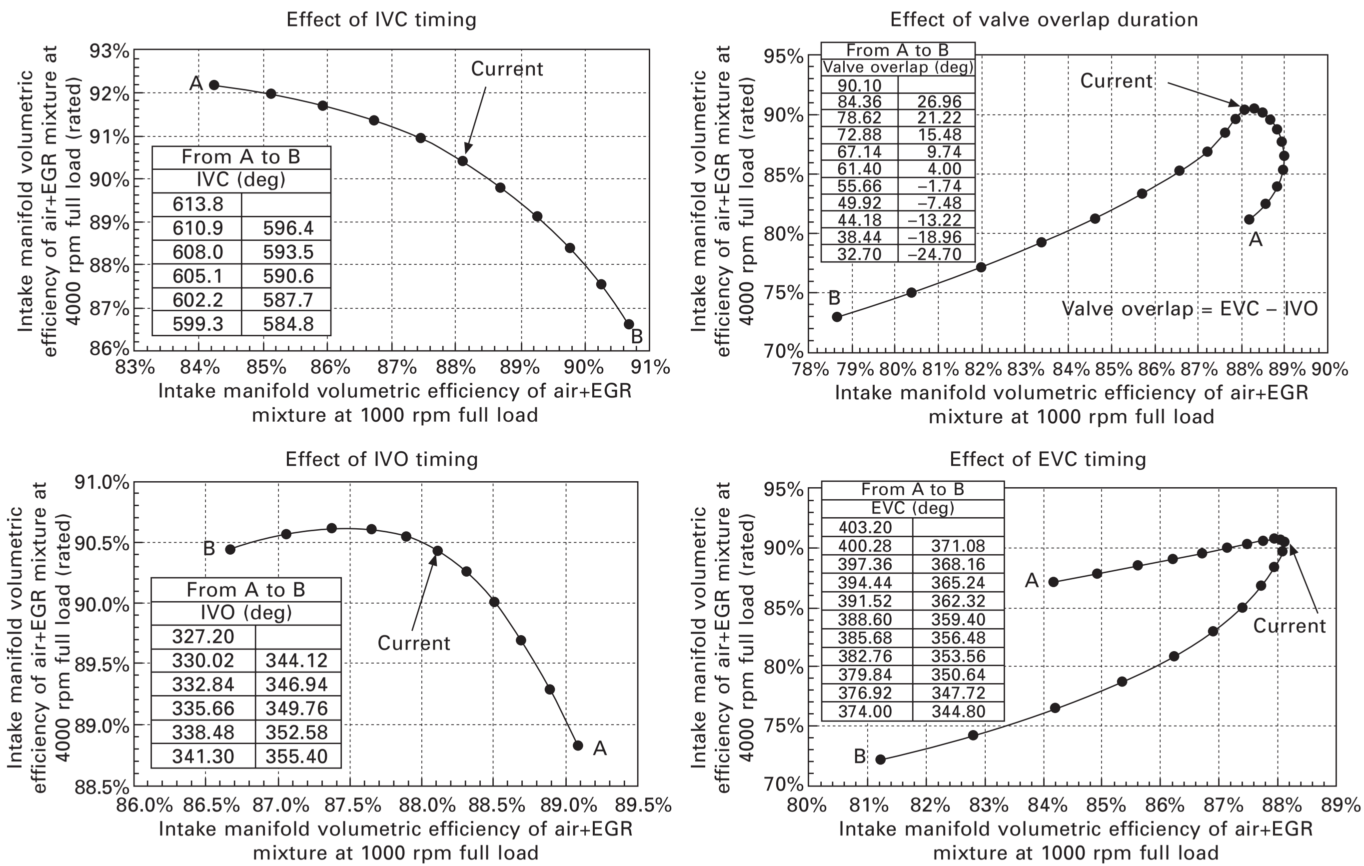
Figure 9.14 presents a sensitivity analysis about the effect of engine valve size on BSFC at a typical low speed and high speed. The analysis shows that at 2000 rpm 30% load, the BSFC is much more sensitive to the change in intake valve diameter (i.e., 0.65% BSFC reduction for every 1 mm increase in diameter) than the change in exhaust valve diameter (i.e., 0.29% BSFC reduction for every 1 mm increase). However, at 3000 rpm full load, the BSFC is more sensitive to the change in exhaust valve diameter (i.e., 0.24% BSFC reduction for every 1 mm increase) than the change in intake valve diameter (i.e., 0.21% BSFC reduction for every 1 mm increase). This indicates that the size of the exhaust valve becomes more important at high speeds where the exhaust manifold pressure and pumping loss are high.



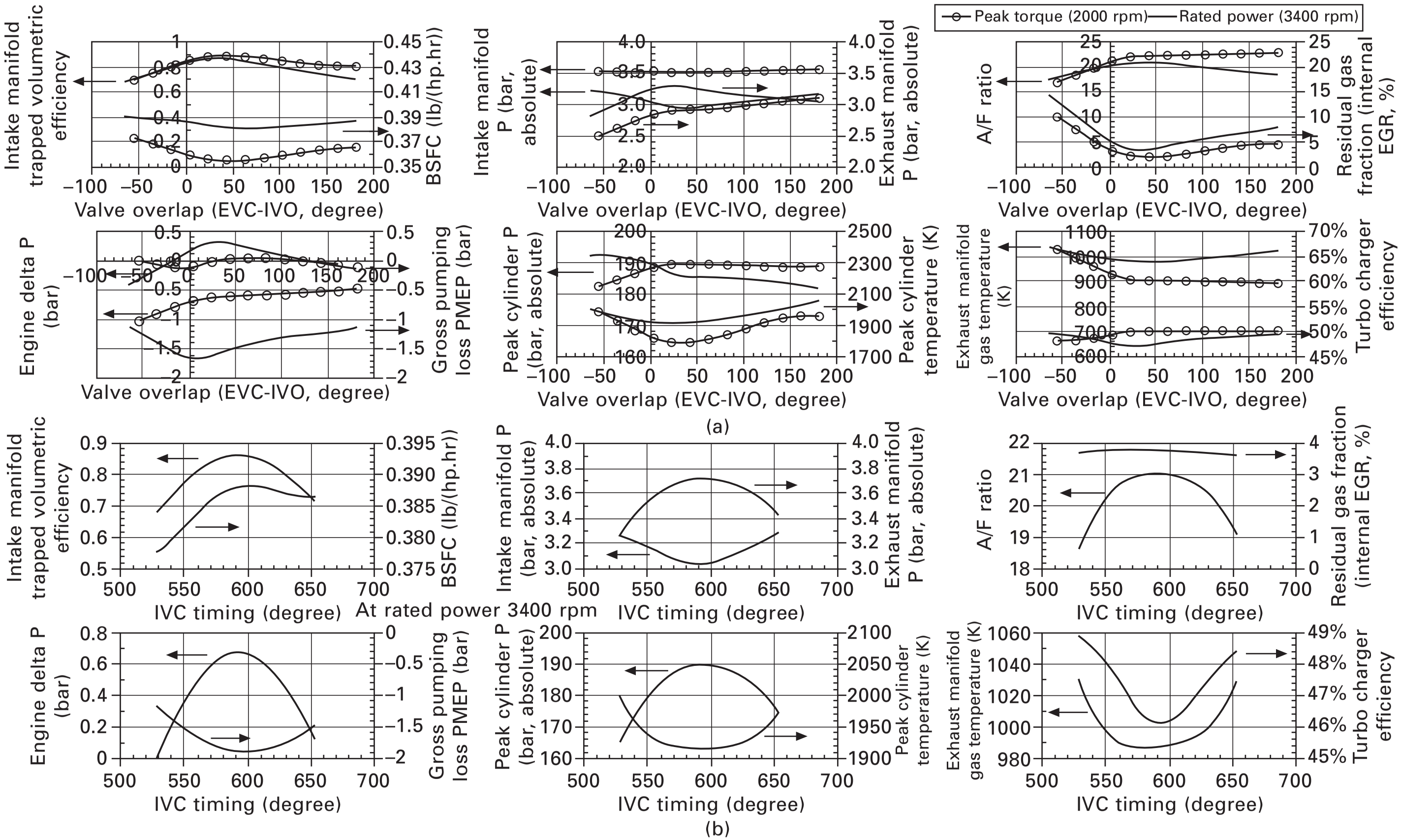
9.9 Intake valve timing effect at high speed (3600 rpm) and low speed (2000 rpm) of a LD diesel engine.



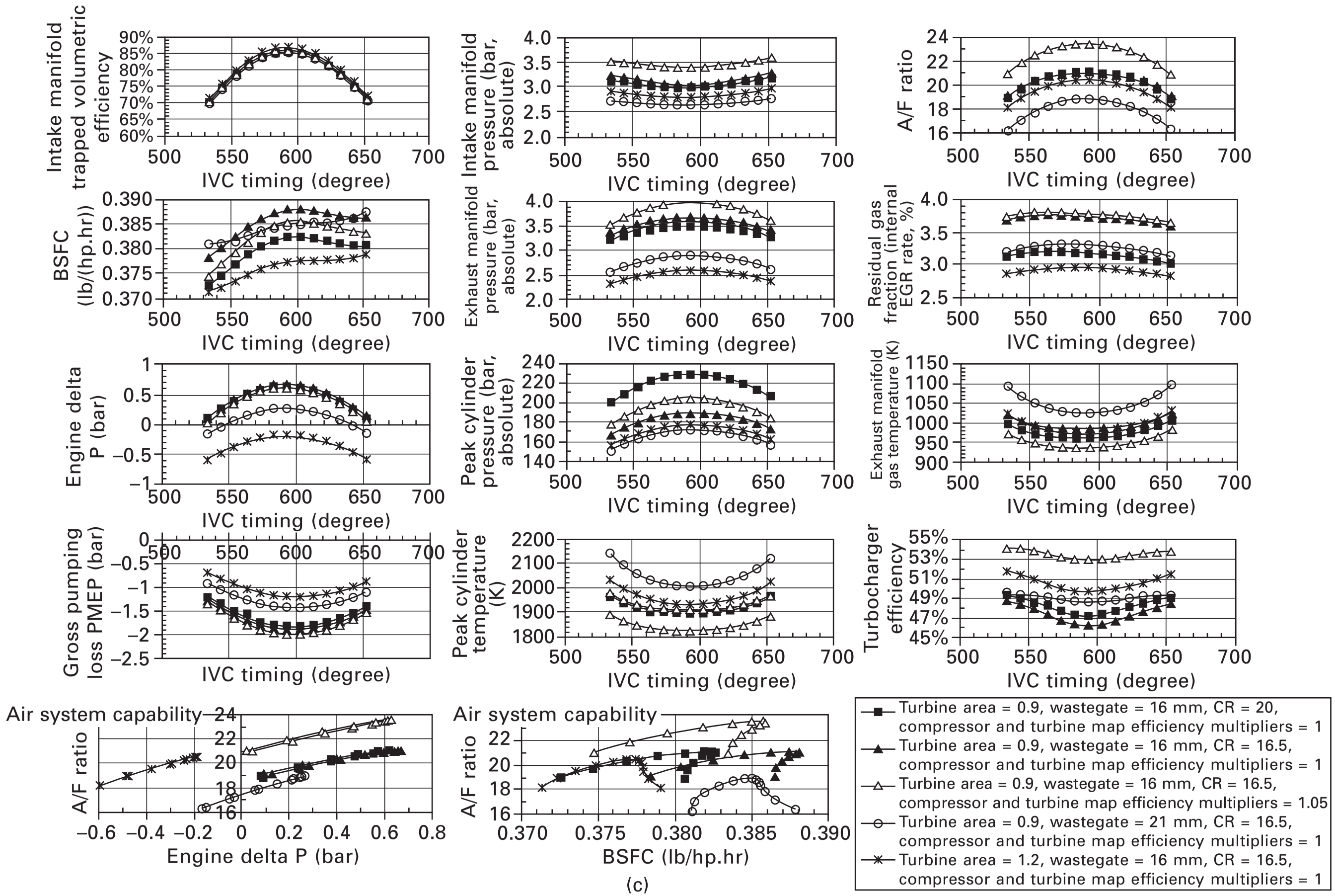
9.10 Exhaust valve timing effect at high speed (3600 rpm) and medium speed (2000 rpm) of a LD diesel engine.



9.11 Valve timing trade-offs of a LD diesel engine between high speed (4000 rpm) and low speed (1000 rpm).

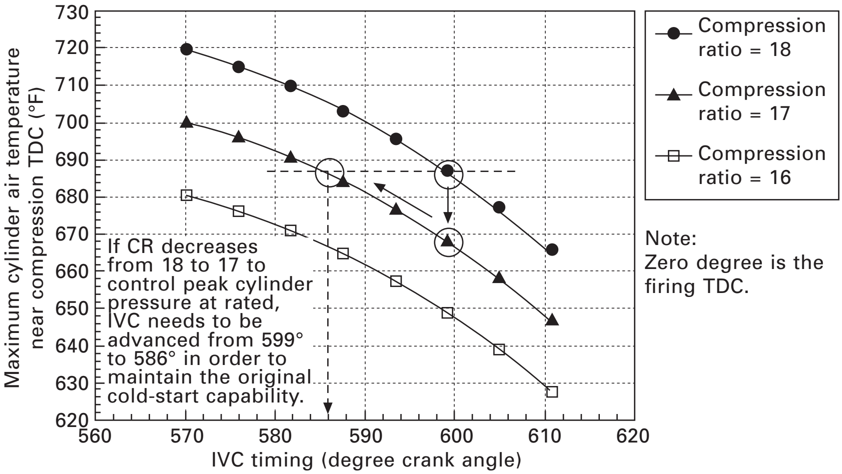


9.12 Effects of valve overlap (a) and IVC timing (b) and (c) on the performance of a non-EGR diesel engine.

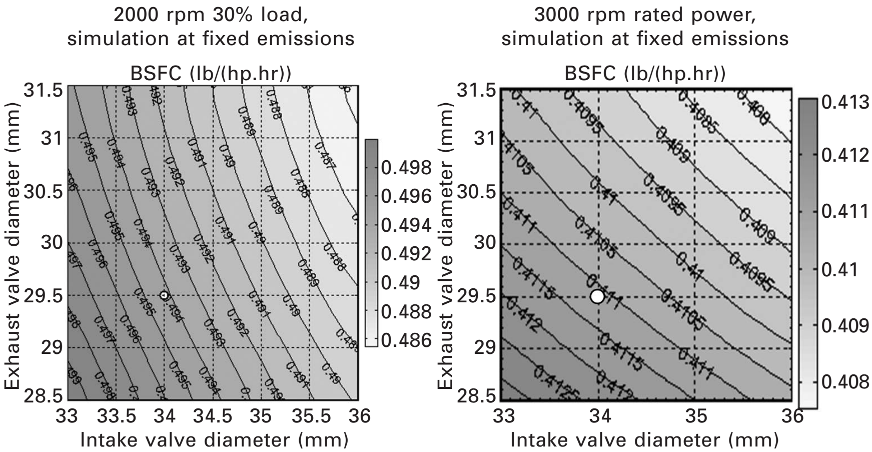


(c)

9.12 Continued



9.13 Effect of IVC timing and engine compression ratio on cold-start performance at 100 rpm cranking in –10°F cold ambient.



9.14 Effect of engine valve size on BSFC.

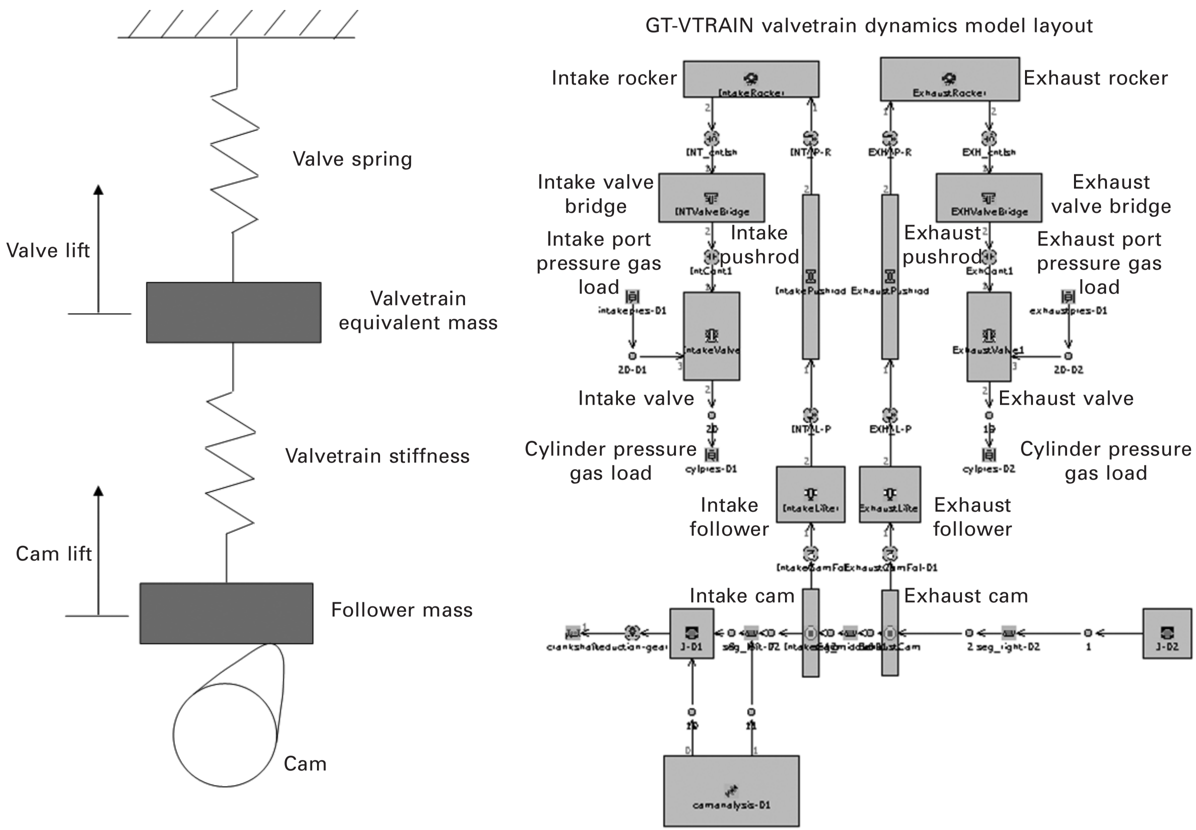
9.3 Valvetrain dynamic analysis

The simplest valvetrain dynamic analysis is the cam harmonic analysis to convert the cam acceleration or cam lift profiles from the time domain (or the crank angle domain) to the frequency domain without the need for using a valvertain dynamics model. The harmonic analysis is an effective and convenient method to compare the relative difference in the dynamic excitation source for the valvetrain, and can be used in cam design to judge the quality and the aggressiveness of the cam acceleration shape. The amplitudes of the harmonics indicate the energy contents of the harmonics. Both harmonic amplitude and harmonic number are important in determining the dominant harmonics of the cam acceleration as a source of vibration excitation for the valvetrain and valve spring surge. The harmonic peaks around approximately the 13th harmonic are important indicators of the tendency for valve spring surge. The higher the peaks, the greater the tendency is. A cam that possesses a good acceleration shape and induces low vibration usually has low harmonic amplitudes. An example of cam harmonic analysis is shown later in Fig. 9.18 in the section on cam design. More information about the frequency analysis of valvetrain vibrations is provided in Norton *et al*. (1998) and Grönlund and Larmi (2004).

Valvetrain dynamics modeling has been motivated by the following two needs: designing the cam and predicting the dynamic response of each valvetrain component. There are generally two types of valvetrain dynamics models (Fig. 9.15): the single-degree-of-freedom model, and the multi-body dynamics model. The single-degree-of-freedom model is a lumped simplification of the valvetrain system by lumping the component masses to a valvetrain equivalent mass and lumping the component stiffness to an overall valvetrain stiffness (Jensen, 1987). The damping and gas loading effects are ignored in the model. Moreover, the model is only valid for the configurations that have one-dimensional translational motion, such as the pushrod valvetrain and the direct-acting OHC valvetrain. The single-degree-of-freedom model can be used to output the Polydyne cam lift profile (Jensen, 1987; Thoren *et al*., 1952; Stoddart, 1953). The Polydyne cam lift profile is computed based on a prescribed input of a smooth non-vibratory dynamic valve lift profile at a 'design speed'. At other engine speeds, the model outputs vibratory dynamic valve lift profiles.

Although the multi-body dynamics model does not have the simplification or restrictions mentioned above, the model only uses cam profile as an input and does not have the ability to output a cam design. An advanced integrated approach requires using the single-degree-of-freedom model to conduct Polydyne cam profile design and then using the multi-body dynamics model to check the dynamic response of the cam by considering the sophisticated effects such as the stiffness and damping of each individual component, gas loading, and the dynamic behavior of the hydraulic lash adjuster (if any). Note that recompression pressure control is especially important for the valvetrain dynamics of modern diesel engines. The recompression pressure analysis is first conducted with a cycle simulation model (e.g., GT-POWER) and then the result is input into the multi-body valvetrain dynamics model.

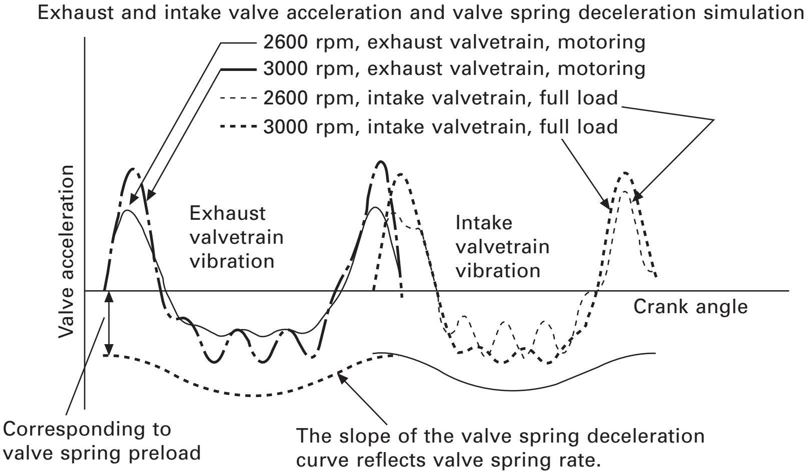
Typical simulation results of valvetrain dynamics are show in Fig. 9.16. At low engine speeds, the pushrod force (or cam force) is dominated by the valve spring force because the dynamic inertia force is low. At high speeds, the vibratory valve acceleration force becomes prominent, and therefore the



(a) Single-degree-of-freedom model

(b) Multi-body dynamics model (GT-VTRAIN)

9.15 Illustration of valvetrain dynamics models.



9.16 Valvetrain dynamics simulation.

pushrod force exhibits an oscillatory pattern within the cam event cycle. Figure 9.16 also illustrates the valve spring deceleration, which is equal to the spring force divided by the valvetrain equivalent mass. The intersection of the spring deceleration on the vertical axis corresponds to the spring preload, and the slope of the spring deceleration curve corresponds to the spring rate. Once the valve acceleration and spring deceleration cross over, valvetrain no-follow occurs. The appropriate choices of valve spring preload and spring rate can be analyzed by visually examining the relative difference between the valve acceleration and the spring deceleration as shown in Fig. 9.16.

The scope of valvetrain dynamics not only includes the evaluation of the valvetrain vibration forces within a cam event cycle (i.e., single-cylinder valvetrain dynamics), but also includes the torsional vibration and the dynamic bearing forces of the camshaft (i.e., multi-cylinder valvetrain dynamics). The single-cylinder valvetrain dynamics simulation is presented by Johnson (1962), Dennis and Neuser (1966), Akiba *et al*. (1981), Pisano and Freudenstein (1983a, 1983b) Pisano (1984, 1986), Chan and Pisano (1987), Akiba and Kakiuchi (1988), Kurisu *et al*. (1991), Hellinger *et al*. (1992), Gast and David (1996), Keribar (2000), Iritani *et al*. (2002), Grönlund and Larmi (2004), Ito (2006), and Xu *et al*. (2007). The dynamic modeling of the multi-cylinder valvetrain and camshaft system is detailed by Reinicke-Murmann and Kreuter (1990), Roß and Arnold (1993), Okamura and Yamashita (1997), Stout (1997), Speckens *et al*. (1999), Du and Chen (2000), Takagishi *et al*. (2004), and Londhe *et al*. (2009).

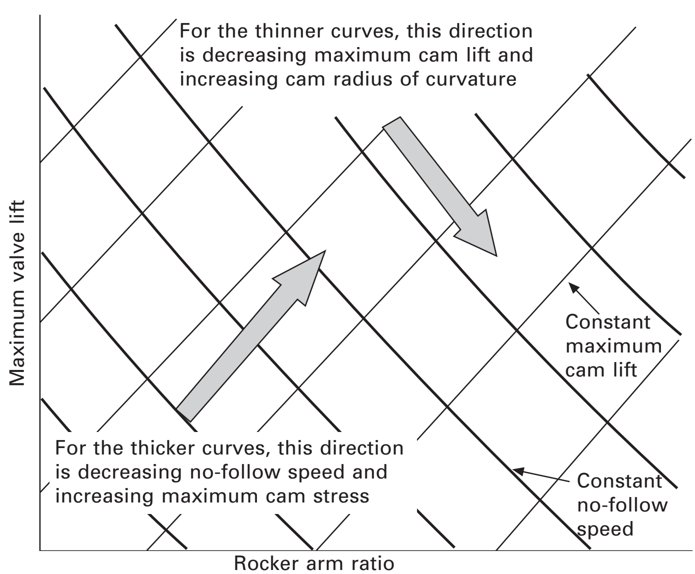
9.4 Cam profile design

9.4.1 Valvetrain system parameters in cam design

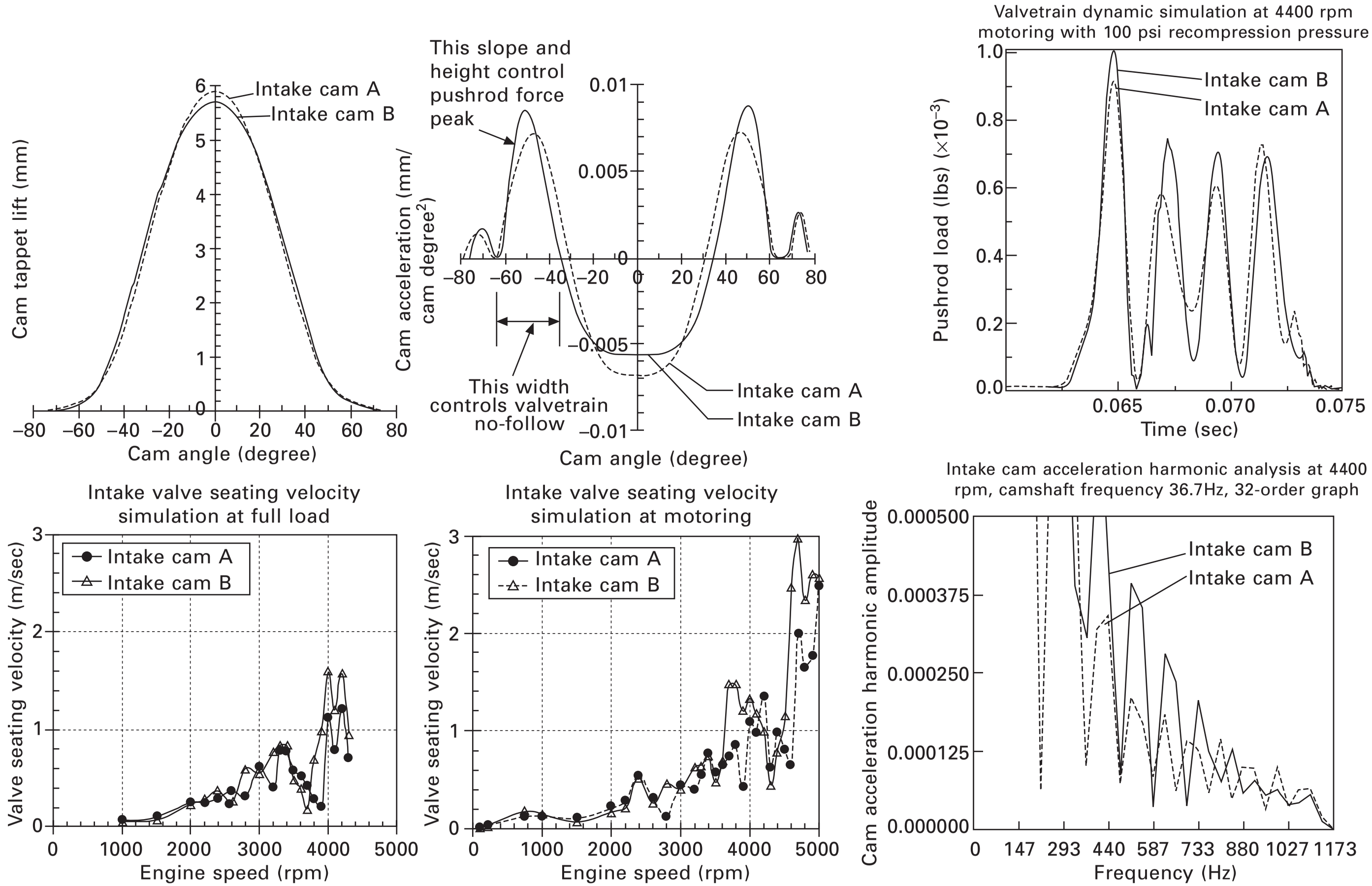
Before the cam lobe profile is designed, several system-level parameters need to be determined as the input data to be used in cam design: maximum valve lift, rocker arm ratio, maximum cam lift, maximum allowable cam stress, no-follow speed, and minimum allowable cam radius of curvature (ROC). Their relationship is illustrated in Fig. 9.17. The radius of curvature is further related to cam base circle radius and the roller radius of the follower.

9.4.2 Impact of cam profile on valvetrain dynamics

The bad design of using discontinuous or non-smooth shape of cam acceleration and jerk should always be avoided simply because they are not optimized from a valvetrain dynamics point of view. Good cam design is a trade-off between the valve flow area and valvetrain vibration. It is an on-target design to achieve the allowable maximum cam stress and no-follow speed target by properly selecting the maximum cam lift and designing the cam acceleration shape. Figure 9.18 illustrates the cam design results of two intake cam lobe profiles for a pushrod valvetrain. The intake cam A is the design that satisfies the no-follow requirement. Although the intake cam B provides more valve flow area at the cam flanks due to its more aggressive cam acceleration profile, it produces unacceptable no-follow conditions



9.17 Parametric design chart for cam design.



9.18 Cam design comparison and impact on valvetrain dynamics.

and higher valve seating velocities than the intake cam A. Note that the intake cam A has lower harmonic amplitudes than the cam B in the cam acceleration harmonic analysis. This difference in the excitation source of the valvetrain directly correlates to the difference in the pushrod vibration forces simulated.

9.4.3 Dynamic cam design

In general there are two categories of methods for cam lobe design: dynamic and static (kinematic). In the kinematic design method, the cam lift profile is obtained by analytically solving the kinematic equations or numerically integrating twice from a prescribed cam acceleration profile, and then a valvetrain dynamics model can be used to check whether the cam lift profile gives satisfactory dynamic response. In the dynamic design method, the cam lift profile is obtained by solving the equations of single-degree-of-freedom valvetrain dynamics (Fig. 9.15) for a prescribed target profile of the dynamic valve lift at a given (selected) 'design speed'. The cam profile obtained by the dynamic method can be checked for valvetrain dynamic response at other speeds by using either a single-degree-of-freedom model or a more accurate multi-body valvetrain dynamics model.

The most widely used dynamic cam design method is the Polydyne theory established by Stoddart (1953). In the Polydyne method, there is an engine speed at which the valvetrain vibration amplitude induced by the cam profile becomes zero theoretically, resulting in a smooth and non-vibratory valve lift profile under the dynamic inertia effect. This unique speed is called the 'design speed' of the cam. Although based on a simplified single-degree-of-freedom dynamic model, the Polydyne method does offer two superior advantages compared with the static cam design approach where any dynamic effect cannot be accounted for directly in the cam profile design. First, the Polydyne method allows the designer to specify a desired target dynamic valve lift profile at the design speed as the cam design input, and then calculate the corresponding cam profile based on the knowledge of valvetrain weight, stiffness, rocker ratio, etc. The cam profile is assembled from the motion desired at the valve and the deflections caused by the spring force and the inertia force. Secondly, the method uses a single-piece polynomial to define the valve lift profile at the design speed for the entire valve event (also called the main cam event) so that the higher-order derivatives of the cam lift profile (i.e., cam acceleration and jerk) can be continuous and much smoother than the multi-piece polynomial cam profile.

The valve motion in the single-degree-of-freedom model can be expressed as

$$m_{VT}\ddot{l}_{VAL} = K_{s,VT}\, l_{CAM,eq} - (K_{s,VT} + K_{s,SP})\, l_{VAL} \qquad 9.5$$

where m_{VT} is the valvetrain equivalent mass, l_{VAL} is the dynamic valve lift as a function of cam angle, $\ddot{l}_{VAL}$ is the valve acceleration, $l_{CAM,eq}$ is the equivalent cam lift at the valve side as a function of cam angle beyond the cam ramps, $K_{s,VT}$ is the valvetrain stiffness, and $K_{s,SP}$ is the valve spring stiffness (i.e., spring rate). Equation 9.5 can be rewritten as follows for the valve motion:

$$f_{dd}\ddot{l}_{VAL} + f_{sd} l_{VAL} = l_{CAM,eq} \qquad 9.6$$

where f_{dd} is the dynamic deflection factor, $f_{dd} = m_{VT}/K_{s,VT}$; and f_{sd} is the static deflection factor, $f_{sd} = (K_{s,VT} + K_{s,SP})/K_{s,VT}$. Once the cam lift is known, the valve lift can be solved analytically with equation 9.6 at each engine speed. The concept of the Polydyne cam design is to prescribe a desired smooth valve lift profile without any vibration at a selected engine speed, the design speed, and compute the cam lift profile by using equation 9.6 at the design speed. At any other speeds, the dynamic valve lift profile calculated by equation 9.6 would be non-smooth or vibratory.

The desired valve lift profile can be assumed in the form of a polynomial as

$$l_{VAL,design} = C_0 + C_1\phi_{CAM} + C_2\phi_{CAM}^2 + \ldots + C_n\phi_{CAM}^n \qquad 9.7$$

where ϕ_{CAM} is the cam angle, and $C_0 \sim C_n$ are the coefficients that need to be determined according to certain constraints or boundary conditions. The more boundary conditions specified, the more coefficients can be constructed in equation 9.7. Commonly used boundary conditions include the following: (1) at the end of the valve event, the valve lift, velocity and acceleration are zero; and the valve jerk and the rate of change of jerk can be prescribed finite values; (2) at the cam nose, the valve lift is equal to the desired maximum valve lift; the valve velocity is zero; the valve acceleration is a prescribed finite value; and the valve jerk is zero. To ensure continuity and smoothness of the valve lift profile, usually the first four time derivatives of the valve lift profile are controlled as the boundary conditions at the end of the valve event. Usually, the desired valve lift at the design speed of the cam has the following form:

$$l_{VAL,design} = l_{VAL,max} + C_2\phi_{CAM}^2 + C_3\phi_{CAM}^{C_p} + C_4\phi_{CAM}^{C_q} + C_5\phi_{CAM}^{C_r} + C_6\phi_{CAM}^{C_s} \qquad 9.8$$

where $l_{VAL,max}$ is the maximum valve lift, and C_p, C_q, C_r and C_s are the exponents of the polynomial power series specified by the designer. The corresponding valve acceleration becomes the following:

$$a_{VAL,design} = \ddot{l}_{VAL,design}$$

$$= C \cdot N_{design}^2 \begin{bmatrix} 2C_2 + C_3 C_p (C_p - 1)\phi_{CAM}^{C_p - 2} \\ + C_4 C_q (C_q - 1)\phi_{CAM}^{C_q - 2} + C_5 C_r (C_r - 1)\phi_{CAM}^{C_r - 2} \\ + C_6 C_s (C_s - 1)\phi_{CAM}^{C_s - 2} \end{bmatrix} \quad 9.9$$

where N_{design} is the design speed and C is a constant for unit conversion. The coefficients C_2, C_3, C_4, C_5 and C_6 are calculated using the boundary conditions described above. Then the equivalent cam lift can be obtained by considering both the dynamic and static deflections of the valvetrain according to the following:

$$l_{CAM,eq} = f_{dd} a_{VAL,design} + f_{sd} l_{VAL,design} \quad 9.10$$

At any other speed than the design speed, the valve lift is solved by equation 9.6.

The actual cam lift is calculated from the equivalent cam lift as follows:

$$l_{CAM} = \frac{l_{CAM,eq}}{f_{RA}} + l_{CAM,ramp} \quad 9.11$$

where f_{RA} is the rocker arm ratio, and $l_{CAM,ramp}$ is the ramp height at the cam side as a function of cam angle. The nominal rocker arm ratio is defined as the maximum valve lift divided by the maximum cam lift. Strictly speaking, the rocker ratio is actually a variable as a function of cam angle or valve lift. Its value varies slightly from the position of valve closed to the position of valve fully open, and reaches the maximum at the cam nose. The cam ramp design is detailed in Section 9.4.4.

Equation 9.11 includes all the related factors that influence the cam profile and the cam acceleration shape. Among them, the polynomial powers and the design speed have the dominant effects. The polynomial powers affect the height of the two positive pulses (humps) in the cam acceleration curve and the shape of the acceleration curve (Fig. 9.18). The width of the positive cam acceleration hump is strongly affected by the design speed, and increases as the design speed increases. The valve lift profile at other speeds is vibratory due to different dynamic inertia effect. The vibration and the no-follow speed are certainly affected by the polynomial powers and the design speed selected. It is desirable to choose a design speed at the frequently encountered engine speed since at the design speed by definition the vibration amplitude is zero theoretically. However, very often it is not possible to satisfy this condition because of the need to achieve an acceptable no-follow speed. Usually, the design speed has to be selected higher than the normal operating speed.

Longer duration of the main cam event results in lower vibration amplitude, higher no-follow speed and lower contact stress between the cam and the follower. The influence of the maximum cam lift is opposite to the effect of the main cam event: higher maximum cam lift results in higher vibration amplitude and higher cam stress.

Cam acceleration profile must be very smooth in order to achieve good dynamic characteristics especially at high engine speeds. The derivative of acceleration, cam jerk, also needs to be continuous and preferably smooth. Any discontinuous and sharp acceleration and its consequent inertia load are equivalent to a hammer blow to the valvetrain system and must be avoided. The width of the positive acceleration hump should be greater than the time of one cycle period corresponding to the valvetrain natural frequency in order to avoid valvetrain separation within the range of operating speeds. The initial compression of the valvetrain when the valve begins to open causes the vavletrain to vibrate or rebound at its resonant frequency. Using a wide acceleration hump ensures a compressive load acting on the valvetrain throughout the period of the rebound. The aggressiveness of cam acceleration can be characterized by a coefficient of cam positive acceleration hump, defined as

$$f_{CAM,ac} \triangleq \frac{f_{n,VT} \cdot \Delta\phi_{CAM,ac}}{N_E} \tag{9.12}$$

where $\Delta\phi_{CAM,ac}$ is the width of cam positive acceleration hump in crank angle, $f_{n,VT}$ is the valvetrain natural frequency, and N_E is the engine speed. Note that higher $f_{CAM,ac}$ results in lower valvetrain vibration and less valve flow area. The aggressiveness or roundness of the valve lift profile can be characterized by a coefficient of valve lift roundness as follows:

$$f_{lift} \triangleq \frac{\bar{l}_{VAL}}{l_{VAL,max}} \tag{9.13}$$

where $\bar{l}_{VAL}$ is the average valve lift over the valve event, and $l_{VAL,max}$ is the maximum valve lift. A larger f_{lift} gives larger valve flow area and less pumping loss.

In order to achieve lower valvetrain vibration and higher no-follow speed, the positive cam acceleration hump must be designed wider, lower, and smooth. There are three ways to achieve this goal: (1) increasing the valve event duration at a given maximum valve lift; (2) reducing the aggressiveness of the valve lift profile; or (3) reducing the maximum cam lift by reducing the maximum valve lift or increasing the rocker arm ratio. Unfortunately, all these measures have negative effects on either engine breathing performance or cam stress. For instance, longer intake valve duration and later IVC timing may reduce the volumetric efficiency at low engine speeds; and using larger rocker arm ratio may cause an increase in cam force and cam stress.

Cam radius of curvature is directly affected by cam base circle radius, roller radius of the follower (in the case of roller followers), cam lift, cam velocity, and cam acceleration. The minimum allowable cam radius of curvature should be determined as a design target since it affects cam stress and the cost or the feasibility of cam manufacturing. A too gentle cam acceleration profile tends to produce a sharp nose in the cam lift profile with a very small radius of curvature and hence increases the cam stress at the cam nose. On the other hand, a more aggressive cam acceleration profile results in excessively small or even negative radius of curvature at the cam flank. The aggressiveness of cam acceleration is often limited by the minimum allowable negative radius of curvature (for roller followers). This is especially true for the overhead cams where the valvetrain no-follow is usually not a limiting factor due to very high stiffness of the OHC valvetrain.

The elastic contact stress (i.e., the Hertzian stress) between the cam and the follower needs to be calculated by either a cylindrical contact (also called line contact) model or a more accurate elliptical contact model. The maximum allowable stress limits are different for different materials (Kitamura *et al.*, 1997), types of followers (flat-faced tappet, spherical follower, roller follower; Turkish, 1946), surface crown designs (flat-faced without a crown, or crowned surface with a finite radius to avoid edge loading), surface finishes, and lubrication conditions (Wang, 2007). Cam flash temperature also needs to be calculated in cam design for the sliding motion with flat-faced followers because cam scuffing and wear failures are associated with the temperature in the contact zone. Usually, the highest flash temperature occurs at the highest combination of cam force and engine speed (sliding speed). Moreover, a good cam design needs to have satisfactory lubrication characteristics (e.g., oil film thickness). More detailed discussions on cam stress and wear are covered in Chapter 2 related to engine durability. Valvetrain lubrication, friction, and flash temperature are discussed in Chapter 10. More in-depth investigations on the cam flash temperature calculations are provided by Dyson and Naylor (1960), Yang *et al.* (1996), and Ito *et al.* (2001).

Detailed information on the mathematical formulations of dynamic cam design are presented by Thoren *et al.* (1952), Stoddart (1953), and Jensen (1987). A cam design approach, 'n-harmonic cam', was proposed by Hanaoka and Fukumura (1973) as a revision to the Polydyne cam design method. More discussions on cam design guidelines or techniques are provided by Roggenbuck (1953), Nourse *et al.* (1961), and Kanesaka *et al.* (1977).

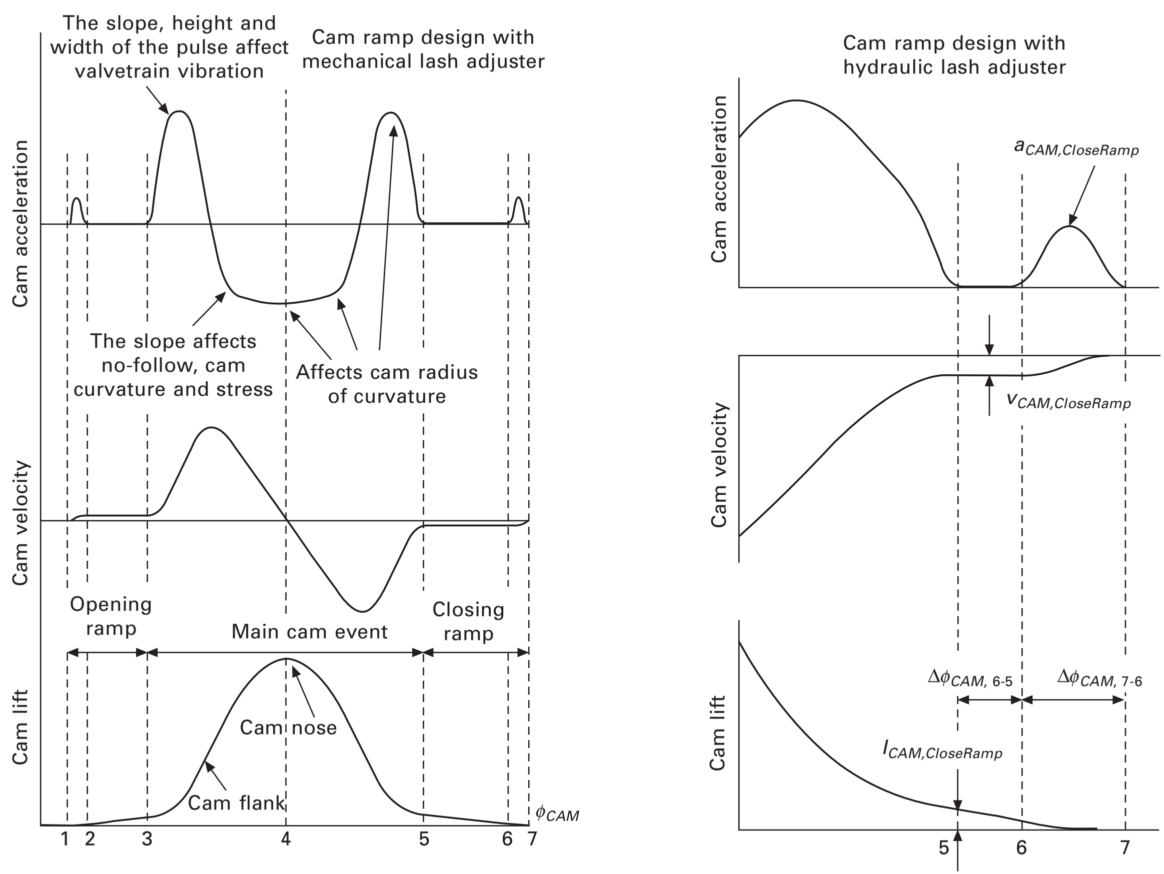
9.4.4 Kinematic cam design

The kinematic cam design method has a vital importance for the following three reasons. First, it connects the main cam event with the cam ramps via kinematic constraints. Secondly, it can produce the cam lift profile by directly integrating a prescribed cam acceleration profile when the dynamic cam design method fails or cannot modify the cam acceleration profile flexibly or locally. For example, when a special cam acceleration or lift profile or some local adjustment is needed, the single-piece polynomial used in the dynamic cam design method may be too restrictive to make the change. Thirdly, the kinematic cam design method is usually required for the end-pivot finger follower OHC valvetrain because there is no dynamic cam design method available to handle the two-dimensional motion of the follower. In this section, the cam ramp design is introduced first, followed by a description of the cam profile design within the valve event duration.

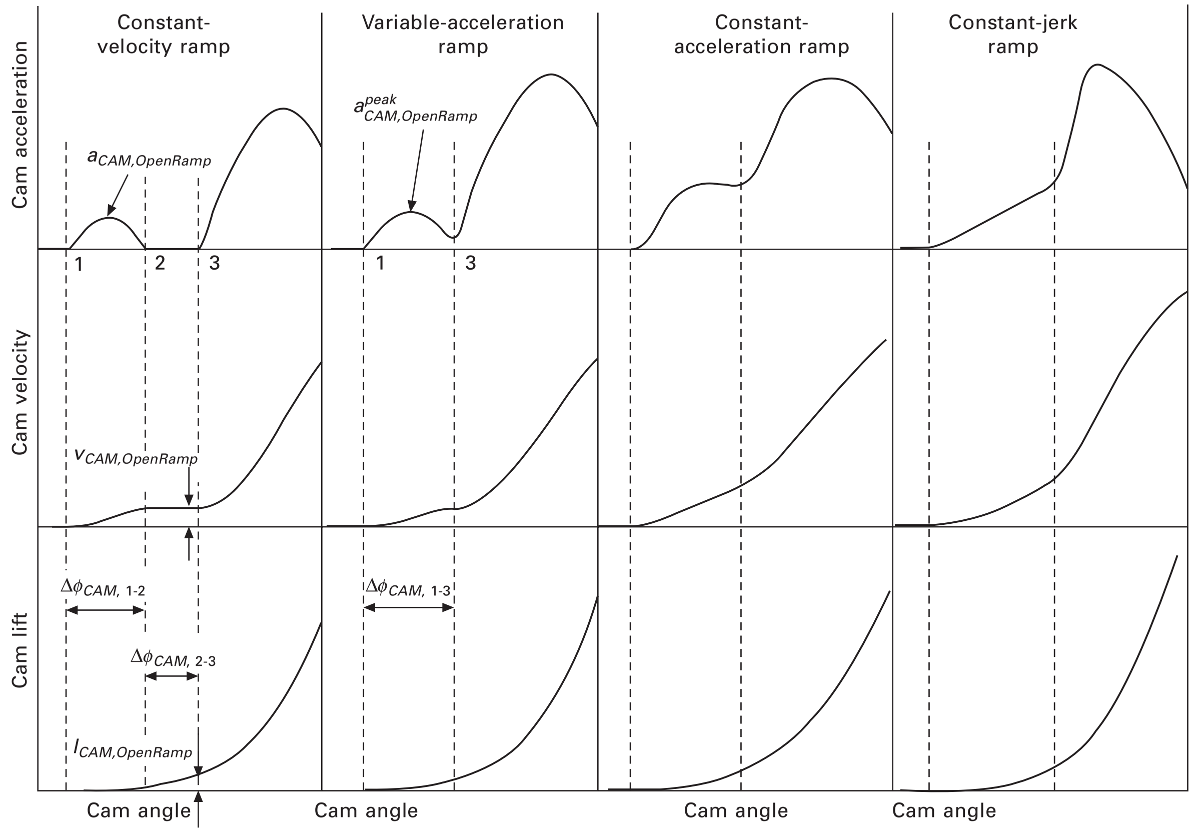
As shown in Figs 9.19 and 9.20, a cam profile consists of three sections: an opening ramp, a main cam event, and a closing ramp. The ramps must provide a smooth transition between the cam base circle and the main event. The function of the ramp is to take up the running clearances (e.g., the valvetrain lash and the elastic compression caused by the valve spring preload) before the valve opens or after the valve closes. Excessively long ramps should be avoided for the exhaust cam because long and slow opening of the valve on the ramp can cause excessive heating of the exhaust valve as the high-velocity exhaust gas passes through the small opening. Long intake cam ramp may lead to difficulties in cold start. Theoretically, the valve is supposed to open and close precisely at the top of the ramp at a prescribed controlled kinematic velocity. A constant-velocity ramp is most widely used at the opening and closing sides (Fig. 9.19), especially with mechanical lash adjusters, in order to control the opening and closing velocities. Accelerated opening ramps (i.e., variable-acceleration, constant-acceleration or constant-jerk opening ramps) are sometimes used with hydraulic lash adjusters to achieve faster lift-off and a shorter ramp because there is no concern about the cold lash and its associated mechanical impact (Fig. 9.20). The accelerated ramps may also reduce valvetrain vibrations because the positive peak of the cam acceleration may be reduced and a part of the dynamic inertia loading can be transmitted to and shared by the opening ramp before the valve opens. The reduced positive cam acceleration peak may enable a lower negative cam acceleration so that a softer spring may be used to reduce the valvetrain friction and the cam stress (Fig. 9.16).

The mathematical formulation of connecting the constant-velocity opening ramp (Fig. 9.20) with the main cam event is given as follows:

$$\begin{cases} v_{CAM,OpenRamp} = \int_0^{\Delta\phi_{CAM,1-2}} a_{CAM,OpenRamp} \mathrm{d}\phi_{CAM} \\ l_{CAM,OpenRamp} = \int_0^{\Delta\phi_{CAM,1-2}} \left(\int_0^{\phi_{CAM}} a_{CAM,OpenRamp} \mathrm{d}\phi_{CAM} \right) \mathrm{d}\,\phi_{CAM} \\ \qquad + v_{CAM,OpenRamp} \Delta\phi_{CAM,2-3} \end{cases} \tag{9.14}$$



9.19 Cam ramp blending in cam design.



9.20 Illustration of different types of cam opening ramps.

In equation 9.14, the opening ramp height, $l_{CAM,OpenRamp}$, and the opening ramp velocity, $v_{CAM,OpenRamp}$, are prescribed known inputs. The cam opening ramp acceleration, $a_{CAM,OpenRamp}$, is a prescribed polynomial (e.g., a second-order polynomial) as a function of cam angle with certain assumed acceleration height. The two unknowns in equation 9.14, $\Delta\phi_{CAM,1-2}$ and $\Delta\phi_{CAM,2-3}$, need to be solved iteratively. For the variable-acceleration ramp (Fig. 9.20), the two unknowns in the formulation of ramp blending are $\Delta\phi_{CAM,1-3}$ and $a^{peak}_{CAM,OpenRamp}$. The ramp blending is also required in the dynamic cam design.

The mathematical formulation for connecting the constant-velocity closing ramp with the main cam event is given similarly as follows (Fig. 9.19):

$$\begin{cases} v_{CAM,CloseRamp} = \int_0^{\Delta\phi_{CAM,7-6}} a_{CAM,CloseRamp}\,\mathrm{d}\phi_{CAM} \\ l_{CAM,CloseRamp} = \int_0^{\Delta\phi_{CAM,7-6}} \left(\int_0^{\phi_{CAM}} a_{CAM,CloseRamp}\,\mathrm{d}\phi_{CAM} \right) \mathrm{d}\,\phi_{CAM} \\ \qquad + v_{CAM,CloseRamp}\,\Delta\phi_{CAM,6-5} \end{cases} \tag{9.15}$$

The $l_{CAM,CloseRamp}$ and $v_{CAM,CloseRamp}$ are the actual cam lift and cam velocity, respectively, exactly at the end of the main cam closing event where the cam acceleration is equal to zero. The $a_{CAM,CloseRamp}$ is a prescribed polynomial as a function of cam angle with certain assumed acceleration height. The two unknowns in 9.15, $\Delta\phi_{CAM,7-6}$ and $\Delta\phi_{CAM,6-5}$, need to be solved iteratively.

There are generally two methods to generate the cam lift profile in the main cam event in the kinematic cam design, namely, analytical solution and numerical integration. The advantage of the analytical solution is that the ending points of the main cam events are precisely defined. The disadvantage is that the choice of the acceleration profile is limited to certain pre-defined mathematical functions (e.g., sine wave, harmonic wave, polynomial) and local adjustment is impossible. The numerical integration method is just the opposite. The coefficients used in the analytical solution that represent the cam acceleration profile are solved by using the design constraints or the boundary conditions described in the previous section. The boundary conditions are usually set at the end of the valve event and the cam nose. Higher-order time derivatives (at least to the cam jerk level) are usually required in order to ensure the continuity and smoothness of the cam acceleration profile. The multi-piece polynomial or multi-segment cam design method belongs to the analytical kinematic cam design method. In this case, more boundary conditions are required at the end of the multiple segments such as on the cam flanks (Heath, 1988; Keribar, 2000).

In the numerical integration method, it should be noted that simply integrating any arbitrary cam acceleration profile without the boundary conditions usually does not produce an acceptable cam lift profile for the following two reasons. First, the target cam ramp height and velocity cannot be guaranteed by the arbitrary cam acceleration. Secondly, the cam lift, velocity, and acceleration at the end of the cam event reach zero at different cam angles due to both the lack of kinematic constraints and the round-off errors accumulated during the numerical integration. The round-off errors exist even if a higher-order integration scheme and a small cam angle step are used (e.g., the fourth-order Runge-Kutta integration with 0.1° cam angle step). Appropriate cam ramp blending must be used to fix the round-off error problem at the closing ramp. The procedure of the numerical integration method is summarized in the following steps:

1. Use the target values of cam opening ramp height and cam opening ramp velocity to establish the opening ramp by solving equation 9.14.
2. Construct a smooth cam acceleration profile for the main cam event.
3. Integrate the cam acceleration with respect to cam angle to obtain cam velocity, and integrate the cam velocity to obtain the cam lift profile.
4. Adjust the cam acceleration until the target maximum cam lift is achieved, and match the closing ramp height and velocity at zero cam acceleration to the target values as closely as possible.
5. Use the actual closing ramp height and velocity to establish the closing ramp by solving equation 9.15 for acceptable ramp duration. Integrate the cam acceleration to obtain the cam ramp velocity and lift profiles to the end of the cam closing ramp.

More information on cam ramps can be found in Norton *et al.* (1999).

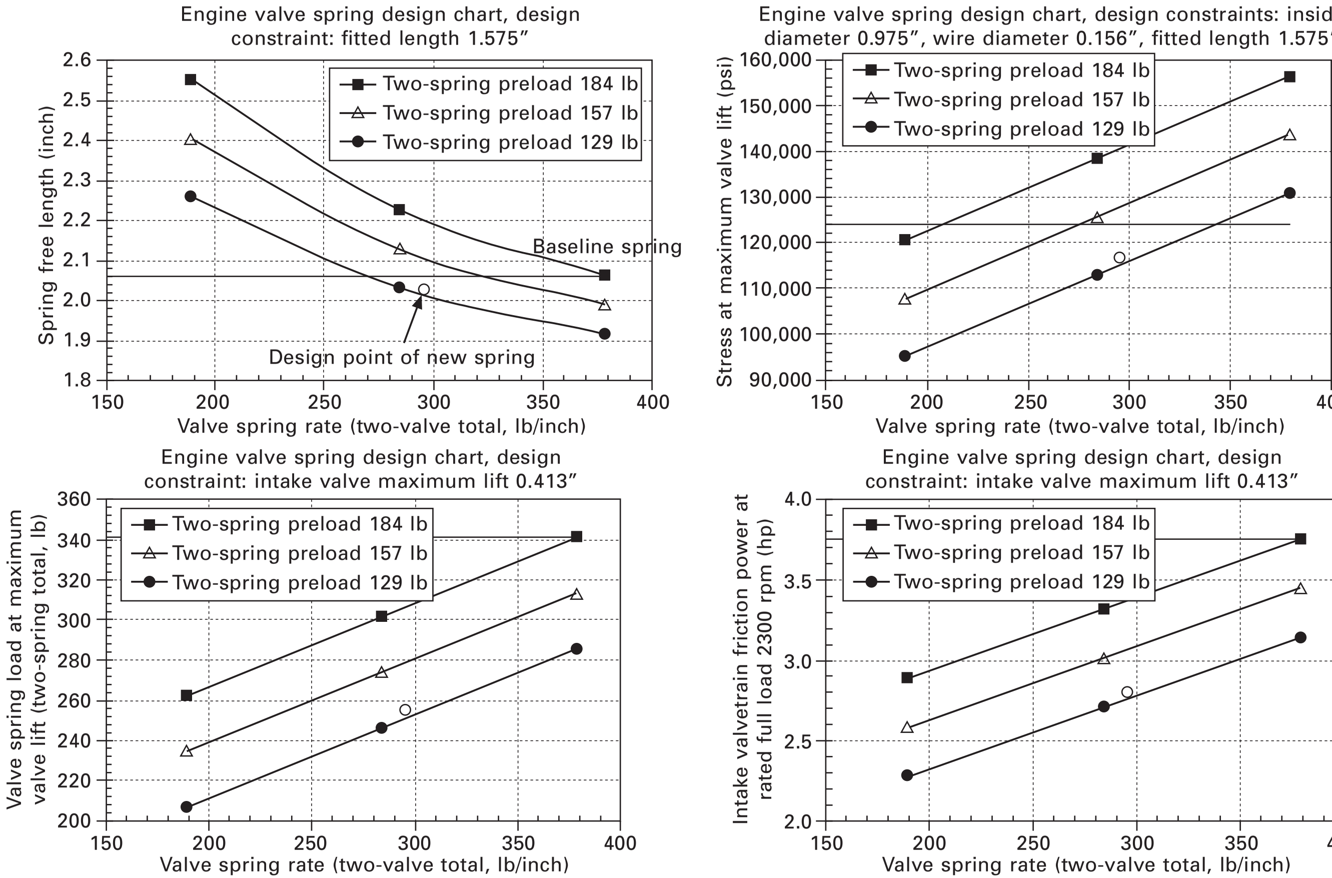
9.5 Valve spring design

9.5.1 Analytical design method for valve springs

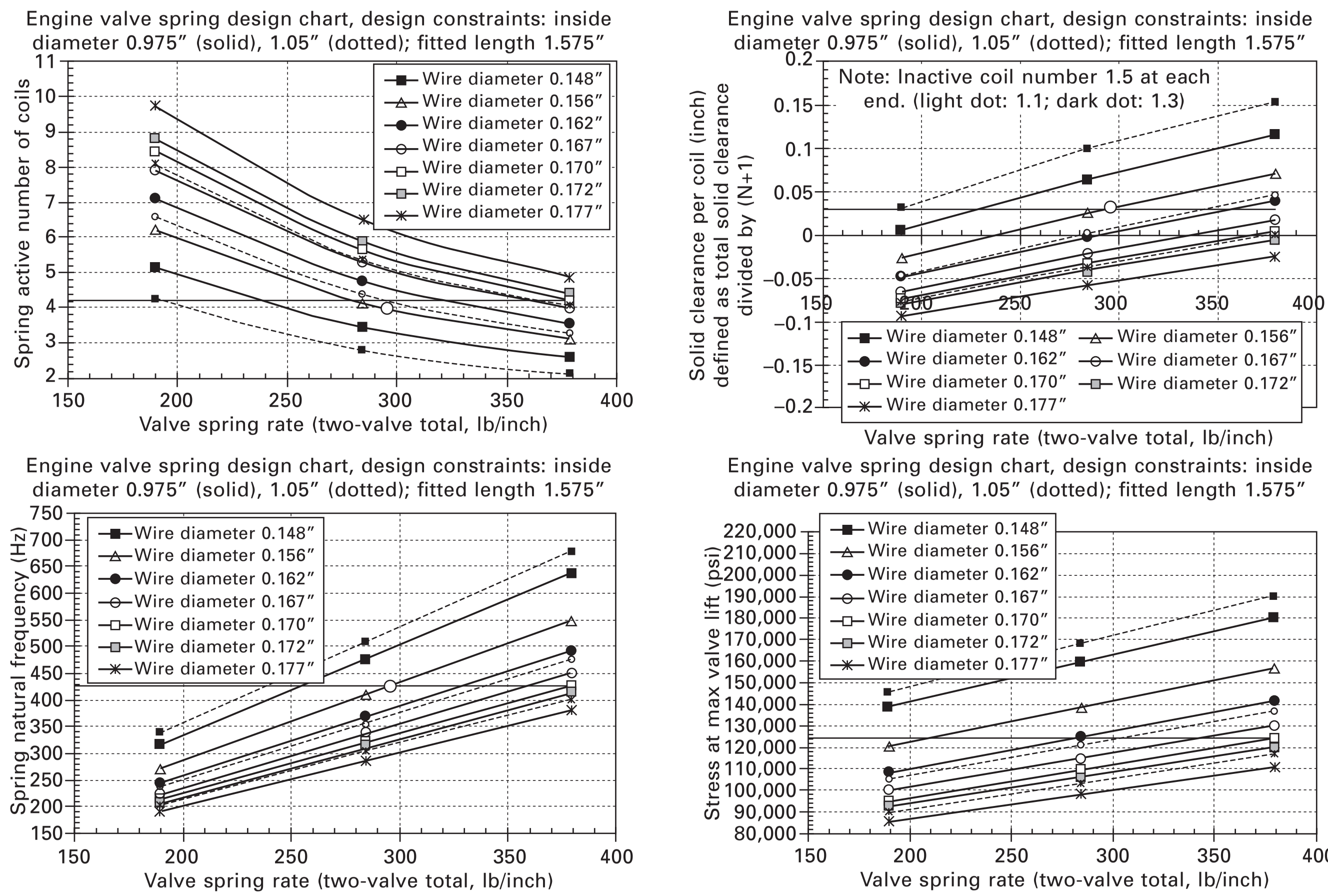
Valve spring design has equal importance as cam design for engine system performance. The functions of the valve spring include preventing the valve from floating off the seat under gas pressure loading and controlling the valve motion to avoid valvetrain separation. Valve spring design affects cam stress, valvetrain friction, and spring surge. Engine valve springs are usually open coil helical compression springs with closed ground ends. Most engines use constant-rate springs although some engines use variable-rate springs. The single-spring design is usually sufficient for diesel engines, but sometimes the double-spring design using a damper spring or an inner spring is required in order to reduce the severity of valve spring surge. Valve spring design is a very complex task. It may serve as a good showcase to illustrate the principles of engine system design for the following three reasons. First, the analytical spring design method illustrates the link between the component design parameters and the system design parameters. Secondly, the analytical spring design method illustrates the mathematical formulations for the same design topic as a deterministic solution or an optimization solution. In the optimization formulation, explicit objective functions and constraint functions are used. Note that in most of other areas of engine system design (e.g., cycle performance, cam design, and valvetrain dynamics) the optimization functions are usually the more complex implicit functions. Thirdly, the analytical spring design method provides an example of using a graphical design approach to construct parametric sweeping maps to handle the multi-dimensional design problems that are frequently encountered in diesel engine system design.

In valve spring design, the input data include the following: (1) the maximum valve lift; (2) the given installed (fitted) spring length; (3) the required spring preload; and (4) the required spring rate. Note that the spring preload and spring rate are system-level design parameters and need to satisfy the requirements of the maximum allowable spring force and cam stress, exhaust valve sealing, and valvetrain no-follow control. There is a strong interaction between valve spring design and cam design. If it is difficult to find a solution to spring design, the input data have to be revised.

The following parameters are calculated as outputs in valve spring design: (1) basic or independent spring design parameters (i.e., spring mean diameter, spring wire coil diameter, the number of active coils); and (2) derived design parameters (e.g., spring free length, maximum compressed length, solid length, the free clearance between the coils, the solid clearance at the maximum compression between the coils, spring natural frequency and surge order, the maximum spring load, and the maximum spring torsional stress). The basic spring design parameters determine the spring rate. Spring design is a multi-dimensional design problem that can be handled by a graphical approach to examine the parametric sensitivity trend (Fig. 9.21). Some of the output parameters are subject to certain design constraints. For example, installed length and spring mean diameter are subject to the constraints of packaging space. The spring torsional stresses at the maximum spring compression and at the solid length are subject to the constraints of spring fatigue life/strength and the maximum allowable stress limit. The constraint of spring surge protection is applied to the solid clearance and the spring natural frequency. The spring surge order is the ratio of the spring natural frequency to the engine operating frequency. In order to assure that the spring will not surge in operation, the natural frequency of the valve spring should usually be at least 13 times the operating frequency of the engine, i.e., a surge order higher than 13th is usually preferred. If the spring natural frequency analysis shows that the spring is responsive to one of the dominant harmonics of the cam profile, the tendency of spring surge definitely exists. In this case, a design revision is required for the cam or the spring. Sometimes variable-rate springs or nested springs can be used to vary the spring frequency in order to help alleviate the spring surge problem.



9.21 Analytical engine valve spring design method (intake valve spring).



9.21 Continued

The objective of valve spring design optimization is to maximize the spring natural frequency to minimize spring surge, subject to the following constraints: (1) the system requirements of spring preload and spring rate; (2) the maximum allowable spring stress; and (3) appropriate solid clearance to control spring surge vibration.

For more information on valve spring design and spring dynamics, the reader is referred to Wang (2007), Turkish (1987), Phlips *et al*. (1989), Muhr (1993), Schamel *et al*. (1993), Lee and Patterson (1997), Takashima *et al*. (2005), SAE J1121 (2006) and J1122 (2004).

9.5.2 Deterministic and optimization solutions of spring design

The valve spring design formulae are summarized as below in order to illustrate the parametric dependency among the valvetrain parameters. Further details on the derivations of the formulae can be found in many handbooks on spring design. The static spring outside diameter is calculated by

$$d_{SP,outer} = d_{SP,inner} + 2d_{SC} \tag{9.16}$$

where $d_{SP,in}$ is the static spring inside diameter, and d_{SC} is the spring coil wire diameter. The static spring mean diameter is defined as:

$$d_{SP} = 0.5(d_{SP,in} + d_{SP,out}) \tag{9.17}$$

The valve spring deflection Δl_{SP} at the spring load F_{SP} is calculated by

$$\Delta l_{SP} = \frac{8F_{SP}d_{SP}^3 n_{SC,active}}{\vartheta_{m,SP}d_{SC}^4} \tag{9.18}$$

where $n_{SC,active}$ is the number of active coils, and $\vartheta_{m,SP}$ is the torsional modulus of rigidity of the spring. Since the spring rate $K_{s,SP} = F_{SP}/\Delta l_{SP}$, the spring rate can be expressed as:

$$K_{s,SP} = \frac{\vartheta_{m,SP}\, d_{SC}^4}{8d_{SP}^3\, n_{SC,active}} \tag{9.19}$$

where d_{SP}, d_{SC} and $n_{SC,active}$ have selected design values. If $K_{s,SP}$ is a known input requirement, the number of active coils can be calculated by re-arranging equation 9.19 as:

$$n_{SC,active} = \frac{\vartheta_{m,SP}\, d_{SC}^4}{8d_{SP}^3\, K_{s,SP}} \tag{9.20}$$

The spring's free length at zero load can be calculated with the given spring preload F_{pre} and the installed spring length $l_{SP,1}$ as follows:

$$l_{SP,0} = l_{SP,1} + \frac{F_{pre}}{K_{s,SP}} \tag{9.21}$$

The maximum compressed spring length at the maximum valve lift is given by

$$l_{SP,2} = l_{SP,1} - l_{VAL,\,max} \tag{9.22}$$

The spring's solid length can be calculated by:

$$l_{SP,3} = d_{SC}(n_{SC,active} + n_{SC,ia1} + n_{SC,ia2} - 0.5) \tag{9.23}$$

where $n_{SC,ia1}$ and $n_{SC,ia2}$ are the number of inactive coils at the two ends of the spring. Note that the solid length needs to be shorter than the maximum compressed spring length. At least one inactive coil is required for each end of the spring. Using more inactive coils at each end (e.g., 1.25) provides a more gradual change from the inactive to the active (working) coils for highly stressed springs. The free clearance between the coils at the free length is largely affected by the installed spring length. The free clearance is given by:

$$c_{SP,0} = \frac{l_{SC,0} - l_{SP,3}}{n_{SC,active} + 1} = \frac{l_{SP,1} + \dfrac{F_{pre}}{K_{s,SP}} - d_{SC}(n_{SC,active} + n_{SC,ia1} + n_{SC,ia2} - 0.5)}{n_{SC,active} + 1} \tag{9.24}$$

The free clearance should not be excessively large in order to minimize the height of the cylinder head. The solid clearance per active coil at the maximum valve lift is given by:

$$c_{SC,2} = \frac{l_{SP,2} - l_{SP,3}}{n_{SC,active} + 1} = \frac{l_{SP,1} - l_{VAL,\,max} - l_{SP,3}}{n_{SC,active} + 1} \tag{9.25}$$

The spring's first natural frequency is calculated for the distributed mass of the spring as follows:

$$f_{n,SP} = \frac{1}{2}\sqrt{\frac{K_{S,\,SP}}{m_{SP}}} = \frac{d_{SC}}{2\pi d_{SP}^2 n_{SC,active}}\sqrt{\frac{\vartheta_{m,SP}}{2\rho_{SP}}} \tag{9.26}$$

where m_{SP} is spring mass and ρ_{SP} is the density of the spring material. The maximum spring load at the maximum valve lift is calculated as follows:

$$F_{SP,\,max} = F_{pre} + K_{s,SP} l_{VAL,\,max} = K_{s,SP}(l_{SP,0} - l_{SP,1} + l_{VAL,\,max}) \tag{9.27}$$

The maximum spring torsional stress can be calculated as follows:

$$s_{SP,\,max} = \frac{8F_{SP,\,max}d_{SP}}{\pi d_{SC}^3} \cdot f_{Wahl}$$

$$\approx \frac{8F_{SP,\,max}d_{SP}}{\pi d_{SC}^3} \cdot \left[\frac{4(d_{SP}/d_{SC})-1}{4(d_{SP}/d_{SC})-4} + \frac{0.615}{d_{SP}/d_{SC}}\right] \qquad 9.28$$

where f_{Wahl} is the Wahl stress correction factor for the helical spring.

From the above calculations it is observed that the spring wire coil diameter, the mean diameter and the number of active coils determine the spring rate and the natural frequency of the spring, according to equations 9.19 and 9.26, respectively. The spring wire coil diameter, the mean diameter, and the spring force determine the spring torsional stress according to equation 9.28. Although both valve spring preload and spring rate can increase the valvetrain no-follow speed, their effects on valve spring design are different. Higher spring preload increases the maximum spring load and stress, and requires higher cam ramp height. Higher spring rate demands larger coil diameter, heavier spring, reduced clearance between the coils, increased natural frequency, higher spring load and spring stress.

The spring design problem can be formulated as a system of three nonlinear equations consisting of equations 9.19, 9.25, and 9.28 for a deterministic solution to solve for three unknowns, d_{SP}, d_{SC}, and $n_{SC,active}$. Then, the spring's natural frequency can be calculated using equation 9.26. Alternatively, the spring design problem can be formulated as a nonlinear constrained optimization topic in the following:

$$\begin{cases} \text{Find } d_{SP}, d_{SC} \text{ and } n_{SC,active} \text{ to maximize } f_{n,SP} \\ \text{subject to } c_{SC,2} \geq c_{SC,2}^{min} \quad c_{SC,0} \leq c_{SC,0}^{max} \\ \qquad n_{SC,ia1} \geq 1.25 \quad n_{SC,ia2} \geq 1.25 \\ \qquad K_{s,SP} = K_{s,SP}^{target} \quad s_{SP}^{max} \leq s_{SP}^{limit} \end{cases} \qquad 9.29$$

The graphical design approach on valve spring is illustrated in Fig. 9.21 and discussed in Section 9.5.3.

9.5.3 Procedure of valve spring design

Valve spring design is a complex system design topic. Good spring design minimizes valvetrain friction and wear. An analytical approach to engine valve spring design based on parametric sensitivity maps is summarized as below.

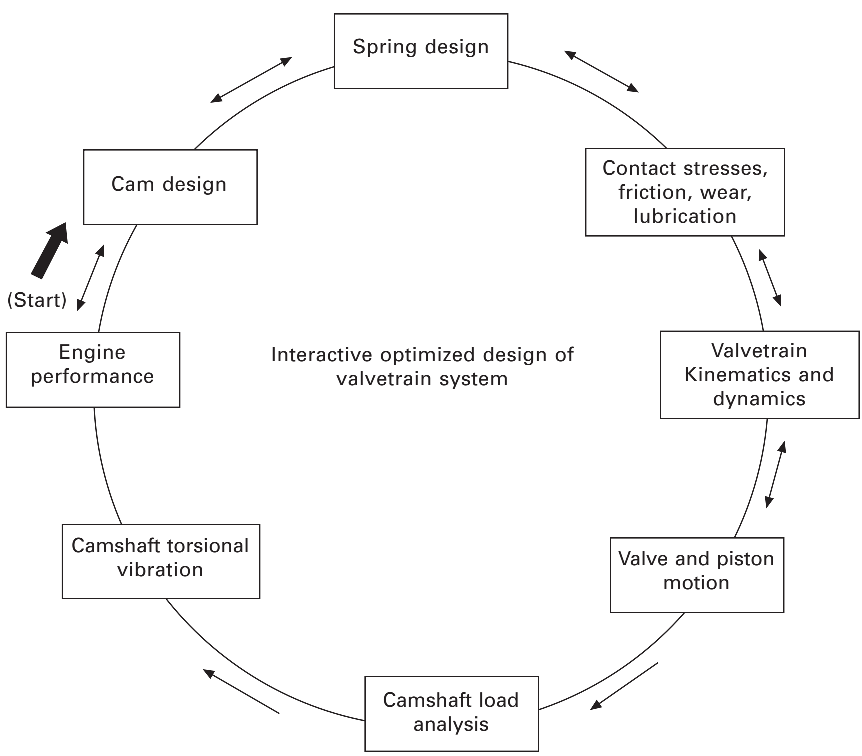
1. Step 1: Determine the target of valvetrain over-speed limit through the analysis of vehicle downhill driving performance and engine braking in order to determine the required valve spring preload and spring rate.
2. Step 2: Build a valvetrain dynamics model in order to accurately predict no-follow. Evaluate the effect of recompression pressure on no-follow.
3. Step 3: Construct parametric maps of valvetrain dynamics by sweeping different levels of spring preload and spring rate in order to check their impact on the valvetrain vibration. Pushrod force, valvetrain acceleration, and spring deceleration need to be plotted to indicate the no-follow design margin in order to facilitate a wise selection of spring preload and spring rate.
4. Step 4: Select exhaust valve spring preload for the engines with and without the exhaust brake. Compute the required spring preload to prevent exhaust valve floating based on the static force balance across the exhaust valve head.
5. Step 5: Use the graphical design method to construct parametric sensitivity charts for spring design by sweeping the design parameters (Fig. 9.21). Select spring mean diameter, wire coil diameter, and the number of coils subject to the design constraints of spring torsional stress, natural frequency, coil clearances, etc. Or, alternatively use the analytical optimization method to solve equation 9.29 directly.
6. Step 6: Estimate the changes in cycle-average valve spring load, engine friction power, coolant heat rejection and BSFC.

As discussed in Chapter 6, vehicle performance is related mainly to vehicle weight, frontal area, tire size, road grade, rear axle ratio, transmission gear ratio, engine retarding power, and engine speed. The most convenient braking happens when the driver does not need to apply the service brakes at a desired downhill control speed. These conditions are represented by the 'ZWB' curves on the vehicle performance maps. The engine over-speed target should be designed well above the 'ZWB' contour of typical vehicle downhill driving. If the no-follow speed target can be reduced, a softer valve spring with reduced spring preload and rate can be used to alleviate the problems of valvetrain wear and lash growth. As shown in the example of Fig. 9.21, for engines either with or without the exhaust brake, a new intake valve spring's preload can be reduced by 28% from the baseline spring's 184 lb to 133 lb; the new intake valve spring's rate is reduced by 22% from the baseline spring's 379 lb/inch to 296 lb/inch. The overall cycle-average intake spring force (an indicator of the valvetrain friction or wear load) is reduced by 26%. For engines with the exhaust brake, a new exhaust valve spring's preload is reduced by 8% from the baseline spring's 184 lb to 170 lb; the new exhaust valve spring's rate is reduced by 16% from the baseline spring's 379 lb/inch to 317 lb/inch. In this case, the overall spring force is reduced by 10.5%.

9.6 Analytical valvetrain system design and optimization

Analytical valvetrain design refers to using advanced simulation tools in each of the following design steps to optimize the relationship between the parameters in the valvetrain system. The scope of the valvetrain system design is illustrated in Fig. 9.22. The key steps in the valvetrain design are summarized as follows.

1. Select an appropriate target of valvetrain no-follow speed by the analysis of vehicle driving and engine motoring or braking operation.
2. Use engine cycle simulation to determine the optimum valve event duration, valve timing, and valve overlap size based on the best trade-off between low engine speeds and high speeds. This needs to be conducted in a coupled analysis with turbocharger matching since engine delta P has a direct impact on gas exchange and the reverse valve flow. The effect of valve recession on the combustion chamber 'K-factor' also needs to be considered.



9.22 Concept of valvetrain system optimization.

3. Determine the gas loading from the cylinder and the port acting on the valvetrain by engine cycle simulation based on the preliminary valve lift.
4. Select a preliminary exhaust valve spring preload based on the engine firing and braking requirements. Select a preliminary valve spring rate based on the dynamic simulation of valve acceleration.
5. Simulate the effect of gas loading on valvetrain vibration.
6. With given valve timing, valvetrain stiffness and weight, valve size and valvetrain lash, optimize the following design parameters all together in the valvetrain system by using a cam design tool and a valvetrain dynamics simulation model to meet all the design criteria: the maximum valve lift, rocker arm ratio, cam acceleration shape, cam base circle radius, roller radius of the follower, cam radius of curvature, valve spring preload, and spring rate. For example, the cam base circle radius should be maximized within the allowable limits of space and weight. This helps reduce the cam stress and maximize the breathing performance. Higher rocker ratio results in higher cam force and cam stress. Lower rocker ratio requires higher cam lift (for a given valve lift), higher cam acceleration, and smaller radius of curvature at the cam nose.
7. Conduct valve spring design optimization on the spring diameter, the coil diameter and the number of coils to maximize the spring natural frequency.
8. Re-iterate from steps 2 to 7 until satisfactory trade-offs between all the design parameters are obtained.

It should be apparent that valvetrain design is very complex due to the large number of design parameters involved and the complex interactions among them. An iterative process or even DoE optimization is often required. Constructing the parametric design charts with a graphical design method is very powerful to enhance the understanding of the parametric relationships (e.g., Fig. 9.17 for the cam and Fig. 9.21 for the valve spring). More information on valvetrain architecture design is provided by Jacques (1997), Clarke and Innes (1997), and Buuck and Hampton (1997). Valvetrain system design optimizations are presented by Seidlitz (1990), Ernst *et al.* (1993), and Keribar (2000).

9.7 Variable valve actuation (VVA) engine performance

9.7.1 The need for variable valve actuation (VVA)

In the conventional fixed-cam valvetrain, the camshaft controls the intake and exhaust valves. Valve timing, valve lift, and event duration are all fixed values specific to the camshaft design. The optimum valve timing is usually a compromise between different engine operating conditions (e.g., different speeds or loads), and the optimum is determined to achieve a high volumetric efficiency or a low BSFC at a selected working condition. Engine valve flow directly affects the breathing of the engine cylinder and the thermodynamic cycle process of the engine. Air and fuel flows consist of the energy flows of the internal combustion engine. Modern diesel engines are equipped with electronically-controlled flexible fuel injection systems and air/gas/EGR control valves. Variable geometry turbochargers are also widely used in production engines to regulate gas flows. The only part in the air system left for revolutionary rather than evolutionary change is the valvetrain. An electronically-controlled air management variable valve actuation (VVA) system is highly desirable for modern diesel engines.

Like the expectation for variable compression ratio for optimum engine performance without the compromise in efficiency at different operating conditions throughout the working range of the engine, a flexible variation in engine valve actuation in crank angle resolution has been a dream of engine designers since the era of steam engines more than a century ago. In fact, the first study on VVA for internal combustion engines dates back to as early as 1902 when Louis Renault conceived a simple VVA device for a spark ignition engine. However, only in the last twenty years has the engine industry experienced a rapid progress in VVA research, evidenced by a drastic increase in the number of publications and patents, and the materialization in production of spark ignition engines. This fast growth is primarily the result of the advent of electronic engine controls applied to VVA devices as well as market pressure on fuel economy improvement.

Various forms of VVA have been commonly used in today's gasoline engines in automotive production since the 1980s. The VVA application to diesel engines has lagged behind. The biggest reason is that the conventional gasoline engine relied on throttled operation at part load where the use of VVA to eliminate or reduce the pumping loss due to intake throttling to improve fuel economy can be easily justified. Intake throttle reduces the cylinder pressure during the intake stroke. The cylinder pressure difference between the exhaust stroke and the intake stroke forms the pumping loss. The diesel engine inherently runs without the throttle and has a narrower engine speed range (e.g., up to 3000 rpm for HD diesel and 4500 rpm for LD diesel). Therefore, the benefit of using VVA for the diesel engine has traditionally been believed to be far more limited than the gasoline engine. The questions on the benefits of VVA for the diesel engine have often focused on the cost–benefit ratio for valve timing optimization and other advanced design features such as engine compression brake.

Future advanced diesel engine technologies require a new look at the potential of VVA. These include pumping loss reduction, HCCI combustion, cylinder deactivation, engine brake, reconciliation between the engine valves (with the control in crank angle resolution) and air system control valves (with the control in engine cycle resolution), and valve timing optimization. The driving forces for diesel VVA are usually fuel economy improvement, cost reduction via consolidating all air flow control features, and customer demands for special features such as low-speed torque, reduced turbocharger lag during low-speed light-load 'tip-in' transient accelerations, and engine compression brake.

9.7.2 Classification of VVA

VVA generally refers to the ability to vary some or all of the following: valve opening and closing timing or valve event duration, maximum valve lift, and the rising rate of valve opening and closing. Although numerous patents exist for VVA, the VVA technology can be classified into two groups with various complexities: (1) the devices using camshafts (e.g., cam phase shifter or phaser, cam lobe switcher, cam-driven mechanical VVA with variable tappets or rockers, cam-based 'lost-motion' VVA, variable cam lobe contour, or the '3-D' cam); and (2) camless devices (e.g., electro-hydraulic, electro-magnetic, electro-mechanical, pneumatic). VVA can also be classified into VVT (variable valve timing) and VVL (variable valve lift) according to the functionality.

An example of cam phase shifter is that the intake and exhaust cams are located on separate camshafts in a double-overhead-cam valvetrain, and the phase relationship and the valve overlap period can be varied between the two sets of cams. Cam phase shifters are widely used in the gasoline engine VVA. Camless VVA is more suitable for diesel engines due to its versatile features in crank-angle-resolution air management. The camless engine can achieve any valve event at any crank angle locations to any valve lift position and hold it for any duration, thus flexibly optimizing the engine performance. In camless VVA, the traditional valvetrain (camshaft, pushrods, lifters, rocker arms, and valve springs) is replaced by small and fast electronically-controlled actuators.

More detailed classifications of VVA systems and the introduction to their design mechanisms are given by Stone and Kwan (1989). Dresner and Barkan (1989a) classified VVA into 15 basic types of concepts.

9.7.3 Design challenges of VVA

The benefits of VVA on the gasoline engine have been well established, and the current development focus is to reduce the cost and the weight, and improve the reliability of the VVA mechanisms. Better reliability can be achieved by reducing complexity and valve seating velocity (e.g., with soft landing), improving the accuracy and the repeatability of valve timing and valve event, and ensuring no collision between the valve and the piston. The performance at low engine oil pressures and temperatures is also another design challenge for the hydraulic-driven VVA systems.

The parasitic loss of the VVA system is generally higher than that of the conventional cam-driven valvetrain. A low friction design is critical for any VVA system in order to avoid negating the fuel economy benefits obtained from the better gas exchange processes of the VVA. For example, in the camless electro-hydraulic valvetrain, the energy consumption is proportional to the maximum valve lift generated. A low lift can be used at low speeds/loads in order to reduce the energy consumption of the valvetrain without a significant negative impact on engine breathing and fuel economy.

9.7.4 Interaction of VVA with other air system components

The valve timing effect on internal combustion engine performance is discussed in detail by Asmus (1982), Stas (1999), Thring (1990), and Leonard *et al.* (1991). VVA technologies for both gasoline and diesel engines are extensively reviewed by Gray (1988), Stone and Kwan (1989), and Dresner and Barkan (1989a, 1989b). Camless engines and their performance are elaborated by Mardell and Cross (1988), Schechter and Levin (1996), Pischinger *et al.* (2000), Salber *et al.* (2001), Tai *et al.* (2002), Schernus *et al.* (2002), and Picron *et al.* (2008). A physics-based volumetric efficiency model has been introduced by Turin *et al.* (2008). Moreover, a thermodynamic second-law analysis applied to VVA has been conducted by Anderson *et al.* (1998). There is a wealthy body of literature addressing engine valve timing and VVA performance. However, no study has been conducted to address the theoretical relationship between VVA and other air system components. The air system in this book refers to the turbocharger, the EGR system, the manifolds, and the valvetrain. In diesel engine system design, it is important to understand the role of valve timing and VVA in the entire air system so that a wise system-level design solution can be selected to reconcile or simplify the functionality between different components and avoid redundancy. This section provides a theoretical analysis to address the relationship among several air system technologies such as VVA, cylinder deactivation, air control valves, EGR, and turbocharging.

The operation of the engine valves affects the number of strokes (e.g., two-stroke or four-stroke operation), effective engine displacement (e.g., via cylinder deactivation by disabling the valve lift), effective engine compression ratio (i.e., via IVC timing change), effective engine expansion ratio (i.e., via EVO), volumetric efficiency (via either valve event duration or effective valve opening area), and eventually the engine in-cylinder thermodynamic cycle process. The air/gas control valves in the engine gas flow network circuit and the turbocharger also affect the engine air flow through their roles outside the cylinders at engine cycle resolution. Their effects are different from VVA on air/EGR flow and pumping loss. However, the valvetrain and the cylinder can be viewed as a lumped element, characterized by volumetric efficiency, in the gas flow network where the valvetrain essentially behaves as flow restriction orifices. The engine air flow rates and pressures are determined by all the flow restriction orifices in the flow network. For example, the intake manifold boost pressure increases when the downstream intake valve flow area reduces by the Miller cycle with early IVC. The exhaust manifold pressure increases when the downstream turbine flow area reduces. The pressure built up is also related to the air/gas flow rate through the orifice. For example, closing the intake throttle reduces the engine air flow rate so that the intake manifold pressure becomes very low in front of the engine intake valve orifice.

The impact of valvetrain and VVA designs on engine performance can best be understood through the roles of volumetric efficiency, pumping loss and EGR, driving capability in the four core equations presented in the air system theory in Chapter 4, namely, equation 4.40 (engine volumetric efficiency), 4.44 (EGR circuit pressure drop), 4.47 (turbocharger power balance) and 4.57 (turbine flow). These four core equations determine the engine air flow rate and the pumping loss. The roles of different air system design and control parameters, including any competing technologies, can be clearly comprehended via the parametric relationships revealed in these four equations.

The intake manifold air–EGR mixture non-trapped volumetric efficiency η_{vol} of a four-stroke internal combustion engine given by equation 4.40 in Section 4.4 is shown as follows:

$$\eta_{vol} = \frac{2(\dot{m}_{air} + \dot{m}_{EGR})T_{2a}R_{gas}}{p_{2a}N_E V_E}$$

This equation can be rearranged in the following form:

$$\frac{2(\dot{m}_{air} + \dot{m}_{EGR})R_{gas}}{N_E} = \frac{\eta_{vol}V_E(p_2 - \Delta p_{IT} - \Delta p_{CAC})}{T_{2a}} \tag{9.30}$$

where N_E is the engine speed, V_E is the effective engine displacement, T_{2a} is the intake manifold gas temperature, p_{2a} is the intake manifold pressure, p_2 is the compressor outlet pressure, Δp_{IT} is the pressure drop associated with the intake throttle, and Δp_{CAC} is the pressure drop of the charge air cooler. Equation 9.30 shows that in order to reach a given air and EGR flow rate requirement, the following five technologies need to match well (shown on the right side of equation 9.30): valvetrain design (η_{vol}), engine displacement (e.g., cylinder deactivation), turbocharging (p_2), intake throttle (Δp_{IT}), and air/EGR charge cooling (T_{2a}).

The pressure drop and the opening duration of the intake and exhaust valves largely affect the volumetric efficiency. Fukutani and Watanabe (1979) indicated that volumetric efficiency could be correlated to a mean intake Mach number and a Mach index. When the Mach number approaches 0.5, the volumetric efficiency decreases drastically because the flow becomes choked during part of the intake stroke. In general, it is always desirable to have a large valve flow area and a high port flow coefficient in order to minimize pumping loss and increase volumetric efficiency.

The four core equations of the engine air system are presented as follows:

$$\begin{cases} \dfrac{2(\dot{m}_{air}+\dot{m}_{EGR})R_{gas}}{N_E} = \dfrac{\eta_{vol}V_E(p_2-\Delta p_{IT}-\Delta p_{CAC})}{T_{2a}} \\ p_{EGRin}-p_{EGRout} = f(C_{d,EGR},\dot{m}_{EGR},T_{EGRcoolerOut}) \approx C_0 + C_1\dot{m}_{EGR}^2 \\ 1-\left(\dfrac{p_2}{p_1}\right)^{(\kappa_c-1)/\kappa_c} + \eta_C\eta_T\eta_{TC,mech}\left(\dfrac{\dot{m}_T}{\dot{m}_C}\right)\left(\dfrac{c_{p,T}}{c_{p,C}}\right)\left(\dfrac{T_3}{T_1}\right)\left[1-\left(\dfrac{p_4}{p_3}\right)^{(\kappa_t-1)/\kappa_t}\right]=0 \\ \dot{m}_T = A_T\cdot\dfrac{p_3}{\sqrt{R_{ex}T_3}}\cdot\sqrt{\dfrac{2\kappa_t}{\kappa_t-1}}\cdot\sqrt{\left(\dfrac{p_4}{p_3}\right)^{2/\kappa_t}-\left(\dfrac{p_4}{p_3}\right)^{(\kappa_t+1)/\kappa_t}} \end{cases} \quad 9.31$$

Note that these four core equations can be used to solve for any four unknowns, either hardware design parameters or performance parameters. Equation 9.31 shows that in order to meet a target requirement of air and EGR flow rates, the following four unknowns can be solved with all the remaining parameters given as known inputs: turbine wastegating $(\dot{m}_T)$, EGR valve opening ($C_{d,EGR}$), intake manifold pressure p_2, and turbine inlet pressure p_3. It is apparent that if the input assumptions of intake throttle (Δp_{IT}) or valve timing (η_{vol}) are changed, a new set of solutions will be obtained and its corresponding engine delta P will be different. Therefore, whether VVA can replace intake throttle or wastegating in functionality depends on the solutions obtained as such.

It should be noted that the change in volumetric efficiency (e.g., via IVC or valve overlap) affects the location of the engine operating point on the compressor map and hence the compressor efficiency because the reciprocal of volumetric efficiency basically reflects the slope of the point on the compressor map. Such an influence of VVA on turbocharger matching can make the technology comparison mentioned above complicated. Usually, full numerical simulation such as using GT-POWER is required to analyze the problem. Gray (1988) reported that the variations in IVC timing and valve overlap helped improve turbocharger matching and reduce fuel consumption throughout the whole load range for non-EGR engines.

If the purpose of VVA is for power/torque capability, turbocharging can be viewed as one of the alternatives to VVA. This is because the valve timing can be optimized for low-speed performance (e.g., with small valve overlap or early IVC), and high turbocharged boost pressure can be used to compensate for the lower volumetric efficiency at high speeds (Stone and Kwan, 1989). However, if the purpose of VVA is to reduce BSFC, turbocharging and VVA are different technologies that cannot substitute each other. The fundamental reason is that the equivalent flow restriction 'orifices' of the VVA and the turbocharger are found at different locations in the engine gas flow network and hence function differently in regulating the gas flows and pressures.

Once the solution to equation 9.31 is obtained, pumping loss needs to be calculated in order to evaluate the impact of the given technology on BSFC. The pumping loss work is given by the integral of the *p–V* diagram in the pumping strokes at 360–720° crank angle as follows (or represented by the area enveloped by the *p–V* curve shown in Fig. 4.4 in Chapter 4):

$$
\begin{aligned}
\int_{360^\circ}^{720^\circ} p_{cyl}\,\mathrm{d}V_{act,E} &= \sum_{\phi=360^\circ}^{720^\circ} (p_{exhaust} - p_{intake})\,\mathrm{d}V_{act,E} \\
&= \sum_{\phi=360^\circ}^{720^\circ} [(p_{EM} + \Delta p_{ex}) - (p_{IM} - \Delta p_{in})]\,\mathrm{d}V_{act,E} \\
&= \sum_{\phi=360^\circ}^{720^\circ} [(p_{EM} - p_{IM}) + (\Delta p_{ex} + \Delta p_{in})]\,\mathrm{d}V_{act,E} \\
&= \sum_{\phi=360^\circ}^{720^\circ} (p_{EM} - p_{IM})\,\mathrm{d}V_{act,E} + \sum_{\phi=EVO}^{EVC} \Delta p_{exV}\,\mathrm{d}V_{act,E} \\
&\quad + \sum_{\phi=IVO}^{IVC} \Delta p_{inV}\,\mathrm{d}V_{act,E} + \sum_{ValveClosed} \Delta p_{PM}\,\mathrm{d}V_{act,E} \\
&= \sum_{\phi=360^\circ}^{720^\circ} (p_{EM} - p_2 + \Delta p_{IT} + \Delta p_{CAC})\,\mathrm{d}V_{act,E} \\
&\quad + \sum_{\phi=EVO}^{EVC} \Delta p_{exV}\,\mathrm{d}V_{act,E} + \sum_{\phi=IVO}^{IVC} \Delta p_{inV}\,\mathrm{d}V_{act,E} + \sum_{ValveClosed} \Delta p_{PM}\,\mathrm{d}V_{act,E}
\end{aligned} \tag{9.32}
$$

where p_{cyl} is the in-cylinder pressure, $V_{act,E}$ is the total engine displacement of the active cylinders, ϕ is the crank angle, $p_{exhaust}$ and p_{intake} are the in-cylinder pressure during the exhaust and intake strokes, respectively; p_{EM} and p_{IM} are the instantaneous exhaust and intake manifold pressures, respectively; Δp_{ex} and Δp_{in} are the pressure drops across the exhaust and intake valves/ports/manifolds, respectively; Δp_{exV} and Δp_{inV} are the pressure drops across the exhaust and intake valves during the valve opening events, respectively; and

Δp_{PM} is the pressure losses in the intake and exhaust ports and manifolds during the period of valve closure.

Equation 9.32 shows that the pumping loss is related not only to the intake throttle pressure drop, but also to the valve flow pressure drops and the valve opening duration (e.g., from IVO to IVC). The valve flow pressure drop is related to the effective valve flow area. A larger valve flow area results in a smaller pressure drop due to less restrictive flow. This is the reason why the cam flow area needs to be maximized with aggressive cam acceleration. When the valve flow area is small, volumetric efficiency becomes low due to high flow losses. The valve opening duration affects the amount of the flow through the engine and thus also directly affects volumetric efficiency. If the valve event duration is too short, the mass flow rate will be too low in equation 9.30 and thus the volumetric efficiency will be low. Moreover, if the valve event duration is too long, reverse flow will occur and the volumetric efficiency will also be low. More information on pumping loss definition on the *p-V* diagram can be found in Shelby *et al.* (2004).

It should be noted that the intake manifold volumetric efficiency is inherently related only to the valvetrain and manifold designs. It is not affected by the operation of intake throttle. When a VVA device with early IVC is used, the volumetric efficiency can become very low due to the very short intake valve duration. However, the pumping loss can become very little if the intake throttle is not closed, according to equation 9.32. In fact, this can be viewed as an example of using an intake VVA to replace the intake throttle at the part-load operation in a naturally aspirated gasoline engine. Although the availability loss of throttling at the throttle valve itself is small, the resulting engine pumping loss due to the throttling can be rather large. Therefore, in principle any throttling in the engine is an indicator of high pumping loss and should be avoided.

Pumping loss consists of two parts: the engine delta P (the first term in equation 9.32) and the flow restrictions related to volumetric efficiency. The second part can be further split into two portions: (1) the losses related to valve event durations (the second and third terms in equation 9.32); and (2) the losses associated with the ports and manifolds. Volumetric efficiency affects pumping loss differently via the valve flow restriction and the valve event duration. If the valve flow area is too restrictive, the large pressure drop across the valve results in a reduction in volumetric efficiency and an increase in pumping loss (i.e., expanding the pumping loss area shown in Fig. 4.4 vertically). If the valve event duration is very short, the resulting low volumetric efficiency leads to a reduction in pumping loss (i.e., shrinking the pumping loss area shown in Fig. 4.4 horizontally). The design modification that reduces the flow restriction for a given valve flow duration (e.g., larger valve flow area in a cam or VVA) results in an increase in volumetric efficiency and a decrease in the valve flow pumping loss. However, the designs affecting the valve event duration (e.g., cam or VVA timing) may increase both volumetric efficiency and the valve flow pumping loss. One extreme example is that in cylinder deactivation the volumetric efficiency of the deactivated cylinders is zero and their pumping loss is also zero.

The link between valvetrain design and turbocharging is the intake manifold pressure. Note that a higher volumetric efficiency means a lower intake manifold pressure and possibly less pumping loss because a lower engine delta P can be created by a larger turbine area in order to reach the same air flow requirement. Therefore, the net effect of valve timing change on BSFC depends on the balance between the pumping losses caused by the engine delta P and the valve flows.

Finally, it should be noted that BSFC (i.e., power loss) and engine gas flow rate (i.e., volumetric efficiency) are two independent different design criteria (e.g., reflected by the trade-off between BSFC and NO_x) in valve timing or VVA optimization. The topic can be a multi-objective optimization. If achieving high flow rate is the design objective (e.g., for meeting emissions or controlling the exhaust manifold gas temperature), different valve timing designs need to be compared on volumetric efficiency at the same intake manifold pressure. On the other hand, if achieving low BSFC is the objective, pumping losses need to be compared at the same engine flow rate.

9.7.5 Gasoline engine VVA performance

Overview of gasoline engine VVA

Before the discussion of diesel engine VVA performance, a summary of the gasoline (spark ignition) engine VVA is beneficial in order to understand their differences. The majority of VVA research has been conducted at the part-load conditions on the gasoline engine with the primary purposes of eliminating or reducing the use of intake throttle, and optimizing the valve timing to increase the volumetric efficiency within a wide range of engine speed to improve the low-speed torque. Both the port-fuel-injection and the direct injection gasoline engines are partially throttled at part load to meet the operational requirement of the exhaust aftertreatment catalyst. In this case, VVA can replace the intake throttle to reduce fuel consumption. At the full load, the engine is usually run with the intake throttle fully open.

Over the next few decades the following three technologies will have decisive impacts on the gasoline engine in air flow controls to improve fuel economy: (1) throttleless operation by using VVA to achieve charge control or load control to replace the intake throttle; (2) stratification by direct injection; and (3) variable displacement by cylinder deactivation.

Important research work on gasoline engine VVA performance has been conducted by Tuttle (1980), Elrod and Nelson (1986), Ma (1988), Payri *et al*. (1988), Saunders and Abdul-Wahab (1989), Ahmad and Theobald (1989), Larsen (1991), Asmus (1991), Kreuter *et al*. (1992), Nuccio and Marzano (1992, a review), Gallo (1992), Ke and Pucher (1996), Anderson *et al*. (1998), Pischinger *et al*. (2000), Pierik and Burkhard (2000), and Ribeiro and Martins (2007).

Gasoline engine valve overlap

In the gasoline engine, valve timing has traditionally been designed to optimize the operation at the high-speed full-load conditions with the intake throttle wide open. The valve overlap size in the gasoline engine is determined based on the considerations of high-speed power, low-speed torque, residue gas quantity at part load for charge composition control and emissions, charge loss to the exhaust at the full load, and idle quality. At part load, if the engine is throttled a large pressure difference between the exhaust manifold and the intake manifold (e.g., high vacuum) is created. Then, a large amount of residue gas (internal EGR) is inducted from the exhaust manifold into the cylinder through the valve overlap.

A small valve overlap in the gasoline engine leads to the following: (1) low volumetric efficiency at high engine speeds due to poor scavenging; (2) high torque at low speeds; (3) low internal EGR at part load; (4) low backflow of unburned mixture of fuel and air to the exhaust system; and (5) good idle quality due to low residue gas fraction. The low backflow to the exhaust duct avoids high BSFC and also protects the catalyst. The improved idle stability and quality may enable a reduction in the idle speed to reduce idle fuel consumption.

Gasoline engine IVC and EVO

Either early or late IVC can reduce the effective compression ratio of the engine and the compression pressure. The load control without throttling in the gasoline engine is achieved by using an early or late IVC (or variable valve lift) to reduce the mass of charge trapped in the cylinder and maintain the intake manifold pressure near the atmospheric pressure rather than a high vacuum. Therefore, the pumping loss can be greatly reduced. Note that Stone and Kwan (1989) pointed out that at low loads the reduced compression ratio due to IVC change may reduce the benefit of throttle loss reduction. Early IVC can be achieved by advancing IVC timing to reduce the intake valve event duration and sometimes accompanied by a reduction in the maximum valve lift to control the valve acceleration for the reduced valve event duration.

Retarding EVO timing may increase the effective expansion ratio. The EVO timing can be especially retarded to around the BDC at low speeds/loads since the pumping loss during the exhaust stroke is not adversely affected at the low gas flow rate. Moreover, if the opening rate of the exhaust valve lift is higher than the cam-driven lift, EVO can be retarded to reduce BSFC even at high speeds by increasing the expansion work without a penalty of pumping loss during the exhaust stroke. From the above discussions on the valve timing, it is obvious that VVA is very important for the gasoline engine.

Performance benefits of gasoline engine VVA

The performance benefits of gasoline engine VVA can be summarized as follows:

- *Improved torque curve*. The gasoline engine usually has a wide speed range. A high volumetric efficiency provides a high air flow rate and hence more fuel can be burned to increase the power. VVA enables optimized valve timing and high volumetric efficiency at all speeds to create higher and flatter torque curve from low to high speeds. VVA is especially useful for non-turbocharged engines.
- *Reduced part-load throttle losses and fuel consumption*. The benefit of pumping loss reduction with VVA to replace the intake throttle is most significant at the low-speed light-load conditions where the throttling is the heaviest. VVA can also reduce the effective compression ratio without a corresponding change in the expansion ratio so that the thermodynamic cycle efficiency can be improved. A range of 15–20% BSFC reduction by VVA on gasoline engines has been widely reported and confirmed (e.g., Dresner and Barkan, 1989b; Pischinger *et al*., 2000).
- *Improved idle stability and fuel consumption*. Low-idle operation largely affects the gasoline engine fuel economy. The highest percentage of fuel economy improvement by VVA usually occurs at low idle, and decreases as the engine speed or load increases. The reduced small valve overlap offered by VVA improves the idle stability and enables an idle speed reduction to reduce fuel consumption by around 30%. Ma (1988) reported that a 200 rpm reduction in idle speed from 800 rpm translated to a 6.1% improvement in the ECE-15 drive cycle fuel economy. Early IVC further reduces the pumping loss at low idle due to throttleless operation.
- *Emissions reduction through the controls on both cylinder charge mass and composition (residue gas)*. VVA largely reduces NO_x by increasing residue gas fraction or internal EGR, for example via large valve overlap. A small valve overlap may reduce HC and CO emissions. VVA can also reduce HC and CO emissions during the cold start and warm-up transients by using late IVO. Moreover, late EVO may reduce emissions and increase the exhaust gas temperature.
- *Emissions reduction through charge motion control*. The objective of charge motion control is to improve mixture formation and increase

internal cylinder flow for better homogeneity of the stoichiometric mixture to accelerate combustion. Variable swirl ratio and high intake valve flow velocity at low speeds can be achieved by low or uneven valve lifts, valve deactivation, or negative valve overlap (i.e., EVC being earlier than IVO around the TDC).

9.7.6 Diesel engine VVA performance

Overview of diesel engine VVA

The history of diesel engine VVA can be traced back to the 1960s when the Jacobs Manufacturing Company used the technology of precision valve actuation to actuate the compression-release brake. The hydro-mechanically created braking exhaust valve event in the compression brake can be seen as an early form of VVA technology for the diesel engine. An example of diesel VVA on the intake valve in the firing operation was Caterpillar's ACERT system created several years ago that used hydraulic actuators to change the intake valve motion and timing to achieve charge control and internal EGR. Since recently, intake VVA has been extensively explored in diesel HCCI combustion. VVA affects effective compression ratio and mixture reactivity with residue gas control and temperature control to enable HCCI combustion.

The need to flexibly regulate the engine valve flows in the diesel engine was recognized by previous researchers decades ago (Mardell and Cross, 1988). Previous research on diesel VVA performance can be categorized in the following areas:

- The effects of valve overlap size and reverse flows at high loads and low loads and for transient load response were simulated by Charlton *et al.* (1990, 1991) for a non-EGR turbocharged diesel engine that had negative engine delta P.
- Hot internal EGR by VVA was researched by Desantes *et al.* (1995), Benajes *et al.* (1996), Schwoerer *et al.* (2004), Parvate-Patil *et al.* (2004), and Millo *et al.* (2007).
- Valve timing and engine breathing performance were analyzed by Ghaffarpour and Baranescu (1995), Lancefield *et al.* (2000), Lancefield (2003), Parvate-Patil *et al.* (2004), and Deng and Stobart (2009). Lancefield *et al.* (2000) and Lancefield (2003) studied cooled external EGR used together with VVA on EVO and IVC timing controls with fixed IVO and EVC timings.
- The diesel Miller cycle with adjustable IVC timing was researched by Ishizuki *et al.* (1985), Bolton and Assanis (1994), Stebler *et al.* (1996), Edwards *et al.* (1998), Mavinahally *et al.* (1996), Kamo *et al.* (1998), Wang *et al.* (2005), Millo *et al.* (2005), Ribeiro and Martins (2007), and

Imperato *et al*. (2009). Other research on the Miller cycle (mainly for gasoline engines) was conducted by Hitomi *et al*. (1995), Okamoto *et al*. (1997), Clarke and Smith (1997), Ge *et al*. (2005), Ribeiro and Martins (2005), Ribeiro *et al*. (2006), Al-Sarkhi *et al*. (2002, 2006), Chen *et al*. (2007, 2008), Junior (2009), and Wik and Hallbäck (2008), for large marine diesel engines).

- The rotary valve used for diesel VVA to achieve early IVC timing was investigated by Ishizuki *et al*. (1985).
- VVA and turbocharger interaction was studied by Ke and Pucher (1996, for a turbocharged gasoline engine), Edwards *et al*. (1998), Tai *et al*. (2002), Lancefield (2003), and Yang and Keller (2008). Gu (1995) and Yang *et al*. (2009, 2010) reported using intake VVA (the Miller cycle) combined with variable exhaust and fuel injection timing on non-EGR diesel engines. Their system was originally called the Gu-system named after Professor Hong-Zhong Gu's name (Gu, 1995).
- General performance benefits of VVA on diesel engines were investigated by Mardell and Cross (1988), Fessler and Genova (2004), Leet *et al*. (2004), and Gehrke *et al*. (2008).
- The effects of VVA (intake valve lift and IVC timing) on combustion, emissions, and turbulence swirl were computationally investigated by Stephenson and Rutland (1995) and Munnannur *et al*. (2005).
- The applications of VVA on diesel engine brake were discussed by Hu *et al*. (1997a, 1997b), Israel (1998), Schwoerer *et al*. (2002), Yang (2002), and Fessler and Genova (2004).
- Diesel VVA design and electro-hydraulic simulation were conducted by Gehrke *et al*. (2008), Bernard *et al*. (2009), Hass and Rauch (2010).
- Intake VVA not only provides benefits on diesel engine air system performance and pumping loss reduction, but also improves combustion efficiency and reduces emissions through higher charge density, colder combustion temperature and better in-cylinder charge motion. The research on the combustion benefits of diesel intake VVA was conducted by Kim and Kim (2002), Su *et al*. (2005, 2009a, 2009b), Kim *et al*. (2009), Murata *et al*. (2010), Su (2010), and De Ojeda *et al*. (2010).

Valve overlap, IVC, and EVO in diesel VVA

The cam phasing VVA devices are widely used in gasoline engines to change the valve overlap. Automotive high-speed diesel engines have a very high geometric compression ratio that prevents the use of large valve overlap offered by the cam phasing devices due to the concern of valve-to-piston contact under the tight clearance at the TDC. Any valve pocket in the piston top or valve recession in the cylinder head that may prevent valve-to-piston contact is generally associated with a sensitive influence on combustion and emissions. At part load or low idle, the engine delta P in the diesel engine is not as great as in the conventional gasoline engine due to throttleless operation. Note that the gasoline engine has a much higher exhaust manifold pressure than the intake manifold pressure (vacuum) at low idle. The diesel engine is operated with a much higher air–fuel ratio than that in the gasoline engine so that it can tolerate high EGR at low idle. The effect of trapped residue mass on diesel combustion is not as detrimental as that in the spark-ignition combustion which is burned stoichiometric. Therefore, a large valve overlap in the diesel engine does not cause the severe problems of high residue fraction and combustion instability that are encountered in the gasoline engine.

The valve overlap in the non-EGR turbocharged diesel engine is a compromise between the low-load reverse residue flow and the full-load component temperatures. The non-EGR diesel engine is usually designed to have large valve overlap in order to enhance gas scavenging with negative engine delta P. Good scavenging and cooling can be achieved to reduce the thermal load on components such as the exhaust valve, the combustion chamber, and the turbine. At lower speeds or loads, the boost pressure is likely lower than the exhaust manifold pressure so that a small valve overlap is desirable in order to prevent or reduce the exhaust reverse flow. The small valve overlap was especially important for old engines that burned heavy diesel fuels. The exhaust residue reverse flow in those old engines was very contaminating for the intake port and manifold. Although the use of VVA for valve overlap control is not as critical as in the gasoline engine, VVA can be helpful for the non-EGR diesel engine in order to achieve short valve overlap at low loads and long overlap at high loads by properly matching the overlap with engine delta P at different loads or speeds. A comprehensive simulation on the effect of valve overlap is given in Fig. 9.12(a).

On the other hand, the modern EGR diesel engine is usually designed to have small valve overlap at any speed and load in order to prevent the residue gas backflow and maximize volumetric efficiency. The valve overlap needs to be small due to the need for a positive engine delta P to drive the external EGR flow. Note that intake manifold pressure, engine delta P and BSFC are affected by valve overlap size (Fig. 9.12(a)). Two special cases of valve overlap in VVA when both the intake and exhaust valves remain open are the intake secondary pre-lift event during the exhaust stroke and the exhaust secondary post-lift event during the intake stroke. They are used to induce internal EGR (residue gas). The secondary valve lift events may offer some opportunities for the diesel engine to regulate air–fuel ratio and engine delta P. Moreover, it should be noted that a negative valve overlap (i.e., EVC earlier than IVO) is generally not desirable due to its negative impact on the reduced valve flow area, high pumping loss, and high recompression pressure. The negative valve overlap can be considered in HCCI combustion control by using hot internal EGR (detailed in Section 9.8).

The IVC timing in diesel VVA is usually related to the so-called Miller cycle (Miller, 1947; Miller and Lieberherr, 1957) with the turbocharging effect. A comprehensive simulation on the effect of IVC timing is given in Fig. 9.12(b) and (c). The EVO timing in diesel VVA is also closely related to the characteristics of pulse turbocharging. These topics are detailed in later sections.

The valve lift rate is limited by cam design, mechanical stresses and other dynamic issues. In general, flow losses across the valve can be reduced by a faster valve opening or closing rate. Camless VVA may offer an advantage over the cam-driven VVA with a faster lift rate. However, it should be noted that excessively fast lift rate for intake valve opening/closing and exhaust valve closing is not necessary due to the fact that near the beginning of the intake stroke and the end of the exhaust stroke the piston velocity and thus the valve flow velocity are very low, and an excessively fast valve lift (as designed in some camless devices) only provides negligible benefit on pumping loss reduction. Therefore, there is no need to pursue a near 'square shape' valve lift profile for the intake valve in the camless design, especially at low engine speeds. A trapezoidal profile is usually sufficient because the shapes of the instantaneous piston velocity and the valve flow rate as a function of crank angle are close to trapezoidal.

On the other hand, unlike EVC, IVO, and IVC, a fast opening lift at EVO is usually desirable in order to achieve retarded EVO timing for better BSFC. Moreover, faster opening of the exhaust valve lift also reduces the valve flow losses during the blow-down process and increases the energy available at the turbine. As pointed out by Stone and Kwan (1989), rapid exhaust valve opening is especially needed at low engine speeds where the turbine power is low.

Performance benefits of diesel engine VVA

The traditional belief was that the diesel engine may benefit less from VVA in engine breathing performance compared with the gasoline engine for the following reasons. First, the load control in the diesel engine is achieved by adjusting the fuel quantity instead of throttling the air. Therefore, the potential of using VVA to reduce pumping loss at part load is much less. Secondly, the fuel injection equipment and the heterogeneous diesel combustion restrict the rated speed of the diesel engine usually below 4500–5000 rpm, which is much lower than the rated speed of the gasoline engine. Therefore, the speed range is much smaller and the trade-off in valve timing is not so severe.

However, it should be noted that the expectation on the percentage fuel economy improvement should be different for gasoline and diesel engines because the diesel engine is largely used in heavy-duty applications where even a small percentage of fuel economy gain is valuable for a large fleet. Therefore, any small BSFC gain offered by VVA should not be simply discarded in diesel engine design. Moreover, most importantly, it is believed that the Miller cycle with VVA can provide significant BSFC improvement for the modern diesel engines that have high engine delta P and high EGR. Beyond that, there are many other benefits that VVA can offer to the diesel engine. A cost–benefit assessment and a comparison with the competing technologies are required to justify the commercial use of VVA for production diesel engines. The competing technologies for fuel economy improvement include hybrid powertrain, waste heat recovery, and advanced turbocharging. The benefits of diesel engine intake and exhaust VVA are summarized as follows.

Increased low-speed torque

Increased low-speed torque can be achieved by the following: (1) retarding EVO timing past the BDC with a penalty in BSFC (Tai *et al*., 2002); (2) using the Curtil exhaust VVA system (Curtil, 1998) to utilize the exhaust reverse flow at the compression BDC (Israel, 1998; Fessler and Genova, 2004); and (3) advancing IVC or EVO timing (Lancefield *et al*., 2000). The approach of retarding EVO can greatly increase the air–fuel ratio due to an increase in turbine power so that more fuel can be burned to obtain higher torque at low speeds. The approach works well only at low speeds (e.g., 1000 rpm). The turbine power increase is due to a higher pressure pulse at the turbine inlet that is caused by both pulse turbocharging and a higher residue gas pressure inside the cylinder. The EVO retarding needs to be conducted with strong pulse turbocharging, and the compressor may surge under vey low flow rate and high boost pressure. Therefore, careful turbocharger matching is required in order to realize the high torque. Moreover, note that the very late EVO gives greater expansion work but more pumping loss and increased residue gas trapped. The associated BSFC penalty can be small in terms of the contribution to the overall driving-cycle fuel economy due to very little time spent at low speed full load (e.g., 1000 rpm) during acceleration.

Reduced pumping loss and fuel consumption

Reduced pumping loss is achieved primarily through the following three measures: (1) use the Miller cycle with early or late IVC to reduce both volumetric efficiency and engine delta P while increasing boost pressure (large benefit); (2) use camless VVA to achieve fast valve lift rate to maximize the effective valve flow area (small benefit); and (3) deactivate both fuel injection and valve flows in cylinder deactivation (large benefit at low loads).

Improved fuel consumption can also be achieved by the following means related to the engine cycle processes: (1) optimize the variable EVO timing (small benefit); and (2) obtain better thermodynamic cycle efficiency by using a very high geometric compression ratio, a relatively larger expansion ratio and a smaller effective compression ratio such as in the Miller or Atkinson cycles (large benefit). Previous research indicated that an improvement of at least a few in BSFC is possible by optimizing the EVO and IVC timings (without using the Miller cycle or cylinder deactivation yet) as well as using fast valve lift rate with VVA, for example, 4% at low idle, 1–3% at light load, 1% at full load, 3–5% in driving cycles. It is found that city driving offers the largest gain in BSFC reduction percentage, and highway driving probably offers the least gain in BSFC reduction.

Reduced emissions

Increased air flow or air–fuel ratio can be achieved to reduce soot by optimizing the volumetric efficiency (mainly through adjusting IVC) across the entire engine speed–load domain. The gain in volumetric efficiency can be several percentage points. Effective compression ratio can be reduced by early or late IVC to obtain lower compression pressure, combustion temperature, and NO_x. Early IVC timing also provides certain limited benefit of internal in-cylinder charge cooling due to the expansion toward the intake BDC. Hot internal EGR (residue gas), although not as effective as cooled external EGR for NO_x reduction, can reduce NO_x in certain applications that are not subject to stringent emissions regulations. Yang and Keller (2008) reported that late IVC could reduce NO_x by 24% and BSFC by 1% at 2000–3000 rpm and 4–5 bar BMEP for a LD small diesel engine. The charge composition control via internal EGR is achieved by an exhaust valve post-lift during the intake stroke (the preferred approach) or an intake valve pre-lift during the exhaust stroke (less desirable). Leet *et al*. (2004) estimated NO_x reduction of 10% or 0.13 g/(hp.hr) by using VVA to achieve the 1.18 g/(hp.hr) NO_x level for HD diesel engines. Millo *et al*. (2007) achieved 13% reduction in NO_x for a small off-road diesel engine by using internal EGR without remarkable detrimental effects on fuel consumption and soot emission. As reported by Fessler and Genova (2004), the Euro-IV emissions limit of NO_x can only be reached with cooled EGR and is not achievable by using VVA alone. As to the particulate matter (PM) and unburned HC emissions, late EVO timing results in longer in-cylinder burning time for PM and HC so that they can be reduced, although the pumping loss may increase with the retarded EVO timing. Late IVO may enhance the air flow velocity and the turbulence to reduce unburned HC emissions. In summary, VVA itself is not a sufficient enabling technology to meet the most stringent NO_x and PM emissions regulations.

Air motion control and swirl/tumble modulation

Increased in-cylinder gas motion can be achieved by using low or unequal intake valve lifts to increase the air flow velocity across the valve seat in order to enhance the swirl formation in the cylinder and reduce PM and BSFC at low speeds. Another alternative is to open the intake valve late after the TDC to create a large pressure differential across the intake valve, which results in high valve flow velocity and also high pumping loss. The VVA can replace the use of swirl flaps and may have comparable influence on the in-cylinder charge motion as the cylinder head design offers. The VVA valve deactivation system may be more effective and obtain a higher swirl ratio than the port deactivation system can do.

Eliminating air control valves

It is possible to eliminate some air control valves by using VVA to achieve a similar or better function of air flow and boost pressure controls. The air control valves include intake throttle, exhaust throttle, turbine wastegate, and EGR valve. Eliminating the redundant air control valves can greatly simplify the engine system design and reduce cost. VVA can regulate the engine air flow and the pressures by using the valves in the cylinder head with crank angle resolution without incurring the additional throttle losses at other air control valves.

Enabled advanced combustion

Intake VVA is an enabler for advanced combustion concepts such as HCCI (detailed in Section 9.8). VVA provides controls on variable compression ratio and charge composition and temperature (via internal EGR) that are required for HCCI ignition and combustion. VVA also provides opportunities for controlling the combustion processes of the engine that use multiple types of fuels.

Improved cold start

Better cold start means quick start, reduced HC emissions and white smoke, and the possibility of eliminating the glow plug to reduce cost. Cold-start quality is affected by the air temperature in the cylinder at the end of the compression stroke, geometric compression ratio, and IVC timing. Quick cold start can be achieved by advancing IVC toward the BDC to avoid the reverse air flow and increase the effective compression ratio, the trapped gas mass, and the compression temperature at the TDC. Engine geometric compression ratio is largely determined by the cold-start requirement, and the compression ratio is usually higher than the ratio that would be optimally required for the best BSFC during hot engine operations after cold start. Unlike the low- or medium-speed diesel engines used in the applications of marine, railway transportation, and power generation sets, automotive high-speed diesel engines use high compression ratios. VVA with early IVC timing also enables a lower geometric compression ratio for better peak cylinder pressure control at rated power. VVA can also tolerate lower quality fuels (e.g., cetane number or rating less than 40) or colder ambient air temperatures in terms of cold-start capability. Other means of assisting quick cold start include the following: (1) using cylinder deactivation to increase the cranking speed; (2) opening the engine valve during the compression stroke to motor the engine with minimum compression at the very early stage of cranking in order to reach the cranking speed more quickly with less energy provided by the starter motor; and (3) increasing the charge temperature by inducting the warm gas from the exhaust port by using a post-lift of the exhaust valve. After cold start in a cold climate, VVA can reduce HC emissions during warm-up by using internal EGR and cylinder deactivation. Because unburned HC emissions are low in the diesel engine, any improvement in HC emissions via VVA is limited to cold start.

Fast warm-up

The cold start portion of the HD transient test cycle becomes more important as the emissions regulations become more stringent. The NO_x aftertreatment system also requires fast warm-up after cold start in order to achieve high conversion efficiency. The following measures can speed up the engine warm-up: (1) throttling the intake or exhaust valve flow by using reduced valve lifts to add load (pumping loss) to the engine; and (2) using hot internal EGR (residue gas) to heat the intake charge. Early EVO timing can help speed up the warm-up of the aftertreatment system.

Reduced turbocharged lag and improved transient response

High low-end torque and fast transient response can improve drivability. Both early EVO and very late EVO can increase turbocharger speed and reduce turbocharger lag during acceleration transients although BSFC usually becomes worse due to the loss of expansion work or increased pumping loss.

Improved aftertreatment performance

Intake or exhaust throttling may reduce air–fuel ratio and increase the exhaust temperature for the aftertreatment system (e.g., DPF regeneration, SCR, LNT). Exhaust throttling generally produces less BSFC penalty than the pre-compressor intake throttling. However, any throttle valves placed in any part of the engine air system increase BSFC significantly, and their use should be avoided or minimized in engine system design. Although all the air control valves can achieve a wide range of controls of exhaust temperature and space velocity by adjusting the air flow rate, only VVA (e.g., early or late IVC, or cylinder deactivation) can achieve them without BSFC penalty. Early EVO and the valve events to obtain internal EGR may also increase the exhaust temperature. The benefit of exhaust temperature increase by VVA is especially important at light loads and during warm-up.

Enhanced engine braking

VVA allows the braking valve timing and the retarding process efficiency to be optimized at each engine speed so that the retarding power of the compression-release brake can be maximized (Hu *et al.*, 1997a, 1997b).

Integrated air compressor

Camless VVA can be used in a designated cylinder to deliver the compressed air to charge the air tank for vehicle use. Such an integrated air compressor may eliminate the external air compressor on the vehicle.

Switching between two-stroke and four-stroke operations for firing and engine braking

The ability to switch the engine operation from four-stroke to two-stroke is attractive due to the high power delivered by the two-stroke operation, for example during vehicle hill-climbing, acceleration, or deceleration. A camless VVA device can achieve such an advanced feature on the engine.

Internal EGR for emissions control with diesel VVA

Internal EGR refers to the mass of the residue gas from the previous engine cycle that is retained in the combustion chamber to participate in the combustion in the subsequent engine cycle. Although engine development practice in the past decade has proven that cooled external EGR is necessary to meet the most stringent emissions regulations for modern diesel engines with the minimum BSFC penalty, internal EGR finds its role as a useful technology for diesel HCCI combustion in low-load operation.

The use of hot internal EGR in conventional diesel combustion generally results in the following: decreased volumetric efficiency; increased compression temperature; reduced oxygen concentration and air–fuel ratio; reduced NO_x;

increased soot, BSFC, and exhaust temperature; and a large cylinder-to-cylinder variation of internal EGR rate due to manifold wave effects. A series of change in internal EGR rate forms a trade-off between NO_x and soot, and another trade-off between NO_x and BSFC. Increasing the boost pressure may improve the trade-offs by simultaneously reducing NO_x and BSFC at the same soot level.

Internal EGR is useful for NO_x reduction when negative engine delta P occurs. The effectiveness of internal EGR on NO_x and BSFC varies with different EGR induction methods in terms of VVA valve events, depending on the maximum available EGR rate, pumping loss, and the heat loss of the burned gas. There are three internal EGR methods: (1) negative valve overlap with early EVC to trap the residue gas; (2) intake valve pre-lift during the exhaust stroke to induct and store the burned gas in the intake port; and (3) exhaust valve post-lift during the intake stroke to draw the burned gas from the exhaust manifold into the cylinder with the exhaust pressure pulses. The third method seems the best for NO_x and BSFC in conventional diesel combustion (Edwards *et al.*, 1998; Millo *et al.*, 2007), while the first method gives the highest cylinder charge temperature which is sometimes desirable for HCCI. Internal EGR methods and their gas exchange processes are discussed in detail by Desantes *et al.* (1995), Benajes *et al.* (1996), and Pischinger *et al.* (2000).

Desantes *et al.* (1995) and Benajes *et al.* (1996) used an inline six-cylinder engine with a divided-entry turbine in their simulation. They found that the internal EGR rate was influenced more by the engine speed than the load. Higher engine speed produced higher EGR rate. They also found that the intake pre-lift event produced lower EGR rate than the exhaust post-lift event did. They obtained 5–20% EGR rate at different engine speeds and loads with different valve timing/event. They discovered that the exhaust post-lift affected swirl, while the intake pre-lift did not. The swirl ratio was greatly reduced by the high internal EGR rate attained with the exhaust post-lift.

The Miller cycle and wastegating reduction with diesel VVA

In addition to hot internal EGR, the Miller cycle (Miller, 1947; Miller and Lieberherr, 1957) is another commonly used feature of VVA application in the turbocharged diesel engine. The Miller cycle has the following characteristics: (1) shorter effective compression stroke than the expansion stroke (i.e., the Atkinson cycle); (2) variable intake valve timing and the resulting variable effective engine compression ratio; and (3) increased boost pressure via supercharging or turbocharging to compensate for the reduction in volumetric efficiency and air–fuel ratio caused by the short intake valve event duration. When the engine geometric compression ratio is reduced, the expansion ratio is also lowered. This results in a reduction in thermodynamic cycle indicated efficiency. The Miller cycle and the Atkinson cycle can lower the compression ratio without reducing the expansion ratio by advancing or retarding the intake valve timing in order to improve the thermodynamic cycle efficiency. The Miller cycle can also reduce peak cylinder pressure and mechanical load via lower effective compression ratio.

When the Miller cycle uses early IVC during the intake stroke before the BDC rather than late IVC during the compression stroke, the effect of internal charge cooling is achieved via cylinder expansion toward the BDC. Internal cooling can effectively reduce the charge temperature at the BDC and hence the charge temperature at the start of combustion, resulting in lower cycle temperatures, thermal loads and NO_x. Alternatively, internal cooling allows the start-of-injection timing to be advanced to achieve better BSFC at the same NO_x level. Previous calculations showed that a 6% reduction in temperature and pressure can be achieved through adiabatic expansion at the start of the compression stroke if the intake valve is closed at 60° crank angle before the intake BDC (Ishizuki *et al.*, 1985). Edwards *et al.* (1998) showed in their research of the combined Miller cycle and internal EGR that internal cooling was achieved to reach the same in-cylinder gas temperature as that given by the cooled external EGR at the Euro-III level (about 8% EGR), and using the expansion cooling for the Euro-IV EGR level (18%) also seemed feasible.

The fuel economy advantages of the Miller cycle result from the following four aspects: (1) greatly reduced engine delta P and pumping loss; (2) increased thermodynamic cycle indicated efficiency due to reduced compression ratio relative to the unchanged expansion ratio; (3) increased geometric compression ratio enabled by the reduced effective compression ratio; and (4) internal charge cooling due to early IVC. The Miller cycle can reduce NO_x due to lower compression temperature, but may increase soot significantly if the air–fuel ratio becomes too low, especially at full load. The NO_x reduction potential of the Miller cycle used with internal EGR is limited to the high-NO_x engines such as HD Euro-III or at LD part load. Cooled external EGR is required for the Miller cycle used in low-NO_x engines. There are six critical parameters involved in designing an effective Miller cycle: IVC timing, turbine area, turbine wastegate opening, turbocharger efficiency, compressor matching, and engine geometric compression ratio.

The retarded or advanced IVC timing creates a shorter intake valve event duration and higher intake manifold boost pressure so that engine delta P is reduced. The effect of engine delta P reduction by the Miller cycle was illustrated by Ishizuki *et al.* (1985). Mavinahally *et al.* (1996) reported that the engine with 50% greater expansion ratio than the effective compression ratio produced 3% higher brake thermal efficiency if compared at the same air–fuel ratio and peak cylinder pressure, although the power capability of the engine was reduced due to the shortage of air–fuel ratio at full load caused by late IVC. Ishizuki *et al.* (1985) reported the benefit of the Miller cycle on BSFC by illustrating a significant BSFC reduction around 3.3% in a large area in the engine speed–load domain. They also reported that there was no gain and no loss in BSFC under the high-speed conditions, but this statement is believed misleading and not conclusive for the Miller cycle. They mentioned that one remarkable improvement by the Miller cycle was the increase in BMEP at low speeds by using a smaller turbine area.

In fact, it is found that if compared with the conventional engine at the same brake power and peak cylinder pressure, the Miller cycle can offer significantly lower fuel consumption (e.g., 7%) but at the expense of a little lower air–fuel ratio due to its low intake manifold volumetric efficiency. If compared at the same air–fuel ratio, EGR rate, brake power, and peak cylinder pressure, the Miller cycle can still offer significantly lower BSFC, usually at least 2% and often in the range of 4–5% at rated power for modern heavy-duty high-EGR diesel engines. The variation of the BSFC benefit is caused by power density (or engine load), EGR level, engine speed, turbocharger configuration, etc. The BSFC benefit of the Miller cycle working with wastegating reduction or elimination is especially prominent at high speeds. The Miller cycle can reduce excess air flow and pumping loss effectively. Although early or late IVC timing can be used to reduce engine delta P and pumping loss as low as possible, it should be noted that an insufficient engine delta P causes the problem of EGR driving, especially at peak torque. Therefore, fixed cam timing usually may not be sufficient to cover the needs of both delta P reduction and EGR driving at both high speeds (e.g., rated power) and low speeds (e.g., peak torque). The design challenge in the Miller cycle is to achieve the variable intake closing timing required at all speeds via VVA.

The turbine area is usually determined by the EGR-driving and/or the air–fuel ratio requirements at peak torque with the EGR valve fully open. The turbine wastegate needs to be adjusted together with the IVC timing in order to reach the desired air–fuel ratio and engine delta P. The turbocharger efficiency can be adjusted in turbocharger design in order to compensate for the gap in air–fuel ratio without affecting the engine delta P. It should be noted that the benefit of the thermodynamic cycle efficiency of the Miller cycle can be offset by the increased engine delta P that is caused by a smaller turbine area, while the engine delta P and BSFC are not negatively affected by the boost pressure increase that is produced by higher turbocharger efficiency.

The compressor size needs to be properly adjusted in order to better fit the engine operating point of the Miller cycle at rated power in the high efficiency region on the compressor map. The boost pressure of the Miller cycle can be much higher than that in the conventional engine, and the compressor flow rate can be similar or slightly reduced. The high boost pressure is not an issue since two-stage turbocharging has been widely used in today's high-

EGR or high-BMEP engines. In fact, compared to the conventional engine, the Miller cycle offers an advantage in two-stage turbocharging in the way that it can split its high boost pressure to the two-stage compressors with a better distribution on the two compressor maps with higher compressor efficiencies.

The engine geometric compression ratio needs to be increased in the Miller cycle to match with the reduced effective compression ratio that is caused by early or late IVC timing. The purpose of increasing the geometric compression ratio is to fully utilize the design limit of peak cylinder pressure while maximizing the engine thermodynamic cycle efficiency.

Both early and late IVC can affect effective engine compression ratio and volumetric efficiency. The ratio is defined as the volumetric ratio of the cylinder volume at IVC to the cylinder volume at the TDC. Other definitions of effective pressure ratio also exist, such as a dynamic pressure-based compression ratio used by He *et al*. (2008). A common feature between early and late IVC is that they both produce very small pumping loss, as illustrated in the p–V diagram by Pischinger *et al*. (2000). The difference between early and late IVC in throttleless operation is that in the p–V diagram early IVC gives decreasing cylinder pressure after IVC toward the BDC during the intake stroke as a closed thermodynamic system for the cylinder, and the cylinder pressure returns along the same path after IVC during the compression stroke. On the other hand, the late IVC gives almost constant cylinder pressure during the intake stroke and the compression stroke as an open thermodynamic system from IVO to IVC.

Usually, late IVC produces slightly higher pumping loss than early IVC because the valve throttling flow loss during the longer valve opening duration is higher. Late IVC consumes work to induct the gas mass that is immediately pumped back out of the cylinder. Early IVC reduces the amount of gas inducted into the cylinder, and thereby reduces the pumping loss associated with the induction process. As pointed out by Gallo (1992), high irreversibility peaks in the irreversibility rate profile of the intake or exhaust process are associated with the backflow through the intake and exhaust valves and the losses during the mixing process of the hot residue gas and the colder air/gas mixture. Moreover, late IVC at light loads may cause an increase in the intake manifold gas temperature due to (1) charge heating by the combustion chamber walls; and (2) inducting the charge back as hot gas. The heating effect of charge refilling causes a reduction in gas density.

On the other hand, there are concerns with early IVC, too. One problem is the reduction of in-cylinder air motion and turbulent mixing as consequences of early IVC (Edwards *et al*., 1998). Moreover, compared to late IVC, the practicality of early IVC can be an issue in mechanical cam design due to the constraints of valvetrain dynamics with short valve event duration at the same maximum valve lift. Certainly, using lower maximum valve lift is feasible for early IVC in order to make the cam design easier. The difference between early and late IVC is discussed in detail by Anderson *et al*. (1998). The effects of early and late IVC on boost pressure and engine delta P are shown in the simulation data in Fig. 9.12(b) for a heavy-duty diesel engine.

In addition to the cam phase shifters and camless devices, rotary valves have been used as a cost-effective means of VVA to achieve early IVC timing or the Miller cycle in the gasoline engine (Anderson *et al*., 1998) and the diesel engine (Ishizuki *et al*., 1985; Kamo *et al*., 1998). The rotary valve is located upstream of the engine intake valve in the intake port or manifold. The rotary valve is different from the intake throttle in the way that it provides a timed open and closed event. The optimum cut-off timing of the rotary valve can be found at each engine speed and load condition. The relatively simple mechanism of the rotary valve assures good reliability and durability, compared with the more complicated modifications to the engine valvetrain in other VVA mechanisms. Ishizuki *et al*. (1985) conducted a comprehensive analysis on the advantages and disadvantages of the rotary valve mechanism. They concluded that its inherent disadvantages such as the dead volume and leakage did not have a significant adverse effect on the Miller cycle performance, although its impact on swirl ratio remained uncertain. Although the rotary valve technology was initiated for non-EGR gasoline engines, it may become important for modern high-EGR low-NO_x diesel engines. The rotary valve can also be used to achieve cylinder deactivation. More information on the rotary valve mechanisms can be found in Sakai *et al*. (1985).

The Miller cycle can also be used for boost pressure control in turbocharged engines. The research conducted by Ke and Pucher (1996) is probably the most important VVA work (although on the gasoline engine) that sheds light for turbocharged diesel VVA research. They investigated the feasibility of using early IVC timing to replace turbine wastegate for boost control in order to prevent over-boosting for engine protection. They showed that under high speed (5600 rpm) full load conditions the same cylinder pressure and trapped air mass at the intake BDC could be achieved by either a conventional cam with IVC timing at 63° after the BDC with wastegating or an early-IVC cam with IVC at 59° before the BDC without opening the wastegate. The in-cylinder temperature was approximately 20 K colder at the intake BDC for the early-IVC cam. Moreover, the pumping loss of the early-IVC case was lower than that given by the wastegating case due to lower engine delta P.

The Curtil system with diesel VVA

The above discussion on the Miller cycle shows that some advanced thermodynamic cycles via valve timing variations may suffer the loss of air–fuel ratio although they can greatly improve pumping loss, engine thermal efficiency and BSFC. The capability of charging fresh air to maintain an acceptable air–fuel ratio is important for diesel engines. The traditional way to compensate for the loss of air–fuel ratio without a negative impact on engine delta P is to increase turbocharger efficiency by better aerodynamic design of the turbine or the compressor. Innovative ways of charging fresh air by using flexible engine valve events and manifold pressure wave dynamics without a penalty in pumping loss are also important and attractive.

It is worth considering the 'Curtil' system (Curtil, 1998), which can back-charge the cylinder with the fresh air stored in the exhaust port. In the Curtil system, the fresh air flows from the cylinder into the exhaust port via a secondary exhaust valve post-lift event near IVC during the intake stroke when the intake manifold boost pressure is higher than the exhaust manifold pressure. Then, the fresh air stored in the exhaust port is pushed back by the exhaust pulses from adjacent cylinders to return into the cylinder when the exhaust pressure pulse is greater than the intake port pressure. Therefore, the trapped air mass in the cylinder can be increased to reach a high air–fuel ratio for soot reduction or allow a higher fueling rate to produce more torque. The Curtil system can greatly increase the transient torque and reduce smoke. Fessler and Genova (2004) reported that 50% higher torque was obtained using the Curtil system at low engine speeds such as 1000–2000 rpm with acceptable smoke and BSFC penalty.

It should be noted that the Curtil system only works well in conditions where the exhaust port pressure oscillates around the intake port pressure during the intake stroke (e.g., typically at high engine speeds and low loads). The Curtil system and adjusting EVO timing with VVA are two primary means of increasing the low-speed torque to improve drivability and transient performance. The Curtil system is also a promising means for fresh air charging to supplement other advanced thermodynamic cycles to compensate for the air–fuel ratio shortage. More detailed information on the Curtil system is provided by Curtil (1998), Israel (1998), and Fessler and Genova (2004).

The theory of using wastegating reduction or elimination and VVA (i.e. WR-VVA or WE-VVA systems) to control pumping loss

Modern high-EGR diesel engines with non-VGT turbines (i.e., fixed geometry turbines) are characterized by high engine delta P or pumping loss at high speeds (from part load to full load) due to the requirement of using a small turbine area to drive sufficient EGR at peak torque or low speeds. Although VGT is probably the best solution at high speeds and high loads to solve the problem of high engine delta P, there are concerns about the VGT technology with regard to the cost, durability, turbine efficiency loss at large turbine area openings (e.g., at fully open vane position), and product availability. Moreover, at engine part load even a good VGT may not have an advantage over the Miller cycle with VVA. Intake VVA with the Miller cycle, VGT, and low-pressure-loop EGR (or the hybrid EGR system) are three competing technologies in terms of engine delta P control (reduction) in the entire engine speed–load domain.

The application of the Miller cycle with VVA on modern high-EGR diesel engines offers great advantages in pumping loss control for fuel economy and air system simplification for cost reduction. It can be foreseen that in the future HCCI, VVA, and VGT will be the three most important technologies for diesel engines, one for advanced combustion and the other two for advanced air systems. Since intake VVA (particularly IVC-VVA) is used as an enabler for HCCI combustion anyway, using the IVC-VVA or the Miller cycle to extend its functionality to improve the air system will become naturally logical for product cost saving (e.g., to avoid the expensive VGT).

The conventional rationale for VVA in diesel engines was to use early IVC to achieve internal charge cooling due to in-cylinder expansion, or to use early/late IVC to reduce the effective compression ratio to reduce the compression temperature and NO_x. With other emissions control technologies widely used on diesel engines (e.g., cooled external EGR system, DPF), the focus of using VVA should be placed on pumping loss reduction and fuel economy improvement. In particular, the VVA with early or late IVC timing is very effective in reducing engine delta P by raising the intake manifold boost pressure. IVC-VVA and VGT are two solutions for the problem of high pumping loss in modern diesel engines. VGT can be more effective than IVC-VVA but VGT does not offer the other advantages of VVA described earlier. Note that the Miller cycle does not specify how to use the wastegate.

The theory of using IVC timing to control engine delta P and pumping loss for turbocharged diesel engines is illustrated by a comprehensive simulation shown in Fig. 9.12(c), which includes the parametric sensitivity effects of the following key design parameters: IVC timing, turbine area, turbine wastegate opening, turbocharger efficiency, and engine geometric compression ratio. The simulation is conducted at 3400 rpm rated power for a high-power-density engine at 20 bar BMEP with a single-stage wastegated turbocharger. The EGR rate can be fixed at any constant value (zero here without losing the validity for illustration). The second legend in the figure represents the baseline system. The air system capabilities of different systems corresponding to five different legends are presented in the domain of ‘air–fuel ratio vs. engine delta P’, and the domain of ‘air–fuel ratio vs. BSFC’. The ideal design target point at high engine speeds should be located at low engine delta P (with EGR valve fully open) and sufficiently high air–fuel ratio in such an air system capability domain. The relationship between different systems and their potentials to reach the ideal design target point are clearly illustrated in the figure, especially in the figure of the air system capability domains.

The principles of the modern diesel engine air system design related to intake VVA and turbine wastegating reduction or elimination are summarized as follows.

1. An intake-valve-closing VVA system (preferably early IVC), in conjunction with using minimum or no turbine wastegating, is highly desirable for reducing engine delta P (by increasing the intake manifold boost pressure), pumping loss, and in-cylinder cycle temperature in order to achieve the best BSFC and the lowest coolant heat rejection for non-VGT engines. The design goal is to minimize engine delta P at any speeds/loads so that the required EGR rate can be achieved with the EGR valve always fully open with minimum EGR circuit flow restriction.
2. The intake manifold volumetric efficiency of the IVC-VVA system needs to be reduced to around 60–70% via valve event duration change with an increase intake manifold boost pressure at high speeds in order to reduce the engine delta P to the minimum level to minimize BSFC.
3. For the IVC-VVA engine, a high geometric compression ratio (CR) can be designed to maximize thermodynamic cycle efficiency and further enhance cold start capability, compared to the wastegate engine.
4. For fixed geometry turbines, the turbine effective area is sized at the peak torque condition based on the requirements of EGR rate and air–fuel ratio with the (also considering the high altitude conditions) EGR valve fully open and usually the intake throttle fully open.
5. The IVC timing at peak torque is optimized with the turbine area for the required air–fuel ratio and engine delta P to drive sufficient EGR flow. The IVC timing at rated power or high speeds is optimized to achieve the best trade-off between air–fuel ratio and engine delta P (shown by the characteristic lines in the air system capability chart in Fig. 9.12(c)).
6. Compared to the required air–fuel ratio for combustion and emissions, if the actual air–fuel ratio delivered by the air system is too low with early or late IVC at the desirable engine delta P with the EGR valve fully open (e.g., 0.1–0.2 bar engine delta P in Fig. 9.12(c)), the turbocharger efficiency needs to be increased in order to raise the air–fuel ratio without affecting the engine delta P significantly. Or, alternatively the turbine wastegate can be closed along with a variation in IVC timing in order to raise the air–fuel ratio in the air system capability chart.
7. If the air–fuel ratio is too high with early or late IVC at the desirable engine delta P with the EGR valve fully open, the turbocharger efficiency needs to be decreased or the wastegate needs to be opened more.
8. With the IVC-VVA, it is possible to delete the turbine wastegate or make the EGR valve a less expensive on-off type rather than the continuously adjustable valve, depending on the air–fuel ratio requirement at different

speeds/loads. In other words, the IVC-VVA may be used as an EGR control device by adjusting engine delta P to replace the EGR valve.

9. The IVC-VVA can replace the intake throttle to reduce air–fuel ratio to a certain extent. Their difference is on the effect on engine delta P. The air–fuel ratio reduction by VVA is accompanied by a large reduction in engine delta P, while the intake throttle does not affect engine delta P significantly.
10. VGT can perform better than VVA in terms of the air system capability of 'air–fuel ratio vs. engine delta P'. If the required air–fuel ratio is not demanding, the IVC-VVA can be used to achieve low pumping loss.
11. The WE-IVC-VVA performs better than wastegating in terms of the capability of 'air–fuel ratio vs. engine delta P' because VVA does not waste the exhaust energy. This advantage is shown by comparing the curves in the capability-domain plot in Fig. 9.12(c) at the 0.1 bar engine delta P. It shows that the IVC-VVA system can deliver significantly higher air–fuel ratio than the wastegating system without VVA at the same engine delta P. The WE-IVC-VVA system also allows an SOI/SOC advance.
12. The WE-IVC-VVA is able to adjust volumetric efficiency by raising the boost pressure and the location of the engine operating points on the compressor map. Therefore, a re-match of the compressor is required in order to fit the operating points as close to the high efficiency region as possible. This is particularly important for two-stage turbo. Compressor outlet air temperature control should be paid attention since the boost pressure can become very high at rated power in the IVC-VVA engine.
13. All the questions related to the functionalities of IVC timing, turbine area, wastegate, turbocharger efficiency, and intake throttle can be well explained by Fig. 9.12(c) and equation 9.31. For example, for the IVC-VVA engine, the four unknowns in equation 9.31 can be p_2, p_3, $\dot{m}_{air}$, and $\dot{m}_{EGR}$, with turbine area, wastegate opening, volumetric efficiency (related to IVC timing), and EGR valve opening (e.g., fully open) set as known inputs.

9.8 Variable valve actuation (VVA) for diesel homogeneous charge compression ignition (HCCI)

9.8.1 Controlled auto-ignition (CAI) and HCCI combustion

Controlled auto-ignition (CAI) is a lean combustion mode that does not require the necessity of a lean exhaust gas aftertreatment on the gasoline engine. CAI engines usually use VVA to achieve the controls on air charge mass, gas composition and charge temperature in addition to the benefit of pumping loss reduction due to the throttleless operation. CAI significantly improves fuel economy and greatly reduces NO_x. CAI research was summarized by Zhao (2007).

Homogeneous charge compression ignition (HCCI) is the counterpart of CAI for diesel and other compression ignition engines. HCCI is a hybrid strategy between the traditional homogeneous charge spark ignition (used in the conventional gasoline engine) and the stratified charge compression ignition (used in the conventional diesel engine). It refers to the auto-ignited combustion of premixed fuel–air–diluent mixture under lean and/or diluted conditions with burned gas. The spontaneous ignition occurs simultaneously at many sites throughout the combustion chamber with an optimum auto-ignition timing close to the TDC in order to maximize the indicated efficiency. Fuel consumption can be improved by HCCI due to its short and efficient heat release rate. The absence of a high-temperature flame front leads to practically negligible formation of nitrogen oxides, and the absence of fuel-rich zones in the homogeneous lean mixture leads to very little soot formation, thus resulting in a simultaneous great reduction in both NO_x and particulate matter. HC and CO emissions of HCCI are usually high. Sometimes they can be at the same levels as the direct injection gasoline engine. The primary objective of diesel HCCI is to reduce NO_x and soot so that the use of NO_x aftertreatment can be avoided. HCCI research is summarized by Zhao *et al.* (2003) and Zhao (2007).

As pointed out by Pastor *et al.* (2007) in their review about conventional and HCCI diesel combustion, how to create a homogeneous mixture, how to ignite such a mixture, and how to control the combustion are three major aspects of HCCI technology. The challenges in HCCI include ignition timing and combustion phasing controls, restricted engine operating load range, and high HC and CO emissions. Unlike the conventional diesel combustion where the start of combustion and combustion phasing are controlled by fuel injection characteristics, the start of combustion and the heat release rate of HCCI combustion cannot be controlled by fuel injection rate. They depend only on the reaction kinetics of the cylinder charge. HCCI ignition is controlled by the cylinder charge temperature during the compression, which is affected by the following three factors: (1) initial temperature and composition (e.g., EGR or residue gas) of the intake charge mixture; (2) the rate of compression (i.e., temperature evolution along the crank angle during the compression); and (3) the extent of compression, which is affected by the effective compression ratio. Note that compared to the gasoline engine, the diesel engine has a higher geometric compression ratio and the diesel fuel has a lower ignition temperature due to its lower octane number. Therefore, the HCCI diesel engine relies on intake charge heating and hot burned exhaust gas to trigger auto-ignition to a lesser extent than the CAI gasoline engine does.

There are different methods of fuel injection to form the mixture in diesel HCCI, such as port injection and in-cylinder fuel injection during the compression stroke prior to ignition. Fuel injection timing can be adjusted to control the homogeneity of the mixture. Once a homogeneous lean mixture is achieved, the start of combustion is triggered by the high temperature obtained near the end of the compression stroke. EGR is effective to retard ignition timing and reduce the rate of heat release. The heat of internal EGR can promote the evaporation of the diesel fuel. However, EGR also usually increases HC and CO emissions. A commonly used measure of combustion timing in HCCI research is the crank angle at which 50% of accumulated heat release occurs. The combustion phasing crank angle at 50% heat release is mainly adjusted through a combination of EGR rate and IVC timing with a VVA device to reach an optimum crank angle slightly after the TDC.

The upper limit of BMEP (engine load) in HCCI operation is bounded by the knocking combustion that is indicated by a very early ignition, a too rapid heat release rate, an excessively high rate of cylinder pressure rise, and loud combustion noise. In this case, a lower compression ratio and a large amount of cooled external EGR are needed in order to control the ignition and the combustion rate. However, the geometric compression ratio in diesel engines is usually very high for good cold start capability. The effective compression ratio often cannot be aggressively reduced by early or late IVC timing in VVA because the air–fuel ratio needs to be sufficiently high. Moreover, a large amount of EGR at high BMEP will demand a very high boost pressure and hence high peak cylinder pressure. These are the primary reasons why HCCI has not been successfully used at high loads to date. The lower limit of BMEP in HCCI operation is bounded by the fact that misfire occurs and the mixture fails to auto-ignite when the temperature of the mixture of air and trapped residue gas is too low. The low-load limit of HCCI is characterized by a large increase in CO and unburned HC emissions as well as combustion instability. Therefore, HCCI currently has to coexist with the conventional diffusion-controlled combustion in the same diesel engine for different loads.

9.8.2 VVA applications for HCCI

There are two major applications of VVA for the purpose of HCCI controls: (1) effective compression ratio control to modulate the mass and the temperature of the charge air/gas; and (2) residue gas control to modulate the composition and the temperature of the internal EGR. Variable compression ratio (by IVC timing) and hot internal EGR (by exhaust valve post-lift, intake valve pre-lift or negative valve overlap) or cooled external EGR are used in HCCI controls of ignition timing and combustion phasing. Regardless using internal or external EGR, intake VVA is a necessary enabler for diesel HCCI.

VVA control for HCCI is reviewed by Tunestål and Johansson (2007). VVA applications in diesel HCCI are reported by Babajimopoulos *et al.* (2002), Milovanovic *et al.* (2005), Shi *et al.* (2005), Helmantel and Denbratt (2006), Kodama *et al.* (2007), Nevin *et al.* (2007), Zhao (2007), Peng *et al.* (2008), Murata *et al.* (2006, 2008), and He *et al.* (2008). Related fundamental research on different VVA strategies used for HCCI has been conducted by Caton *et al.* (2005a, 2005b). VVA controller design for HCCI is presented by Strandh *et al.* (2005) and Bengtsson *et al.* (2006).

Compression ratio control with VVA for HCCI

A variable compression ratio, from the ratio as high as in the conventional diesel engine to the ratio as low as in the gasoline engine, is desirable for the HCCI diesel engine as the operating condition changes from cold start to HCCI. A lower compression ratio than that in the conventional diesel engine is usually required in HCCI in order to prevent premature or too early auto-ignition of the diesel fuel. A true variable compression ratio (VCR) system (Ryan *et al.*, 2004) with sufficiently fast response is ideal in HCCI controls at different speeds and loads. However, a VCR engine is not mature enough to be production-viable at the current time (Roberts, 2003; Cao, 2007). VVA can be used to achieve a similar effect by reducing the effective compression ratio at the price of decreased air–fuel ratio. (Additional information on VCR engine designs, mainly gasoline engines can be found in Rabhi *et al.*, 2004; Tomita *et al.*, 2007; Tsuchida *et al.*, 2007; and Kobayashi *et al.*, 2009).

VVA can change the cylinder pressure and temperature by early or late IVC and therefore increase the usable operational range of diesel HCCI. The IVC timing is adjusted to lower the compressed gas temperature near the TDC to prevent too early auto-ignition and to increase the ignition delay to enhance fuel–air mixing with prolonged premixing time to reduce smoke. However, an excessively low effective compression ratio causes high unburned HC emissions and fuel consumption. For example, the compression ratio cannot be less than 5:1, as reported by Helmantel and Denbratt (2006). Note that late IVC instead of early IVC has been successfully used in many HCCI engines (e.g., Nevin *et al.*, 2007; He *et al.*, 2008).

Residue gas control with VVA for HCCI

HCCI combustion control can be achieved by adjusting the EGR rate and temperature. The EGR gas is able to suppress the knocking combustion. Most diesel HCCI concepts use high levels of EGR (e.g., 40–60%) and EGR cooling along with low compression ratios. A large amount of hot internal EGR is helpful for the low-load operation of HCCI by heating the fresh charge and thus reducing the need for compression heating. VVA can achieve the controls of gas composition and EGR temperature via internal EGR mechanisms. VVA also offers the advantage of faster response compared to the external EGR system. The thermodynamic performance of internal EGR by using VVA for CAI gasoline engines is reviewed by Zhao (2007) and Furhapter (2007).

There are generally two types of VVA strategies that are used to obtain internal EGR: (1) exhaust gas re-breathing (i.e., re-induction); and (2) residue gas compression (i.e., trapping or retention, also called negative valve overlap with early EVC and late IVO). The re-induction strategy refers to introducing the exhaust gas from the exhaust port into the cylinder. The trapping strategy refers to purposefully trapping the residue gas in the cylinder. The losses of the re-induction strategy occur during the gas flowing process across the intake and exhaust valves and in the manifold. The trapping strategy has the following disadvantages: thermal losses through the cylinder wall, blow-by and extra pumping loss during the recompression. Previous experimental research (Caton *et al.*, 2005) revealed that the re-induction strategy gave significantly higher engine thermal efficiency and lower NO_x emissions than the trapping strategy at the same engine speed and load, and had better tolerance to the marginal combustion or misfire at the lower load. Furhapter (2007) reported that at low loads the trapping method had lower pumping loss than the re-induction method, while at high loads the re-induction method showed slightly lower pumping loss. He also concluded that for CAI gasoline engines the internal EGR supply method with the re-induction strategy via a secondary exhaust valve post-lift event during the late intake stroke is the most favorable operation strategy to ensure stable auto-ignition. The re-induction strategy with exhaust post-lift gives hotter gas than the intake pre-lift does.

In addition to the considerations of residue gas fraction and pumping loss during the gas exchange processes, the effect of residue gas temperature is also important in the internal EGR strategy for HCCI controls because it strongly affects auto-ignition. Negative valve overlap is used to initiate auto-ignition by trapping a proportion of the exhaust gas in the cylinder. Injecting diesel fuel in the negative valve overlap interval is another method of forming the homogeneous mixture to achieve HCCI combustion (Shi *et al.*, 2005). Shi *et al.* (2005) pointed out that increasing the negative valve overlap had a positive impact on the combustion stability at low loads but impair the stability at high loads. Peng *et al.* (2008) discovered that a larger negative valve overlap results in higher cylinder gas temperature, higher temperature homogeneity, stronger turbulence intensity, and improved fuel vaporization and air–fuel mixing to assist HCCI combustion.

Overall, although hot internal EGR is helpful for HCCI, it seems not likely that the hot internal EGR achieved by VVA can replace cooled external EGR in HCCI for knock intensity control (Helmantel and Denbratt, 2006).

9.9 Cylinder deactivation performance

9.9.1 Introduction of cylinder deactivation

Cylinder deactivation (abbreviated as CDA by some authors) was an old concept originated a century ago. It is also called cylinder cutout, cylinder-on-demand, displacement-on-demand, variable displacement, and variable cylinder management. It refers to the technology that deactivates fuel injection and valve operation of selected cylinders and thus flexibly varies the effective engine displacement at low loads (e.g., usually 0–5 bar BMEP).

The following can be deactivated individually or in a combined manner for selected cylinders in different cylinder deactivation methods: spark ignition or fuel injection, intake valve, exhaust valve, port or manifold flow. The simplest form of cylinder deactivation is to only turn off the spark ignition and/or fuel injection to stop the combustion process. However, the pumping loss in this method cannot be significantly reduced since the deactivated cylinders still continue consuming power to pump the air through. Moreover, the deactivated cylinders will cool quickly as the fresh air flows through them, thus creating difficulties in subsequent re-firing and emissions problems. Although disabling either the intake or exhaust valve can prevent the engine from pumping the air through, some pumping loss will still incur as the air flows through the operating valves. For example, only deactivating the intake valve causes the exhaust gas to be repeatedly inducted into and expelled out of the deactivated cylinders.

The most effective and commonly used method of cylinder deactivation to minimize pumping loss is to shut off both fuel injection and all the valves (intake and exhaust) of the deactivated cylinders (Watanabe and Fukutani, 1982; Dresner and Barkan, 1989b; Falkowski *et al*., 2004). Sandford *et al*. (1998) reported that when only the intake valve in a gasoline engine was deactivated the fuel consumption reduction was 7%. But when both the intake and exhaust valves were deactivated the fuel consumption reduction was 14–17%.

The active cylinders must have higher fueling rate than they would if the engine were run without cylinder deactivation in order to maintain the same engine brake power. The higher load (fueling rate) means higher thermal efficiency and fewer HC emissions for the cylinder. For the conventional gasoline engine, the higher load also means a higher intake manifold pressure and reduced throttling and hence lower pumping loss for the firing cylinders.

The power loss in the deactivated cylinder is low, only due to a little pumping loss of recompression, heat transfer and blow-by. In addition to the pumping loss reduction, BSFC improvement with cylinder deactivation comes from the following three other sources: (1) the decrease of the camshaft friction power used to activate the intake and exhaust valves (as measured by Watanabe and Fukutani, 1982); (2) less cylinder heat transfer loss for the whole engine because only some of the cylinders are firing; and (3) possible piston ring friction power reduction due to decreased cylinder pressure in the deactivated cylinders.

Variable engine displacement is a very attractive feature for advanced diesel engines due to its advantages of fuel economy at part load. On the other hand, design challenges in engine balance and NVH exist for cylinder deactivation due to its deviated excitation forces in the cylinders compared to those designed for the non-deactivation operation. Cylinder deactivation is normally used in the engines that have an even number of cylinders since the engine can operate in firing with half of its cylinders while retaining an even firing interval. It is usually used in six or more cylinders in order to avoid excessively high vibration. For example, in a V8 engine with the firing order 1–5–2–6–4–8–3–7, either the cylinders 1, 2, 3, and 4 or the cylinders 5, 6, 7, and 8 can be deactivated to fulfill the requirement of a constant firing interval of 180° crank angle. The constant firing interval is achieved by deactivating the two inner cylinders on one bank of the engine and the two outer cylinders on the other bank. In general, from the vibration perspective, V8, V10 or V12 engines are more suitable for cylinder deactivation than V6 or I4 engines.

Cylinder deactivation has been used in limited production engines since the 1980s (e.g., General Motor's 1981 Cadillac Eldorado V8 gasoline engine with two- and four-cylinder deactivation, denoted as V8–6–4; Mitsubishi's Orion-MD engine as presented by Fukui *et al*., 1983; and Mitsubishi's two-cylinder deactivation in an 1.6 liter I4 gasoline engine in 1992).

Almost all the published work on cylinder deactivation was conducted on the gasoline engine. Cylinder deactivation performance was investigated by Bates *et al*. (1978), Watanabe and Fukutani (1982), Dresner and Barkan (1989a, 1989b), Hatano *et al*. (1993), Sandford *et al*. (1998), Leone and Pozar (2001), and Douglas *et al*. (2005). The research conducted by Leone and Pozar (2001) provides an excellent comprehensive summary on cylinder deactivation for gasoline engines. The NVH of cylinder deactivation is discussed by Falkowski *et al*. (2004), Lee and Rahbar (2005), and Bemman *et al*. (2005). The details of design and transient controls are introduced by Bates *et al*. (1978), Kreuter *et al*. (2001), Zheng (2001), Falkowski *et al*. (2004), Stabinsky *et al*. (2007), and Rebbert *et al*. (2008). Only Watanabe and Fukutani (1982) included a brief description of the testing results of cylinder deactivation for a naturally aspirated diesel engine in their study.

9.9.2 Mechanisms and performance benefits of cylinder deactivation

Influential factors of cylinder deactivation benefits

The performance benefits and issues of cylinder deactivation compared to the non-deactivation operation are measured primarily by BSFC, drivability, vibration, noise, emissions, and exhaust temperature (for aftertreatment). The key factors in diesel engine cylinder deactivation design include the following:

- number of cylinders deactivated
- strategy of valve switching timing for deactivated cylinders, which is related to the trapped gas mass in the cylinders
- effective area of the fixed-geometry turbine used on the engine, which is related to the trade-off between the low-load and high-load operations in terms of the required air–fuel ratio in the cylinder deactivation operation at part load and the required engine power capability in the non-deactivation operation at full load
- VGT vane control if a VGT is used on the engine – VGT can better fit the needs of variable engine air flow rate caused by variable engine displacement by flexibly adjusting the turbine area
- turbine wastegate opening control capability – for example, an electronically-controlled wastegate is able to minimize the pumping loss and the air–fuel ratio at part load by fully opening the wastegate in the non-deactivation operation as a low-BSFC baseline for BSFC comparison
- acceptable air–fuel ratio for soot control in cylinder deactivation
- in-cylinder heat transfer loss
- piston ring friction caused by cylinder pressure
- valvetrain friction power consumed to drive the intake and exhaust valves.

The factors to determine the effectiveness of cylinder deactivation for fuel economy improvement include engine displacement, vehicle weight, driving cycle, and the engine and VVA designs. The commonly used test cycles or operating modes include the EPA City and Highway cycles, the New European Driving Cycle (NEDC), and the Japanese 10–15 Modes. The BSFC benefit of cylinder deactivation reaches the greatest value for high performance vehicles such as those using large engine displacement with six or more cylinders and having relatively low vehicle weight and high parasitic losses. Large engine displacement places the vehicle operating point at lower BMEP on the engine map, and therefore more benefit of BSFC with cylinder deactivation can be obtained compared to the smaller engines. The fuel economy improvement at low loads (e.g., 0–5 bar BMEP) is important for real world driving, especially in LD applications. Numerous studies (e.g., Kreuter *et al.*, 2001) show that in LD applications the majority of the operating points in the NEDC and at constant-speed cruising fall in the very low load region on the engine map (i.e., 0–5 bar BMEP, at low and medium speeds). Such a light-load condition is very suitable for cylinder deactivation.

Valve switching strategies in cylinder deactivation

There are three methods of trapping the gas mass in the deactivated cylinders, depending on the timing of shutting off the valves within an engine cycle: (1) keep the minimum gas mass and almost vacuum in the cylinder; (2) induct the cool fresh air into the cylinder; and (3) trap the hot residue gas in the cylinder. It is desirable to stop the valve motion of the selected valves at a specific point in the engine cycle with minimum impact on design, durability, packaging, and cost. The shut-off timing is critical in the way that the pressure history inside the cylinder determines the lube oil consumption and the lubrication condition of the power cylinder components, as well as engine vibration. Although a minimal amount of gas mass trapped produces the minimum piston ring friction and pumping loss of the deactivated cylinders, some gas mass and cylinder pressure need to be maintained in order to achieve an acceptable level of lube oil consumption and lubrication condition for the cylinder by minimizing the possibility of sucking lubricant oil via the piston ring gaps into the combustion chamber. In addition to lube oil consumption, another important consideration in selecting the valve switching strategy is NVH. The engine needs to compress an appropriate amount of air mass trapped in the deactivated cylinders to act as a damper to smooth the vibration and speed variation of the engine.

Rapid switching of cylinder deactivation events is important for synchronizing the cylinders in order to ensure reliable and repeatable switching between the operating modes without uncontrolled transient conditions that could disturb emissions or drivability. Electro-mechanical actuation is generally faster than electro-hydraulic actuation. Fast valve deactivation and reactivation can be realized within one engine cycle up to high engine speeds such as 5000 rpm with electro-mechanical systems in today's design (Kreuter *et al.*, 2001).

In the valve actuation sequence, either the intake or exhaust valve can be first deactivated. There are generally three valve switching strategies of deactivation:

1. Deactivate the intake valve first immediately at the end of the exhaust stroke without refilling the cylinder with fresh air and external EGR.
2. Deactivate the intake valve first during or at the end of the intake stroke to refill the cylinder with a certain amount of fresh air and external EGR.
3. Deactivate the exhaust valve first before the burned cylinder gas is expelled out of the cylinder in order to trap the hot exhaust gas in the cylinder.

In the first valve switching strategy, when the piston moves down during the intake stroke, the cylinder pressure becomes lower than the atmospheric pressure, thus creating a high vacuum with a possible negative effect on lube oil consumption. Oil consumption is caused mainly by the oil loss via the piston rings and the valve guides. The oil can be drawn up into the combustion chamber through the joint gap areas of the top compression ring and the second ring by the high vacuum generated in the deactivated cylinders during the intake stroke. Note that the conventional gasoline engine actually runs with high vacuum in the cylinders all the time at part load under throttled operation. In the deactivated cylinders shown in some studies, the peak cylinder pressure was usually very low (only 2–3 bar). The cylinder pressure at the intake BDC was only around 0.2 bar absolute pressure, and the cylinder pressure stayed below the atmospheric pressure with vacuum during the majority of the engine cycle (Hatano *et al.*, 1993; Leone and Pozar, 2001). It should be noted that the piston ring pack design may be different between the naturally aspirated gasoline engine and the turbocharged diesel engine.

Saito *et al.* (1989) confirmed in their experimental work that when the intake throttle was closed the high vacuum in the cylinder resulted in an increase in oil consumption through the piston ring pack in a gasoline engine. They found that the amount of lubricant oil drawn into the combustion chamber in the engine braking condition with the intake throttle closed was about six times as much as that at high load with the intake throttle fully open. The oil drawn was evidenced by a large amount of oil gathered in the top land above the top ring joint and in the second and third lands, as well as a thick oil film on the piston skirt. As the engine speed increased, the oil flow was drawn up faster.

In the third valve switching strategy where the exhaust valve is deactivated before the intake valve, the deactivated cylinder has a warmer charge due to the trapped hot exhaust gas. The charge is gradually cooled down as heat transfer occurs from the gas to the cylinder walls. This strategy may result in excessively high peak cylinder pressures in the deactivated cylinders at relatively higher engine load during the first few seconds immediately after the valve deactivation. The cylinder pressure quickly decays cycle by cycle due to the heat transfer loss within several seconds to a much lower stabilized level (Sandford *et al.*, 1998), which is slightly higher than the pressure level given by the second valve switching strategy. It should be noted that recompressing the trapped gas in the deactivated cylinders produces some pumping loss. The pumping loss of trapping hot combustion gas is higher than that of trapping the fresh air.

A good guideline for the deactivation sequence is to find the best balance among the following: avoiding undesirable vacuum in the cylinder, keeping the cylinder as warm as possible, maintaining acceptable peak cylinder pressure for durability and engine vibration, and minimizing the pumping loss. In many gasoline engines, the exhaust valve was often deactivated and reactivated first before the intake valve (e.g., Falkowski *et al.*, 2004).

The reason is that the gasoline engine runs at high vacuum in the cylinder during the intake stroke at part load with intake throttle. Therefore, if a large amount of mass (or especially hot gas mass) is desired, shutting off the exhaust valve first is the only solution. In the diesel engine, the situation is different because it does not run with vacuum under throttled operation. Therefore, the second valve switching strategy of first shutting off the intake valve during the intake stroke to induct cold charge can be considered.

Normally, during the extended operation of cylinder deactivation the cylinders need to be reactivated for a short period of time in order to keep the cylinders warm and prevent the build-up of oil deposits in the combustion chamber (Falkowski *et al.*, 2004). When reactivating the cylinder, the exhaust valve usually needs to be opened before the intake valve is opened. Reactivating the exhaust valve first can avoid the exhaust gas from the active cylinders being pushed back into the intake port to cause noise problems and misfire during the following combustion cycle (Kreuter *et al.*, 2001). Reactivating the exhaust valve first can also avoid high recompression pressures acting on the intake valvetrain. Moreover, the cylinders can be alternatively activated or deactivated in order to keep the deactivated cylinders warm or prevent them from being excessively cooled down.

BSFC improvement of cylinder deactivation

In the gasoline engine, indicated specific fuel consumption (ISFC) increases slightly when the engine load decreases, while BSFC increases sharply due to the large increase in pumping loss associated with intake throttling and the larger proportion of mechanical friction (Bates *et al.*, 1978). Even simple fueling cut-off in the deactivated cylinders can improve fuel consumption because it increases the fueling in the firing cylinders so that the intake throttle opening is increased and the pumping loss is reduced. In the diesel engine, the effect of fuel injection cut-off is negligible (Watanabe and Fukutani, 1982).

Because a throttle valve is not necessary for load control in the diesel engine, the effect of cylinder deactivation was believed to be less than that in the gasoline engine by some people in the diesel engine industry. It was believed that the benefit would only come from the reduction in valvetrain friction power and fuel injection parasitic losses. In fact, there is a significant benefit in fuel economy with cylinder deactivation in the diesel engine, as shown in the simulation results presented later. Watanabe and Fukutani (1982) reported that when half of the cylinders were shut off by deactivating the valves in a diesel engine in their experimental work, 30% of BSFC reduction was achieved under low idle conditions. It should be noted that if an inappropriate fuel injection compensation strategy is used at low idle to compensate for the larger vibration amplitude and the resulting instantaneous speed reduction caused by cylinder deactivation, the vibration and engine speed instability can become worse and the BSFC reduction can be very little. In such a case, sometimes the BSFC can become even worse with cylinder deactivation.

The BSFC reduction mechanisms in the deactivated cylinders are similar or the same between the gasoline engine and the diesel engine. But the mechanisms of pumping loss reduction in the firing cylinders are different. In the naturally aspirated gasoline engine, the pumping loss is reduced via an increased intake manifold pressure due to less throttling. In the turbocharged diesel engine, the pumping loss is reduced due to the lower exhaust manifold pressure caused by the drastically reduced engine air flow rate flowing through the given turbine area. The reduction in air flow rate is caused by the reduced effective engine displacement as shown in equation 9.30.

Leone and Pozar (2001) found that a universal steady-state correlation of the percentage benefit of BSFC exists as a strong function of BMEP (in the range of 0–5 bar) for different gasoline engines with cylinder deactivation. The percentages of BSFC gains are all almost the same at the same BMEP. They found the BSFC benefit had little sensitivity to engine speed. The average benefits of fuel consumption reduction at the steady-state operation across the speed range in the gasoline engine cylinder deactivation are summarized below based on the reported data in the literature: 30–40% BSFC reduction at low idle; 22% at 1 bar BMEP; 15% at 2 bar BMEP; 10% at 3 bar BMEP; 6% at 4 bar BMEP; and 0% at 5 bar BMEP. The fuel economy improvement by cylinder deactivation in vehicle driving cycles ranges from 6% to 15% with an average of 10%, depending on vehicle weight, engine size, and design constraints such as NVH, gear numbers used for deactivation, the usage of cylinder deactivation at low idle, and the low idle time. Leone and Pozar (2001) and Sandford *et al*. (1998) reported large variations in the percentage benefit of BSFC for different vehicle applications and driving cycles. The fuel economy benefit varies significantly between different test cycles, especially between the city and highway driving cycles. City driving usually has a greater BSFC benefit than highway driving. If the driving cycle has a large proportion of part-load operating conditions, the BSFC benefit of cylinder deactivation will become more prominent.

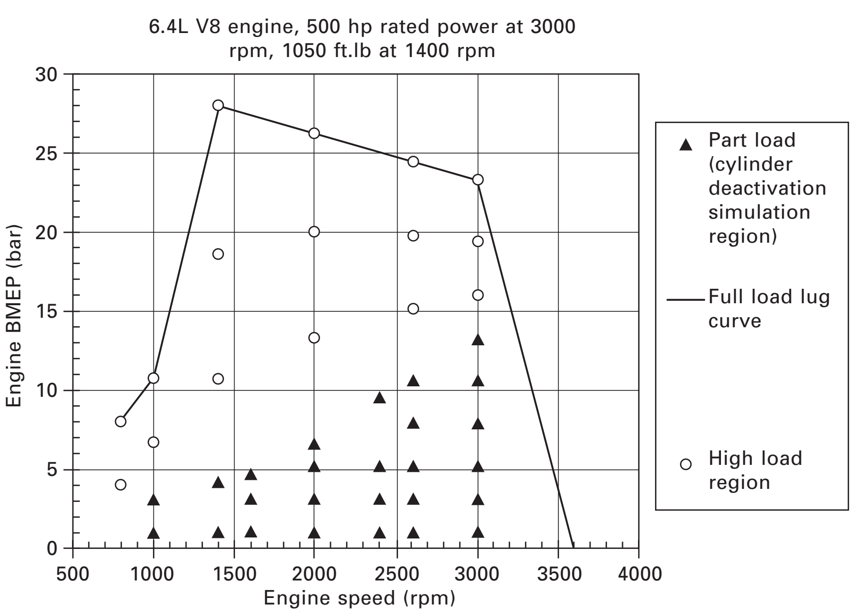
Emissions and exhaust temperature in cylinder deactivation

The cylinder deactivation in the gasoline engine tends to increase brake specific NO_x and reduce HC and CO emissions. HC spikes exist at the transition from the deactivation to the activation modes. Very little data exist in the published literature about the emissions in diesel engine cylinder deactivation. The BMEP level in cylinder deactivation in the diesel engine is primarily limited by soot emissions when the air–fuel ratio becomes low in the firing cylinders.

Another major benefit of cylinder deactivation by VVA is the great increase in the exhaust gas temperature for aftertreatment operation at light loads, low idle and warm-up without significant BSFC penalty. Cylinder deactivation is very effective in increasing the exhaust temperature to improve the light-off time for the catalysts such as SCR during warm-up. Although cylinder deactivation itself is usually not sufficient to quickly light off most aftertreatment systems, a combined means with early or late IVC timing and retarded fuel injection timing can be used to further increase the exhaust temperature.

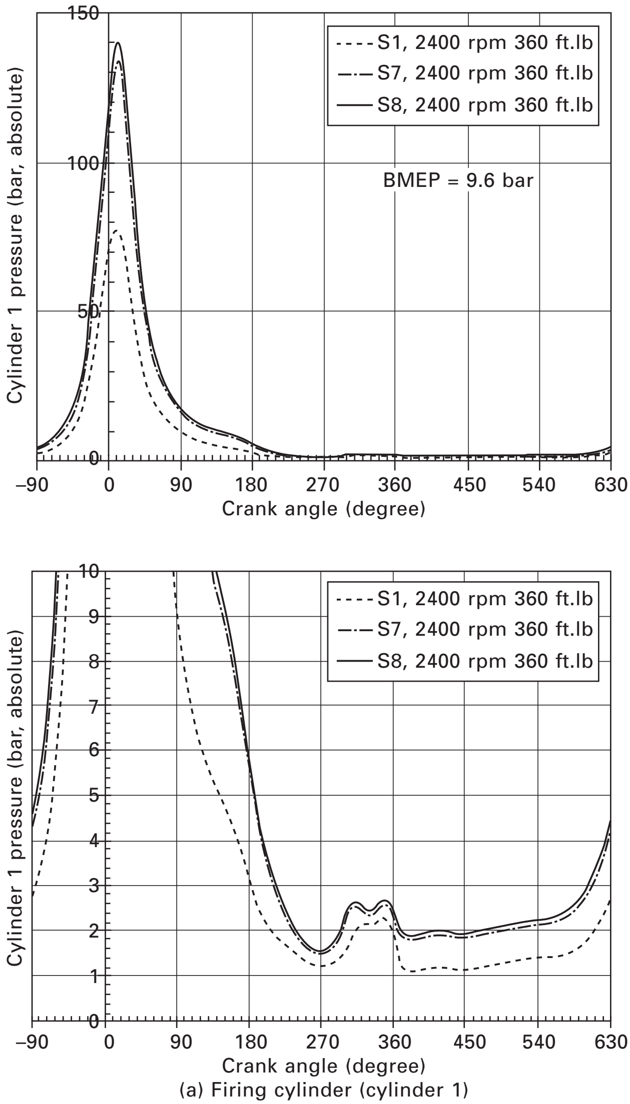
9.9.3 Diesel cylinder deactivation performance simulation

The benefit of cylinder deactivation in turbocharged diesel engines has not been published to date. Figures 9.23–9.29 present a comprehensive simulation analysis of cylinder deactivation for a V8 heavy-duty diesel engine with high power density. There are fourteen scenarios (S1–S14) representing different conditions in engine load, fixed geometry turbine area, wastegate opening, valve deactivation and shut-off timing strategies, as listed in Table 9.2. In the analysis, both the BMEP and PMEP are defined for the entire engine,

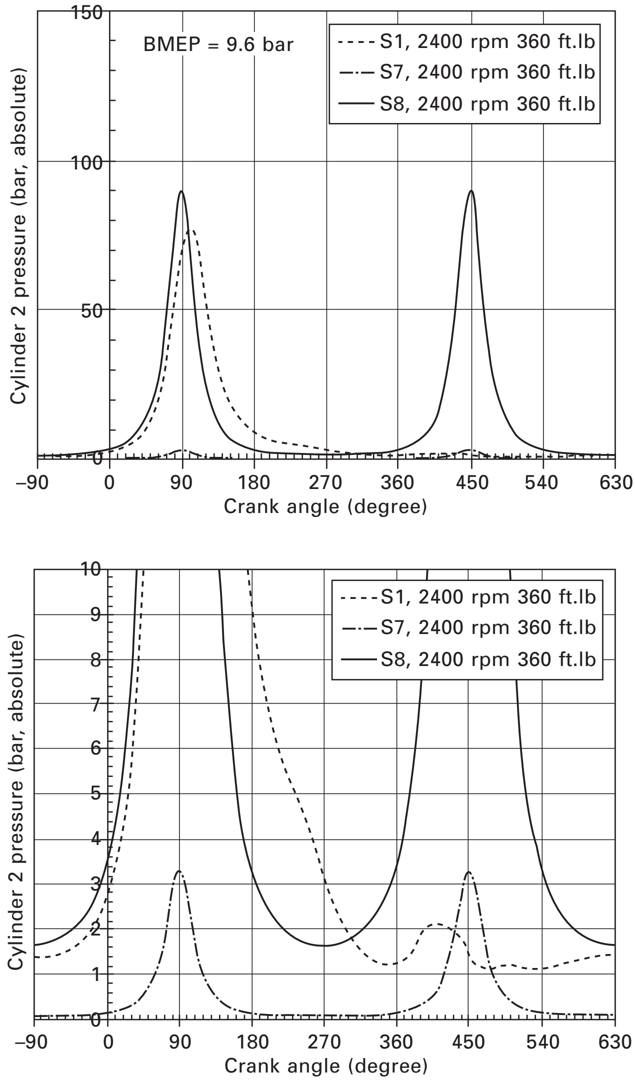


9.23 Engine speed–load domain in cylinder deactivation analysis.

i.e., defined as the work per cycle divided by the total engine displacement. In the simulation, four cylinders are deactivated (i.e., the V4 operation). A two-stage fixed geometry turbine system is used in the simulation, and there is a wastegate valve in the high-pressure (HP) stage turbine. EGR is assumed to be zero when the fixed-geometry turbine area is selected to



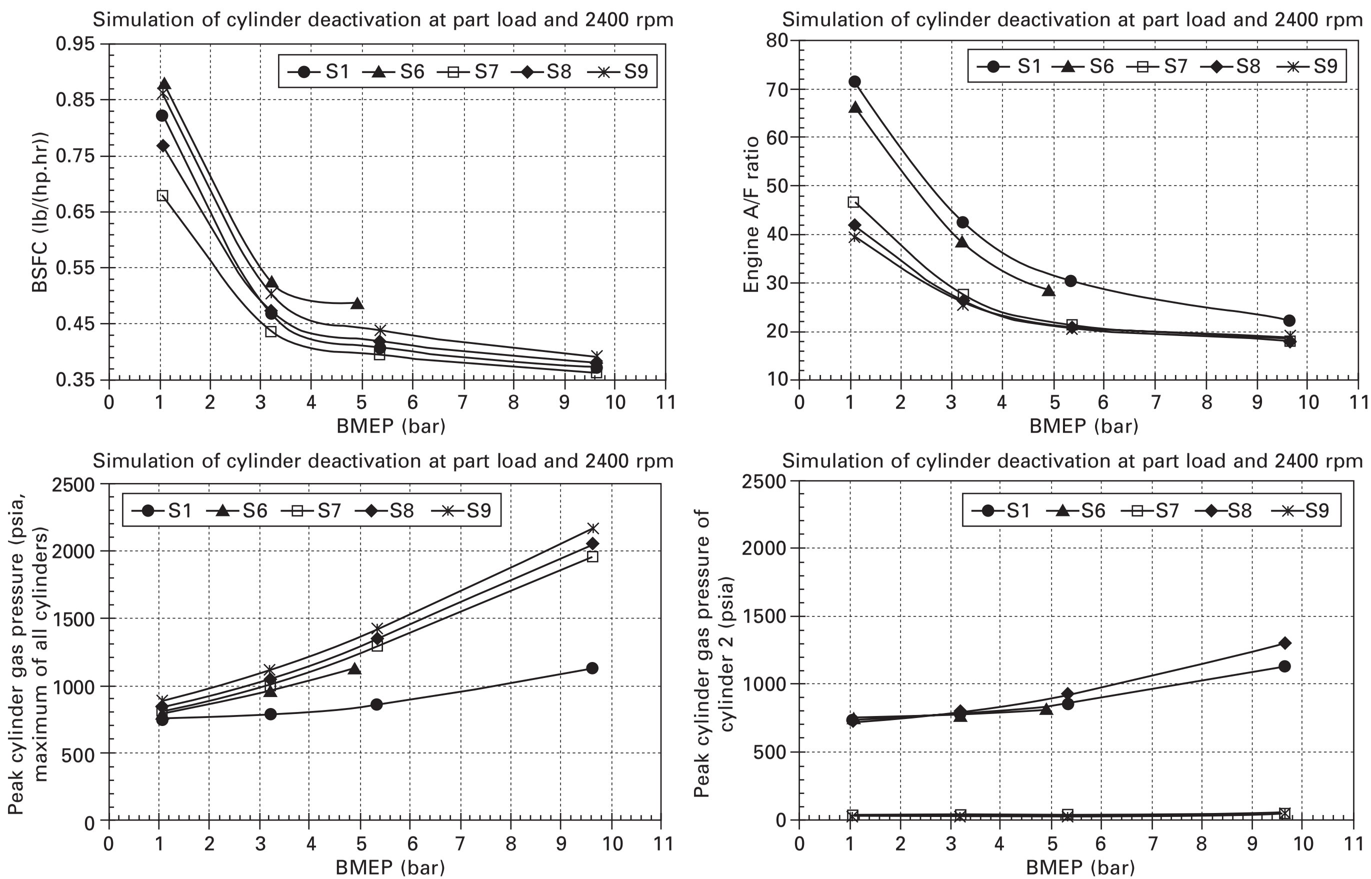
9.24 Effect of cylinder deactivation valving strategy on cylinder pressure.



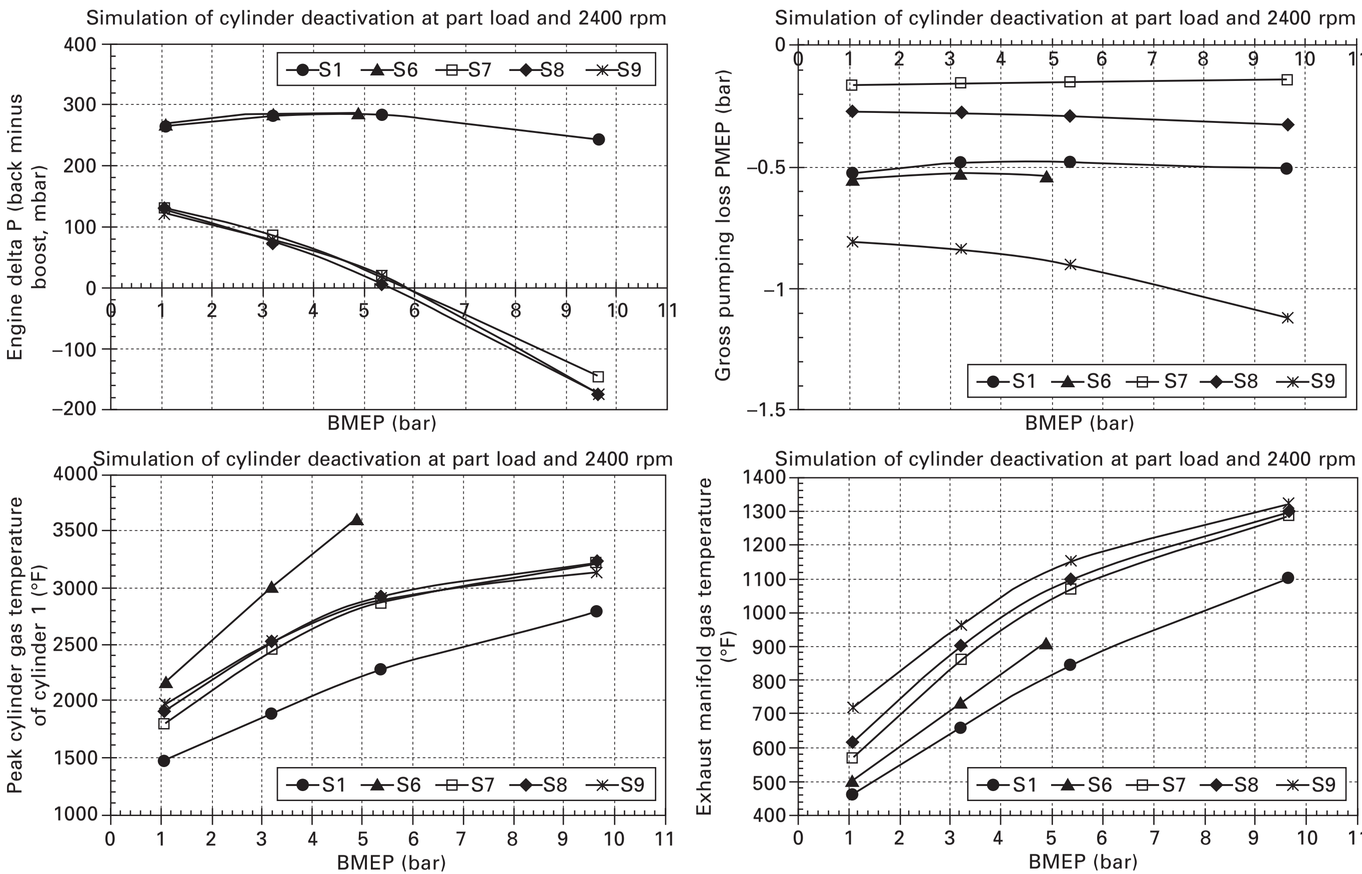
(b) Deactivation cylinder (cylinder 2)

9.24 Continued

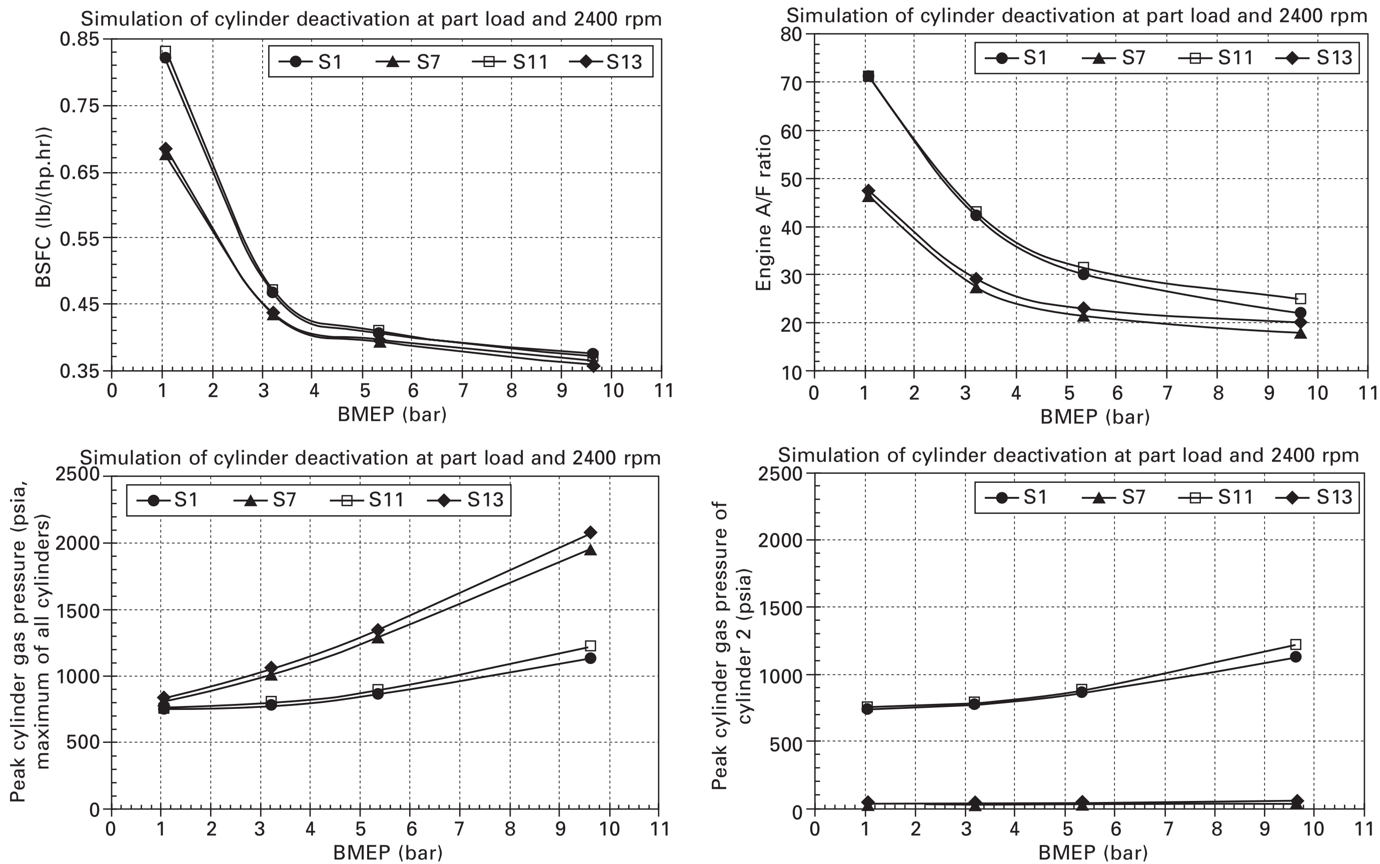
deliver sufficient air–fuel ratio for both peak torque and rated power in the V8 operation. If a certain amount of EGR is needed to reduce NO_x, a smaller turbine can be selected to deliver sufficient air–fuel ratio at both full load and part load at the given required EGR rate with a resulting higher engine delta P and BSFC than the zero-EGR case. Therefore, with an appropriate turbine area matched with an assumed EGR rate, the cylinder deactivation analysis conducted here is directionally valid for other EGR rates.



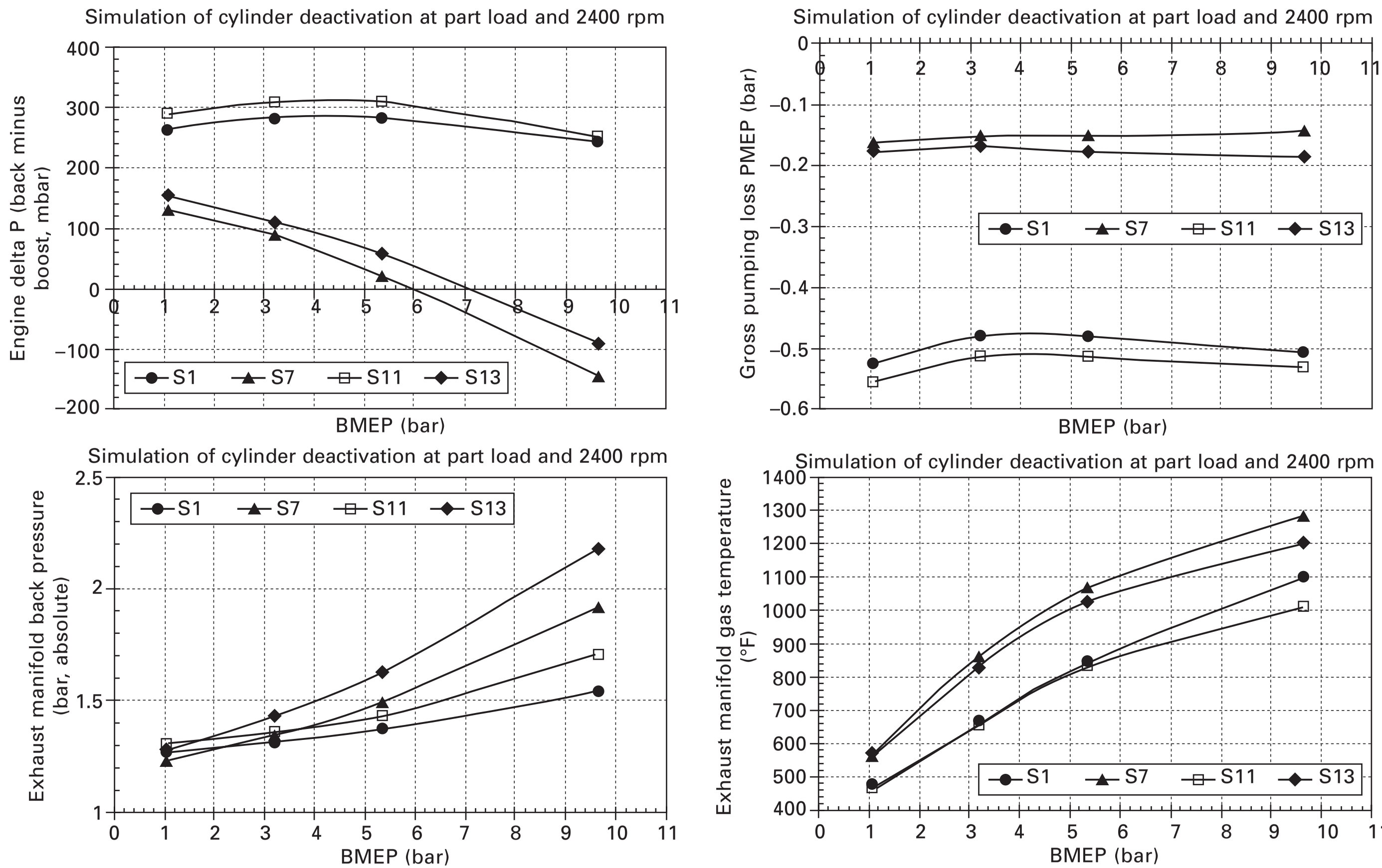
9.25 Investigation of cylinder deactivation strategies.



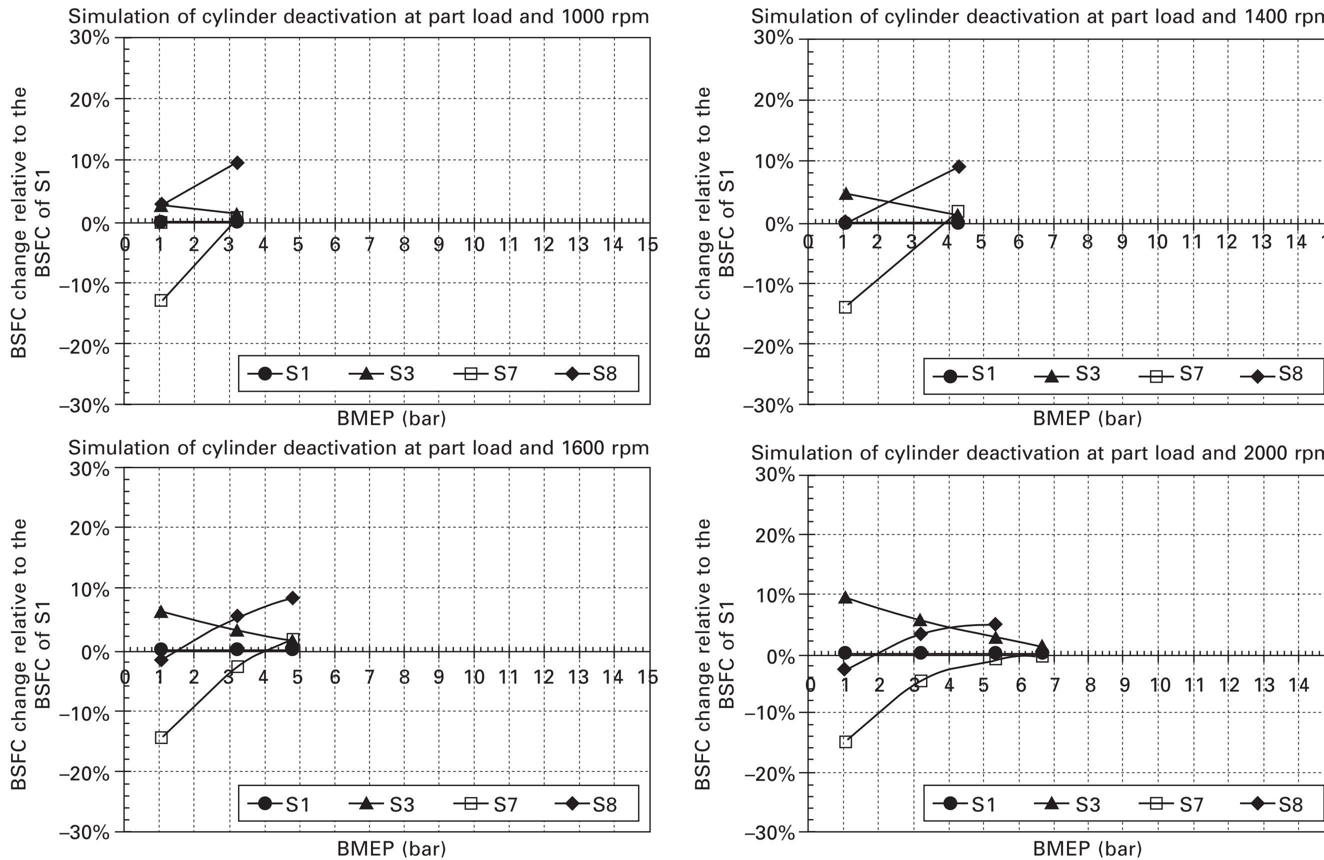
9.25 Continued



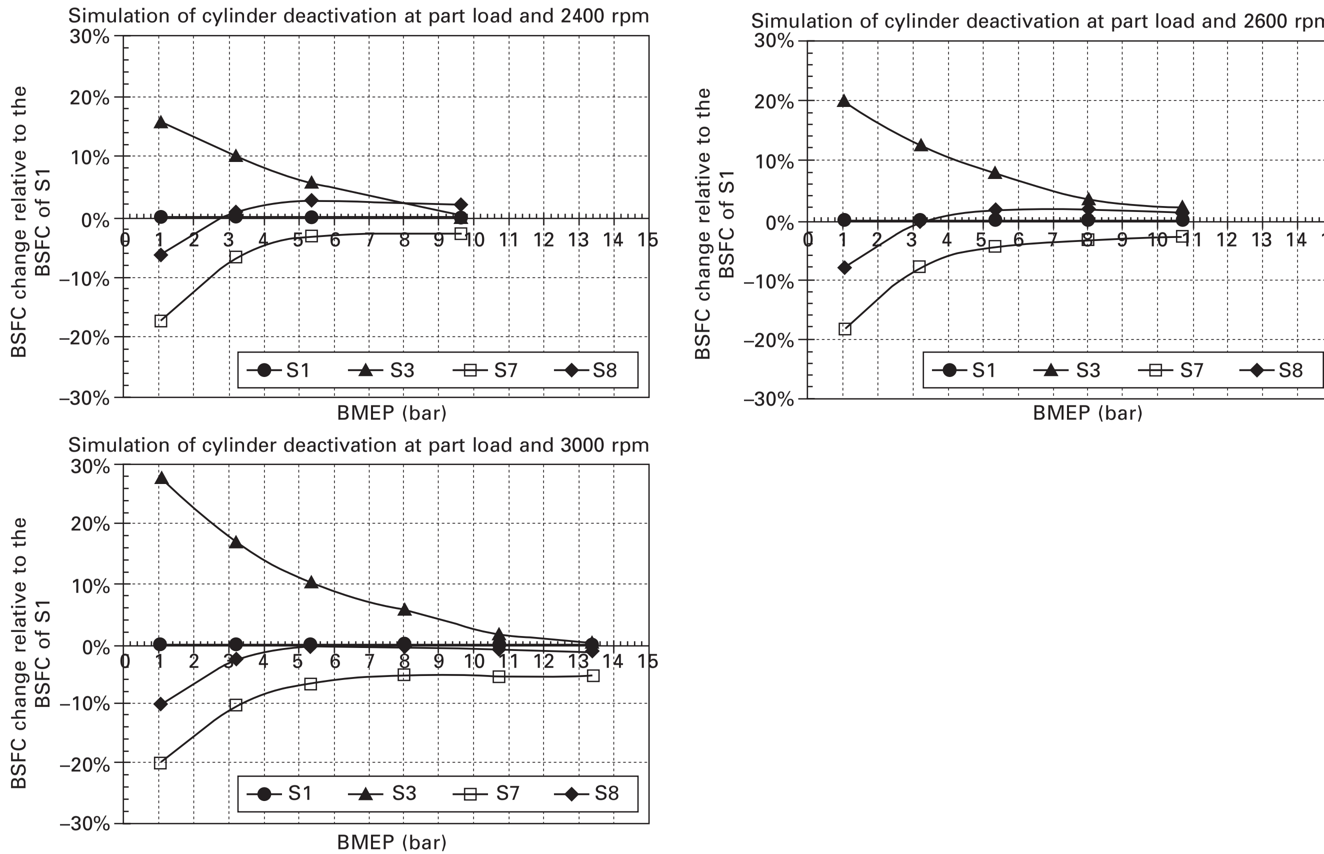
9.26 Effect of turbine area on cylinder deactivation.



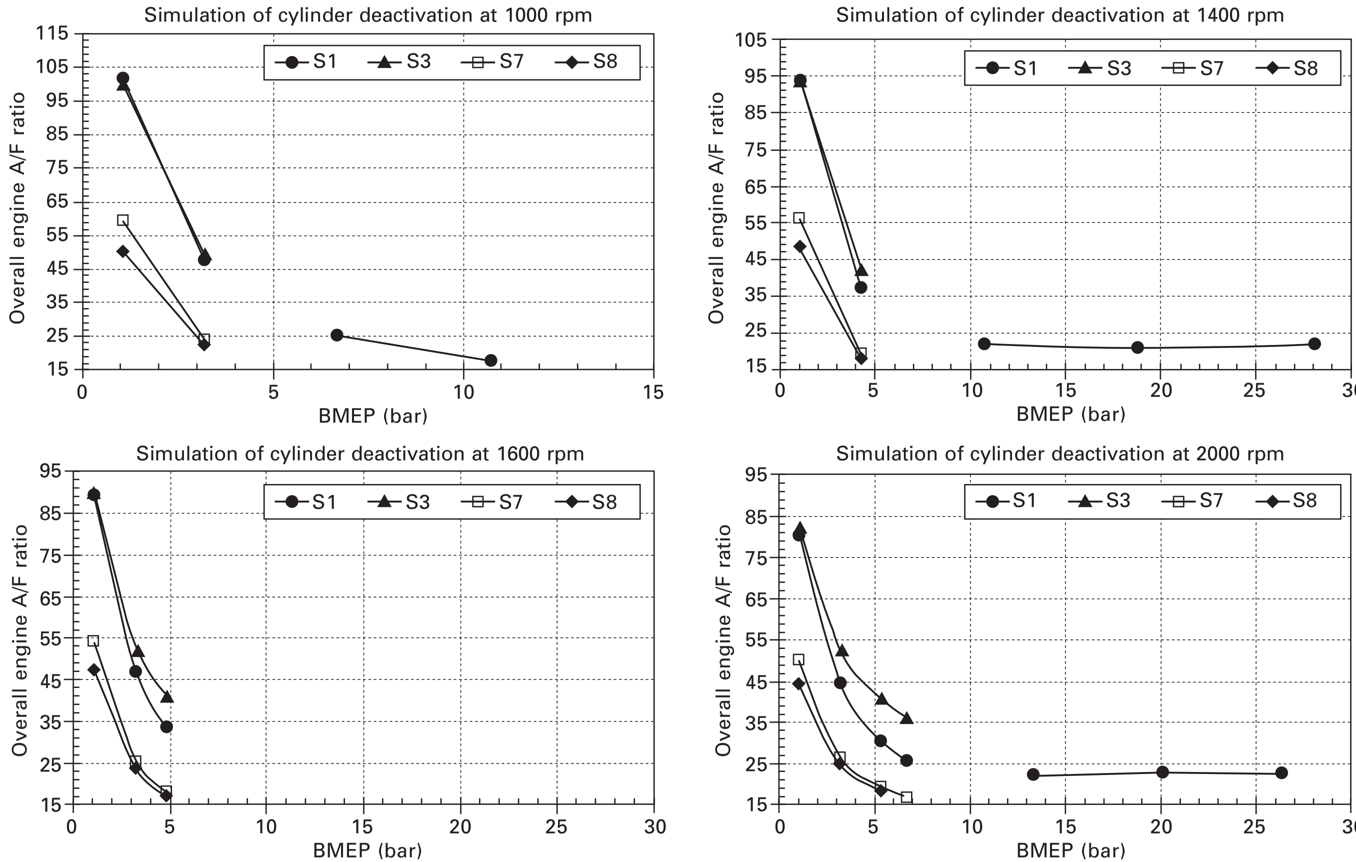
9.26 Continued



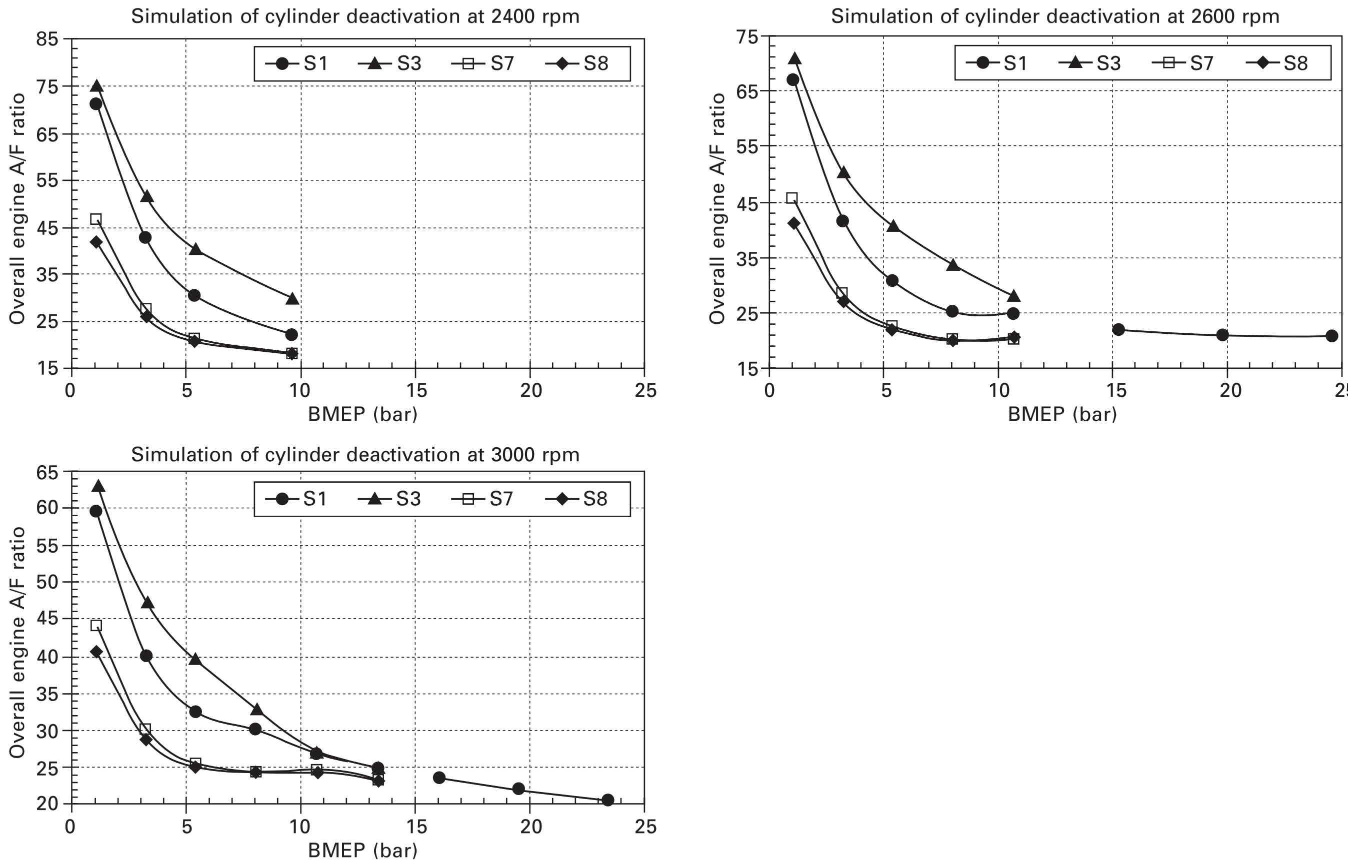
9.27 BSFC of cylinder deactivation at different engine speeds and loads.



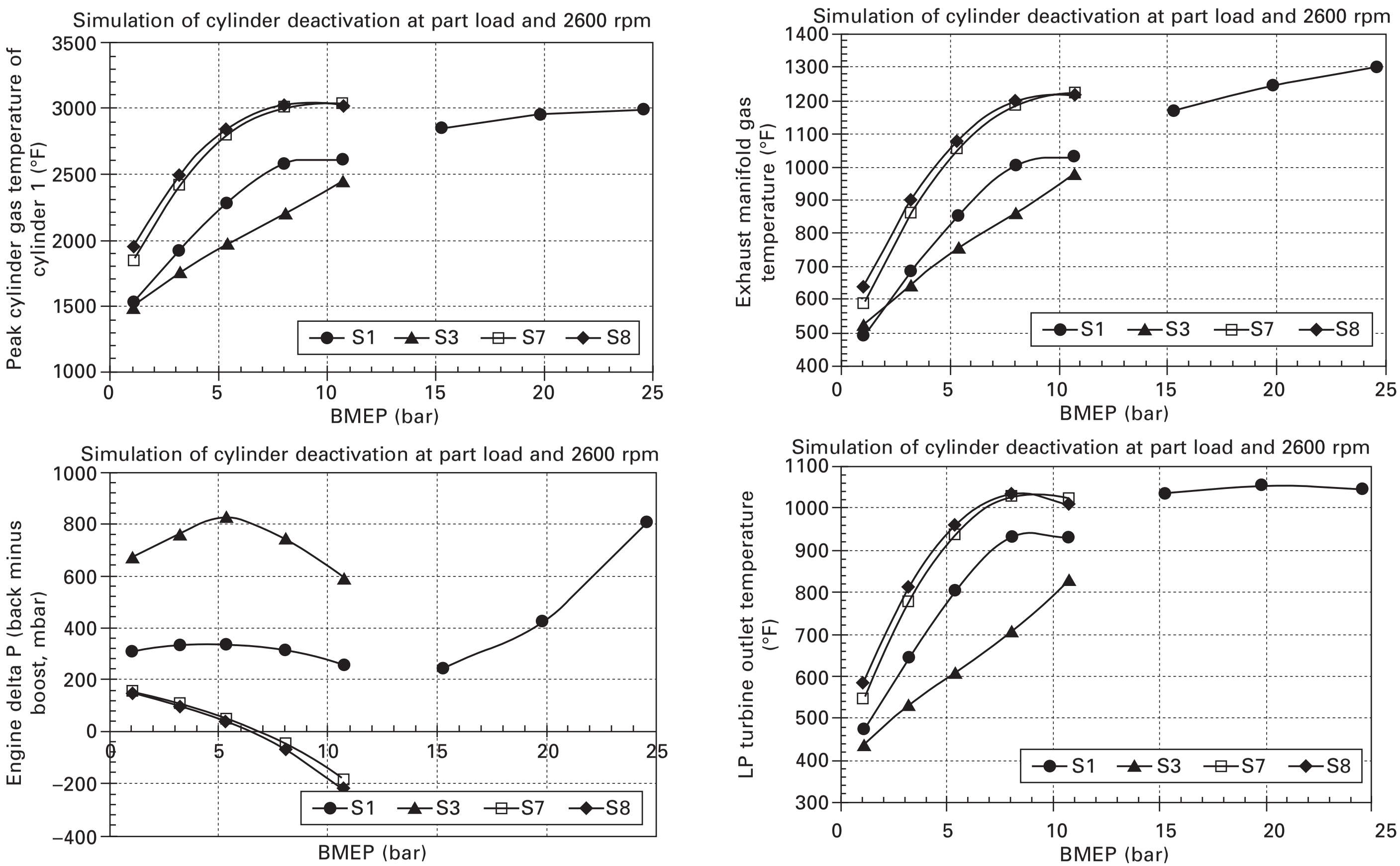
9.27 Continued



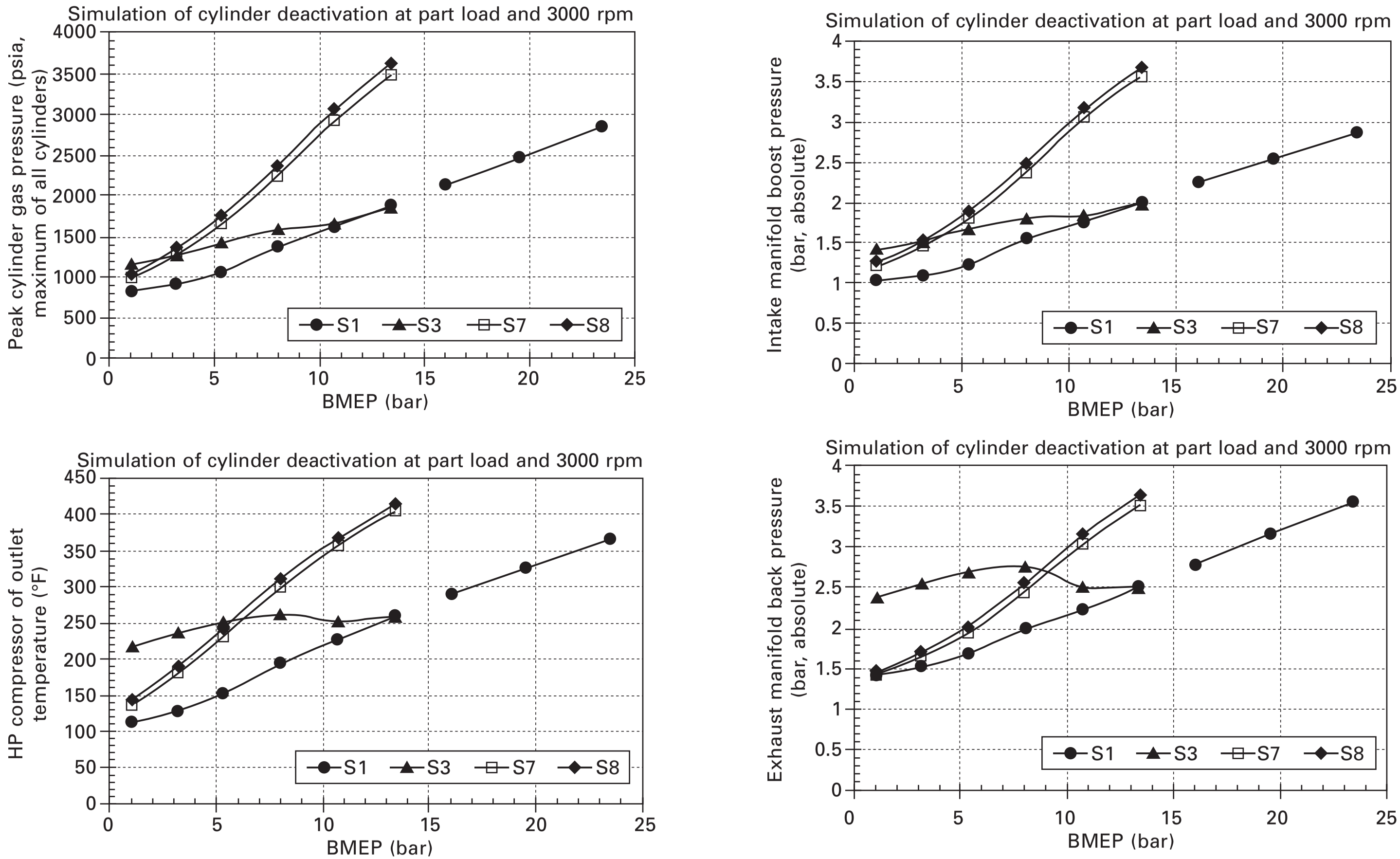
9.28 Air–fuel ratio of cylinder deactivation at different engine speeds and loads.



9.28 Continued



9.29 Cylinder deactivation performance at 2600 rpm and limiting constraints at 3000 rpm.



9.29 Continued

Table 9.2 Simulation cases of cylinder deactivation

Scenario number	Mode	Turbine area	HP-stage turbine wastegate opening	Valvetrain operation	Strategy of engine valve shut-off timing for deactivated cylinders
S1	Part load	Baseline turbine	Fully open	No cylinder deactivation	
S2	High load	Baseline turbine	Open as needed	No cylinder deactivation	
S3	Part load	Baseline turbine	Fully closed	No cylinder deactivation	
S4	High load	Baseline turbine	Closed	No cylinder deactivation	
S5	Part load	Baseline turbine	Partially closed	No cylinder deactivation	
S6	Part load	Baseline turbine	Fully closed	Cylinder deactivation by shutting off fuel only	
S7	Part load	Baseline turbine	Fully closed	Cylinder deactivation by shutting off fuel, intake valve and exhaust valve (shut-off strategy 1)	Shut-off strategy 1: Shut off intake valve before each cylinder's IVO
S8	Part load	Baseline turbine	Fully closed	Cylinder deactivation by shutting off fuel, intake valve and exhaust valve (shut-off strategy 2)	Shut-off strategy 2: Shut off intake valve after each cylinder's IVC

S9	Part load	Baseline turbine	Fully closed	Cylinder deactivation by shutting off fuel and only intake valve (shut-off strategy 1)	Shut-off strategy 1: Shut off intake valve before each cylinder's IVO
S10	Part load	Baseline turbine	Fully closed	Cylinder deactivation by shutting off fuel and only intake valve (shut-off strategy 2)	Shut-off strategy 2: Shut off intake valve after each cylinder's IVC
S11	Part load	10% smaller turbine area in both HP and LP stages	Fully open	No cylinder deactivation	
S12	High load	10% smaller turbine area in both HP and LP stages	Open as needed	No cylinder deactivation	
S13	Part load	10% smaller turbine area in both HP and LP stages	Fully closed	Cylinder deactivation by shutting off fuel, intake valve and exhaust valve (shut-off strategy 1)	Shut-off strategy 1: Shut off intake valve before each cylinder's IVO
S14	Part load	10% smaller turbine area in both HP and LP stages	Fully closed	Cylinder deactivation by shutting off fuel, intake valve and exhaust valve (shut-off strategy 2)	Shut-off strategy 2: Shut off intake valve after each cylinder's IVC

Note: S9 and S10 have the same steady-state performance.

Figure 9.23 illustrates the speed–load modes used in the GT-POWER simulation. Note that usually the load range in cylinder deactivation research is 0–5 bar BMEP, but this simulation extends the load range up to 10–13 bar BMEP at high speeds greater than 2500 rpm. Figure 9.24 shows the cylinder pressure traces of different valve switching strategies in deactivation.

Figure 9.25 compares the engine performance given by different deactivation strategies at 2400 rpm. It can be seen that, compared to the V8 operation (S1), shutting off only fueling (S6) and shutting off only fueling and intake valve (S9) do not provide any benefit in BSFC reduction. In fact, the BSFC of S6 and S9 is even higher than that of S1. In S7 and S8, the exhaust valve is shut off after the intake valve is shut off. Some BSFC benefit occurs at very low load (0–2 bar BMEP) when both the intake and exhaust valves are closed and a certain amount of air is trapped in the deactivated cylinders (S8). The largest benefit of BSFC reduction occurs with scenario S7 in a wide range of BMEP (0–10 bar) when very little air is trapped in the deactivated cylinders.

Figure 9.26 illustrates the effect of 10% smaller turbine area on the air–fuel ratio. The smaller turbine can increase the air–fuel ratio in the cylinder deactivation operation in order to extend its operating range to higher BMEP level. However, the turbine area reduction in a fix-geometry turbine that is sized for achieving higher air–fuel ratio during cylinder deactivation results in a pumping loss penalty in the non-deactivation operation due to the excessively high air–fuel ratio. VGT can help alleviate this trade-off between the cylinder deactivation operation at low loads and the non-deactivation operation at full load.

Figure 9.27 summarizes the BSFC changes at different speeds and loads for different turbine wastegate openings (S1 and S3) and deactivation valve switching strategies (S7 and S8). The lower and upper bounds of the BSFC reduction benefit range bounded by S7 and S8 indicate the two extreme limits of different valve switching strategies. It is observed that the BSFC benefit is a strong function of both engine load and speed for the diesel engine. Very large benefit of BSFC reduction can be obtained by cylinder deactivation. Note that the BSFC benefit computed here is with respect to a baseline engine that already has very low BSFC with the turbine wastegate set fully open at part load by using electronic controls (i.e., S1 rather than S3). At high engine speeds, the BSFC benefit can extend to a very high BMEP level with acceptable air–fuel ratio (e.g., 10 bar BMEP at 2600 rpm) because the turbocharger can deliver more air than at lower speeds to maintain a minimum required air–fuel ratio.

Figure 9.28 summarizes the air–fuel ratios corresponding to the cases in Fig. 9.27. Note that due to reduced air–fuel ratio the cylinder deactivation operation generally will produce more soot than the non-deactivation operation, especially when the air–fuel ratio is close to the smoke limit. Other measures (e.g., retarding EVO and advancing IVC with VVA in the firing cylinders) can be used to reduce the soot at relatively high BMEP level and at low to medium speeds.

Figure 9.29 illustrates other key performance parameters at 2600 rpm and the limiting design constraints at 3000 rpm in cylinder deactivation. It is important to note the effect of cylinder deactivation on the following two parameters: (1) engine delta P, which is the driving force for EGR flow in EGR engines and can also be affected by the turbine area selection; and (2) the low-pressure (LP) stage turbine outlet gas temperature, which can help the operation and regeneration of aftertreatment devices (e.g., DPF). Finally, note that the BMEP limit of cylinder deactivation can be restricted by the peak cylinder pressure in the firing cylinders and the compressor outlet air temperature at high engine speeds (e.g., 3000 rpm), in addition to the minimum required air–fuel ratio.

The performance of diesel engine cylinder deactivation is summarized as follows.

1. The optimum cylinder deactivation method is to shut off fueling and all the intake and exhaust valves in the deactivated cylinders.
2. The deactivation valve switching timing and gas mass retained in the deactivated cylinders greatly affect the BSFC benefit of the deactivation.
3. The optimum number of deactivated cylinders largely depends on the balance between the air–fuel ratio for combustion/emissions and the pumping loss for the lowest BSFC. The optimum number is also affected, although secondarily, by the design features of cylinder heat transfer loss and piston ring friction of the engine. Deactivating half of the cylinders may not always be the optimum for achieving the lowest BSFC.
4. Cylinder deactivation provides very large BSFC improvement under low-load conditions. The benefit of BSFC reduction strongly depends on both BMEP and engine speed for turbocharged diesel engines.
5. BSFC improvement is difficult to achieve with cylinder deactivation at medium to high loads due to the limit of air–fuel ratio at low to medium speeds and due to other design constraints at high speeds such as peak cylinder pressure and compressor outlet air temperature.
6. The extent of BSFC improvement by cylinder deactivation is dependent upon the following factors: the number of cylinders deactivated, valve shut-off strategy, valve switching timing, turbine area (variable or fixed, and the area size), turbine wastegate control capability and flexibility at low loads, cylinder heat transfer, and the pressure-dependent portion of the piston ring friction.
7. As the cylinder heat transfer and piston ring friction loss increase in the baseline engine design, more benefit of BSFC reduction can be achieved

by using cylinder deactivation because a larger portion of the energy losses can be avoided in the deactivated cylinders.

9.9.4 Design challenges of cylinder deactivation

Engine vibration

NVH and drivability are important concerns for cylinder deactivation. Cylinder deactivation reduces the frequency and increases the amplitude of engine vibration at the crankshaft, which may not be acceptable at all engine speeds in terms of NVH and comfort. The success of cylinder deactivation and the benefit in fuel economy improvement can only be achieved if the NVH issues are resolved in mass production vehicles. As pointed out by Leone and Pozar (2001), the benefit of BSFC reduction in cylinder deactivation is constrained or limited by NVH. For example, NVH and drivability concerns are particularly severe in the first and second gears of the transmission. If cylinder deactivation is constrained only to the gears higher than the first and second, 2–4% of BSFC reduction benefit will not be gained. If cylinder deactivation is not used during warm-up, another 1–2% of BSFC reduction benefit will not be gained (Leone and Pozar, 2001). The major challenge of cylinder deactivation is to maximize the fuel economy benefit while meeting the requirements of NVH and drivability in both steady-state operation and transient transitioning of turning the cylinders on and off (i.e., smooth ramp-out and ramp-in behavior).

The gas flow through the firing cylinders and the gas pressure pulses in the manifolds affect the intake and exhaust noises. The cylinder pressure in the deactivated cylinders plays an important role in engine vibration. As long as the cylinder pressure is not uniform between the cylinders, engine vibration and speed variability will increase. The cylinder pressure is affected by the valve switching strategy in deactivation and is dependent on the air or gas mass retained in the deactivated cylinders. The regular compression without shutting off the valves in the non-fueling cylinders balances better the engine vibration caused by the firing cylinders. With the extreme case of vacuum compression with both the intake and exhaust valves shut off and without air mass retained, the engine speed will exhibit higher variability due to more violent vibration and torque pulsation at the crankshaft. The higher dynamic torque and the lower frequency associated with the halved basic engine order in cylinder deactivation cause the vibration amplitude of the powertrain and drivetrain to increase (Lee and Rahbar, 2005). In particular, cylinder deactivation used at low idle may create the problem of excessive NVH due to the very low frequency and high amplitude of engine vibration at low speed. Watanabe and Fukutani (1982) found that the amplitude increase of engine vibration in diesel cylinder deactivation at low idle was much larger than that in gasoline cylinder deactivation. The much higher peak cylinder pressure in the diesel engine was believed to be the reason.

Potential design improvement in vibration control for cylinder deactivation includes optimizing the engine mounts, using additional torsional dampers or isolators in the driveline, increasing the inertia of the rotating engine components, and increasing torque converter slip (Hatano *et al*., 1993; Falkowski *et al*., 2004), as well as active vibration control measures such as active tuned absorbers or active engine mounts (Lee and Rahbar, 2005). However, all these measures have some negative effects. In general, there is no easy solution for the vibration problem caused by cylinder deactivation. New technologies are needed to solve the NVH and fuel injection compensation problems encountered in cylinder deactivation.

Exhaust noise

The engine operating with different number of active cylinders in cylinder deactivation exhibits different exhaust sound characteristics due to the change in exhaust pressure waves. The design objective for cylinder deactivation is to avoid perceivable subjective changes in the tailpipe acoustics. The traditional active noise control device for this problem in the gasoline engine is usually to use a flap valve located in the exhaust pipe. The more advanced passive noise control design is to use passive resonators to attenuate the noise during the cylinder deactivation operation. Bemman *et al*. (2005) successfully used Helmholtz resonators to attenuate the low frequencies inherent to the V4 cylinder deactivation mode for a V8 gasoline engine. Their system is more cost-effective than the traditional approach of using active valves in the exhaust system to tune the exhaust noise.

9.9.5 Cylinder deactivation matched with other technologies

Cylinder deactivation is essentially a technology of variable engine displacement achieved by VVA to try to optimize fuel consumption at low loads and obtain high power at high loads. A fixed cylinder displacement downsizing via turbocharging does not have the NVH issues brought by cylinder deactivation, and may also produce lower parasitic losses. However, the displacement reduction in fixed cylinder displacement downsizing is severely limited by the achievable power density at full load.

The competing engine technologies for cylinder deactivation in pumping loss reduction or fuel economy improvement include VVA, VGT, HCCI/CAI, continuously variable transmission, hybrid electric, and hybrid hydraulic. Both HCCI/CAI and cylinder deactivation have the potential to decrease fuel consumption at light loads. Note that the primary objective of diesel HCCI is to reduce NO_x and soot emissions to eliminate NO_x aftertreatment, while the primary goal of gasoline CAI is to reduce fuel consumption.

The synergy between CAI and cylinder deactivation (with conventional spark ignition combustion in the firing cylinders) for an I4 gasoline engine was researched by Douglas *et al*. (2005). In their cylinder deactivation, the fueling was shut off for two cylinders, the fresh air was blocked at the intake manifold, and the engine intake and exhaust valves were operated with low-lift cam profiles to pump the hot exhaust gases. The CAI was operated in the low to medium load region, while the cylinder deactivation was operated to cover the even lower load region. They concluded that CAI and cylinder deactivation were complementary for the gasoline engine so that they could significantly improve the fuel consumption over a broadened area on the engine map. In their study, for the city driving or light-duty applications, a majority of the time (75%) in the driving cycle was spent in the low load and idle region where CAI and cylinder deactivation were used to reduce fuel consumption by 10.2% and reduce NO_x emissions by 28%. They predicted 4.9% reduction in fuel consumption (i.e., from 33 mpg to 34.7 mpg) during the New European Driving Cycle (NEDC) for a 1.6 liter gasoline engine in a light-duty vehicle when CAI was used. They also predicted 10.2% reduction in fuel consumption (i.e., from 33 mpg to 36.7 mpg) when both CAI and cylinder deactivation were used. It is expected that HCCI and cylinder deactivation can also be complementary for the diesel engine in order to maximize the fuel economy improvement in a broadened load range on the engine map from the very low load to the medium load.

9.10 References and bibliography

Abell R F (1969), 'The operation and application of hydraulic valve lifters', SAE paper 690347.

Ahmad T and Theobald M A (1989), 'A survey of variable-valve-actuation technology', SAE paper 891674.

Akiba K and Kakiuchi T (1988), 'A dynamic study of engine valving mechanisms: determination of the impulse force acting on the valve', SAE paper 880389.

Akiba K, Shimizu A and Sakai H (1981), 'A comprehensive simulation of high speed driven valve trains', SAE paper 810865.

Al-Sarkhi A, Akash B A, Jaber J O, Mohsen M S and Abu-Nada E (2002), 'Efficiency of Miller engine at maximum power density', *International Communications in Heat and Mass Transfer*, 29(8), 1159–1167.

Al-Sarkhi A, Jaber J O and Probert S D (2006), 'Efficiency of a Miller engine', *Applied Energy*, 83(4), 343–351.

Anderson M K, Assanis D N and Filipi Z S (1998), 'First and second law analyses of a naturally-aspirated, Miller cycle, SI engine with late intake valve closure', SAE paper 980889.

Asmus T W (1982), 'Valve events and engine operation', SAE paper 820749.

Asmus T W (1991), 'Perspectives on applications of variable valve timing', SAE paper 910445.

Babajimopoulos A, Assanis D N and Fiveland S B (2002), 'An approach for modeling the effects of gas exchange processes on HCCI combustion and its application in evaluating variable valve timing control strategies', SAE paper 2002-01-2829.

Bates B, Dosdall J M and Smith D H (1978), 'Variable displacement by engine valve control', SAE paper 780145.

Bemman Y-J, Frei T, Jones C and Keck M (2005), 'Passive exhaust system with cylinder deactivation', SAE paper 2005-01-2351.

Benajes J, Reyes E and Lujan J M (1996), 'Intake valve pre-lift effect on the performance of a turbocharged diesel engine', SAE paper 960950.

Bengtsson J, Strandh P, Johansson R, Tunestål P and Johansson B (2006), 'Multi-output control of a heavy duty HCCI engine using variable valve actuation and model predictive control', SAE paper 2006-01-0873.

Bernard L, Ferrari A, Micelli D, Perotto A, Rinolfi R and Vattaneo F (2009), 'Electro-hydraulic valve control with MultiAir technology', *ATZ autotechnology*, 9(6), 32–37.

Bolton B K and Assanis D N (1994), 'Optimum breathing strategies for turbocharged diesel engines based on the Miller cycle concept', *ASME Petroleum Division (publication) PD*, 64(8–2), 253–262.

Buuck B and Hampton K (1997), 'Engine trends and valvetrain systems for improved performance, fuel economy, and emissions', *Proceedings of the International Symposium on Valvetrain System Design and Materials*, April 14–15, Dearborn, MI.

Cao Y (2007), 'Thermodynamic cycles of internal combustion engines for increased thermal efficiency, constant-volume combustion, variable compression ratio, and cold start', SAE paper 2007-01-4115.

Caton P A, Song H H, Kaahaaina N B and Edwards C (2005a), 'Strategies for achieving residual-effected homogeneous charge compression ignition using variable valve actuation', SAE paper 2005-01-0165.

Caton P A, Song H H, Kaahaaina N B and Edwards C F (2005b), 'Residual-effected homogeneous charge compression ignition with delayed intake-valve closing at elevated compression ratio', *International Journal of Engine Research*, 6, 399–419.

Chan C and Pisano A P (1987), 'Dynamic model of a fluctuating rocker-arm ratio cam system', *Transactions of the ASME, Journal of Mechanisms, Transmissions, and Automation in Design*, 109, 356–365.

Charlton S J, Keane A J, Leonard H J and Stone C R (1990), 'Application of variable valve timing to a highly turbocharged diesel engine', IMechE paper C405/044.

Charlton S J, Stone C R, Leonard H J, Elliott C and Newman M J (1991), 'Transient simulation of a highly turbocharged diesel engine with variable valve timing', *Conference of Computers in Engine Technology*, September, Cambridge, UK, Institution of Mechanical Engineers, C430/016.

Chen L, Zhang W and Sun F (2007), 'Power, efficiency, entropy-generation rate and ecological optimization for a class of generalized irreversible universal heat-engine cycles', *Applied Energy*, 84(5), 512–525.

Chen L-G, Ge Y-L and Sun F-R (2008), 'Unified thermodynamic description and optimization for a class of irreversible reciprocating heat engine cycles', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 222, 1489–1500.

Clarke J R and Innes R C (1997), 'Valvetrain system design at Ford Motor Company', *Proceedings of the International Symposium on Valvetrain System Design and Materials*, April 14–15, Dearborn, MI.

Clarke D and Smith W J (1997), 'The simulation, implementation and analysis of the Miller cycle using an inlet control rotary valve', SAE paper 970336.

Curtil R (1998), 'Method for improving the operation of an air-scavenged supercharged heat engine', US Patent 5,819,693, French patent FR 93/13922.

Dammers W A (1997), 'A chronology of hydraulic lash compensation in the United States', *Proceedings of the International Symposium on Valvetrain System Design and Materials*, April 14–15, Dearborn, MI.

De Ojeda W, Cornelius D and Kumar R (2010), 'Effect of variable valve timing on diesel combustion characteristics', SAE paper 2010-01-1124.

Deng J and Stobart R (2009), 'BSFC investigation using variable valve timing in a heavy duty diesel engine', SAE paper 2009-01-1525.

Dennis R C and Neuser C J (1966), 'Computer-assisted valve train design and development', SAE paper 660348.

Desantes J M, Benajes J, Lujan J M and Fremond E (1995), 'Production of internal EGR in turbocharged diesel engines by means of intake valve pre-lift and exhaust valve post-lift', *The 5th International Congress*, Strasbourg, June 21–23, SIA 9506A17.

Douglas K J, Milovanovic N, Turner J W G and Blundell D (2005), 'Fuel economy improvement using combined CAI and cylinder deactivation (CDA) – an initial study', SAE paper 2005-01-0110.

Dowson D and Higginson G R (1977), *Elasto-Hydrodynamic Lubrication*, SI edition, Pergamon Press, New York.

Dowson D, Harrison P and Taylor C M (1986), 'The lubrication of automotive cams and followers', *The Twelfth Leeds-Lyon Symposium on Tribology–Mechanisms and Surface Distress*, Oxford, Butterworths, pp. 305–322.

Dresner T and Barkan P (1989a), 'A review and classification of variable valve timing mechanisms', SAE paper 890674.

Dresner T and Barkan P (1989b), 'A review of variable valve timing benefits and modes of operation', SAE paper 891676.

Du H Y I and Chen J S (2000), 'Dynamic analysis of a 3D finger follower valve train system coupled with flexible camshafts', SAE paper 2000-01-0909.

Dyson A and Naylor H (1960), 'Application of the flash temperature concept to cam and tappet wear problems', *Proceedings Institution of Mechanical Engineers (A.D.)*, No. 8, pp. 255–280.

Edwards S P, Frankle G R, Wirbeleit F and Raab A (1998), 'The potential of a combined Miller cycle and internal EGR engine for future heavy duty truck applications', SAE paper 980180.

Elrod A C and Nelson M T (1986), 'Development of a variable valve timed engine to eliminate the pumping losses associated with throttled operation', SAE paper 860537.

Ernst R, Schamel A R and Meyer J (1993), 'Advanced optimization techniques in valvetrain design', SAE paper 932004.

Falkowski A, McElwee M and Bonne M (2004), 'Design and development of the DaimlerChrysler 5.7L HEMI engine multi-displacement cylinder deactivation system', SAE paper 2004-01-2106.

Fessler H and Genova M (2004), 'An electro-hydraulic "lost motion" VVA system for a 3.0 liter diesel engine', SAE paper 2004-01-3018.

Fujiwara H and Kawase T (2007), 'Logarithmic profiles of rollers in roller bearings and optimization of the profiles', *NTN Technical Review*, No. 75, pp. 140–148.

Fukui T, Nakagami T, Endo H, Katsumoto T and Danno Y (1983), 'Mitsubishi Orion-MD – a new variable displacement engine', SAE paper 831007.

Fukutani I and Watanabe E (1979), 'An analysis of the volumetric efficiency characteristics of 4-stroke cycle engines using the mean inlet Mach number Mim', SAE paper 790484.

Furhapter A (2007), 'Four-stroke CAI engines with internal exhaust gas recirculation (EGR)', in Zhao H (ed.), *HCCI and CAI Engines for the Automotive Industry*, Woodhead Publishing, Cambridge, chapter 6.

Gallo W L R (1992), 'Thermodynamic evaluation of variable valve timing', SAE paper 921479.

Gast G J and David J W (1996), 'Pushrod modeling and valvetrain dynamics of high speed IC engines', SAE paper 960352.

Ge Y, Chen L, Sun F and Wu C (2005), 'Reciprocating heat-engine cycles', *Applied Energy*, 81(4), 397–408.

Gehrke S, Weiskirch C and Eilts P (2008), 'Development and implementation of a variable valve actuation system to a HD diesel engine', SAE paper 2008-01-1359.

Ghaffarpour M R and Baranescu R (1995), 'Breathing analysis using analytical techniques in diesel engines', ASME paper, Internal Combustion Engine Division (Publication) ICE, Vol. 23, *Engine Modeling*, pp. 1–10.

Gheorghiu V (2009), 'CO_2-emission reduction by means of enhanced thermal conversion efficiency of ICE cycles', SAE paper 2009-24-0081.

Gray C (1988), 'A review of variable engine valve timing', SAE paper 880386.

Grönlund T and Larmi M (2004), 'Valve train design for a new gas exchange process', SAE paper 2004-01-0607.

Gu H-Z (1995), 'The Gu-system that enhances the performance of turbocharged diesel engines', *Ship Engineering* (in Chinese), 4, 13–16.

Haas M and Rauch M (2010), 'Electro-hydraulic fully variable valve train system', *ATZ autotechnology*, 10(2), 39–43.

Hatano K, Lida K, Higashi H and Murata S (1993), 'Development of a new multi-mode variable valve timing engine', SAE paper 930878.

Hanaoka M and Fukumura S (1973), 'A study of valve train noises and a method of cam design to reduce the noises', SAE paper 730247.

He X, Durrett R P and Sun Z (2008), 'Late intake valve closing as an emissions control strategy at Tier 2 Bin 5 engine-out NO_x level', SAE paper 2008-01-0637.

Heath A R (1988), 'Valve train design for multivalve automotive gasoline engines', SAE paper 885133.

Hellinger W, Priebsch H H, Landfahrer K and Mayerhofer U (1992), 'Valve train dynamics and its contribution to engine performance', Paper C389/129, FISTIA, London.

Helmantel A and Denbratt I (2006), 'HCCI operation of a passenger car DI diesel engine with an adjustable valve train', SAE paper 2006-01-0029.

Herrin R J (1982), 'Measurement of engine valve train compliance under dynamic conditions', SAE paper 820768.

Hitomi M, Sasaki J, Hatamura K and Yano Y (1995), 'Mechanism of improving fuel efficiency by Miller cycle and its future prospect', SAE paper 950974.

Hu H, Israel M A and Vorih J M (1997a), 'Variable valve actuation and diesel engine retarding performance', SAE paper 970342.

Hu H, Vorih J M and Israel M A (1997b), 'The integrated lost motion VVT diesel engine retarder', SAE paper 973180.

Imperato M, Antila E, Sarjovaara T, Kaario O, Larmi M, Kallio I and Isaksson S (2009),

'NO_x reduction in a medium-speed single-cylinder diesel engine using Miller cycle with very advanced valve timing', SAE paper 2009-24-0112.

Iritani T, Shozaki A, Sheng B, Sugimoto M, Okazaki T and Aketa M (2002), 'Prediction of the dynamic characteristics in valve train design of a diesel engine', SAE paper 2002-32-1839.

Ishizuki Y, Shimizu Y, Hikino H and Kawashima Y (1985), 'A new type of Miller supercharging system for high speed engines: part 2 – realization of high BMEP diesel engines', SAE paper 851523.

Israel M A (1998), 'Jacobs variable valve actuation system for diesel powering and retarding', *GT-Suite Users' Conference Proceedings*, November.

Ito A (2006), 'A study on the measurement and estimation method of skew motion of roller-tappets in an OHV type DI diesel engine', SAE paper 2006-01-3348.

Ito A, Yang L and Negishi H (1998), 'A study on cam wear mechanism with a newly developed friction measurement apparatus', SAE paper 982663.

Ito A, Yang L and Negishi H (2001), 'A study of the measurement of surface temperature at a cam/tappet contact', *Tribotest Journal*, 7(3), pp. 203–218.

Jacques R (1997), 'Marketplace requirements and their impact upon valvetrain system architecture', in Bolton H A and Larson J M (eds), *Proceedings of the International Symposium on Valvetrain System Design and Materials*, April 14–15, Dearborn, MI, pp. 11–17.

Jensen P W (1987), *Cam Design and Manufacture*, 2nd edn, Marcel Dekker, New York.

Johnson G I (1962), 'Studying valve dynamics with electronic computers', SAE paper 620289.

Junior O M F (2009), 'Impact of the Miller cycle in the efficiency of an FVVT (fully variable valve train) engine during part load operation', SAE paper 2009-36-0081.

Kamo R, Mavinahally N S, Kamo L, Bryzik W and Reid M (1998), 'Emissions comparisons of an insulated turbocharged multi-cylinder Miller cycle diesel engine', SAE paper 980888.

Kanesaka H, Akiba K and Sakai H (1977), 'A new method of valve cam design – HYSDYNE cam', SAE paper 770777.

Ke Y and Pucher H (1996), 'Controlling the load and the boost pressure of a turbocharged SI engine by means of early intake-valve closing', SAE paper 960588.

Keribar R (2000), 'A valvetrain design analysis tool with multiple functionality', SAE paper 2000-01-0562.

Kim K-D and Kim D-H (2002), 'Improving the NO_x-BSFC trade off of a turbocharged large diesel engine using performance simulation', *ASME Internal Combustion Engine Division (publication) ICE*, 39, 159–165.

Kim H, Lee J, Kim K and Lee K (2009), 'Effect of the Atkinson cycle combined with calibration factors on a two-stage injection-type premixed charge compression ignition engine', *Energy and Fuels*, 23(10), 4908–4916.

Kitamura K, Takebayashi H, Ikeda M and Percoulis H M (1997), 'Development of ceramic cam roller follower for engine application', SAE paper 972774.

Kobayashi M, Tanabe T, Ushijima K and Aoyama S (2009), 'A lubrication analysis of multi link VCR engine components using a mixed elasto-hydrodynamic lubrication theory model', SAE paper 2009-01-1062.

Kodama Y, Nishizawa I, Sugihara T, Sato N, Iijima T and Yoshida T (2007), 'Full-load HCCI operation with variable valve actuation system in a heavy-duty diesel engine', SAE paper 2007-01-0215.

Korte V, Barth R, Kirschner R and Schulze J (1997), 'Camshaft/follower-design for different stress behavior in heavy duty diesel engines', SAE paper 972776.

Korte V, Glas T, Lettmann M, Krepulat W and Steinmetz C (2000), 'Cam roller follower design for heavy duty diesel engines', SAE paper 2000-01-0525.

Krepulat W, Dusik M and Korte V (2002), 'Advanced calculation method of the contact stress in roller follower valve train systems', SAE paper 2002-01-0852.

Kreuter P and Maas G (1987), 'Influence of hydraulic valve lash adjusters on the dynamic behavior of valve trains', SAE paper 870086.

Kreuter P and Pischinger F (1985), 'Valve train calculation model with regard to oil film effects', SAE paper 850399.

Kreuter P, Heuser P and Schebitz M (1992), 'Strategies to improve SI-engine performance by means of variable intake lift, timing and duration', SAE paper 920449.

Kreuter P, Heuser P, Reinicke-Murmann J, Erz R, Stein P and Peter U (2001), 'Meta – CVD system, an electro-mechanical cylinder and valve deactivation system', SAE paper 2001-01-0240.

Kurisu T, Hatamura K and Omoti H (1991), 'A study of jump and bounce in a valve train', SAE paper 910426.

Lancefield T (2003), 'The influence of variable valve actuation on the part load fuel economy of a modern light-duty diesel engine', SAE paper 2003-01-0028.

Lancefield T, Methley I, Räse U and Kuhn T (2000), 'The application of variable event valve timing to a modern diesel engine', SAE paper 2000-01-1229.

Larsen G J (1991), 'Research engine for evaluating the effects of variable compression ratio (VCR) and/or variable valve timing (VVT)', SAE paper 910053.

Lee J and Patterson D J (1997), 'Nonlinear valve train dynamics simulation with a distributed parameter model of valve springs', *Transactions of the ASME*, *Journal of Engineering for Gas Turbines and Power*, 119, 692–698.

Lee P and Rahbar A (2005), 'Active tuned absorber for displacement-on-demand vehicles', SAE paper 2005-01-2545.

Leet J A, Simescu S, Froelund K, Dodge L G and Roberts C E (2004), 'Emissions solutions for 2007 and 2010 heavy-duty diesel engines', SAE paper 2004-01-0124.

Leonard H J, Stone C R and Charlton S J (1991), 'Parametric investigation of variable valve timing applied to a turbocharged diesel engine', SAE paper 910453.

Leone T G and Pozar M (2001), 'Fuel economy benefit of cylinder deactivation – sensitivity to vehicle application and operating constraints', SAE paper 2001-01-3591.

Londhe A, Yadav V and Patil A (2009), 'Evaluation of a 4-stroke 4-cylinder diesel engine valve-train for replacement of a solid camshaft with a hollow camshaft', SAE paper 2009-01-0405.

Ma T H (1988), 'Effect of variable valve timing on fuel economy', SAE paper 880390.

Mardell J E and Cross R K (1988), 'An integrated, full authority, electrohydraulic engine valve and diesel fuel injection system', SAE paper 880602.

Mavinahally N, Ramo R, Bryzik W, Reid M and Wong V (1996), 'Insulated Miller cycle diesel engine', SAE paper 961050.

Miller R H (1947), 'Supercharging and internal cooling cycle for high output', *Transactions of the ASME,* 69(5), 453–457.

Miller R and Lieberherr H U (1957), 'The Miller supercharging system for diesel and gas engine operating characteristics', *CIMAC Congress*, Zurich.

Millo F, Mallamo F and Mego G G (2005), 'The potential of dual stage turbocharging and Miller cycle for HD diesel engines', SAE paper 2005-01-0221.

Millo F, Mallamo F, Arnone L, Bonanni M and Franceschini D (2007), 'Analysis of different internal EGR solutions for small diesel engines', SAE paper 2007-01-0128.

Milovanovic N, Turner J W G, Kenchington S A, Pitcher G and Blundell D W (2005), 'Active valvetrain for homogeneous charge compression ignition', *International Journal of Engine Research*, 6, 377–397.

Morel T and LaPointe L A (1994), 'Concurrent simulation and testing concept in engine development', SAE paper 940207.

Muhr T H (1993), 'New technologies for engine valve springs', SAE paper 930912.

Munnannur A, Kong S C and Reitz R D (2005), 'Performance optimization of diesel engines with variable intake valve timing via genetic algorithms', SAE paper 2005-01-0374.

Murata Y, Kawano D, Kusaka J, Daisho Y, Suzuki H, Ishii H, Goto Y and Odaka M (2006), 'Achievement of medium engine speed and load premixed diesel combustion with variable valve timing', SAE paper 2006-01-0203.

Murata Y, Kusaka J, Daisho Y, Kawano D, Suzuki H, Ishii H and Goto Y (2008), 'Miller-PCCI combustion in an HSDI diesel engine with VVT', SAE paper 2008-01-0644.

Murata Y, Nishio Y, Kusaka J, Daisho Y, Kawano D, Suzuki H, Ishii H and Goto Y (2010), 'Numerical analysis of Miller-premixed charge compression ignition combustion on a dynamic φ-T map', *International Journal of Engine Research*, 11(2), 89–98.

Nevin R M, Sun Y, Gonzalez M A and Reitz R D (2007), 'PCCI investigation using variable intake valve closing in a heavy duty diesel engine', SAE paper 2007-01-0903.

Newton J A and Allen C H (1954), 'Valve gear fundamentals for the large-engine designer', *Transactions of the ASME*, 76(2) 79–95.

Norton R L, Stene R L, Westbrook J and Eovaldi D (1998), 'Analyzing vibrations in an IC engine valve train', SAE paper 980570.

Norton R L, Eovaldi D, Westbrook J and Stene R (1999), 'Effect of valve-cam ramps on valve train dynamics', SAE paper 1999-01-0801.

Nourse J H, Dennis R C and Wood W M (1961), 'Recent developments in cam design', SAE paper 610048.

Nuccio P and Marzano M R (1992), 'A review of variable valve actuations and their effects on SI engine performance', International Seminar on High-Performance Spark Ignition Engines for Passenger Cars, November 5–6, Cinisello Balsamo, Milan, SAE paper 92A211.

Okamoto K, Zhang F-R, Shimogata S and Shoji F (1997), 'Development of a late intake-valve closing (LIVC) Miller cycle for stationary natural gas engines – effect of EGR utilization', SAE paper 972948.

Okamura H and Yamashita K (1997), 'Influence of the valve and accessory gear train on the crankshaft three-dimensional vibrations in high speed engines', SAE paper 971967.

Okarmus M, Keribar R, Oliva M and Tonin N (2008), 'Application of an integrated valvetrain and hydraulic model to characterization and retuning of exhaust valve behavior with a DPF', SAE paper 2008-01-0292.

Parvate-Patil G B, Hong H and Gordon B (2004), 'Analysis of variable valve timing events and their effects on single cylinder diesel engine', SAE paper 2004-01-2965.

Pastor J V, Lujan J M, Molina S and Garcia J M (2007), 'Overview of HCCI diesel engines', in Zhao H (ed.), *HCCI and CAI Engines for the Automotive Industry*, Woodhead Publishing, Cambridge, Chapter 10.

Payri F, Desantes J M and Corberan J M (1988), 'A study of the performance of a SI engine incorporating a hydraulically controlled variable valve timing system', SAE paper 880604.

Peng Z, Jia M and Lu L (2008), 'CFD analysis of VVT/VVA on the gas exchange and fuel-air mixing in a diesel engine', SAE paper 2008-01-1635.

Phlips P and Schamel A (1991), 'The dynamics of valvetrains with hydraulic lash adjusters and the interaction with the gas exchange process', SAE paper 910071.

Phlips P J, Schamel A R and Meyer J (1989), 'An efficient model for valvetrain and spring dynamics', SAE paper 890619.

Picron V, Postel Y, Nicot E and Durrieu D (2008), 'Electro-magnetic valve actuation system: first steps toward mass production', SAE paper 2008-01-1360.

Pierik R J and Burkhard J F (2000), 'Design and development of a mechanical variable valve actuation system', SAE paper 2000-01-1221.

Pisano A P (1984), 'Coulomb friction in high-speed cam systems', *Transactions of the ASME*, *Journal of Mechanisms, Transmissions, and Automation in Design*, 106, 470–474.

Pisano A P (1986), 'Coulomb friction and optimal rocker arm ratio for high-speed cam systems', *Transactions of the ASME*, *Journal of Mechanisms, Transmissions, and Automation in Design*, 108, 340–344.

Pisano A P and Freudenstein F (1983a), 'An experimental and analytical investigation of the dynamic response of a high-speed cam-follower system, part 1: experimental investigation', *Transactions of the ASME*, *Journal of Mechanisms, Transmissions, and Automation in Design*, 105, 692–698.

Pisano A P and Freudenstein F (1983b), 'An experimental and analytical investigation of the dynamic response of a high-speed cam-follower system, part 2: a combined, lumped/distributed parameter dynamic model', *Transactions of the ASME*, *Journal of Mechanisms, Transmissions, and Automation in Design*, 105, 699–704.

Pischinger M, Salber W, van der Staay F, Baumgarten H and Kemper H (2000), 'Benefits of the electromechanical valve train in vehicle operation', SAE paper 2000-01-1223.

Porot P and Trapy J (1993), 'A numerical and experimental study of the effect of aeration of oil on valve trains equipped with hydraulic lash adjusters', SAE paper 930997.

Rabhi V, Beroff J and Dionnet F (2004), 'Study of a gear-based variable compression ratio engine', SAE paper 2004-01-2931.

Rebbert M, Kreusen G and Lauer S (2008), 'A new cylinder deactivation by FEV and Mahle', SAE paper 2008-01-1354.

Reinicke-Murmann J and Kreuter P (1990), 'Analysis and development of camshaft drive systems using a computer simulation model', SAE paper 900448.

Ribeiro B and Martins J (2005), 'Miller cycle analysis using EGM', *Proceedings of 2005 Fall Technical Conference of the ASME Internal Combustion Engine Division*, pp. 441–446.

Ribeiro B and Martins J (2007), 'Direct comparison of an engine working under Otto, Miller and Diesel cycles: thermodynamic analysis and real engine performance', SAE paper 2007-01-0261.

Ribeiro B, Martins J and Kothari N (2006), 'Otto and VCR Miller cycle engine performance during the European driving cycle', SAE paper 2006-01-0440.

Roberts M (2003), 'Benefits and challenges of variable compression ratio (VCR)', SAE paper 2003-01-0398.

Roggenbuck R A (1953), 'Designing the cam profile for low vibration at high speeds', *SAE Transactions*, 61, 701–705, SAE paper 530262.

Roß J and Arnold M (1993), 'Analysis of dynamic interactions in valve train systems of IC-engines by using a simulation model', SAE paper 930616.

Ryan T W, Challahan T J and Mehta D (2004), 'HCCI in a variable compression ratio engine – effects of engine variables', SAE paper 2004-01-1971.

SAE Surface Vehicle Recommended Practice J1121 (2006), 'Helical compression and extension spring terminology'.

SAE Surface Vehicle Recommended Practice J1122 (2004), 'Helical springs: specification check lists'.

Saito K, Igashira T and Nakada M (1989), 'Analysis of oil consumption by observing oil behavior around piston ring using a glass cylinder engine', SAE paper 892107.

Sakai H, Noguchi H, Kawauchi M and Kanesaka H (1985), 'A new type of Miller supercharging system for high-speed engines – part 1: fundamental considerations and application to gasoline engines', SAE paper 851522.

Salber W, Kemper H, van der Staay F and Esch T (2001), 'The electro-mechanical valve train – a system module for future powertrain concepts', *MTZ Motortechnische Zeitschrift* (English version), 62(2), 20–30.

Sandford M H, Allen J and Tudor R (1998), 'Reduced fuel consumption and emissions through cylinder deactivation', IMechE paper S490/014, SAE paper 984014.

Saunders R J and Abdul-Wahab E A (1989), 'Variable valve closure timing for load control and the Otto Atkinson cycle engine', SAE paper 890677.

Schaefer S K, Larson J M, Jenkins L F and Wang Y (1997), 'Evolution of heavy duty engine valves – materials and design', in Bolton H A and Larson J M (eds), *Proceedings of the International Symposium on Valvetrain System Design and Materials*, April 14–15, Dearborn, MI, pp. 129–139.

Schamel A R, Hammacher J and Utsch D (1993), 'Modeling and measurement techniques for valve spring dynamics in high revving internal combustion engines', SAE paper 930615.

Schechter M M and Levin M B (1996), 'Camless engine', SAE paper 960581.

Schernus C, van der Staay F, Janssen H, Neumeister J, Vogt B, Donce L, Estlimbaum I, Nicole E and Maerky C (2002), 'Modeling of exhaust valve opening in a camless engine', SAE paper 2002-01-0376.

Schwoerer J, Huang S and Trzaska G (2002), 'Simulation tools for development of advanced engine braking and variable valve actuation systems', SAE paper 2002-01-0375.

Schwoerer J, Dodi S, Fox M, Huang S and Yang Z (2004), 'Internal EGR systems for NO_x emission reduction in heavy-duty diesel engines', SAE paper 2004-01-1315.

Seidlitz S (1990), 'An optimization approach to valve train design', SAE paper 901638.

Shelby M H, Stein R A and Warren C C (2004), 'A new analysis method for accurate accounting of IC engine pumping work and indicated work', SAE paper 2004-01-1262.

Shi L, Deng K and Cui Y (2005), 'Study of diesel-fuelled homogeneous charge compression ignition combustion by in-cylinder early fuel injection and negative valve overlap', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 219, 1193–1201.

Speckens F M, Buck J and Lach R (1999), 'Marked progress in both techniques and handling of valve train and valve train drive calculation on commercial platforms', SAE paper 1999-01-0560.

Stabinsky M, Albertson W, Tuttle J, Kehr D, Westbrook J, Karbstein H and Kuhl M (2007), 'Active fuel management™ technology: hardware development on a 2007 GM 3.9L V-6 OHV SI engine', SAE paper 2007-01-1292.

Stas M J (1999), 'Effect of exhaust blowdown period on pumping losses in a turbocharged direct injection diesel engine', SAE paper 1999-01-0188.

Stebler H, Weisser G, Horler H-U and Boulouchos K (1996), 'Reduction of NO_x emissions of D.I. diesel engines by application of the Miller-system: an experimental and numerical investigation', SAE paper 960844.

Stephenson P W and Rutland C J (1995), 'Modeling the effects of valve lift profile on intake flow and emissions behavior in a DI diesel engine', SAE 952430.
Stoddart D A (1953), 'Polydyne cam design', *Machine Design*, January–March, pp. 175–216.
Stone C R and Kwan E (1989), 'Variable valve actuation mechanisms and the potential for their application', SAE paper 890673.
Stout J L (1997), 'Valvetrain unbalance and its effects on powertrain NVH', SAE paper 971993.
Strandh P, Bengtsson J, Johansson R, Tunestål P and Johansson B (2005), 'Variable valve actuation for timing control of a homogeneous charge compression ignition engine', SAE paper 2005-01-0147.
Su W (2010), 'Alternative combustion system for heavy-duty diesel engines', in Zhao H (ed.), *Advanced Direct Injection Combustion Engine Technologies and Development, Volume 2: Diesel Engines*, Woodhead Publishing Limited, Cambridge, UK.
Su W H, Lin T J, Zhao H and Pei Y Q (2005), 'Research and development of an advanced combustion system for the direct injection diesel engine', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 219, 241–252.
Su W, Lu Y, Pei Y and Zhao F (2009a), 'Effect of charge density on emissions in a HD-LTC diesel engine by retarding intake valve timing and rising boost pressure', *IMechE Proceedings of Internal Combustion Engines: Performance, Fuel Economy and Emissions*, December 8–9, London, UK, Chandos Publishing, pp. 89–97.
Su W, Lu Y, Yu W, Wang C, Pei Y, Sun Y, Li Y, Yu Z and Wang J (2009b), 'High density-low temperature combustion in diesel engine based on technologies of variable boost pressure and intake valve timing', SAE paper 2009-01-1911.
Tai C, Tsao T C, Schörn N A and Levin M B (2002), 'Increasing torque output from a turbodiesel with camless valvetrain', SAE paper 2002-01-1108.
Takagishi H, Shimoyama K and Asari M (2004), 'Prediction of camshaft torque and timing chain load for turbo direct injection diesel engine', SAE paper 2004-01-0611.
Takashima Y, Mizuno T, Goto T, Wakita M and Hasegawa K (2005), 'Parametric design with valve train dynamic analysis and digital data based manufacturing of the valve spring', SAE paper 2005-01-0515.
Tauschek M J (1958), 'Valve train engineering – fundamentals for the engine designer', SAE paper 580381.
Taylor C M (1991), 'Valve train lubrication analysis', *the Seventeenth Leeds-Lyon Symposium on Tribology – Vehicle Tribology, Tribology Series* 18, Elsevier, Oxford, pp. 119–131.
Thoren T R, Engemann H H and Stoddart D A (1952), 'Cam design as related to valve train dynamics', *SAE Quarterly Transactions*, 6(1), 1–14.
Thring R H (1990), 'The flexible diesel engine', SAE paper 900175.
Tomita M, Aoyama S, Ushijima K, Tanaka Y, Takahashi N, Moteki K and Takaba T (2007), 'Compact and long-stroke multiple-link VCR engine mechanism', SAE paper 2007-01-3991.
Trajkovic S, Tunestål P and Johansson B (2008), 'Investigation of different valve geometries and valve timing strategies and their effect on regenerative efficiency for a pneumatic hybrid with variable valve actuation', SAE paper 2008-01-1715.
Tsuchida H, Hiraya K, Tanaka D, Shigemoto S, Aoyama S, Tomita M, Sugiyama T and Hiyoshi R (2007), 'The effect of a longer stroke on improving fuel economy of a multiple-link VCR engine', SAE paper 2007-01-4004.
Tunestal P and Johansson B (2007), 'HCCI control', in Zhao H (ed.), *HCCI and CAI Engines for the Automotive Industry*, Woodhead Publishing, Cambridge.

Turin R C, Zhang R and Chang M F (2008), 'Volumetric efficiency model for variable cam-phasing and variable valve lift applications', SAE paper 2008-01-0995.

Turkish M C (1946), *Valve Gear Design: A Handbook for Designers and Engineers to Aid in the Design of Cams, Tappets, and Springs for the Valve Gear of Internal Combustion Engines*, Eaton Manufacturing Company, Detroit, MI.

Turkish M C (1987), 'Dynamic stresses in valve springs', *Springs*, 26(2), 51–55.

Tuttle J H (1980), 'Controlling engine load by means of late intake-valve closing', SAE paper 800794.

Wang Y (2007), *Introduction to Engine Valvetrains*, SAE International, Warrendale, PA.

Wang Y, Zeng S, Huang J, He Y, Huang X, Lin L and Li S (2005), 'Experimental investigation of applying Miller cycle to reduce NO_x emission from diesel engine', *Proc. IMechE, Part A: Journal of Power and Energy*, 219, 631–638.

Watanabe E and Fukutani I (1982), 'Cylinder cutoff of 4-stroke cycle engines at part-load and idle', SAE paper 820156.

Wik C and Hallbäck B (2008), 'Reducing emissions using 2-stage turbo charging', *Wärtsilä Technical Journal*, 2008 (1), pp. 35–41.

Xu T, Tang C, Shen H, King M, Nowak M and Xin Q (2007), 'Modeling, validation and dynamic analysis of diesel pushrod overhead bridged valve train', SAE paper 2007-01-1256.

Yang B and Keller P (2008), 'Analysis of diesel engine emissions reduction by late intake valve close and VTG turbocharger using 1-D simulation', SAE paper 2008-01-2444.

Yang L, Ito A and Negishi H (1996), 'A valve train friction and lubrication analysis model and its application in a cam/tappet wear study', SAE paper 962030.

Yang S, Deng K, Cui Y and Gu H (2009), 'A study on an automatically variable intake exhaust injection timing turbo-charging system for diesel engines', *Proceedings of the Spring Technical Conference of the ASME Internal Combustion Engine Division*, pp. 711–720.

Yang S, Deng K, Cui Y and Gu H (2010), 'A study on an automatically variable intake exhaust injection timing turbocharging system for diesel engines', *Transactions of the ASME*, *Journal of Engineering for Gas Turbines and Power*, 132(5), 1–7.

Yang Z (2002), 'A study of two-stroke engine braking for HD diesel engines', SAE paper 2002-01-0487.

Zhao F, Asmus T W, Assanis D N, Dec J E, Eng J A and Najt P M (eds) (2003), *Homogeneous Charge Compression Ignition (HCCI) Engines – Key Research and Development Issues*, SAE International, Warrendale, PA.

Zhao H (ed.) (2007), *HCCI and CAI Engines for the Automotive Industry*, Woodhead Publishing, Cambridge.

Zhao Y, Tong K and Lu J (1999), 'Determination of aeration of oil in high pressure chamber of hydraulic lash adjuster in valve train', SAE paper 1999-01-0646.

Zheng Q (2001), 'Characterization of the dynamic response of a cylinder deactivation valvetrain system', SAE paper 2001-01-0669.

Zhu G (1993), 'Valve trains – design studies, wider aspects and future developments', in Taylor C M (ed.), *Engine Tribology*, Elsevier Science, Amsterdam, pp. 183–211.

Zou D and McCormick H E (1996), 'Dynamic modeling and computer simulation of valve train assemblies with hydraulic lash adjuster', SAE paper 960351.

10
Friction and lubrication in diesel engine system design

Abstract: This chapter addresses engine friction and lubrication dynamics modeling in diesel engine system design. It starts by introducing important fundamental principles of engine tribology and builds up a three-level system modeling approach of engine friction. The chapter summarizes the friction characteristics and friction-reduction design measures for both the overall engine system and individual subsystems such as the piston assembly, the piston rings, the bearings, and the valvetrain.

Key words: engine tribology, friction, lubrication, cavitation, fuel economy, piston assembly dynamics, piston ring, bearing, valvetrain, stiff ordinary differential equation (ODE), numerical computation.

10.1 Objectives of engine friction analysis in system design

10.1.1 Definition of engine friction

The friction work is defined as the total of mechanical (rubbing) friction work and accessory work consumed. The rubbing friction is the work to overcome the resistance to relative motion of all the moving parts of the engine, including the piston assemblies, the bearings (for the crankshaft, the connecting rod, the camshaft, the rocker arm, and the balance shaft, if any), the front and rear oil seals and other bearing seals, the valvetrain, the gears, the pulleys, the drive belts, and the chains. The accessories can include part or all of the auxiliary devices driven by the engine, such as the water pump, the oil pump, the fuel pump, the alternator (the generator), the air pump or the EGR pump for emissions control, the air compressor for the service brakes, the radiator cooling fan, the power-steering pump, and the air conditioner.

Unlike the definition used by other authors which includes pumping work as part of the total engine friction, the pumping work is not defined as part of the friction work in this book for the following reasons:

- In four-stroke engines, the pumping work is defined as the work in the exhaust and intake strokes. It is difficult to define pumping work in two-stroke engines due to the lack of these two pumping strokes.
- The pumping work can be either a positive quantity (i.e., a pumping

'gain') or a negative quantity (i.e., a pumping 'loss'). The pumping work consists of two parts: the part due to engine delta P, and the part related to engine volumetric efficiency, as analyzed in the air system theory in Chapter 4. The latter part is always a loss or dissipation work due to the air flow restrictions through the valves and other restrictive orifices/ducts. However, the engine delta P can be either positive or negative, depending on turbine area and turbocharger efficiency. When the pumping work is a positive gain, if it were included in the definition of the engine friction work, the total engine friction could be completely offset to zero or even a positive gain. This can be very misleading for comparing the friction levels between different engines and for the design efforts of reducing the engine mechanical friction.

- The pumping work is highly variable. Significant differences can be seen between the naturally aspirated engines and the turbocharged engines, with large differences existing among the engines on the following aspects: (1) turbine areas (e.g., in a VGT or wastegated turbine); (2) turbocharger efficiencies; (3) intake/exhaust restriction levels (related to the intake/exhaust system design, the aftertreatment devices, or the exhaust brake); and (4) engine operating conditions (speed, load, ambient condition, etc.). If the pumping work were included in the definition of engine friction, the friction work would become highly uncertain and inconvenient for comparison.
- The pumping loss eventually dissipates in the form of exhaust heat discharged to the ambient through turbulent mixing and non-isentropic compression. The mechanical friction eventually dissipates in the form of heat rejection to the cooling system. Their mechanisms are different.
- The theoretical background and the mechanisms are very different for the pumping work and the mechanical (rubbing) friction. Modern engines become increasingly complex, from turbocharger and EGR systems to the exhaust brake. The pumping work belongs to the territory of the air system theory, while the mechanical friction loss belongs to the areas of lubrication and tribology. Their completely different academic backgrounds make it necessary to separate the pumping work from the scope of engine friction.
- Modern in-cylinder pressure measurement techniques and engine cycle simulation software allow an accurate determination of the pumping work so that there is no reason/need to mix it with the mechanical friction work.

Engine motoring tests are often used to obtain engine friction data. It is necessary to subtract the pumping work from the total motoring work in order to obtain the 'true friction work' (i.e., the rubbing friction plus the accessory work; Gish *et al.*, 1958). The pumping work can be calculated by using $\int p dV$ for the intake and exhaust strokes, where the cylinder pressure p

can be obtained from either engine testing or cycle simulation. Alternatively, a less accurate but simpler approach can be used to estimate the pumping loss, given as follows:

$$\varpi_{PMEP} = (p_{intake} - p_{exhaust}) + C_1 v_{mp}^2 + C_2 v_{mp}^2 \tag{10.1}$$

where C_1 and C_2 are the constants reflecting the valve flow restrictions at the intake and exhaust valves, respectively; v_{mp} is the mean piston speed, $p_{exhaust}$ is the exhaust manifold pressure, and p_{intake} is the intake manifold pressure.

It is important to clearly specify the auxiliary components included in the accessory work. Usually, the oil, water, and fuel pump power are included, but the power from the cooling fan, the alternator, the power steering pump and other vehicle accessories may not be included in the engine motoring friction power.

The above engine friction definition can be expressed as follows:

$$\varpi_{FMEP} = (\varpi_{IMEP} + \varpi_{PMEP}) - \varpi_{BMEP} \tag{10.2}$$

$$\varpi_{FMEP} = \varpi_{MFMEP} + \varpi_{AFMEP} \tag{10.3}$$

where ϖ_{FMEP} is the friction mean effective pressure (defined here as positive value for convenience), ϖ_{IMEP} is the indicated mean effective pressure, ϖ_{PMEP} is the pumping mean effective pressure, and ϖ_{BMEP} is the brake mean effective pressure. The ϖ_{MFMEP} and ϖ_{AFMEP} represent the mean effective pressure of the mechanical (rubbing) friction and the accessories, respectively. The engine mechanical efficiency is defined by

$$\eta_{E,mech} = \frac{\varpi_{BMEP}}{\varpi_{BMEP} + \varpi_{FMEP}} \tag{10.4}$$

The rubbing friction power $\dot{W}_f$ and the friction force F_f are related by

$$\dot{W}_f = F_f v \tag{10.5}$$

where v is a characteristic velocity. Moreover, ϖ_{MFMEP} is proportional to engine design parameters according to the following relationship:

$$\varpi_{MFMEP} \propto \frac{\dot{W}_f}{N_E V_E} = \frac{F_f v}{N_E V_E} = \frac{F_f v}{N_E n_E B_E^2 S_E} \tag{10.6}$$

More engine friction fundamentals are described by Heywood (1988).

10.1.2 The need to study friction in engine system design

Friction, as one of the engine performance attributes, directly affects fuel economy, and is closely related to noise (e.g., piston slap and bearing impact) and wear (through the mixed and boundary lubrication). Internal combustion engine friction has been constantly reduced over the past several decades, nearly 10% per decade (Sandoval and Heywood, 2003). Among the total mechanical losses in the diesel engine, the piston assembly may account for 45–50%, bearings 20–30%, valvetrain 7–15%, and accessories 20–25% (Taylor, 1993a). In order to completely address diesel engine system design specifications, engine friction needs to be researched by system engineers for the following reasons:

- Engine friction modeling is important for the accuracy and effectiveness of the engine system design tool – the cycle simulation model. Friction affects the prediction of engine power, fuel economy, and heat rejection. A reliable estimation of engine friction is also important for comparing different design concepts related to configuration, component geometry, and weight. The friction model needs to accurately reflect the overall engine friction at different operating conditions, and reflect the parametric dependency that a system engineer is concerned about (e.g., the impact of engine bore, stroke, weight, and cylinder pressure).
- The system engineer is responsible for coordinating various performance attributes such as fuel economy, noise, durability (e.g., wear), and cold start capability. A thorough understanding of engine friction of each related component is necessary.
- The system engineer is in charge of planning the fuel economy improvement roadmap and friction reduction roadmap for the engine. A good understanding of the contribution of each friction component and their relative importance is necessary in order to decide the overall system-level design strategy and coordinate the activities for component engineers.
- Engine transient/controls simulation and fault diagnostics are important topics in diesel engine system design. Mechanical friction prediction is important for the torque-based controls at a system level. The friction torque in the engine control unit is usually presented as a lookup table with two input variables, the engine speed and the indicated torque. The accuracy of the friction torque at steady-state and transient operations is important for engine controls and advanced real-time diagnostic functions. Friction models are also important for cold start analysis. Not only is an accurate overall engine friction model needed, but an instantaneous crank-angle-resolution friction model is also required in order to accurately predict the engine speed fluctuation. (The importance of instantaneous friction model for fault diagnostics was mentioned by Ciulli, 1993.)

10.1.3 The approach of engine system design to friction

The focus of engine friction research for the system engineer is different from that for the component engineer. In component design, the most sophisticated friction models are required. This may include the detailed models for the piston assembly (the piston skirt and the rings), the valvetrain, and the engine bearings. The topics of concern in component design usually include the effect of design parameters on friction, lubrication, wear, motions, blow-by, oil consumption, and noise. These requirements often require the simulation model to possess high fidelity on hydrodynamic lubrication or even thermo-elastohydrodynamic lubrication with advanced sub-models of cavitation, lubricant viscosity, and surface roughness. These models tend to be computationally expensive and highly specialized. At a system level, the focus must be less detailed. Engine system design requires the system-level engine friction model to possess the following three progressive levels of complexity:

1. *The level-1 engine friction model*: This type of model focuses on the overall engine friction mean effective pressure (FMEP) on a cycle-average basis without the instantaneous (crank-angle-resolution) details of the friction forces. The level-1 model lumps all different friction components together and does not differentiate their lubrication regimes. However, it does include a range of major design parameters such as engine bore, stroke, bearing length, engine speed, and peak cylinder pressure.
2. *The level-2 engine friction model*: This type of model is more detailed than the level-1 model, and focuses on the overall engine friction torque on an instantaneous basis. It predicts the friction coefficients and the friction forces for major components based on lubrication regime, lubricant property and instantaneous loading of different components. The level-2 model can be used in engine transient and fault diagnostic simulations or for more detailed design parameter studies, while the level-1 model ignores the instantaneous fluctuations, underestimates actual friction force around the firing TDC, and does not include so many design parameters as inputs. Both the level-1 and level-2 models need to be computationally fast and real-time capable.
3. *The level-3 engine friction model*: This type of model is the most complex. With limited complexity, the level-3 model solves the lubrication Reynolds equation analytically or numerically for lubricating oil film thickness and pressure distribution in a lubrication dynamics formulation. Then the friction shear forces can be computed. This type of model is usually not real-time capable and may have a broader range of applicability extended for component designs by containing more fundamental physics and more detailed parameters (e.g., surface topography). The level-3 model may be solved to generate the so-called Stribeck curves in order

to look up the coefficient of friction for the level-1 and level-2 models. It can also help a system design engineer to thoroughly understand the mechanisms involved in engine friction, lubrication, and wear.

10.2 Overview of engine tribology fundamentals

10.2.1 Friction fundamentals

In general, friction can be classified into static friction and dynamic friction. Dynamic friction can be further divided into Coulomb's friction and viscous friction by different mechanisms. The static friction model is the classic model of friction of Leonardo Da Vinci: friction force is proportional to normal load, opposes the direction of motion, and is independent of contact area. The break-away force necessary to initiate motion from rest in static friction is often greater than the kinetic/dynamic (or Coulomb's) friction force. In Coulomb's friction law, the friction force is only dependent on the direction of velocity, not on the magnitude of the velocity. Moreover, the Coulomb friction force is directly proportional to the applied normal load and independent of the apparent area of contact. A typical example of the Coulomb's friction is the dry sliding or the boundary lubrication sliding contacts between two metal pieces. The Coulomb friction force is calculated by the following:

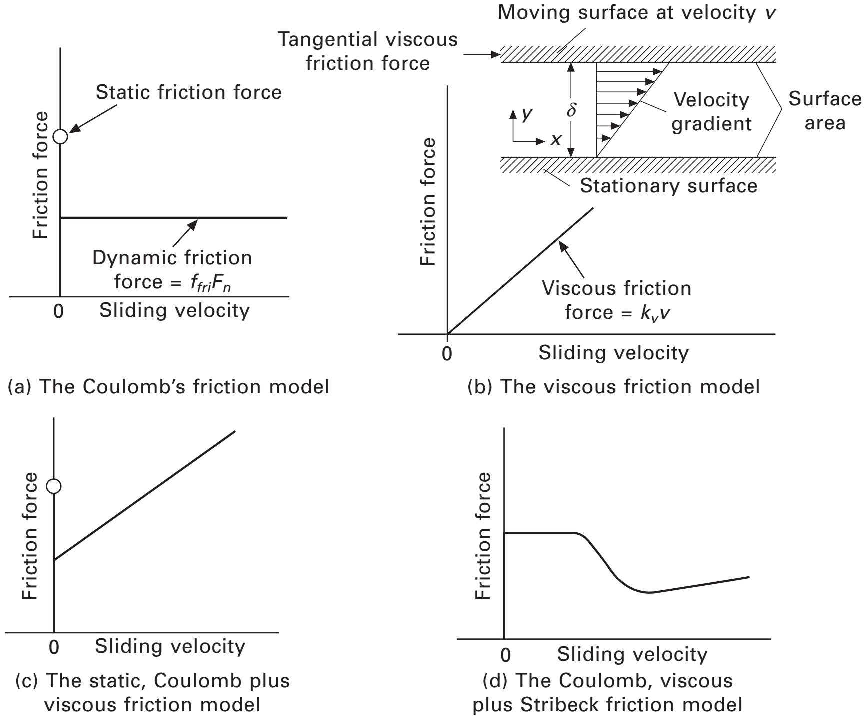
$$F_{f,Coulomb} = f_{fri} F_n \quad \text{or} \quad f_{fri} = \frac{F_{f,Coulomb}}{F_n} \qquad 10.7$$

where f_{fri} is the coefficient of friction and F_n is the force applied in the normal direction. Coulomb's friction model, along with other fundamental friction models commonly used in engineering, is shown in Fig. 10.1.

Viscous friction is another type of friction, caused by the viscosity of the fluid. In viscous friction, the friction force is proportional to the sliding velocity and goes to zero at zero velocity. The viscous friction force comes from the shear stress in the fluid flow. The ratio of the shear stress to the velocity gradient is a measure of the viscosity of the fluid and is called the coefficient of viscosity. These relationships can be expressed as follows, in the case of the planar Couette flow and based on Newton's law of laminar flow viscosity:

$$F_{f,v} = \int \tau \mathrm{d}A = \int \left(\mu_v \frac{\mathrm{d}v}{\mathrm{d}y} \right) \mathrm{d}A_c \approx \left(\frac{\mu_v A_c}{h_o} \right) v = k_v v \qquad 10.8$$

where $F_{f,v}$ is the tangential viscous friction force, τ is the shear stress, μ_v is the dynamic viscosity of the fluid, v is the relative velocity, A_c is the contact area, h_o is the fluid film thickness (such as the piston-to-bore or



10.1 Commonly used fundamental friction models.

bearing clearance), k_v is a defined viscous coefficient, and dv/dy is the velocity gradient across the thickness of the oil film (Fig. 10.1). Viscosity is the primary factor resisting the motion in the laminar flow, such as in fluid lubrication. However, when the velocity increases to the level of turbulent flow, the pressure differences resulting from the eddy currents rather than the viscosity may provide the major resistance to the motion. The viscous friction occurs in the hydrodynamic lubrication. It should be noted that unlike Coulomb friction where the friction force is proportional to the normal load, the viscous force may exist even at zero normal load (i.e., unloaded viscous friction, for example sliding a rod within a lubricated clearance inside a sleeve). Accordingly, there is no defined friction coefficient, *per se*, like that in Coulomb friction. A typical example of the unloaded viscous friction is the friction force on engine bearings at certain moments when the dynamic loading becomes zero but the journal rotates at a constant speed. Another example of unloaded viscous friction is the friction force on the piston skirt near the middle of the intake or exhaust stroke where the piston side thrust becomes zero but the piston velocity is high. When the dynamic loading is greater than zero the viscous friction becomes loaded viscous friction. When the velocity becomes zero (e.g., for piston skirt and piston rings at the TDC and the BDC), the viscous friction becomes zero. Figure 10.2 shows the reaction force to the piston side thrust force, calculated viscous shear force by using the two-dimensional Reynolds equation (to be detailed later) and the calculated 'coefficient of friction' (defined as the ratio of the shear friction force to the normal side thrust). It is observed that at the moments where the side thrust becomes zero, the coefficient of friction becomes extremely large (or infinity) and hence very misleading. This shows that such a concept of 'friction coefficient' should not be used to calculate the unloaded viscous friction. It should be handled with great care when applied to the modeling of the dynamically loaded viscous friction.

Both Coulomb friction and the viscous friction may exist at the same time for an engine component, for example in the mixed lubrication regime. Stribeck observed from the use of fluid lubrication in 1902 that the friction force decreases continuously with increasing velocities at low velocities. This phenomenon of a decreasing friction at increasing but relatively low velocities is called the 'Stribeck effect' (Fig. 10.1).

In engine lubrication and friction, it is important to determine which friction mechanism occurs and use the corresponding friction models appropriately. Sometimes, in the case of loaded viscous friction, especially in the mixed lubrication, the following forms of the definition of the viscous friction coefficient are seen in the literature:

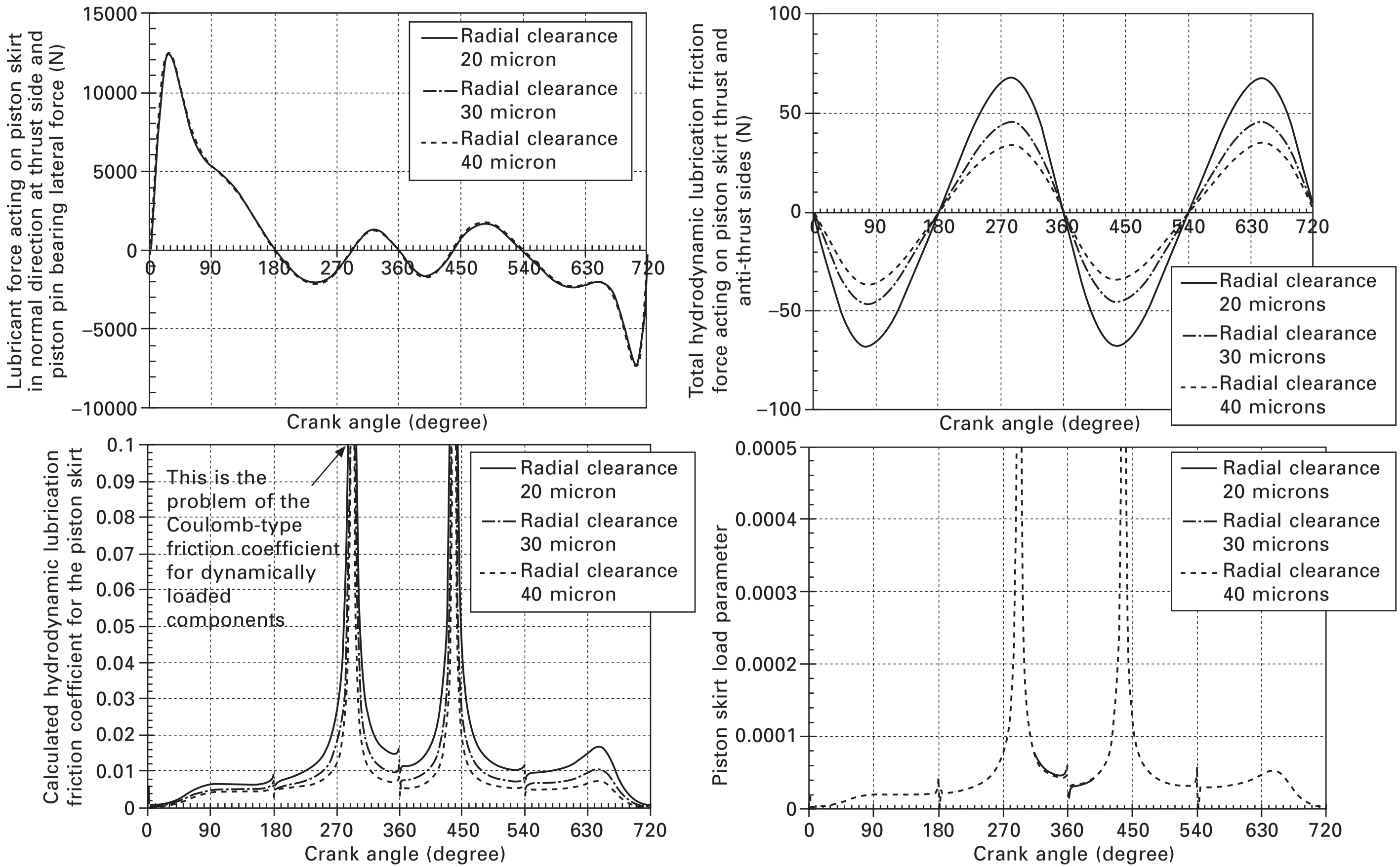
$$F_{f,v} = f_{fri} F_n v \quad \text{or} \quad f_{fri} \triangleq \frac{F_{f,v}}{F_n v} \qquad 10.9$$

$$F_{f,v} = f_{fri} F_n \quad \text{or} \quad f_{fri} \triangleq \frac{F_{f,v}}{F_n} \qquad 10.10$$

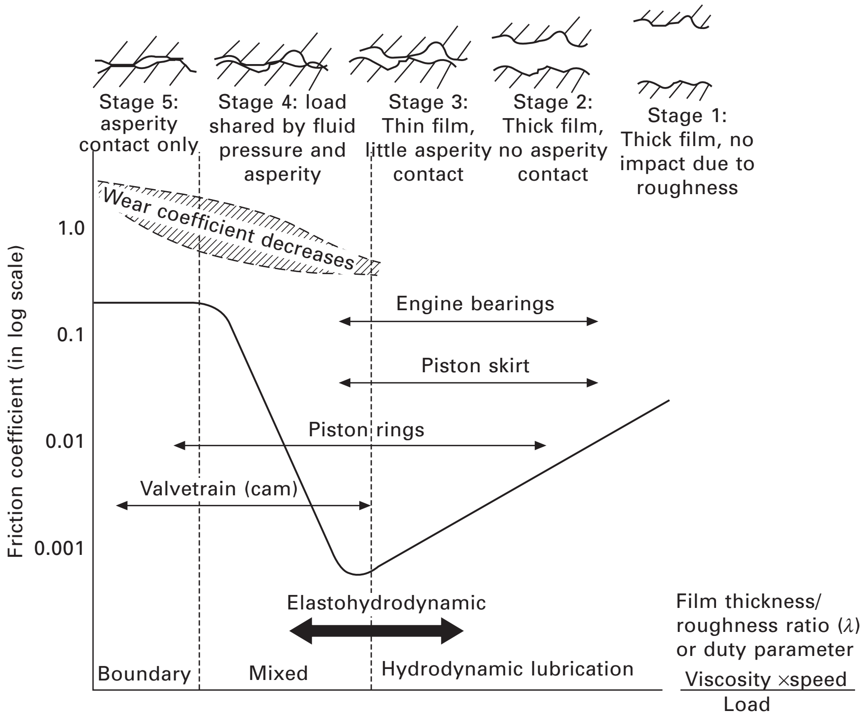
These forms of definition were created in order to be close to or the same as the form used in Coulomb friction because the Coulomb friction model (equation 10.7) and the concept of friction coefficient are so widely used due to their simplicity. These coefficients are often used as if the contacts exhibit the Coulomb friction behavior although in reality they do not. These definitions of the friction coefficients for the viscous friction may be acceptable for the constantly loaded components such as the piston rings, but they are not always valid for the case of the dynamically unloaded viscous friction such as for the piston skirt and the engine bearings. Andersson *et al.* (2007) provide more details on various fundamental friction models.

10.2.2 Lubrication regimes and the Stribeck diagram

Different lubrication regimes have different friction mechanisms, and lead to distinct friction and wear responses. Engine tribology modeling includes three



10.2 Piston skirt dynamic normal load, viscous friction force, and calculated Coulomb-type 'coefficient of friction'.



10.3 The Stribeck diagram with lubrication regimes of engine tribological components.

fundamental aspects: (1) lubrication regimes and friction types; (2) surface topography; and (3) lubricant property. There are basically two approaches to calculate the friction force on the lubricated engine components: (1) for the constantly loaded viscous friction or Coulomb friction under a non-zero normal load, a coefficient of friction multiplied by the normal load can be used; and (2) for the dynamically unloaded viscous friction, the viscous shear friction force needs to be calculated using equation 10.8, where the lubricating oil film thickness can be estimated using the component clearance or solved by an empirical equation or the Reynolds equation. The level-1 friction model, which ignores the instantaneous dynamically loaded or unloaded details, uses the first approach (a Coulomb-like method) as an approximation. The level-2 and level-3 friction models should use the second approach of the true viscous friction model because the instantaneous dynamic loading needs to be calculated by using engine dynamics models.

As reviewed by Spikes (1997), there are five stages of rough surface effect as the lubricating oil film thins:

- Stage 1: The surface roughness has no effect on the oil film thickness.

- Stage 2: The surface roughness influences the oil film thickness, but there is no surface asperity contact.
- Stage 3: Some asperity contact occurs, but the load is mainly carried by the fluid lubrication pressure.
- Stage 4: The load is shared by both the fluid pressure and the asperity contact.
- Stage 5: All the load is carried by the asperities.

The lubrication regimes are described by the well-known Stribeck diagram, where the 'coefficient of friction' is defined as the ratio of the total friction force to the normal load (Fig. 10.3),

$$f_{fri} \triangleq \frac{F_f}{F_n} \qquad 10.11$$

Note that equation 10.11 has the same form as Coulomb friction in appearance. When $F_n = 0$ and the viscous shear force $F_f \neq 0$, $f_{fri} = \infty$, where equation 10.11 becomes invalid for unloaded viscous friction. Mayo Hersey showed approximately one century ago that the friction due to viscous shear was a unique function of the product of viscosity and bearing rotational speed then divided by the average load (i.e., the Hersey number or the duty number) for a loaded lubricated contact (Dowson, 1998). Such a curve with the friction coefficient plotted against the Hersey number is commonly known as the 'Stribeck diagram'. The friction coefficient is a critical parameter related to the lubrication regime. Usually, the lubrication regime is characterized by a duty parameter

$$S_{ub} = \frac{\mu_v v}{\tilde{F}_n} \qquad 10.12$$

where μ_v is the dynamic viscosity of the lubricant oil, v is the relative sliding velocity of the two surfaces, and $\tilde{F}_n$ is the unit normal loading force per length. There are three lubrication regimes, namely (1) hydrodynamic lubrication (including elastohydrodynamic lubrication; e.g., the piston skirt and the engine bearings); (2) mixed lubrication (e.g., the piston rings); and (3) boundary lubrication (e.g., the cam–tappet interface). The behavior of the coefficient of friction is different in each lubrication regime. The coefficient of friction can be expressed as (Heywood, 1988):

$$f_{fri} = C_b f_{fri,b} + (1 - C_b) f_{fri,h} \qquad 10.13$$

where $f_{fri,b}$ is the friction coefficient of the boundary lubrication involving metal-to-metal contact. The coefficient $f_{fri,b}$ is affected by the surface finish and the properties of the materials in contact. The coefficient $f_{fri,h}$ is the friction coefficient of the hydrodynamic lubrication (modeled in detail in 10.15), and C_b is the metal-to-metal contact constant, varying between 0 and 1.

In the hydrodynamic lubrication regime, the friction force from the shearing of the viscous lubricant between the surfaces is inversely proportional to the lubricating oil film thickness, given by (for a Newtonian fluid)

$$F_{f,v} = \frac{\mu_v A_c v}{h_o} \tag{10.14}$$

where A_c is the lubricated contact area and h_o is the oil film thickness. A thinner oil film may develop a higher fluid film pressure and thus carry higher loads. As the applied load increases, the oil film thickness becomes thinner and the coefficient of friction decreases, but the oil film shear gradient increases. The normal load increases by a greater percentage than the percentage drop of the friction coefficient. Therefore, the friction force increases as a net result. The oil film thickness has a spatial distribution on the component surface and exhibits a certain pattern with respect to crank angle for the engine components. For instance, the minimum oil film thickness on a piston ring usually behaves like a half-sine wave as

$$h_o(\phi) = C_1 + C_2 \,|\sin(\phi - C_3)|$$

where ϕ is crank angle, C_1 and C_2 are constants (C_1 is usually below 1 micron), and C_3 is a shift ($C_3 \approx 10°$ crank angle). The actual thickness reaches a minimum value a little after the TDC and the BDC and has the largest value at the mid-stroke (shown by Arcoumanis *et al.*, 1997).

Both the theoretical lubricating oil film thickness (assuming smooth surfaces) and the friction force increase proportionally with $\sqrt{\mu_v v/\tilde{F}_n}$ (Spikes, 1997) in the isothermal hydrodynamic lubrication. When $\mu_v v/\tilde{F}_n$ increases, the coefficient of friction basically increases in a linear fashion if using the logarithmic scale on both axes (Fig. 10.3). The friction force is the viscous drag of the oil film for conformal contacts. The coefficient of friction in the fluid film hydrodynamic lubrication regime typically ranges from 0.0005 to 0.005. In the elastohydrodynamic lubrication, the contact is more concentrated than the hydrodynamic lubrication and the oil film is thinner, resulting in elastic distortion of the surfaces. As pointed out by Spikes (1997), for the elastohydrodynamic lubrication in non-conformal contacts (e.g., in gears or roller bearings), the oil film thickness varies only marginally with the load so that $\mu_v v$ is in fact a better duty parameter than $\mu_v v/\tilde{F}_n$. In this case, the coefficient of friction slightly decreases with an increasing $\mu_v v$, as shown in Fig. 10.3.

When the duty parameter $\mu_v v/\tilde{F}_n$ becomes smaller, the fluid film becomes thinner and consequently the coefficient of friction decreases to a minimum value. At this point, metal-to-metal surface asperity contact and wear start to occur. The low value of $\mu_v v/\tilde{F}_n$ makes the hydrodynamic lubricating oil film pressure insufficient to support the load by itself, and the load must be carried by the asperity contact as well. From this point, the lubrication enters the mixed lubrication regime. The coefficient of friction increases as $\mu_v v/\tilde{F}_n$ reduces. In the mixed lubrication regime, the total friction consists of the contributions from both the hydrodynamic element and the asperity contact. The asperity contact can be modeled by the average shear stress model developed by Patir and Cheng (1979).

Finally, a further reduction in $\mu_v v/\tilde{F}_n$ makes the oil film thickness much smaller than the height of the surface asperities. This brings the lubrication into the boundary lubrication regime where the metal-to-metal contact dominates and there is no hydrodynamic lubrication pressure. In the boundary lubrication, the physical and chemical actions of the thin oil films attached to the surfaces determine the tribological performance. The surface actions that determine the behavior of the boundary lubricant include the following: the physically adsorbed layers of the lubricant, the chemically adsorbed layers, and the oil films formed by chemical reactions. Friction modifier additives in the lubricant have a major influence by modifying the shear strength of the boundary lubricant film. When the load or the sliding speed at the contact interface area is high, significantly high contact temperatures may occur. The lubricating oil film formed by the physical and chemical adsorption layers cease to be effective above a certain temperature threshold. However, some additives such as sulfur, zinc, and phosphorus in the lubricant may help to maintain effective boundary lubrication. Special boundary lubricant can be effective in the most arduous operating conditions in the engine. Therefore, the friction is affected by the solid materials and the lubricant additives in the boundary lubrication. The coefficient of friction is essentially independent of oil viscosity, load, speed, and apparent area of the contact. The properties of the bulk lubricant such as viscosity are of minor importance. The coefficient of friction in the boundary lubrication is equal to the shear strength of the material (or more realistically, the shear strength of the oil film on the surface) divided by the yield pressure of the material (Rosenberg, 1982). The coefficient of friction in the boundary lubrication is basically a constant, typically at 0.08–0.12. The mixed and boundary lubrication regimes are responsible for the wear on the engine components.

The Stribeck diagram reflects the fundamental tribological characteristics of a friction pair with a given geometry. For example, in the hydrodynamic lubrication regime, different geometries (e.g., profile, contact area, bearing aspect ratio, clearance) will have different curves in the diagram. It can be used to look up the coefficient of friction based on the duty parameter. The Stribeck diagram is the foundation of the level-1 and level-2 engine friction models used in engine system design. It can be obtained either experimentally or numerically. When solved numerically, the friction shear force of the hydrodynamic or elastohydrodynamic lubrication is calculated by solving the Reynolds equation for the lubricating oil film thickness and the pressure distributions (e.g., by a level-3 friction model). The numerical result is dependent on certain theoretical assumptions or simplifications, such as the 'lubrication approximation' introduced by Reynolds, surface roughness, elastic deformation of the surfaces, lubricant properties (e.g., Newtonian or non-Newtonian), oil starvation, or partially flooded lubrication. Generating the Stribeck curves numerically for different component geometries with a level-3 model can facilitate the friction calculations with the level-2 model. Stanley *et al*. (1999) used the Reynolds equation to solve for different Stribeck curves for different piston ring geometries.

The coefficient of friction in the hydrodynamic lubrication can be expressed in the following form based on empirical experimental evidence (Fig. 10.3):

$$f_{fri,h} = C_0 S_{lub}^{C_1} = C_0 \left(\frac{\mu_v v}{p_l L} \right)^{C_1} \tag{10.15}$$

where C_0 and C_1 are the constants varying widely with component geometry (e.g., $C_1 = 0.3 \sim 1$, often $C_1 \approx 0.5$ for the piston skirt and the piston rings for approximate analysis), p_l is the loading pressure acting on the sliding component, L is the width of the component in the direction of motion (e.g., the thickness of the piston ring or the length of the piston skirt), and $p_l L$ is the load per unit transverse length. McGeehan (1978) provided detailed discussions on the coefficient of friction. By substituting Newton's law of viscosity (10.14) in the following

$$f_{fri,h} \triangleq \frac{F_{f,h}}{F_n} = \frac{\dfrac{\mu_v v A_c}{h_o}}{p_l L L'} = \frac{\dfrac{\mu_v v L L'}{h_o}}{p_l L L'} = \frac{\mu_v v}{h_o p_l} \tag{10.16}$$

where L' is the transverse length, and by equating 10.15 and 10.16, the lubricating oil film thickness can be derived with the following parametric dependency:

$$h_o = \frac{\mu_v v}{C_0 \left(\dfrac{\mu_v v}{p_l L} \right)^{C_1} p_l} = \frac{L^{C_1}}{C_0} \left(\frac{\mu_v v}{p_l} \right)^{1-C_1} \tag{10.17}$$

Multiplying the coefficient of friction in equation 10.15 by the normal load, the friction force can be obtained as follows:

$$F_{f,h} = C_0 \left(\frac{\mu_v v}{p_l L} \right)^{C_1} (p_l L L') = C_0 L' (\mu_v v)^{C_1} (p_l L)^{1-C_1} \tag{10.18}$$

The friction power is given by the following:

$$\dot{W}_{f,h} = F_{f,h}v = C_0 L' \mu_v^{C_1} v^{1+C_1} (p_l L)^{1-C_1} \tag{10.19}$$

Equation 10.19 indicates that the friction power is proportional to $p_l^{1-C_1}$. However, note that equation 10.19 is only applicable for the loaded surfaces (e.g., the piston rings) and not applicable to the unloaded viscous friction where p_l is zero (e.g., the unloaded side of the piston skirt, and the journal bearings at certain moments during an engine cycle). Note that the friction power of the entire journal bearing is in fact insensitive to load.

Stanley *et al.* (1999) and Taraza *et al.* (2000) found that C_0 and C_1 are essentially independent of the piston ring thickness but very sensitive to the curvature of the ring face profile. Moreover, partially flooded lubrication changes C_0 and C_1, giving a higher friction coefficient than that in the fully flooded ideal case. The effect of partially flooded lubrication can be modeled by reducing the lubricated length of the component.

The coefficient of friction was found to vary with other defined duty parameters similar to the Stribeck type, such as a Sommerfeld number for the piston rings (Ting, 1993a, 1993b). Stanley *et al.* (1999) proposed a characteristic curve instead of the Stribeck curve for a piston skirt by correlating a non-dimensional force parameter with a non-dimensional skirt characteristic parameter.

The coefficient of friction in the boundary lubrication can be regarded as a constant. The coefficient of friction in the mixed lubrication regime can be approximately modeled as varying with the duty parameter S_{lub} linearly as follows (Fig. 10.3, also see equation 10.13 for comparison):

$$f_{fri,m} = f_{fri,b}\left(1 - \frac{S_{lub}}{S_{lub,h-m}}\right) + f_{fri,h-m}\frac{S_{lub}}{S_{lub,h-m}} \tag{10.20}$$

where $f_{fri,b}$ is the friction coefficient in the boundary lubrication, $f_{fri,h-m}$ is the minimum friction coefficient at the transition between the hydrodynamic and the mixed lubrication. $S_{lub,h-m}$ is the duty parameter at that transition.

The Stribeck diagram does not contain the oil film thickness and the surface roughness as the explicit controlling parameters when determining the coefficient of friction. A parameter, lambda (λ), called 'film thickness to roughness ratio', which is the ratio of the calculated oil film thickness to the statistical root mean square roughness of the two surfaces, has generally been used to judge the safety of the operating limits for a lubricated contact. The theory on the transition between different lubrication regimes based on λ will be detailed in Section 10.2.3.

10.2.3 Mixed lubrication theory

Mixed lubrication, also called partial lubrication, is an important lubrication regime in internal combustion engines. Both elastohydrodynamic lubrication and metal-to-metal contact occur in mixed lubrication. The load is supported partly by the fluid film and partly by the surface asperities. Many engine components operate in mixed lubrication, for example the piston rings and the cams. The engine bearings may also operate in mixed lubrication under severe instantaneous loading. Understanding mixed lubrication is particularly important to a system engineer for the following reasons. First, it is the lubrication regime in which an accurate prediction of the friction is the most difficult, due to the interaction between the complex surface topography and the fluid pressure (or the oil film thickness). Compared to mixed lubrication, hydrodynamic and elastohydrodynamic lubrication regimes are relatively simpler. The calculation of mixed lubrication is also more complicated than that of the boundary lubrication regime. Secondly, mixed lubrication is a bridge between the hydrodynamic (or the elastohydrodynamic) and the boundary lubrication regimes for a system design engineer to fully understand all the links between them. Thirdly, engine lubricating oil film breakdown and wear (a durability concern) start from mixed lubrication.

Analysis of mixed lubrication is closely related to the ratio of oil film thickness to the combined (composite) surface roughness,

$$\lambda = h_o/\sigma_{sr}$$

which can be used in a Stribeck-type diagram to replace the duty parameter. The oil film thickness h_o can be calculated with the hydrodynamic or elastohydrodynamic lubrication models. The composite surface roughness σ_{sr} can be determined by the combination of the average values of both surfaces' asperity heights,

$$\sigma_{sr} = (\sigma_{sr1}^2 + \sigma_{sr2}^2)^{0.5}$$

or by a more complex calculation of topography. A high λ value ($\lambda > 3$) indicates the hydrodynamic lubrication where no metal-to-metal asperity contact happens. Generally, the smooth surface approximation in the oil film thickness prediction is valid when λ is large and when there is no oil starvation. Mixed lubrication occurs when $\lambda = 1 \sim 3$ (most authors believe it takes place at about $\lambda = 3$). Probably when $\lambda < 0.5$, boundary lubrication occurs, where both friction and wear are high. The λ ratio was discussed in further detail by Cann *et al.* (1994).

In addition to using the duty parameter $\mu_v v/\tilde{F}_n$ or the λ ratio, a third way, probably the best way, to characterize the lubrication regimes is by using a lubrication parameter, which is usually defined in the form of

$$S_{lub,\,sr} \triangleq \frac{\mu_v v}{p_l \sigma_{sr}} \qquad 10.21$$

where p_l is the mean Hertzian contact pressure (Schipper *et al.*, 1991). The lubrication parameter is essentially similar to the λ ratio, since $\mu_v v/p_l$ is proportional to the fluid film thickness. Schipper *et al.* (1991) suggested using the lubrication parameter to replace the λ ratio for the following reasons. Although the duty parameter is most widely used, it has the disadvantage that the surface roughness is not taken into account. The film thickness-to-roughness ratio requires the calculation of oil film thickness. The lubrication parameter can be used in situations where the oil film thickness is not *a priori* known. This situation matches the need of engine system design where the oil film thickness is not known in the level-1 and level-2 friction models.

Schipper *et al.* (1991) concluded that the transition between the mixed and the boundary lubrication is controlled by the product of lubricant viscosity and sliding velocity, and is independent of the mean Hertzian contact pressure (or the load). On the contrary, they found that the transition between the elastohydrodynamic and the mixed lubrication depends on the contact pressure or the load. Therefore, they believe that the λ ratio, which is commonly used to characterize the regime transitions, is pressure-dependent and should not be a constant as suggested in the literature. An interesting conclusion from Schipper *et al.* (1991) is that because the transition between the mixed and the boundary lubrication is pressure independent, a contact operating in the mixed lubrication regime at a constant product of viscosity times sliding velocity will not enter the boundary lubrication only when the contact pressure is increased. Schipper *et al.* (1991) provided the rules (Fig. 10.4) of the lubrication parameter as a function of the contact pressure for the regime transitions. With these rules it is possible to predict in which lubrication regime a particular lubricated concentrated contact operates.

Schipper and de Gee (1995) further concluded that the transition from the elastohydrodynamic to the mixed lubrication takes place at

$$\frac{\sigma_{sr}^{1.5} p_{l,mean}^{0.5}}{\mu_v v} = 6.5\times10^{-5} \qquad 10.22$$

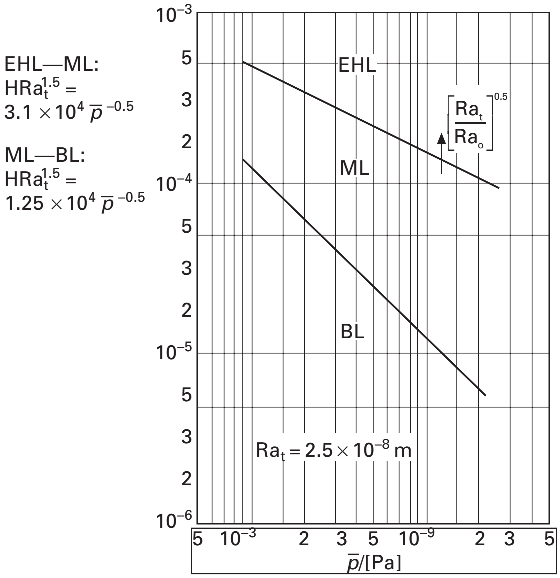
and the transition from the mixed to the boundary lubrication takes place at

$$\frac{\sigma_s}{\mu_v v} = 1.6 \times 10^{-5} \qquad 10.23$$

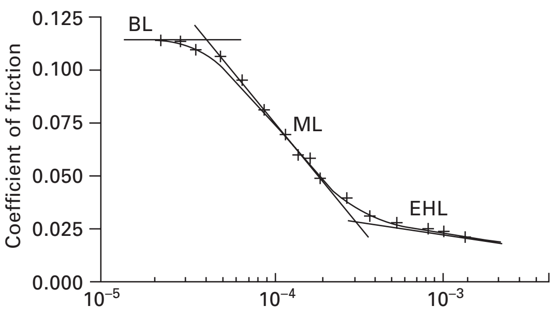
where v is the mean rolling speed (m/s) and μ_v is the inlet viscosity (Pa.s).

For a rolling contact (e.g., a tappet roller on a cam), the rolling friction coefficient (usually in the range of 0.001–0.003) is much lower than the sliding friction coefficient (e.g., 0.1 in boundary lubrication). The rolling friction coefficient actually rises with the λ ratio (Spikes, 1997).

The experimental coefficients of friction for the light-duty and heavy-duty engine bearings were reported by Kapadia *et al.* (2007). In their work,



(a) Lubrication parameter ($H/Ra_t = L$) as a function of mean contact pressure ($\overline{p}$) (from Schipper *et al.*, 1991)

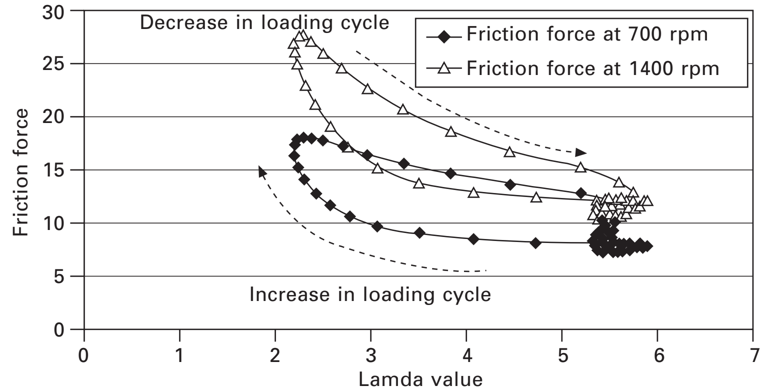


(b) Experimental friction speed diagram for a concentrated contact (from Schipper and de Gee, 1995)

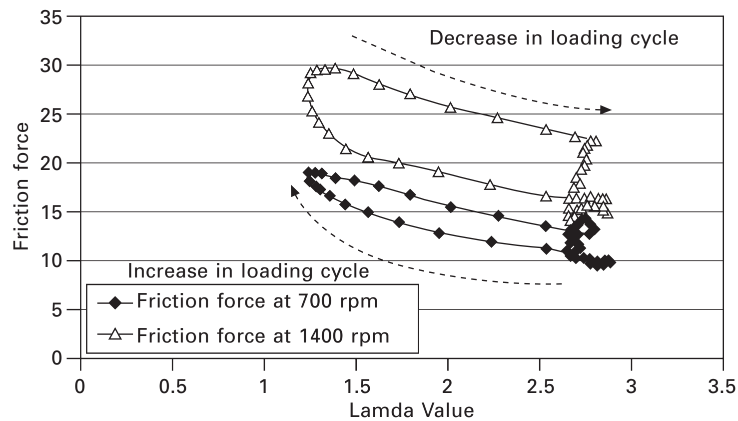
10.4 Use of a lubrication parameter to characterize lubrication regime transitions (from Schipper *et al.*, 1991 and Schipper and de Gee, 1995).

they also reported the measurement of friction forces vs. calculated λ ratio at different engine speeds and loads (Fig. 10.5).

The fundamentals of hydrodynamic lubrication are introduced in numerous books (e.g., Cameron, 1981; Heywood, 1988; Taylor, 1993a). The analysis of thick-film hydrodynamic lubrication for engine bearings is reviewed by Tanaka (1999). Thin-film lubrication and wear is reviewed by Jacobson (1997). Mixed lubrication is reviewed by Spikes (1997). The fundamentals



(a) Friction force versus lambda ratio for 15W-40 low-viscosity oil using the PCMO bearing set-up.



(b) Friction force versus lambda ratio for 15W-40 low-viscosity oil using the HDD bearing set-up.

10.5 Measured friction force vs. the lambda ratio for LD and HD engines (from Kapadia *et al.*, 2007).

on the boundary lubrication and scuffing are introduced in Ling *et al*. (1969), Spikes (1995), Taylor (1993a), and Ludema (1984).

10.2.4 Surface topography

Surface topography refers to both the profile shape and the surface roughness (including the waviness and the asperity or the finish). Surface topography affects the film thickness to roughness ratio and the lubrication regime. In the thin-film hydrodynamic or mixed lubrication calculations, the influence of surface roughness on the lubricating oil film thickness and pressure distribution obtained by solving the Reynolds equation is not negligible. The orientations of surface roughness have different effects. The transverse roughness (i.e., at right angles to the sliding direction) should enhance the load support and the oil film thickness, while the longitudinal roughness would reduce them (Spikes, 1997). Surface topography has an important influence on the load carrying capability of the oil film. On the other hand, the oil film pressure in the elastohydrodynamic lubrication has an impact on the surface roughness to cause it to deform and affect the load supported by the asperity contacts in the mixed lubrication. These interactions become more important as the thickness-to-roughness ratio reduces. The widely used models related to surface topography for the mixed lubrication regime include the following: (1) the Patir and Cheng (1979) average flow model, which accounts for the influence of surface roughness on the lubricant flow and on the load carrying capacity of the lubricating oil film; and (2) the Greenwood and Tripp (1971) asperity contact model, which accounts for the load carrying capacity of the asperity contacts in the mixed and boundary lubrication.

Surface topography also changes after engine running-in (break-in) and gradually changes during the engine life due to wear. This affects the friction power of all the sliding surfaces, especially the power cylinder. During break-in, some of the asperities are worn off and the sliding surfaces become smoother. This reduces the metal-to-metal contact and the boundary friction. As the break-in progresses, the engine friction reduces rapidly for an initial period, then stabilizes or changes very slowly.

The importance of surface topography on engine friction and wear was emphasized by Taylor (1998). The surface topography of cylinder liner, piston skirt or piston rings and the effect on engine friction and wear are discussed by Zhu *et al.* (1992), Pawlus (1996, 1997), and Rao *et al.* (1999).

10.2.5 Lubricant properties and impact on engine friction

Lubricant oil reduces friction and wear between the components, serves as a coolant to remove the frictional heat, and also removes impurities/debris. Lubricant oil properties have a direct influence on engine friction and wear. Lubricant oil additives include friction modifiers, viscosity index improvers (VI), anti-wear agents, detergents, anti-rust agents, oxidation inhibitors, and antifoam agents. Some important effects of the lubricant on engine friction include oil starvation or partially flooded (related to oil feed designs), cavitation, thermal effects, and the changes in viscosity with temperature, pressure and shear rate.

The engine oil viscosity classification is given in SAE J300 (2004). Common monogrades of viscosity include the SAE 5, 10, 30, 40, 45, and 50. The higher the grade, the higher the viscosity. The multigrade oils introduced since the 1950s use viscosity index improvers (the additives of temperature-sensitive polymers) to stabilize the viscosity at high temperatures so that the viscosity does not reduce as much as the monograde oils over the engine operating temperature range. Common multigrade oils include the SAE 10W-30, 10W-40, 10W-50, 15W-40, 15W-50, and 20W-50. The 10W-30 means that the oil has the viscosity of the SAE 10 when it is cold (at 15°C) and the viscosity of the SAE 30 when it is hot (at 90°C). Multigrade oil may reduce the engine start friction at very cold temperatures without having the problems of an unacceptably low viscosity and metal contact at high temperatures. Monaghan (1988) indicated that the multigrade oil may achieve a reduction in viscosity equivalent to about one grade at intermediate temperatures, hence reducing the viscous friction by about 20%. Multigrade oils contain the polymetric additives with relatively high molecular weight, and exhibit both temporary and permanent shear thinning as the non-Newtonian flow behavior, which affects their viscosity. Newtonian fluid obeys a linear relationship between the shear stress and the shear rate, and its viscosity is independent of the shear rate. Under high pressures, the lubricant exhibits non-Newtonian behavior and the viscosity decreases as the shear rate increases. Moreover, lubricant oil viscosity increases with the level of soot and dispersant in the oil (George *et al.*, 2007). Other fundamental information on engine oils can be found in SAE J357 (2006), J2227 (2006), J1423 (2003), and J183 (2006).

Lubricant viscosity is a strong function of temperature (e.g., Vogel's equation) and pressure (e.g., Barus' equation). The viscosity decreases with increasing temperature, and increases with increasing pressure (important for the elastohydrodynamic conditions). The non-Newtonian and thermal effects can be important in oil film thickness calculations, especially for rough surfaces and in the elastohydrodynamic lubrication. Under the non-Newtonian condition the shear thinning and viscoelasticity effects (i.e., the influence of shear rate and pressure on viscosity) need to be considered. The shear rate impact (e.g., the Cross equation) is especially important for multigrade oils. In the highly dynamically loaded lubricated contacts, the effects of temperature thinning, pressure thickening (the viscoelastic effect) and shear thinning can all be important. The lubricant behavior in the elastohydrodynamic lubrication contacts is reviewed by Jacobson (1996). Coy (1997) and Taylor (1997) provide the details on the models of lubricant rheology, friction and wear based on the Stribek-type diagrams. Note that these models are suitable for diesel engine system design as the level-2 models (to be detailed later).

As seen from the earlier discussion of the Stribeck diagram, the friction coefficient in the hydrodynamic lubrication regime decreases as the lubricant viscosity decreases. However, in the boundary lubrication regime (e.g., for some cam–follower contact and piston oil rings), lower viscosity results in higher friction. Therefore, the bearings and the piston skirt usually give lower friction when the lubricant viscosity is reduced, while the valvetrain gives increased friction. The effectiveness of using the low viscosity oil to reduce engine friction depends largely on how much the boundary lubrication accounts for the total engine friction. Note that the valvetrain friction can be reduced by using the oil containing a friction modifier additive. Since the engine has both hydrodynamic and mixed/boundary lubrication components, an optimum viscosity that provides a minimum engine friction exists at a given engine speed and load.

The fuel economy sensitivity to lubricants is different for heavy-duty and light-duty engines. Taylor (2000) indicated that because the proportion of the valvetrain friction losses is much higher in light-duty engines than in heavy-duty engines, the latter can be regarded more 'hydrodynamically' overall in the lubrication regime. Because the valvetrain is the main boundary lubricated component in an engine, the friction modifier additives can appear to be more effective for light-duty than heavy-duty engines. Taylor (2000) claimed that the fuel consumption savings for heavy-duty diesel engines under mixed driving conditions (low, medium, and high loads) by using carefully formulated engine and transmission lubricants (with reduced viscosity and optimized additive packages) can reach the order of 5%. Another investigation into the lubricant effects on light-duty and heavy-duty vehicle fuel economy was conducted by Kapadia *et al.* (2007).

Reducing viscosity moves the lubrication regime from hydrodynamic toward boundary. This increases the risk of wear and scuffing due to oil film thickness reduction. Therefore, there is a balance between reducing friction and maintaining acceptable durability. In fact, the Association des Constructeurs Européens d'Automobiles (ACEA) has mandated that heavy-duty diesel engine lubricants should have a HTHSV (high-temperature high-shear viscosity) of at least 0.0035 Pa.s (Taylor, 2000). Lubricant specifications also affect emissions and aftertreatment performance (McGeehan *et al.*, 2007).

10.3 Overall engine friction characteristics

10.3.1 Engine friction characteristics in different lubrication regimes

Engine friction has been measured using various methods, including the indicator diagram method, the motoring and teardown method, the pressurized motoring method (i.e., simulating the firing cylinder pressure effect with high intake manifold pressure without actually firing), the Morse test or electronic cylinder disablement method, the Willan's Line method, the hot shutdown method, the free Deceleration Curves method, the instantaneous IMEP method (Uras and Patterson, 1983), and the instantaneous friction torque method (P–ω method, Rezeka and Henein, 1984). The advantages and disadvantages of each experimental method are reviewed by Wakuri *et al.* (1995) and Richardson (2000). Because the diesel engine and the gasoline engine, either heavy-duty or light-duty, may share similar characteristics of engine friction within a common range, the references in this chapter also include the gasoline engine.

Different engine components operate in different lubrication regimes at different operating conditions due to their various load and speed characteristics. Their friction losses respond in different ways to the changes in design parameters and operating conditions. The understanding and modeling of engine friction characteristics should be built on each individual component. Recall that the duty parameter or the lubrication parameter in the Stribeck-type diagrams consists of lubricant viscosity, loading, and sliding velocity to characterize the lubrication. The proportionality between the hydrodynamic and the mixed/boundary lubrication regimes depends on many factors such as engine speed, load, lubricant viscosity, state of break-in, etc.

The piston skirt has a periodic sliding velocity between the TDC and the BDC, and carries the side thrust load generated from the cylinder gas pressure and the piston-assembly inertia force (Munro and Parker, 1975). The piston skirt predominantly operates in the hydrodynamic or elastohydrodynamic lubrication regime, especially in the middle of each stroke where the piston speed is higher. In the expansion stroke and the late compression stroke when the gas loading is high, some pistons may operate in mixed lubrication due to surface asperity contact. Under dynamic conditions when sudden increases in load or decreases in relative speed occur, a 'squeeze film' effect is developed to maintain a certain oil film thickness to separate the surfaces within a short moment. Near the TDC and the BDC where the piston sliding velocity approaches zero, there is a tendency to transition to the mixed lubrication regime although the large 'squeeze film' effect of the entire piston skirt tries to overcome such a tendency. Similar 'squeeze film' effects may exist for the piston rings. Note that the piston side thrust can reach zero at some crank angle locations within one engine cycle (Fig. 10.2).

The piston compression rings (the top and second rings) have the same periodic sliding velocity as the skirt, and their loads basically include the ring tension and the gas pressure acting on the back of the rings. The ring tension is usually defined as the diametrical compression force required to compress the free end of the ring to a specified clearance gap. The compression rings operate primarily in the hydrodynamic or elastohydrodynamic lubrication regime during the mid-stroke and in the mixed lubrication regime near the TDC and the BDC, especially in the mixed or even the boundary lubrication at the firing TDC where the cylinder gas pressure loading acting on the rings is high and the lubricant viscosity is low due to the hot wall temperature. Under that condition, the lubricating oil film breaks down and surface asperities come into contact, causing higher rubbing friction and wear. The second compression ring is commonly designed to have a tapered face (i.e., smaller ring diameter at the top than the bottom) to help control oil through a scraping action in the mixed or boundary lubrication on the down stroke. Note that the piston ring load in the diametrical direction is always greater than zero.

Among the piston rings, the oil control ring contributes much higher friction than the compression rings during the most of the strokes except around the firing TDC where the friction force of the compression rings may be higher due to the high cylinder pressure gas loading. The oil ring usually has two very thin rails and very high tension. Each rail of the oil control ring is handled as a separate single ring in the analysis. The rails usually are too thin to generate a hydrodynamic lubricating oil film when the tension is high and the speed is low. In that case, they operate primarily in the boundary lubrication regime with strong surface asperity contact between the ring face and the cylinder bore.

The instantaneous friction force acting on the piston and the rings may experience a sudden increase at the TDC or a sudden decrease at the BDC when the piston reverses direction due to the abrupt change between the dynamic and the static friction coefficients. This step change in the piston friction force is impulsive in nature and contains a broad range of frequencies which can excite the crankshaft resonant frequencies and produce a knocking noise at low speeds. Such a piston 'stick slip' noise was discovered in gasoline engines (Beardmore, 1982; Werner, 1987). Low-friction coating materials and friction modifiers in the lubricant were found to be effective to depress the abrupt changes between the static and the dynamic friction coefficients in order to reduce the 'stick slip' noise.

The engine bearings (the crankshaft main bearings, the connecting rod bearings at the big end and the small end, the camshaft bearings, and the rocker shaft bearings) carry different types of loading, and operate at different sliding speeds. For example, the crankshaft main bearing has a basically constant sliding velocity at a given engine speed, and it operates mainly in the hydrodynamic or elastohydrodynamic lubrication regime. At low speeds the heavily loaded engine bearings may operate in the mixed lubrication regime. The crankshaft main bearing seals operate in the boundary lubrication regime due to the direct contact between the seal lips and the crankshaft surface. The instantaneous resultant force of the connecting rod small-end bearing can reach zero. The instantaneous resultant force of the connecting rod big-end bearing is usually greater than zero. The instantaneous resultant force of the crankshaft main bearing is greater than zero most of the time during an engine cycle but occasionally it can reach zero. The camshaft resultant force does not reach zero within an engine cycle.

The main friction interface in the valvetrain is between the cam and the follower (tappet). The cam loading consists of the valve spring force and the valvetrain dynamic inertia load. The contact can be sliding (for a flat-faced follower) or rolling (for a roller follower). The cam–follower contact is mainly in the severe mixed lubrication or boundary lubrication regime (including elastohydrodynamic) due to the very high load acting on a small contact area with a relatively slow speed.

The frictional losses in the oil pump consist of bearing friction, internal fluid friction/restriction and intermeshing friction (Baba and Hoshi, 1986). The pump friction torque is proportional to the third power of the gear's outside diameter and the pump speed which is related to the drive ratio (Kovach *et al.*, 1982).

Note that in the hydrodynamic lubrication regime for all the above engine components, elastohydrodynamic lubrication may occur in some portion of the engine cycle, because of the elastic deformation of the contact surfaces due to very high unit load, and because of the influence of the lubricating oil film pressure on viscosity. Also note that mixed or boundary lubrication happens for almost all the components during engine start-up and shut-down because of the very low engine speed, oil starvation, and less effective oil film. Moreover, there is a difference between friction force and friction power. Although poor lubrication at the ends of the stroke results in a large friction force, wear and even scuffing between the piston rings and the cylinder bore, it has little effect on the friction power because the piston sliding velocity is very low there. In contrast, the hydrodynamic drag at the mid-stroke, although lower in terms of friction force, is the major contributor to the friction power due to the high piston velocity.

10.3.2 Engine friction characteristics at different speeds and loads

In boundary lubrication, friction force is independent of speed. In hydrodynamic lubrication, the viscous friction force increases with increasing speed. The fluid pumping torque usually increases with speed squared. Moreover, some component loading changes with engine speed, for example the valvetrain. Basically, the friction torque or the FMEP of all the components increases with engine speed, except for the valvetrain friction torque which decreases as the engine speed increases. The reason is that the boundary lubrication dominates in the valvetrain and the valvetrain load decreases with increasing engine speed. As the engine speed decreases, the proportion of the boundary lubrication friction increases.

The trend of overall engine friction at different loads is more complex, primarily because of the complex trend of the power cylinder friction. When the diesel engine load or fueling rate changes (with an extreme of comparing motoring with full-load firing), the following three aspects change in the power cylinder: (1) gas pressure loading; (2) metal/wall temperature; and (3) the clearances in the power cylinder. Higher gas loading increases the piston assembly friction during the expansion stroke and the late compression stroke in the boundary, mixed and even hydrodynamic lubrication regimes, especially near the firing TDC. The temperature of the oil on the cylinder liner is controlled mainly by the coolant temperature. At higher load, the coolant temperature is higher. The higher metal/wall temperature reduces the lubricant viscosity and the hydrodynamic viscous friction force, particularly in the mid-stroke where the gas pressure effect is small. The smaller clearances due to components' thermal expansion tend to increase the friction force. Another effect to consider is that the high friction force (evidenced by the wear at the top ring reversal) does not translate to high friction power near the TDC because the piston speed is low there. During the mid-stroke of an engine cycle the high temperature effect plays a major role due to the low cylinder pressure. This affects the friction power significantly since the piston velocity is high. The net result depends on which factor plays a dominant role.

It is generally believed that the bearing forces increase with engine load. Accessory friction losses also increase with load. Most authors believe that the overall engine friction torque increases with load, based on experimental findings. Accurate modeling of the engine friction at various loads is important because the fuel economy benefit of friction reduction is most prominent at the low-load conditions, as encountered in many real-world usage conditions, where the proportion of the friction is high compared to the brake power.

A typical argument on the engine friction trend vs. load is whether the motoring friction torque is lower than the friction torque in firing. This topic is important because the motoring test has been widely used as a convenient way to estimate the engine friction at firing. As mentioned above, the mechanisms of the friction losses in motoring and firing are very different due to the changes in cylinder pressure loading, wall temperatures and clearances, even if the overall engine friction is similar. The pressurized motoring tests with high intake manifold pressures without firing at Cummins (Richardson, 2000) confirmed that an increase in the peak cylinder pressure increases the piston-assembly friction force. Although controversial experimental findings have been published regarding the friction power difference between motoring and firing, it appears that most people believe the friction force in a firing engine is greater than that in a motored engine around the firing TDC under the combined influence of higher cylinder pressure and temperature. Richardson (2000) reviewed this topic and concluded that the piston ring pack friction power in hot motoring was lower than that in firing. In general, the cold motoring friction power was close to the hot firing power. He concluded that the overall average power cylinder friction of a firing engine was 0–20% more than the friction of a motored engine.

10.3.3 Design measures for engine friction reduction

Effective design measures to reduce engine friction should be planned for each individual component according to their different lubrication regimes.

The general guidelines for friction reduction are as follows:

- Durability and reliability take precedence over friction reduction and fuel economy. Any friction reduction should not jeopardize durability.
- Always minimize the load on the frictional interface. This is difficult to achieve in many situations because of the nature of the high load (e.g., caused by the demand of high power rating and the corresponding high cylinder pressure) and the compromises involved (e.g., using a shorter stroke to reduce the piston side thrust may cause an unfavorable stroke-to-bore ratio for the combustion performance). However, in certain scenarios the design optimization to reduce the load is achievable. For example, reducing the reciprocating mass of the piston assembly may reduce mechanical noise and even friction, although the magnitude of friction reduction depends on the particular application (e.g., engine speed). Offsetting the piston pin may change the side thrust and the moments acting on the piston so that less violent piston secondary motions can be achieved to reduce the friction loss. Another example is to reduce the valve spring force through valvetrain optimization.
- For boundary lubrication (e.g., cam–follower interface, oil control ring), reducing the normal load acting on the contact is effective. However, this needs to be done carefully. For example, excessively decreasing the tension of the oil control ring will increase oil consumption, especially when the cylinder bore distortion is large. Other design measures need to be taken to ensure the ring–bore conformability for adequate oil control (e.g., by reducing the ring stiffness through a reduction in the width or the radial wall thickness of the oil ring).
- Reduce friction coefficient by replacing sliding friction with rolling friction, especially in situations where high loads must be carried at relatively low sliding velocities (e.g., the follower and the rocker arm pivot in the valvetrain). One approach is to use rolling-element or needle bearings instead of plain bearings. A roller design may increase the contact stress and require a better material to withstand the stress.
- For hydrodynamic lubrication (e.g., the piston skirt, the bearings, the compression rings), reducing the lubricated area or the lubricant viscosity is the most effective way to reduce the shear friction force. Examples include using shorter piston skirt, smaller bearing width and diameter, and fewer and thinner piston rings. However, this needs to be done carefully without decreasing the oil film thickness or increasing the bearing temperature to an unacceptable level. Decreasing the number of compression rings tends to increase blow-by. As shown in the Stribeck diagram, there is a minimum achievable friction coefficient that is located at the verge before the transitioning from the hydrodynamic to the mixed lubrication regime. Reducing the load carrying area increases the unit load and decreases the oil film thickness. This reduction in oil

film thickness may offset the benefit of reduced area, as Kovach *et al.* (1982) showed in their piston skirt experiment. Moreover, if the design change causes the oil film thickness to fall into the mixed lubrication regime, the coefficient of friction starts to increase again. For example, if the piston skirt length reduction causes a drastic decrease in the side-thrust carrying capacity so that the skirt-to-bore asperity contact happens, then there is no friction reduction. In fact, the friction may even increase due to wear or scuffing. Moreover, although reducing the bearing size (especially the diameter) may reduce the hydrodynamic lubrication friction, journal misalignment and edge loading should be avoided in order to prevent structural failures or scuffing.

- Using multigrade oils with high viscosity indexes can reduce friction. When lower viscosity oil is used to reduce friction, care needs to be taken on the improved oil formulation/additives and the wear resistant materials to prevent degradation in wear.
- For hydrodynamic lubrication, reducing the clearance at the frictional interface may reduce friction, especially at low speeds. However, this may cause mechanical noise problems, cylinder liner cavitation due to worse piston slap and a larger oil flow requirement that result in a higher oil pump power. Moreover, changing the clearance usually affects the operating oil viscosity.
- For hydrodynamic lubrication, in order to increase the overall oil film thickness on the lubricated contact (i.e., for the purpose of reducing the viscous friction force) or promote a better lubricating oil film pressure distribution (e.g., through the lubrication 'wedge' effect or the 'squeeze film' effect), component profile design is an important measure for friction reduction and tribological motion control. Examples include the piston skirt's barrel profile and the piston ring's off-centered barrel profile.
- For mixed or boundary lubrication, if the load cannot be reduced, in order to reduce the coefficient of friction, design measures to change the lubrication regime to hydrodynamic should be considered. This can be done by changing the component clearance or profile or by decreasing the surface roughness to increase the film thickness-to-roughness ratio.
- For boundary lubrication, although the lubricant viscosity is not important for the friction characteristics, friction-modifier additives in the lubricant (e.g., the extreme pressure additive) may establish a crucial thin reaction film in the boundary lubrication, and this may reduce the boundary friction coefficient hence reduce friction. When the lubricant viscosity is reduced, the frictions from the piston assembly and the bearings can be reduced, but the valvetrain friction may increase because the valvetrain (e.g., cam–follower) operates mainly in the boundary or mixed lubrication regime. Friction-modifier oils may reduce the boundary friction in the engine, if the oil is supplied effectively into the lubricated interface.

- Using special face materials or coatings on the piston rings and the piston skirt can reduce friction. Changing the bearing material may substantially reduce friction in the mixed or boundary lubrication regime.

A comprehensive list of detailed design methods for power cylinder friction reduction are provided by Richardson (2000).

10.3.4 System design effects on engine friction

An engine system design engineer often faces the challenges of selecting appropriate system-level designs and operating parameters with a minimum penalty on engine friction. The system design parameters usually include the following: engine displacement, the number of cylinders, cylinder centreline distance, cylinder bore, engine stroke, connecting rod length, the ratio of connecting rod length to crank radius, valve size, valvetrain configuration parameters, crankshaft offset, piston compression height, piston pin offset, piston assembly weight, engine weight, engine compression ratio, cold-start friction torque, peak cylinder pressure, and engine speed range. An accurate modeling of the impact of these parameters on engine friction has a direct influence on system design results.

Engine testing results indicated that the engine friction reduces as the cylinder size or the number of cylinders increases (Monaghan, 1988). Therefore, for a given engine displacement, the choice of the number of cylinders may have a complex impact on engine friction. Ciulli (1993) supported the previous published experimental work (e.g., Bishop, 1964) by stating that fewer larger cylinders result in lower friction. In general, a shorter engine stroke (Millington and Hartles, 1968), a lower compression ratio, or lower peak cylinder pressure can reduce overall engine friction. Bishop (1964) showed that the stroke-to-bore ratio has only a very small effect on the tested engine friction. Patton *et al*. (1989) used a friction model to find that the total engine friction decreases with decreasing stroke-to-bore ratio. Millington and Hartles (1968) showed that the ratio of connecting rod length to crank radius has little effect on the piston friction. Larger bore distortion usually increases friction (Monaghan, 1988). Offsetting the crankshaft may significantly affect the piston side thrust and the sliding speed of the piston, hence reducing the piston skirt friction force.

Engine friction fundamentals are systematically introduced by Heywood (1988), Ferguson and Kirkpatrick (2001), and Taylor (1993a). The mechanisms of engine mechanical friction for each major component are introduced by Comfort (2003). Overviews on engine friction are provided by Rosenberg (1982), Parker and Adams (1982), Kovach *et al*. (1982), Furuhama (1987), Monaghan (1988), Parker (1990), Wakuri *et al*. (1995), Ciulli (1992, 1993), and Richardson (2000).

10.4 Piston-assembly lubrication dynamics

10.4.1 Friction characteristics and piston assembly design

From the perspective of multi-body dynamics modeling, the piston assembly includes the piston skirt, the rings, the piston pin (SAE J2612, 2002), and the connecting rod. The discussion in this section focuses on the lubrication dynamics of the piston body and the skirt. The piston rings will be discussed in Section 10.5. The piston pin bearing, the connecting rod big-end bearing, and the crankshaft main bearing are discussed in Section 10.6.

The piston in modern diesel engines needs to have strong thermo-mechanical strength to withstand the peak cylinder pressure and the temperature. It also needs to have low friction and wear, low piston slap noise, good control on skirt distortion, proper cooling and lubrication, and light weight (for high-speed operation), and match with the combustion chamber shape. Piston rings seal the in-cylinder gas with little blow-by, transfer heat from the piston to the cylinder liner, and control the lube oil consumption. Low-friction piston assembly design is very important for fuel consumption. Both analytical and experimental methods play critical roles in piston tribological design.

Piston slap has been identified as the main cause of the cylinder liner cavitation in heavy-duty diesel engines due to the high impact energy of the slap. Piston slap noise can also be the most prominent mechanical noise during warm-up. Mechanical noises emitted from the engine surface are caused by the impact between the components and the resulting vibration. They become louder as the engine speed increases. In piston tribological design, the piston skirt friction power, the dynamic minimum lubricating oil film thickness, and the cold piston slap kinetic energy (or noise) are the three most important performance parameters that should be optimized simultaneously.

While undergoing a primary reciprocating motion inside the cylinder, the piston is pushed laterally by the alternating side thrust force from the thrust side to the anti-thrust side, and vice versa, several times within the skirt-to-bore clearance during an engine cycle. The side thrust force is generated by the reaction force from the connecting rod small end, which is a force generally not in the piston's sliding direction, to resist the combined gas and inertia forces. The lateral (transverse) motion is accompanied by a small tilting motion around the piston pin due to the moments acting on the piston from various forces. These secondary motions cause the piston to slap on the cylinder bore at skirt top and bottom, accompanied by a sliding motion changing from mixed lubrication at the TDC or the BDC to hydrodynamic lubrication at the mid-stroke. Piston-assembly dynamics is an area related to piston skirt secondary motions (lateral and tilting) and their associated piston slap, friction, and wear. Its research started in the 1960s, evolving from the single-body dynamics without lubrication to the more complex multi-body dynamics coupled with elastohydrodynamic lubrication models.

The piston secondary motions not only affect the piston slap noise (to be detailed in Chapter 11), but also affect the piston ring operation and wear. In order to achieve low blow-by and oil consumption, the piston lateral movement and titling need to be minimized to provide a good platform for the piston rings to operate (Wacker *et al.*, 1978). Large piston tilting may cause the rails of the oil control ring to move away from the cylinder wall so that the ring loses its oil control ability. The piston secondary motions also affect the diesel cylinder bore polishing which is caused by the top land carbon deposit (Guertler, 1986). Moreover, the piston secondary motions affect piston skirt friction, and this topic is detailed below.

Piston skirt lubrication and friction are affected by the skirt-to-bore clearance and the oil film thickness distribution on the skirt which is related directly to the side thrust and the piston secondary motions. The piston-assembly friction of diesel engines can account for 40–55% of the total engine mechanical friction (Richardson, 2000), with the piston skirt contributing about 15–20%, the piston rings 15–20%, and the connecting rod 10–14%.

Piston skirt friction characteristics have been researched extensively. Feuga and Bury (1984) measured the friction force and power loss of a gasoline piston–ring–liner assembly at various engine speeds and loads, and with different lubricant oil grades. They found that the piston-assembly FMEP (including the rings) increased with engine speed and load. At higher loads the frictional loss in the expansion stroke was higher than that in other strokes, and this characteristic was especially prominent at low speeds. Wakabayashi *et al.* (2003) found from measurement that the effect of gasoline engine load on the piston skirt friction during the expansion stroke was small. Nakayama *et al.* (1997) measured the piston skirt friction force of a gasoline engine and found that the piston skirt could operate in the mixed lubrication regime in the expansion stroke, and the friction force in the first half of the expansion stroke could be reduced by changing the piston pin offset from the thrust side to the anti-thrust side.

Many design parameters affect piston-assembly dynamics. For example, the secondary movement can be reduced or diminished by using smaller clearance, piston pin offset or optimized skirt profile. An early comprehensive summary of the design guidelines was given by Winship (1967). He pointed out that a thin thickness of the major thrust side of the upper skirt was especially important for noise control, and a shorter skirt length (measured from the piston pin centerline) affected piston tilting, blow-by, and oil consumption. Oetting *et al.* (1984) illustrated that a reduced reciprocating mass and a long connecting rod (i.e., reduced crank–conrod ratio) reduced the piston side thrust and achieved low piston slap noise. Uras and Patterson (1987) measured piston friction and found that friction did not increase with increased piston weight.

Mansouri and Wong (2004) used numerical simulation for a natural gas engine to find that the piston skirt friction power was inversely proportional to the oil film thickness and approximately directly proportional to the surface waviness. The 'waviness' refers to the circumferential machining tracks of saw-tooth shape for oil retention on the surface to prevent scuffing. It is two orders of magnitude higher than the surface 'roughness'. They also found that a flatter skirt profile and a less rigid skirt could reduce the proportion of the boundary lubrication relative to the hydrodynamic lubrication friction. It should be noted that significant boundary lubrication scraping may occur during the expansion stroke, especially for the skirts having large waviness.

Crankshaft offset is another important parameter extensively investigated by many researchers (Haddad and Tjan, 1995; Nakayama *et al*., 2000; Wakabayashi *et al*., 2003; Shin *et al*., 2004). Crankshaft offset may reduce the piston side thrust and skirt tilting; and it may reduce piston friction during the expansion stroke if the friction occurs in the boundary or mixed lubrication regime. In fact, as long as the piston skirt is designed to stay within the hydrodynamic lubrication regime, any side thrust change basically would not affect the skirt friction appreciably. Crankshaft offset has a strong interaction with piston pin offset and piston pin vertical position. Crankshaft offset may either promote or destroy the hydrodynamic lubrication on the skirt, resulting in different conclusions on piston skirt friction accordingly.

Piston skirt profile design was addressed using an experimental method by Yagi and Yamagata (1982). They used a composite material (epoxy resin) to cover the piston skirt surface and run the engine to naturally wear off the material at highly stressed areas. Finally the smoothly curved, barrel-shape piston skirt profile defined by the composite material remaining after the marking test was adopted as the final optimized cold profile. With this method, different profiles to minimize wear and scuffing at the thrust and anti-thrust sides can come out of a real engine test. It should be noted that the profile produced by such a marking method minimizes the wear and scuffing only for a given piston design (e.g., with a given pin offset). Different piston designs may produce different resulting profiles. This method cannot identify which design gives the desirable minimum piston tilting. Piston dynamics simulation is able to complement that aspect and is an important tool in piston design.

Teraguchi *et al*. (2001) conducted experimental work on a small diesel engine and found that a forced oil supply remarkably reduced the skirt friction force by 20% in the expansion stroke on the thrust side and the latter half of the exhaust stroke, without a significant penalty on oil consumption. They found that the effectiveness was equivalent to the friction reduction achieved by using a MoS_2 coated skirt.

10.4.2 History of piston-assembly lubrication dynamics

Although piston slap can be modeled effectively with multi-body dynamics to calculate the side thrust and the titling moment without using a skirt lubrication model, such a model cannot be used to predict skirt friction accurately because in reality the skirt operates mainly in the hydrodynamic lubrication regime and hence the oil film thickness simulation becomes critical. Dry contact modeling with Coulomb friction yields unrealistically high skirt friction force compared to the normal engine operation with lubrication. Piston-assembly dynamics modeling without lubrication will be discussed in more detail in Chapter 11 for the topic of piston slap noise.

Unlike the piston ring lubrication which can be reasonably simplified as a one-dimensional problem of lubricating oil film pressure distribution only along the axial direction because the ring can be regarded circumferentially uniform, the piston skirt lubrication must be modeled as a two-dimensional problem for oil film pressure distribution. The lubrication modeling for the piston skirt with the rigid-body assumption and the numerical solution of the two-dimensional Reynolds equation was conducted by Knoll and Peeken (1982), Li and Ezzat (1983), Zhu *et al.* (1992), Chittenden and Priest (1993), Nakada *et al.* (1997), and Livanos and Kyrtatos (2006). Oil starvation and cavitation modeling is important for predicting the piston secondary motions. Keribar and Dursunkaya (1992a) showed in simulation that the piston secondary motions under fully flooded and partially flooded skirt lubrication were significantly different.

Piston thermal deformation was simulated with a finite element model by Li (1982). Piston skirt elastic deformation was simulated by Kimura *et al.* (1999). The piston skirt experiences significant deformations caused by thermal expansion, mechanical loading, and lubricating oil film pressure, especially for the thinner, more flexible light-duty or articulated piston skirts. Typical skirt deformations are of the same order of magnitude as, or larger than, the skirt-to-bore clearances. Elastohydrodynamic lubrication for the piston skirt has been modeled by Oh *et al.* (1987), Blair *et al.* (1990), Goenka and Meernik (1992), Keribar and Dursunkaya (1992a, 1992b), Dursunkaya and Keribar (1992), Keribar *et al.* (1993), Zhu *et al.* (1993), Dursunkaya *et al.* (1993, 1994), Wong *et al.* (1994), Knoll *et al.* (1996), Scholz and Bargende (2000), Offner *et al.* (2001), and Shah *et al.* (2007).

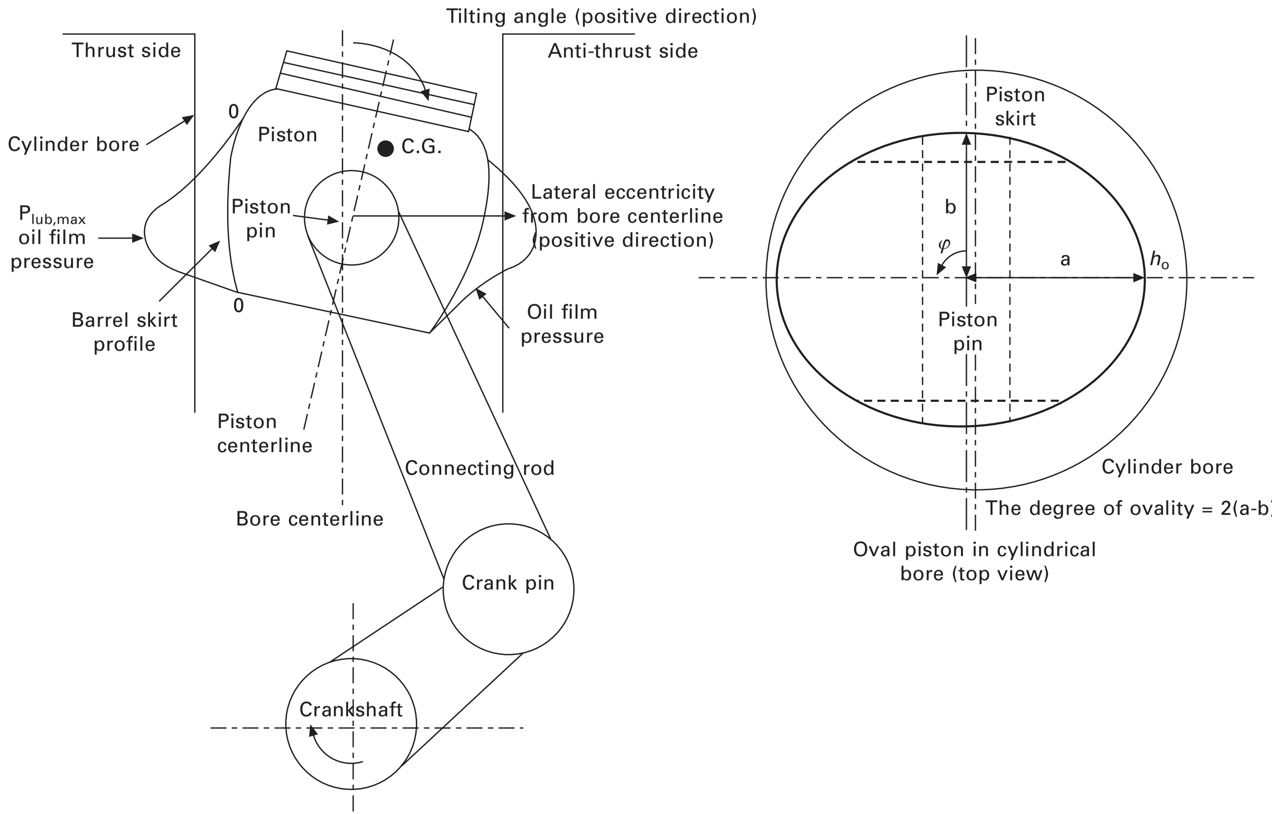
Goenka and Meernik (1992) compared three lubrication models with experimental data. The three models are: (1) a simple model considering only the ‘squeeze effect’ and the lateral motion by ignoring the ‘wedge effect’ and the tilting motion; (2) a rigid-body hydrodynamic lubrication model by ignoring the thermal expansion and the compliance; and (3) a mixed-elastohydrodynamic lubrication (DEHD) model. They concluded that both the rigid-body and DEHD models could predict piston skirt friction reasonably well, while the simple model was only acceptable for trend predictions. The DEHD model was recommended for more accurate analysis used in component design. Dursunkaya *et al.* (1993) pointed out that, compared to the elastohydrodynamic lubrication simulation, the rigid skirt hydrodynamic lubrication model could yield large discrepancies in the boundary friction. Knoll *et al*. (1996) used a finite element method to present small differences in piston secondary motions between the rigid and the elastic skirt simulation results for a diesel engine. Carden *et al*. (2006) concluded that a relatively simple model could be used to provide credible predictions of piston assembly friction loss even if the model did not have the advanced features such as elastohydrodynamic lubrication, finite element, and bore distortion.

Sophisticated piston skirt lubrication dynamics modeling can be conducted with commercial software packages, such as Ricardo's PISDYN (Keribar *et al*., 1993; Carden *et al*., 2006) and AVL's GLIDE. The simulation of deformable-piston dynamics coupled with elastohydrodynamic lubrication can also be conducted with commercial finite-element software packages, as performed by several authors mentioned above (Knoll *et al*., 1996; Scholz and Bargende, 2000; Offner *et al*., 2001).

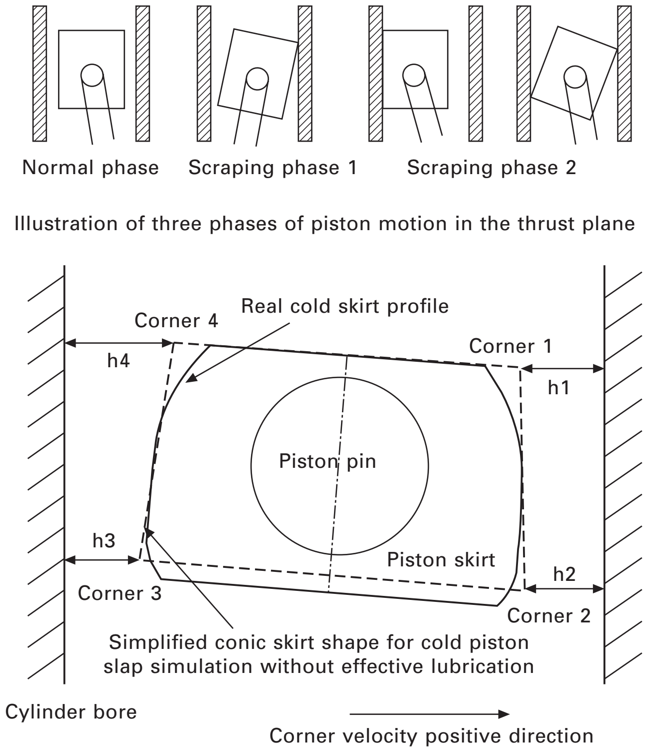
10.4.3 Formulation of piston-assembly lubrication dynamics

An analytical model of piston lubrication dynamics is shown in Fig. 10.6. The following design and operating parameters affect the piston dynamics: (1) piston mass (including the piston pin); (2) connecting rod mass; (3) piston tilting moment of inertia; (4) connecting rod rotating moment of inertia; (5) piston skirt-to-bore clearance; (6) piston center of gravity positions (lateral and vertical); (7) piston pin positions (lateral and vertical); (8) piston skirt length; (9) piston skirt lubrication wetted arc angle; (10) piston skirt axial profile and ovality; (11) lubricant oil viscosity; and (12) engine speed and load. The outputs of the model include the piston primary motion (i.e., the reciprocating sliding motion) and the piston skirt secondary motions (i.e., the lateral and tilting motions within the skirt-to-bore clearance), the lubricating oil film thickness and the pressure distribution on the piston skirt, the friction forces in the hydrodynamic lubrication regime and the metal-to-metal contact scraping regime.

The analytical model includes piston assembly multi-phase multi-body dynamics coupled with piston skirt lubrication. The side thrust (the lateral force acting on the piston pin) results from the cylinder gas pressure and the inertia forces, and induces the piston lateral motion. Piston tilting is caused by the moments acting on the piston skirt from various forces (i.e., the side thrust, the vertical force acting on the piston pin, the cylinder pressure, the lubricating oil film forces (also called lubricant forces), the friction forces between the skirt and the rings, and the piston pin friction force). The multi-phase used in the dynamic modeling includes a normal phase without scraping, a single-corner or single-location scraping phase, and a two-corner or multi-location scraping phase (Fig. 10.7). The scraping refers



10.6 Illustration of piston-assembly lubrication dynamics model.

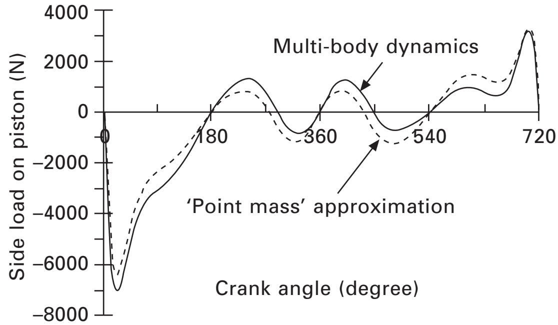


10.7 Multi-phase dynamics model of the piston and definition of piston skirt corners.

to the boundary lubrication. In each phase, the dynamics equations can be formulated for each component of the piston assembly (i.e., the piston skirt and the connecting rod) and the crankshaft based on the force balance using Newton's second law $F = ma$ for the lateral and vertical motions and the moment balance for the rotation. More details about the dynamic equations in piston-assembly dynamics modeling are provided in Shiao and Moskwa (1993) and Xin (1999).

The multi-body dynamics modeling approach is much more accurate than the commonly used 'point mass' simplification of the connecting rod for piston side thrust calculations. The 'point mass' method treats the connecting rod as two point masses concentrated at the small end and the big end, respectively, rather than as a rigid body. An example of the importance of accurate formulation of the piston skirt side thrust calculation is given in Fig. 10.8. It shows that the error due to the simplified 'point mass' approximation can be as great as 20%.

The transition between different phases of the motion is handled by rigid-body impact dynamics based on the impulse and momentum balances. For



10.8 Piston side thrust calculated with different methods (1800 rpm, 70% load).

example, when the oil film thickness becomes lower than a certain threshold value, piston-to-liner impact is assumed to occur with a non-zero piston lateral impact velocity. Rebound occurs after each impact, and a coefficient of restitution is assumed for calculating the piston lateral velocity after the rebound. When the piston impact velocity becomes sufficiently small after a series of impact and rebound events and if there is still a non-zero normal force pushing the piston against the bore, a scraping motion is assumed to occur with boundary lubrication. When the normal force vanishes as the piston side thrust changes, the piston may leave the scraping phase and switch to a normal phase. Such a sophisticated high-fidelity model may not only predict the friction during the scraping phases, but also simulate the transient piston slap behavior before the scraping occurs.

The lubrication model is based on the viscous fluid Reynolds equation. The equation can be solved with the finite-difference numerical method for the lubricating oil film pressure distribution after the piston skirt secondary motions are computed at each time step. The three-dimensional oil pressure distribution is then integrated over the skirt surface to obtain the lubricant force and moment, which are used in the piston dynamics model at each time step. The Reynolds equation governing the lubricating oil film pressure on the piston skirt is given as (Xin, 1999):

$$\frac{\partial}{\partial y}\left(h_o^3 \frac{\partial p_{lub}}{\partial y}\right) + \frac{1}{r_P^2}\frac{\partial}{\partial \varphi}\left(h_o^3 \frac{\partial p_{lub}}{\partial \varphi}\right) = -\,6\mu_v v_P \frac{\partial h_o}{\partial y} + 12\mu_v \frac{\partial h_o}{\partial t} \qquad 10.24$$

where y represents the axial direction of the piston skirt, φ represents the circumferential direction of the skirt, r_P is the piston radius, h_o is the oil film thickness, p_{lub} is the lubricating oil film pressure, μ_v is the dynamic oil viscosity, v_P is the piston sliding velocity, and t is time. Equation 10.24 shows that the oil film pressure, the gradient of the oil film pressure and hence the viscous shear friction force all increase linearly with piston sliding velocity and oil viscosity.

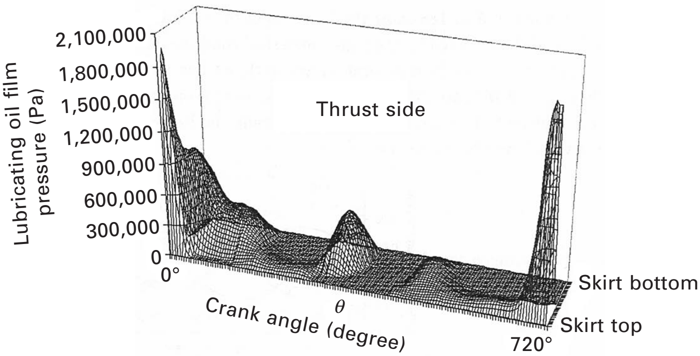
The viscous shear friction force is calculated by integrating the viscous shear stress over the lubricated area on the piston skirt. The viscous friction force increases with the skirt length and the piston diameter, and decreases when the skirt-to-bore clearance increases. The hydrodynamic friction force can be calculated as the summation of the viscous shear term and the hydrodynamic pressure term, ignoring the translation or squeeze term. The total piston skirt friction force is equal to the sum of the hydrodynamic friction force and the boundary lubrication friction force if any, given as:

$$\begin{aligned} F_{f,skirt} &= F_{f,skirt,h} + F_{f,skirt,b} \\ &= \iint \left(\mu_v \frac{v_P}{h_o} + \frac{h_o}{2} \frac{\partial p_{lub}}{\partial y} \right) \mathrm{d}A_c + \int_{A_c} f_{fri,b} \, p_{asperity} \, \mathrm{d}A_c \end{aligned} \qquad 10.25$$

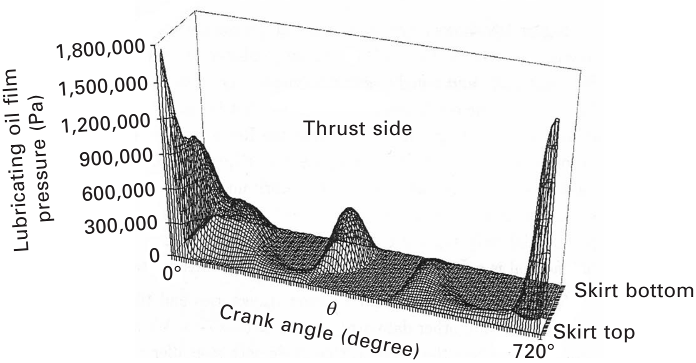
where $p_{asperity}$ is the asperity contact pressure used in the load-carrying asperity models of the mixed lubrication. For more comprehensive coverage about the squeeze term in the friction power of bearings, the reader is referred to Martin (1985) and Taylor (1993a). Note that on the cavitation side of the skirt, although the friction is usually reduced by a certain extent due to the air/vapor pockets in the ruptured lubricant film streams, the viscous shear friction force is still significant and cannot be completely neglected. The piston skirt friction power is equal to the friction force multiplied by the piston sliding velocity. The calculation of the friction torques at the piston pin bearing and the conrod big-end bearing can often be simplified by multiplying the resultant force with an assumed friction coefficient and the bearing radius.

Lubrication boundary conditions are important for the prediction of the lubricating oil film thickness and the pressure distribution on the piston skirt. The boundary conditions include fully flooded or partially flooded on the skirt surface, and also include a lubricant cavitation condition once a negative lubricating oil film pressure is calculated. The cavitation conditions usually include the non-mass-conserving half-Sommerfeld boundary condition and the mass-conservation Reynolds or Jakobsson–Floberg–Olsson (JFO) boundary condition. The comparison between different cavitation boundary conditions for the piston skirt lubrication is provided in Fig. 10.9. It shows that the half-Sommerfeld condition generally gives more conservative results in the minimum oil film thickness, i.e., smaller oil film thickness than that given by the Reynolds condition. However, it seems the Reynolds cavitation boundary condition is still the most appropriate cavitation condition for piston skirt lubrication. More sophisticated lubrication modeling such as the elastohydrodynamic lubrication, including the effects of bore distortion and skirt deformation as well as the more complex models of lubricant viscosity, may be used but the computing time will increase exponentially.

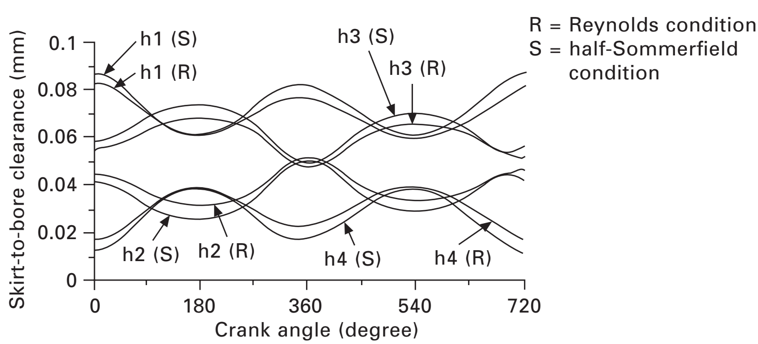
The piston skirt lubrication model can simulate both a cylindrical skirt and a non-cylindrical skirt, e.g., a barrel shape in the axial direction with



Pressure-time map on the thrust line (φ = 90°)
with the Reynolds cavitation condition



Pressure-time map on the thrust line
(φ = 90°) with the half-Sommerfeld condition



Trajectories of piston secondary motions (the
half-Sommerfeld condition and the Reynolds
condition)

10.9 Comparison between half-Sommerfeld and Reynolds boundary conditions of cavitation for piston skirt lubrication dynamics (2000 rpm 70% load).

ovality in the circumferential direction. Usually, a piston skirt has a barrel profile at cold. At hot conditions the barrel shape and ovality change as the piston expands thermally. As the piston moves laterally and tilts within the clearance, the lubricating oil film thickness changes dynamically at each moment during an engine cycle, and the oil film thickness distribution also changes spatially on the piston skirt. The piston skirt wear durability can be analytically characterized by the dynamic minimum oil film thickness within an engine cycle at the thrust side and the anti-thrust side. When different piston designs are compared, a larger minimum oil film thickness usually reflects lower risk of wear and scuffing.

The solution algorithm of the piston lubrication dynamics is briefly described as follows. Denoting the location of the crank pin center as 1 and the piston pin center as 2 in the piston-assembly dynamics model, the dynamics model of the piston skirt can finally be reduced to two second-order nonlinear ordinary differential equations (ODEs) for the piston skirt tilting angle β_2 and the piston pin lateral displacement x_2 as follows (Xin, 1999):

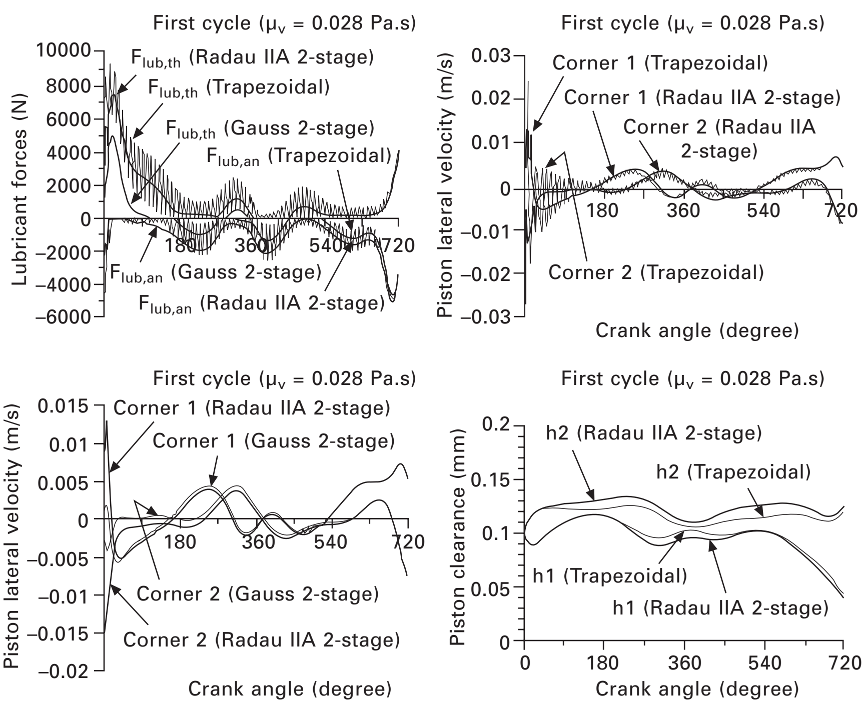
$$\begin{cases} \ddot{\beta}_2 = f_1(t, \beta_2, \dot{\beta}_2, x_2, \dot{x}_2) \\ \ddot{x}_2 = f_2(t, \beta_2, \dot{\beta}_2, x_2, \dot{x}_2) \end{cases} \tag{10.26}$$

Because of the nonlinear nature of the functions f_1 and f_2 in equation 10.26 (especially due to the lubrication force model), this ODE system cannot be solved analytically. Time-marching numerical integration is necessary. The accelerations $\ddot{x}_2$ and $\ddot{\beta}_2$ are calculated after the lubricant forces are calculated by using the lubrication model 10.24 at every time step. The piston lateral and tilting displacements and velocities must be solved by a time-marching implicit integration. The primary motion of the piston assembly is calculated based on the kinematic constraints between the components. The lateral and vertical forces acting on the piston pin and the crank pin are calculated by using the above-mentioned approach of force/moment balance.

10.4.4 Stiff ODE characteristics of lubrication dynamics

The stiff ODE feature is one of the most fundamental and important characteristics in lubrication dynamics. The governing equations of any lubrication dynamics problems are often stiff ordinary differential equations. The stiff ODE features are generated from the Reynolds equation coupled with the dynamics equations that include the acceleration terms (e.g., the piston mass inertia or moment of inertia). The stiffness refers to the largest eigenvalue $\lambda_{ODE,max}$ of the Jacobian matrix of the ODE system for the piston motion, equation 10.26, and to a stiffness ratio that is defined as the ratio of the largest eigenvalue to the smallest eigenvalue of the ODE system. The stiffness affects the numerical stability of the time-marching integration. The stiffness is affected by the design and dynamic parameters. The following parameters can make the $\lambda_{ODE,max}$ and the stiffness ratio larger (i.e., more stiff): smaller mass of the piston; smaller moment of inertia of the piston; smaller skirt-to-bore clearance; longer length of the skirt; larger radius of the skirt; higher oil viscosity; and larger dynamic eccentricity.

The high stiffness of the ODE system causes two types of numerical instability, one for all time-marching explicit integration schemes, and the other for some implicit integration schemes. The instabilities usually exhibit large high-frequency oscillations in the lubricant force trace as a function of time. These oscillations are numerical errors rather than true physical signals (Fig. 10.10). The instabilities are related to the following: the stiffness of the ODE system, the type of integration scheme (explicit or implicit), the stability property of the implicit integration scheme, the time-step size, and the deviation of the initial condition from the final periodic convergent solution. Explicit integration schemes usually cannot solve the stiff ODE problems due to their extremely large numerical instabilities. The larger the eigenvalue of the ODE, the smaller the time step that has to be used with the explicit integrators in order to avoid the numerical instability. The time step



10.10 Stiff ODE and numerical instabilities of implicit integration methods with weak stability properties (1800 rpm, 70% load).

required is approximately the reciprocal of the largest eigenvalue. Lubrication dynamics problems are generally highly stiff so that the time step which can be used in the explicit integrators becomes impractically small. This results in extremely long computing time and large accumulation of the round-off errors during the course of the integration.

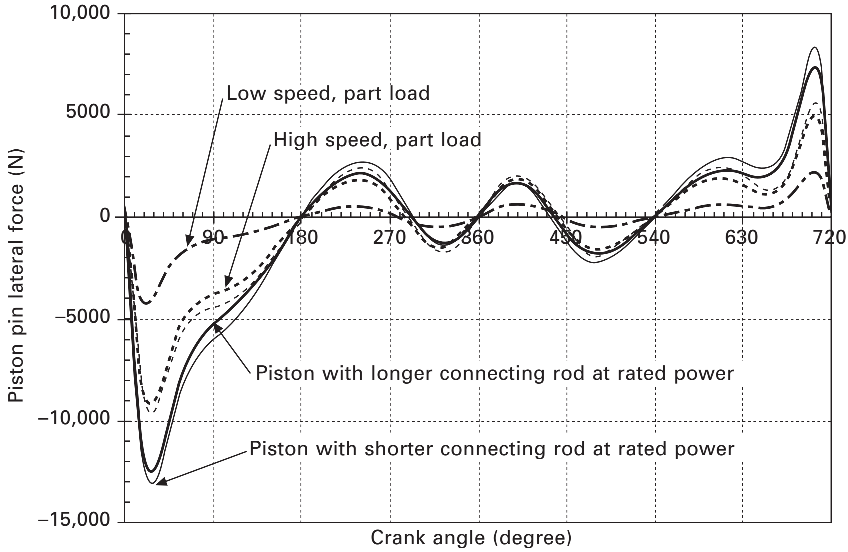
It should be noted that some implicit integration schemes with weak stability properties (e.g., the Gauss or Lobatto IIIA type implicit Runge-Kutta integration methods) also exhibit numerical instabilities during transient simulations of the lubrication stiff ODE problems. The instability exhibits in the lubricant force trace as oscillations (Fig. 10.10). Only the strongly B-stable implicit integrators (e.g., the Radau IIA type implicit Runge-Kutta methods) can compute both the lubricant force and the piston motions without numerical instabilities. Moreover, the order of accuracy is also important for time-marching integration. Higher stiff-order implicit integration schemes with strong B-stability should be used to compute the dynamic response of such stiff ODE systems. The stiff ODE features of piston lubrication dynamics are summarized in Fig. 10.10. The stiff ODE characteristics of lubrication dynamics have been extensively researched by Xin (1999).

10.4.5 Effect of piston design on skirt lubrication and friction

Figure 10.11 shows the calculated piston side thrust within one engine cycle at three speed and load conditions. It is observed that a longer connecting rod reduces the piston side thrust. Figure 10.12 illustrates the computed lubricating oil film pressure distributions on the piston skirt. The oil pressure is generated by a combination of lubrication 'wedge' and 'squeeze film' effects caused by the piston secondary motions. The zero pressures (i.e., equal to the ambient gauge pressure) are caused by oil film cavitation. Note that different piston skirt profiles may generate distinct patterns of the oil pressure distribution on the thrust and anti-thrust sides. For example, during the expansion stroke, a cylindrical skirt profile produces the oil pressure only at the thrust side, while a barrel skirt profile generates the oil pressure at the lower portion of the skirt on both the thrust and anti-thrust sides.

A comparison on the piston secondary motions between using the lubrication model and ignoring lubrication in piston dynamics is provided in Fig. 10.13. It shows the great difference in predicted piston lateral velocities.

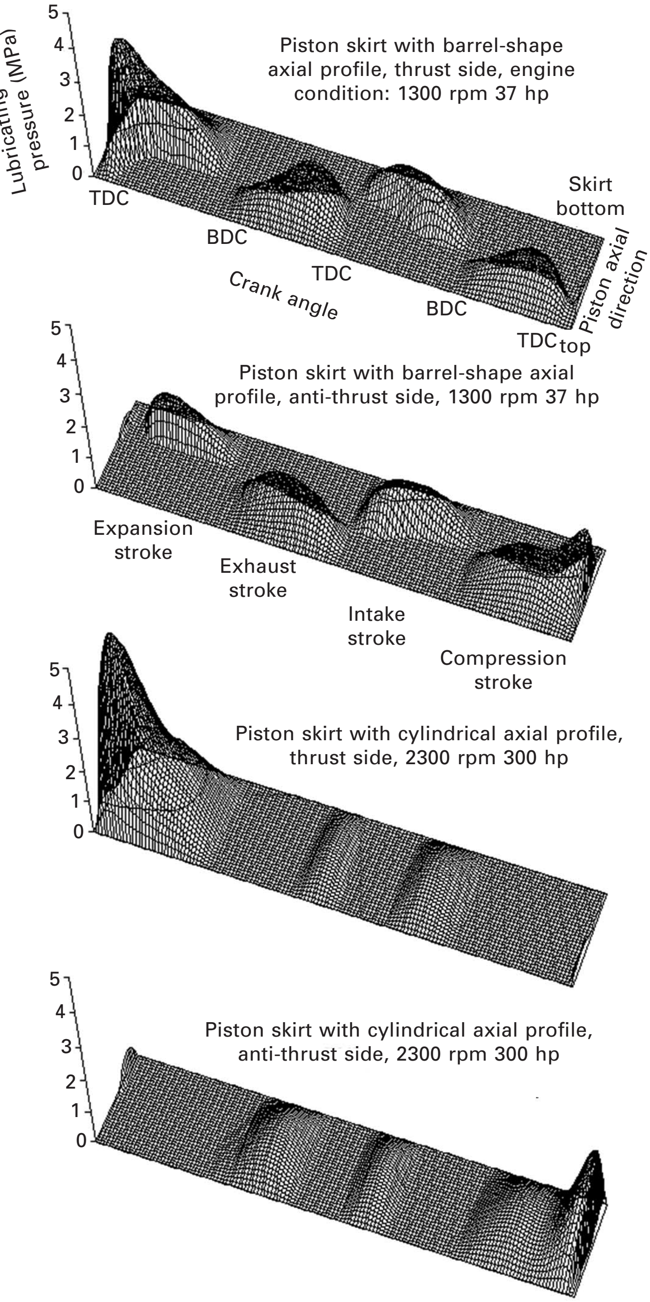
The effects of the piston pin lateral offset and the vertical position of the piston's center of gravity are shown in Figs 10.14 and 10.15, respectively. It is observed that they both have a great effect on piston skirt tilting or scraping. The effect of lubricant oil viscosity is shown in Fig. 10.16. The piston skirt hydrodynamic friction power increases with oil viscosity. Moreover, when piston scraping occurs, the piston friction power increases significantly due



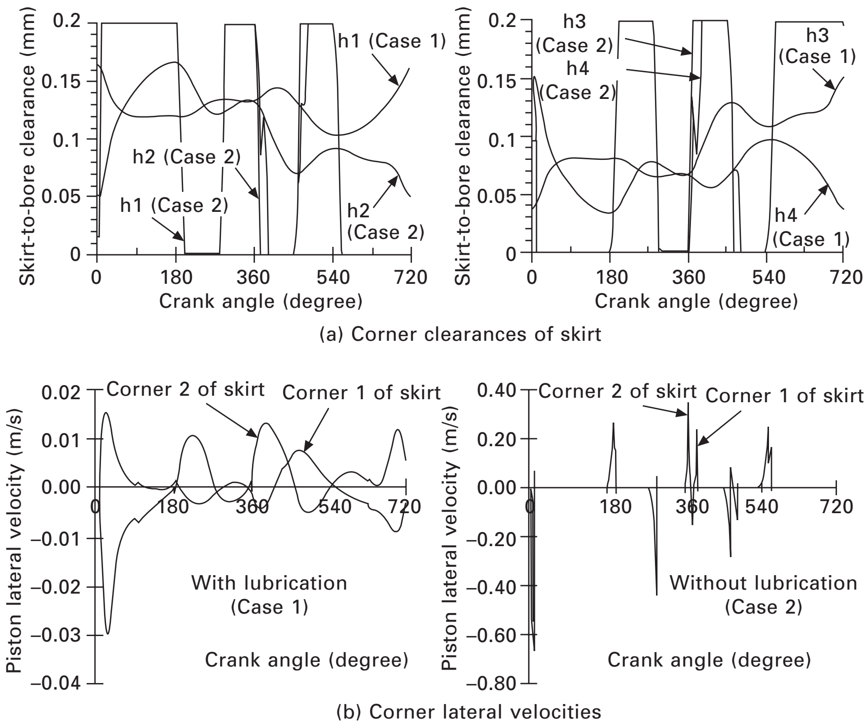
10.11 Illustration of piston side thrust at different engine speeds and loads.

to the higher friction coefficient of the boundary lubrication in the surface asperity contact during the scraping.

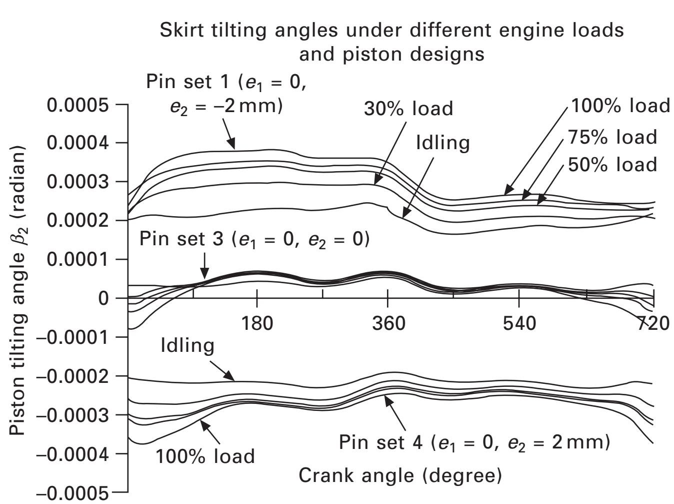
Moreover, simulation shows that as long as the piston skirt is designed to operate within the hydrodynamic lubrication regime, at the same engine speed but different loads, the piston skirt hydrodynamic lubrication friction power varies only in a range of 0.2–2%, depending on the piston skirt profile. However, different piston pin positions can result in a difference of up to 10% in the hydrodynamic lubrication friction power of the piston skirt. Higher friction corresponds to more violent secondary motions given by the unfavorable piston pin position. Piston weight has a very small influence on piston skirt friction power. It also has a negligible effect on the oil film thickness. But piston weight reduction may decrease the kinetic energy and the noise of piston slap in cold conditions. Simulation analysis has revealed that an effective way to reduce piston skirt friction is to reduce the skirt area (i.e., the skirt length or the lubricating wetted arc angle along the circumferential direction of the piston). The reduction in skirt friction power is basically linearly proportional to the reduction of the lubricated skirt area. Figure 10.17 simulates the complex nature of the piston secondary motions within the piston-to-bore clearance at a high engine speed. The piston motions, the minimum oil film thickness, and the skirt friction are affected by piston design parameters.



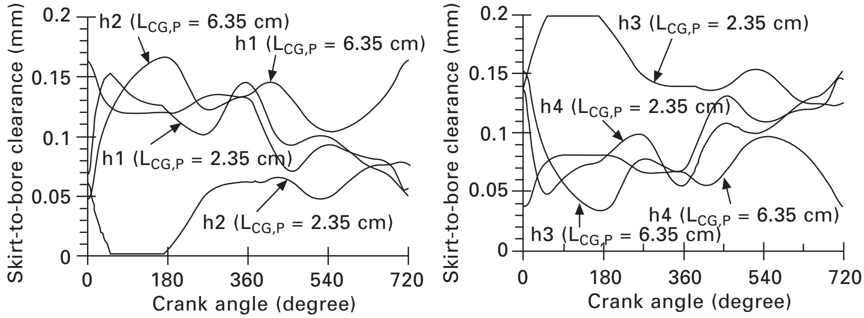
10.12 Lubricating oil film pressure distribution on piston skirt.



10.13 Illustration of piston secondary motions with and without skirt lubrication at 1800 rpm 70% load.



10.14 Effect of piston pin lateral offset and engine load on piston tilting at 1500 rpm.

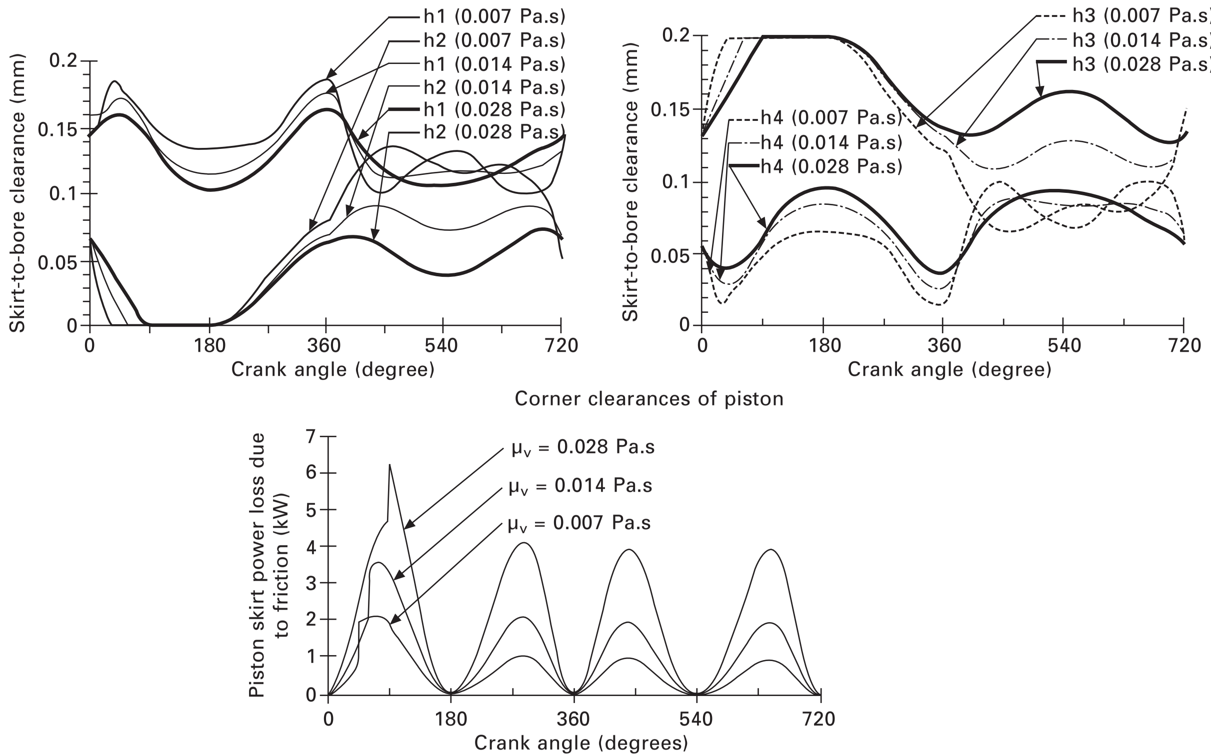


10.15 Effect of the vertical position of piston center of gravity on piston secondary motions at 1800 rpm 70% load.

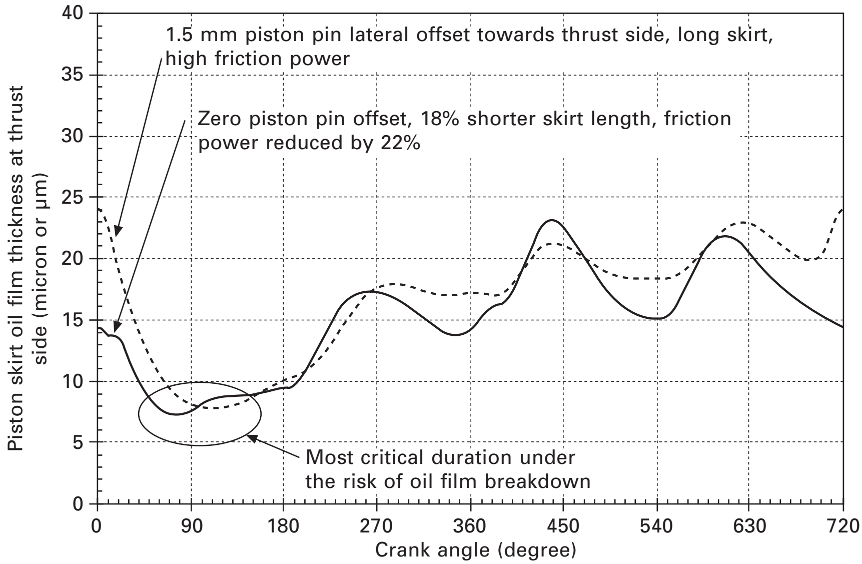
10.5 Piston ring lubrication dynamics

10.5.1 Friction characteristics and piston ring design

Piston rings have the functions of sealing the cylinder gas to control blow-by, transferring the heat, and controlling the oil consumption that are related to smoke and particulate emissions. The blow-by refers to the escape of the compression and combustion gases past the piston and the rings into the crankcase, and is related to power loss and crankcase emissions. High blow-by and high oil consumption often occur simultaneously. Sealing is achieved by the gas forces acting on the top and the back of the rings and also by the ring tension. Piston rings usually account for the largest portion in mechanical friction and are subject to wear. Compared to the compression rings, the oil ring usually has greater friction force due to its higher tension and thinner thickness. A modern piston ring pack has usually two compression rings and one oil-control ring. The top ring's function is primarily sealing the gas pressure. The oil ring's function is oil control. The function of the second ring is usually in between and more for oil control. The compression rings may have various ring face profiles (e.g., barrel – most common, plain, taper), different cross-sectional shapes (e.g., taper – most common for oil control to scrape the oil off the cylinder wall during the downstrokes, keystone, rectangular, bevel step, bevel cut), and different types of twist for sealing (i.e., no-twist, positive twist, negative twist). As pointed out by Tian *et al.* (1996a, 1997), a positive static twist can stabilize the ring when the land immediately above the ring has a higher gas pressure, while a negative twist can generate ring flutter if the gas force and the axial inertia force are comparable in magnitude. For example, the second compression ring may flutter during the late part of the compression stroke and the early part of the expansion stroke. The top compression ring usually has a barrel profile. The second compression ring is commonly



10.16 Effect of lubricant oil viscosity on piston secondary motions and skirt friction power at 1800 rpm 70% load.



10.17 Effect of piston design (skirt length and piston pin offset) on skirt friction power and oil film thickness.

designed to help control the oil by a scraping motion during the downstroke. The piston rings may have different circumferential free shapes with distinct characteristics of the tension force when they are placed inside the cylinder bore. Some ring design takes advantage of the combustion chamber pressure for sealing so that the ring tension can be reduced. The size of the end gap affects blow-by, the gas pressure between the rings and ring dynamics. The oil-control ring usually has two thin rails separated by a spring/expander. The oil-control ring usually has the highest tension and contributes most to the engine friction. All these design features, together with the piston skirt motions and cylinder bore distortion, affect the lubrication dynamics and the friction force of the ring. Effective lubrication has to be achieved with minimum oil consumption, friction, and wear.

The piston rings develop hydrodynamic lubrication (with a friction coefficient of 0.001–0.002) in the middle of the engine strokes and encounter mixed to boundary lubrication (with a friction coefficient of 0.08–0.12) as they approach the TDC and BDC. The ring–liner wear is usually most severe immediately after the firing TDC, when the cylinder gas loading and temperature are high and the ability to develop protective hydrodynamic lubrication oil film is hampered by low sliding speeds.

Piston ring tribology is a highly sophisticated and specialized area. There are numerous publications and many review papers in this area. An informative review on the historical development of piston ring technology can be found in Economou *et al.* (1979). Literature reviews on piston ring tribology are provided by McGeehan (1978), Ting (1985), and Andersson *et al.* (2002).

Excessively thick oil film and high oil consumption cause particulate emissions problems. The oil consumption via the piston rings is believed to be the largest portion of total oil consumption in an engine, compared with other sources such as the oil loss via the valve guides/seals, the oil leakage through various seals (e.g., a turbocharger seal), and evaporative oil consumption from the cylinder wall. The oil consumption via the rings is generated from the throw-off of accumulated oil above the top ring, oil blow through the top ring end into the combustion chamber, and the oil scraping by the edge of the piston top land. The oil consumption rate due to piston ring pack and blow-by usually increase with engine speed and load. Inter-ring pressure control to stabilize the ring motions is important in reducing oil consumption. The compression rings are often pushed against the upper side of the groove during the intake stroke and stay on the lower side of the groove during other strokes. The second compression ring often lifts up around the firing TDC. The axial motion can easily become much more complex for a particular design and engine operating condition. Piston ring axial motions and their effect on oil consumption were studied experimentally by Furuhama *et al.* (1979) and Curtis (1981). The oil consumption mechanisms were researched by De Petris *et al.* (1996), Hitosugi *et al.* (1996), Thirouard *et al.* (1998), and Herbst and Priebsch (2000). The effects of piston skirt lubrication on the mechanism of oil supply to the oil ring and oil consumption were explored by Ito *et al.* (2005) and Nakamura *et al.* (2005, 2006). The effects of piston ring design, cylinder bore distortion and ring conformability on oil consumption were discussed by Wacker *et al.* (1978), McGeehan (1979), and Essig *et al.* (1990). Simulation models of oil consumption were developed by Maekawa *et al.* (1986) and Audette and Wong (1999).

Piston ring friction is affected by many operating and design parameters, such as lubricant viscosity, piston sliding speed, gas load, ring tension, ring face profile, ring width, number of rings, ring static and dynamic twist, piston tilting and ring inclination, bore distortion, ring conformability, surface roughness of the ring and cylinder liner, break-in, and the quantity of lubricant available (oil starvation). The piston ring pack friction force increases with engine speed mainly due to the increase in hydrodynamic lubrication friction. As the engine load increases, the higher gas pressure on the back of the compression rings increases the friction force around the TDC and the BDC in the mixed lubrication regime, but the higher oil temperature and lower viscosity decrease the friction force at the mid-stroke in the hydrodynamic lubrication regime. Oil film thickness measurement at different loads was reported by Tamminen *et al.* (2006).

In EGR diesel engines, the piston friction force in the second half of the compression stroke increases as EGR rate, or essentially the soot amount in the EGR gas, increases, while the piston friction in the mid of each stroke reduces at high EGR rate (Urabe *et al*., 1998). The reason was believed that the soot deposited in the top ring groove made the piston rings slide on the soot particles attached to the cylinder bore surface and caused an increase in the boundary friction coefficient and wear on the components.

Piston ring axial dynamic motions (ring lifting, twisting, and fluttering) and groove design significantly affect blow-by and oil consumption. They may also impact ring friction and wear via the changes in inter-ring pressure and ring tilting. The ring motion is affected by the ring gap. Too large a ring gap results in excessively large blow-by, power loss, and emissions. Too small a ring gap leads to ring butting problems at high operating temperatures. The gap in the second ring can be properly designed to increase the top ring's sealing ability by preventing the inter-ring pressure from building up and lifting the top ring off the bottom of the groove.

Ring thickness significantly affects oil film thickness and ring friction. Furuhama *et al*. (1981) conducted experiments and theoretical calculations to find that the decrease of the ring thickness may not necessarily reduce friction because it decreases the oil film thickness at the same time. The reduced film thickness can increase the friction force and wear around the TDC and the BDC. Off-centered barrel shape profile has been found as the optimum for the top compression ring. The optimum choice of the degree of curvature of the barrel profile is a trade-off between the mid-stroke and end-stroke performance. A highly curved profile gives a large oil film thickness in mid-stroke due to the strong 'wedge effect', resulting in low friction. However, the oil film thickness falls rapidly around the TDC and the BDC, and the 'squeeze film effect' is weak there, resulting in high wear. A flatter ring face profile behaves oppositely.

Piston ring tension is another important parameter for friction. A piston ring design procedure to define a free shape producing circumferentially uniform contact force distribution was provided by Ma *et al*. (1996). Oil ring tension is an important design parameter since it determines whether the oil ring operates in the hydrodynamic or boundary lubrication regime. It also controls the amount of oil that is available for other rings. Uras and Patterson (1985) used the instantaneous IMEP experimental method to find that the friction of a 70 N high-tension oil ring decreased significantly after break-in, but still higher than the friction of a 17.8 N low-tension oil ring.

When the surface roughness increases, the ratio of the piston ring oil film thickness to roughness decreases. When the ratio is less than a threshold value, mixed lubrication occurs. A higher roughness results in a larger proportion of mixed lubrication and a higher friction power loss.

Oil viscosity is also important for ring friction. Friction modifiers in the oil may reduce boundary lubrication friction, with the degree of friction reduction varying based on the oil formulation. Oil starvation usually occurs more severely at full load firing than motoring. This contributes to thinner oil film thickness and higher friction at full load (Sanda *et al*., 1997). Uras and Patterson (1984, 1985) found that the piston assembly friction force increased when the oil viscosity became either too high or too low. As viscosity increases, the hydrodynamic lubrication regime is promoted and the boundary lubrication is depressed. Glidewell and Korcek (1998) conducted experimental work to find that oil starvation increased the friction coefficient, most noticeably in the mid-stroke region. They also found that the effectiveness of the friction modifier significantly degraded with aged engine oil. Richardson and Borman (1992) found that the temperature rise in the oil film due to viscous heating was negligibly small. They indicated that the cylinder inner wall temperature should be used to determine the oil viscosity for modeling.

10.5.2 Piston ring lubrication dynamics

Piston ring pack lubrication modeling was overviewed by Dowson (1993). The calculations of inter-ring gas pressure and blow-by (Kuo *et al*., 1989) provide key input data of gas pressure for ring dynamics and friction calculations. The gas mass flow rate between the adjacent volumes (at the ring gap, circumferential gap, and the ring side clearance) is calculated by using the orifice flow equations. Ring motion (e.g., ring radial collapse and axial fluttering) is very sensitive to the inter-ring pressures. There is a strong coupling between the inter-ring pressure and the ring motion. The characteristics of inter-ring gas pressures and ring lift motion were researched by Furuhama *et al*. (1979), Curtis (1981), Dursunkaya *et al*. (1993), Chen and Richardson (2000), and Herbst and Priebsch (2000) using experimental and numerical investigations.

The piston ring lubrication dynamics has several levels of complexity, from low to high as follows:

- one-dimensional Reynolds equation for a ring pack with the closed-form analytical solution with various lubrication cavitation boundary conditions (e.g., half-Sommerfeld, the mass-conservation Reynolds or JFO condition) and under fully or partially flooded boundary condition
- one-dimensional Reynolds equation with the closed-form analytical solution including surface asperity and mixed lubrication models
- one-dimensional Reynolds equation with numerical solutions
- one-dimensional Reynolds equation coupled with ring axial and twist dynamics
- two-dimensional Reynolds equation with numerical solutions
- two-dimensional Reynolds equation including surface asperity and complex mixed lubrication models

- two-dimensional mixed lubrication coupled with ring axial and twist dynamics
- including the features of elastohydrodynamic lubrication
- including some of the models of bore distortion, ring conformability, ring groove deformation, ring–groove contact pressure distribution, and lubricant rheological properties
- including some of the models of blow-by, oil consumption, ring face–liner wear, and ring groove wear.

For engine system design to estimate piston ring friction, usually the first and second levels of model mentioned above are sufficient. They also offer a real-time fast computation for an instantaneous crank-angle-resolution solution. Other higher level models are suitable for component designs.

A simple one-dimensional Reynolds equation for a piston ring can be written as

$$\frac{\mathrm{d}}{\mathrm{d}y}\left(h_o^3\frac{\mathrm{d}p_{lub}}{\mathrm{d}y}\right) = -\,6\mu_v v_P\frac{\mathrm{d}h_o}{\mathrm{d}y} + 12\mu_v\frac{\mathrm{d}h_o}{\mathrm{d}t} \qquad 10.27$$

where the instantaneous oil film thickness $h_o = h_{o,min} + h_{profile}(y)$, $h_{o,min}$ is the minimum oil film thickness, y refers to the axial direction of the ring, $h_{profile}$ (y) is the ring face profile as a function of the axial distance of the ring, and p_{lub} is the lubricating oil film pressure. The detailed derivation of the Reynolds equation was provided by Richardson and Borman (1992).

The lubricating oil film force per unit length of the ring is given as:

$$F_{lub,\,ring} = \iint_{inlet}^{outlet} p_{lub}\,\mathrm{d}A_c \qquad 10.28$$

The ring diametrical dynamics is given by a force balance as:

$$\tilde{m}_{ring}\ddot{x}_{ring} = \tilde{F}_{gas} + \tilde{F}_{tension} - \tilde{F}_{lub,\,ring} - \tilde{F}_{groove} \qquad 10.29$$

where x_{ring} is the diametrical displacement of the ring at each moment of time, and F_{groove} is the lateral friction force between the ring and the ring groove per unit length of the ring. With the mass inertia term $\tilde{m}_{ring}\ddot{x}_{ring}$ included in the dynamic formulation, equations 10.27 and 10.29 generate a stiff ODE system which requires a higher stiff-order implicit numerical integration scheme with strong B-stability (for B-stability, see Hairer and Warner (1996)).

Usually, the inertia term and the ring groove force are small and can be neglected for simplicity. Then, with the load balanced by the lubricant force in the normal (diametrical) direction, equation 10.29 becomes

$$\tilde{F}_{gas} + \tilde{F}_{tension} = \tilde{F}_{lub,ring} \qquad 10.30$$

which determines the required oil film thickness and the 'squeeze film' term to satisfy this force balance. The friction force of the piston ring can be calculated as

$$F_{f,ring} = F_{f,ring,h} + F_{f,ring,b}$$
$$= \iint \left(\mu_v \frac{v_P}{h_o} + \frac{h_o}{2} \frac{\partial p_{lub}}{\partial y} \right) \mathrm{d}A_c + \int_{A_c} f_{fri,b}\, p_{asperity} \mathrm{d}A_c \qquad 10.31$$

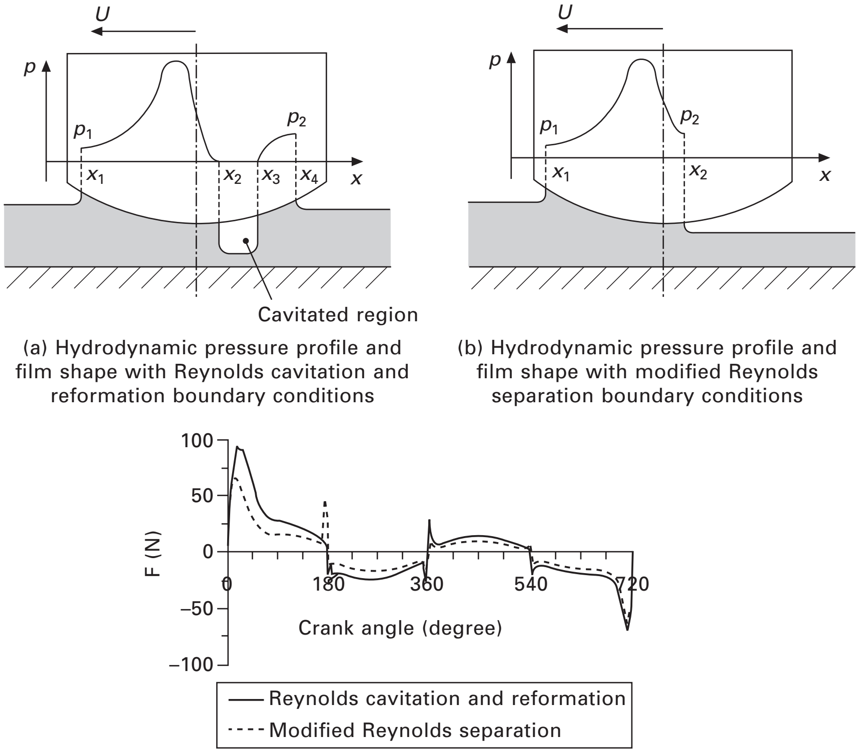
where $p_{asperity}$ is the asperity contact pressure of the ring. The first term in the hydrodynamic friction force is the viscous shear term due to the Couette flow, and the second term is the hydrodynamic pressure term due to the Poisseuille flow. The friction power loss due to the translation or squeeze term in the lubrication is neglected here.

The closed-form analytical solution for the one-dimensional Reynolds equation 10.27 was provided by Ting and Mayer (1974a, 1974b), Dowson *et al.* (1979), Wakuri *et al.* (1981), Furuhama *et al.* (1981), Jeng (1992a, 1992b, 1992c), and Sawicki and Yu (2000).

Oil starvation of the piston ring means the oil does not cover the entire ring face. Oil starvation modeling at the leading edge (i.e., inlet) of the ring can be as simple as assuming an effective ring thickness as input (e.g., assuming 50% of the ring face covered by oil), or can be as complex as predicting the oil starvation boundary with oil transport and flow continuity models. Oil starvation happens often on piston rings. Modeling the starvation is important for the prediction of oil film thickness and friction loss. The oil supplied to each ring is dependent upon the amount of the oil left on the cylinder wall by the preceding ring. In the downstroke the leading ring can be assumed fully flooded. Usually, fully flooded inlet condition can be assumed for the second rail of the oil control ring during the downstroke. Starved lubrication should be assumed for all other rings during the downstroke. Partially flooded lubrication should be assumed for all the rings during the upstroke. The compression rings have more severe oil starvation conditions during firing than motoring. During firing in the upward strokes, the inlet of the two compression rings and the upper rail of the oil control ring have oil starvation. In the downward strokes, the inlet of the two compression rings has oil starvation. Ma and Smith (1996) suggested that the bottom ring in a ring pack can be treated as fully flooded on the downstroke in the model. Oil starvation with the one-dimensional Reynolds equation in mixed lubrication was modeled by Sanda *et al.* (1997). They found the predicted oil film regions with starvation were as small as half of the total thickness of the compression rings, and the predicted oil film thickness with starvation was less than half of the oil film thickness predicted by the fully flooded condition for the compression rings, resulting in much greater friction force. They also found that the starved lubrication simulation for the oil control ring matched the measured data much better than the fully flooded simulation, and the starved simulation gave about 35% higher friction force than the fully flooded simulation data.

Cavitation and separation modeling at the trailing edge of the piston ring is important for the oil pressure distribution. The assumption of the cavitation boundary condition at the trailing edge of the ring in the modeling significantly affects the calculated hydrodynamic lubrication pressure distribution profile, load capacity, oil film thickness, and friction force of the ring. When negative pressures in the lubricating oil film occur, gas bubbles are formed as cavities from the dissolved gases or ventilation to the surroundings. The oil film ruptures to lose any hydrodynamic lubrication pressure. The cavitated oil film has the atmospheric pressure. The viscous shear force in a cavitated/ruptured oil film, although not as high as in a non-cavitated oil film, cannot be neglected. The full-Sommerfeld boundary condition unrealistically requires the lubricating oil film to continuously sustain large negative pressures. The half-Sommerfeld condition simply discards the negative pressures hence violates flow continuity and mass conservation. The Reynolds (Swift–Stieber) cavitation condition obeys mass conservation by requiring the pressure gradient equal to zero at the cavitation boundary. The Reynolds condition is the most commonly used condition for piston rings although other cavitation assumptions have been proposed in tribology (e.g., the flow separation boundary condition, the Floberg condition, the Coyne and Elrod separation condition, and the JFO boundary condition).

Unlike the piston skirt, a large gas pressure may exist at the outlet of the lubricated region for piston rings. This results in the oil film reforming after the cavitation by generating a hydrodynamic pressure gradually rising to the outlet gas pressure at the trailing edge of the ring. This cavitation–reformation sequence forms a cavitated 'pocket' within the oil film near the outlet. Pressure reformation may have strong effects on the peak oil film pressure in the full-film region and on the cavitation location (Yang and Keith, 1995). Some authors believe that when a negative pressure happens the oil film completely separates from the ring face and does not reform at the outlet. As a result, the outlet gas pressure is applied in the entire original negative pressure region all the way back to the separation boundary. Richardson and Borman (1992) reported that there was no indication of oil film reforming at the rear of the ring in their measurement. This indicated that separation occurred rather than ventilation cavitation. It seems that the JFO or Elrod cavitation model used in the journal bearings cannot be used in the piston ring and skirt. Ma and Smith (1996) proposed that it is more reasonable to assume an 'open cavitation' without oil film reformation. Piston ring lubrication cavitation modeling with different boundary conditions was reviewed by Priest *et al.* (1996, Fig. 10.18). They showed that the most commonly used Reynolds cavitation and reformation boundary condition produced thinner



(a) Hydrodynamic pressure profile and film shape with Reynolds cavitation and reformation boundary conditions

(b) Hydrodynamic pressure profile and film shape with modified Reynolds separation boundary conditions

(c) Predicted cyclic variation of friction force

10.18 Piston ring lubrication dynamics, cavitation, and friction modeling (from Priest *et al.*, 1996).

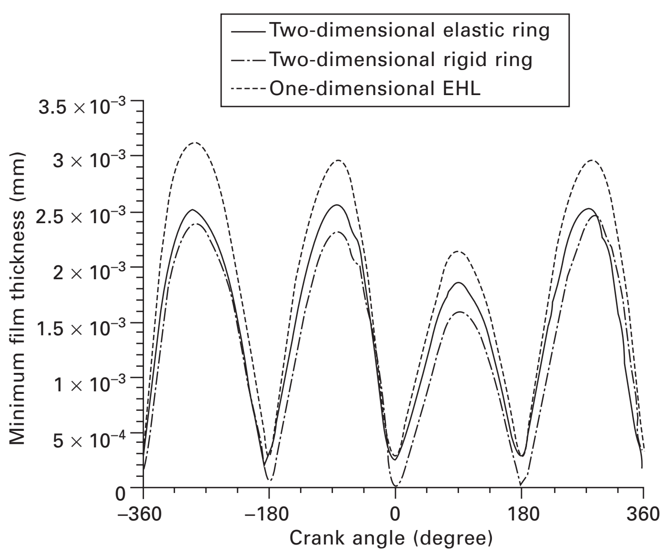
oil film thickness and higher shear friction force than the modified Reynolds reformation condition of separation without reformation. They showed that the Reynolds condition gave 34% higher FMEP than the modified one. Arcoumanis *et al*. (1995) and Sawicki and Yu (2000) compared different cavitation conditions for piston ring lubrication. Although the calculation results from these authors are very sensitive to the choice of cavitation boundary condition, there is neither consensus nor solid experimental evidence on which condition is more appropriate. The half-Sommerfeld condition predicts thinner oil film thickness than the Reynolds condition does. It seems the Reynolds cavitation and reformation boundary condition is still the most effective to date for piston ring lubrication. Moreover, it should be noted that the difference in the friction force caused by the different cavitation boundary conditions is usually much smaller for the piston skirt than the piston rings. This is partly because the two sides (thrust and anti-thrust) of the skirt tend to offset or balance each other on the overall oil film thickness and shear friction around the piston when one side produces thinner film thickness than the other side.

Surface topography plays a dominant role in mixed lubrication. Both surface roughness pattern (oriented in the transverse, isotropic, or longitudinal direction) and roughness magnitude have significant influence on piston ring oil film thickness, friction and wear. A rougher surface increases the proportion of boundary friction, in the mixed lubrication. The piston ring surface topography effect in the mixed lubrication was modeled by Sui and Ariga (1993) and Michail and Barber (1995). Sui and Ariga (1993) concluded that the friction of the oil ring and the second compression ring is the most sensitive to surface roughness variations, while the top compression ring is less affected by surface roughness. Tian *et al*. (1996b) studied the effect of surface roughness on oil transport in the top liner region. Arcoumanis *et al*. (1997) developed mixed lubrication models for Newtonian and non-Newtonian shear thinning fluids on rough surfaces. Gulwadi (2000) introduced a model to calculate the ring–liner wear.

The forces and moments due to gas pressures, axial inertia, hydrodynamic normal and shear forces and the reaction and friction forces at the ring–groove pivot positions cause the ring to move axially and twist in the groove. Ruddy *et al*. (1979), Keribar *et al*. (1991), Tian *et al*. (1996a, 1997, 1998), and Gulwadi (2000) extended the axisymmetrical one-dimensional Reynolds equation lubrication analysis by including the ring dynamics of the radial, axial, and twist motions within the groove so that blow-by, oil consumption and ring–groove wear can be analyzed in addition to a more accurate prediction of ring friction. Tian *et al*. (1996a) also introduced a lubrication and asperity contact model for the oil film pressure between the ring and its groove. Piston ring hydrodynamic lubrication and friction are significantly affected by the dynamic twist of the ring and the inter-ring gas pressure loading that is influenced by the ring axial motion. The influences of particles on the tribological performance of piston ring packs were numerically studied by Meng *et al*. (2007b, 2010). Piston ring dynamics modeling can be conducted with commercial software packages such as Ricardo's RINGPAK (Keribar *et al*., 1991; Gulwadi, 2000) and AVL's EXCITE Piston&Rings and GLIDE (Herbst and Priebsch, 2000).

The lateral dynamic friction force between the piston ring and the ring groove is believed to be significant, and this partly contributes to a circumferential variation of the oil film thickness of the ring. Non-axisymmetrical oil film distribution is also caused by other factors such as circumferentially non-uniform ring elastic pressure (e.g., caused by improper design of the ring free shape), bore distortion, the circumferential variation of the ring face profile, dynamic ring twist, the non-uniform static twist caused by ring groove deformation, circumferential ring gap position, and the different/asymmetrical inter-ring gas pressures at the thrust and anti-thrust sides due to the piston secondary motions. The modeling work by Das (1976) was one of the earliest efforts to solve the two-dimensional Reynolds equation for piston rings. Two-dimensional piston ring lubrication was modeled in detail by Hu *et al*. (1994), Yang and Keith (1996a, 1996b), and in a series of research conducted by Ma and Smith (1996) and Ma *et al*. (1995a, 1995b, 1997), particularly modeling the effects of bore out-of-roundness and ring eccentricity on ring friction and oil transport. The non-axisymmetrical modeling gives circumferentially non-uniform hydrodynamic lubricating oil film pressure distribution and uneven oil film thickness. Yang and Keith (1996b) found that the circumferential flow lowers the load-carrying capacity of the ring, hence the minimum oil film thickness in the non-axisymmetrical modeling is smaller than that in the axisymmetrical modeling. Ma *et al*. (1995b) found that better overall ring performance can be achieved when the ring face barrel profile has a small offset (i.e., asymmetric barrel).

Piston ring elastohydrodynamic lubrication was modeled by Yang and Keith (1995, 1996b) with one- and two-dimensional Reynolds equations, respectively, by considering the pressure–viscosity relation and ring elastic deflection and deformation. They found that the elastohydrodynamic effect is a strong factor in piston ring lubrication because their calculated minimum oil film thickness around the TDC is thicker than that in the rigid ring case (Fig. 10.19). They also found that under high cylinder pressures the elastohydrodynamic oil film thickness tends to be uniform circumferentially because the elastic deformation of the ring tends to reduce the gap caused by the noncircular bore.



10.19 Comparisons of predicted oil film thickness of a piston ring (from Yang and Keith, 1996b).

10.6 Engine bearing lubrication dynamics

Engine bearing tribology has been extensively researched and there is a large body of literature available. A broad discussion in this highly specialized field is outside the scope of this book. Excellent reviews have been provided by Campbell *et al*. (1967), Martin (1983, 1985), Goenka and Paranjpe (1992), Taylor (1993a), and Tanaka (1999). The following discussion focuses on the key characteristics of engine bearing lubrication dynamics, general analysis methods, and bearing friction calculation.

10.6.1 Characteristics of engine bearing lubrication dynamics

The engine bearings are dynamically loaded bearings, including the crankshaft main bearings, the connecting rod bearings, the piston pin bearings, the camshaft bearings, and the balancer shaft bearings. The analytical solutions of the Reynolds equation for bearing lubrication were usually based on the 'short bearing' approximation. Booker's mobility method (Booker, 1965) to calculate the journal trajectory and the oil film thickness of dynamically loaded bearings was very popular and is still being used in many commercial software packages for the calculations of oil film thickness and friction. The mobility method determines the journal eccentricity and the attitude angle for a given dynamic force.

A more accurate solution of the journal trajectory and friction loss requires a numerical algorithm such as finite difference or finite element to solve the two-dimensional Reynolds equation. The mass inertia effect of the journal within the bearing clearance may become important for some bearings. It may impact the journal trajectory, for example in the crankshaft main bearings near the flywheel. The journal bearing dynamics including the mass inertia or the moment of inertia term (or the acceleration term) coupled with the Reynolds equation possesses a fundamental characteristic of stiff ODE. Stiff ODE requires a time-marching implicit numerical integration algorithm with superior numerical stability in order to avoid (1) the round-off error going out of control in any explicit integration algorithm, or (2) an artificial oscillation of the lubricant force with respect to time caused by a less stable implicit integration algorithm. The stiffer the ODE system, the more difficult the numerical integration. A more detailed discussion of the stiff ODE feature is summarized in Section 10.4.4. One important indicator of the stiffness of the ODE system is the largest eigenvalue of the Jacobian matrix of the ODE, which is related to the lubrication and dynamic parameters of journal bearings as follows (Xin, 1999):

$$\lambda_{ODE,\,max} \propto \frac{\mu_v r_B L_B^3}{m_J c_B^3} \qquad 10.32$$

where μ_v is the lubricant dynamic viscosity, r_B is the bearing radius, L_B is the bearing length, m_J is the journal mass, c_B is the bearing radial clearance. Equation 10.32 is derived using the short-bearing assumption. A larger $\lambda_{ODE,max}$ means higher stiffness of the ODE. Equation 10.32 shows the major parameters affecting the stiffness. For example, a large value of oil viscosity, a very light journal mass or a very small bearing clearance can all increase the severity of the numerical instability in the time-marching integration, especially when an explicit integrator is used.

Stiff ODE is one of the most important fundamental characteristics of any lubrication dynamics problem when the mass inertia or moment of inertia term is included in the dynamics equation of the motion coupled with the Reynolds equation. Correctly handling stiff ODE is critical for the robustness and the computational efficiency of the model. Fortunately, in many practical applications of lubrication dynamics, the mass inertia effect is small hence the inertia term may be neglected for simplicity. However, the inertia term should be included if a rigorous dynamic formulation is adopted. In this case the stiff ODE feature is inevitable.

The complexity of the bearing models can vary from a simple level of the rigid isothermal two-dimensional hydrodynamic lubrication with the half-Sommerfeld (non-mass-conserving) cavitation boundary condition to a sophisticated level of thermo-elastohydrodynamic three-dimensional lubrication including the effects of journal tilting, surface topography, and non-Newtonian fluid with the JFO mass-conserving cavitation boundary condition, and the Elrod algorithm to solve the boundary condition. Although a true mass-conserving cavitation boundary condition has a secondary effect on the calculations of oil film thickness, oil film pressure, and friction loss, it is important for the oil flow and temperature predictions (Goenka and Paranjpe, 1992). The half-Sommerfeld and Reynolds cavitation conditions do not satisfy the condition that the oil inflow should equal the outflow, whereas the JFO condition does. For engine system design calculations, usually a simplified short-bearing approximation is sufficient to estimate the bearing friction. Paranjpe *et al.* (2000) provided a comparison between the theoretical calculations and the oil film thickness measurements for the engine crankshaft main bearings and the connecting rod big-end bearings. The relationship between the typical instantaneous bearing load and the oil film thickness behavior of those bearings was illustrated. An example of the instantaneous bearing friction torque simulation with a mixed lubrication model for dynamically loaded journal bearings was provided by Ai *et al.* (1998).

Bearing oil operating temperature and operating viscosity have a large impact on the accuracy of friction calculations. The friction generated inside the bearing in turn heats the oil as it flows through the bearing. Therefore, the estimation of the bearing oil temperature and viscosity needs to account for the viscous heating due to the fluid shear friction. The rate of the viscous heat generated per unit volume of the lubricant is proportional to the product of oil viscosity and the square of bearing shear rate. Thinner oil films generate a higher shear rate and a larger amount of viscous heat. In a component-level analysis, the energy equation may be introduced along with the true mass-conserving cavitation condition to conduct a coupled thermohydrodynamic lubrication analysis in order to predict the oil temperature and the viscosity variation. However, for engine system-level estimation, a fast and simplified approach is preferred to estimate the operating oil temperature in the bearing. Spearot *et al.* (1989) found that the effective temperature of the oil in the bearing is generally higher than that in the sump. To estimate the bearing operating oil temperature, an empirical formula was provided as a function of sump oil temperature, oil viscosity, minimum oil film thickness, engine rotational speed, and brake torque.

10.6.2 Engine bearing friction calculation

Assuming the journal has no eccentricity within the bearing and the circumferential clearance (2π) is full of lubricant, ignoring cavitation, the viscous shear friction force acting on the bearing can be calculated by using the law of viscous shear force as follows:

$$F_{f,B} \approx \frac{v\mu_v A_c}{h_o} = \frac{2\pi r_B N_B \mu_v \cdot 2\pi r_B L_B}{c_B} = \frac{4\pi^2 \mu_v N_B r_B^2 L_B}{c_B} \qquad 10.33$$

where N_B is the journal relative rotational speed (rev/s) and c_B is the radial clearance. Equation 10.33 is sometimes called the Petroff equation. Obviously, equation 10.33 is an over-simplification that does not allow for any load. Therefore, it can only be used for trend estimates (e.g., serving as a foundation for the level-1 system friction model) or low-load engine operating conditions. When the lubricant viscosity is low or the load is high, the Petroff equation can dramatically underestimate friction losses (Taylor, 1997).

The bearing friction torque is given by

$$J_{f,B} = F_{f,B} r_B \qquad 10.34$$

The friction power loss in the bearing is given by

$$\dot{W}_{f,B} = N_B J_{f,B} \qquad 10.35$$

In the hydrodynamic lubrication regime, the bearing friction force does not vary significantly with bearing load. The friction coefficient is given by

$$f_{fri,B} \approx \frac{4\pi^2 \mu_v N_B r_B^2 L_B}{c_B F_{n,B}} = \left(\frac{2\pi^2 r_B}{c_B}\right) \frac{\mu_v N_B}{\left(\dfrac{F_{n,B}}{2 r_B L_B}\right)} \qquad 10.36$$

The first part of equation 10.36 indicates the friction coefficient is dependent on bearing size and clearance. The second part of equation 10.36 is in the form of the duty parameter used in the Stribeck diagram. When the duty parameter decreases to the mixed lubrication regime, the bearing friction coefficient increases rapidly due to the surface asperity contacts.

For constant-loaded bearings with a large eccentricity, a more sophisticated calculation of the hydrodynamic lubrication friction force acting on the bearing is given by the short bearing theory developed by Dubois and Ocvirk in 1953 (Cameron, 1966) as follows, assuming a 2π film (i.e., a full circumference oil film):

$$F_{f,B} = \frac{4\pi^2 \mu_v N_B r_B^2 L_B}{c_B\sqrt{1-\varepsilon_B^2}} + \frac{c_B \varepsilon_B F_{n,B}}{2r_B}\sin\varphi_B \qquad 10.37$$

where ε_B is the dimensionless bearing eccentricity and $\varepsilon_B = 1 - (h_{o,min}/c_B)$, where $h_{o,min}$ is the minimum oil film thickness. $F_{n,B}$ is the bearing load, and φ_B is the attitude angle between the load line and the line connecting the journal and bearing centers. The attitude angle is given by

$$\varphi_B = \tan^{-1}\left(\frac{\pi\sqrt{1-\varepsilon_B^2}}{4\varepsilon_B}\right) \qquad 10.38$$

The first term in 10.37, the 'shear term', represents the shearing force of the oil film in the entire 2π circumference (without considering the cavitation impact). The second term in equation 10.37, the 'hydrodynamic pressure term', is a circumferential pressure gradient $\int_{A_c} (h_o/2)(\partial p_{lub}/\partial\varphi)\mathrm{d}A_c$ term of the hydrodynamic lubrication, reflecting the work done on the oil film due to the pressure gradient created by the journal rotation and the load. The 'hydrodynamic pressure term' can be either positive or negative, depending on the minimum oil film thickness position and the lubricating oil film pressure distribution shape along the circumferential direction. The second term is small or negligible compared to the first shear term for engine bearings (Martin, 1985; Taraza *et al.*, 2007). Note that the short-bearing theory provides a good approximation to the full solution of the two-dimensional Reynolds equation if the ratio of the axial length of the bearing to the journal diameter is less than approximately 0.7; however, it is inaccurate at large eccentricities (Dowson *et al.*, 1996).

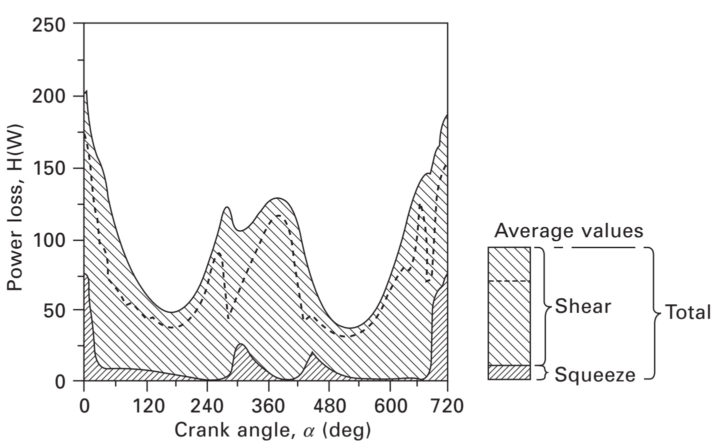
For dynamically variable-loaded bearings (e.g., the connecting rod big-end bearing, the crankshaft main bearing, and the camshaft bearing), a third term, the 'translation or squeeze term', is added as follows to the bearing friction force, which corresponds to the work done on the oil film due to the translation motion of the journal center with respect to the bearing center,

$$F_{f,B} = \frac{4\pi^2 \mu_v N_B r_B^2 L_B}{c_B\sqrt{1-\varepsilon_B^2}} + \frac{c_B \varepsilon_B F_{n,B}}{2r_B}\sin\varphi_B + \frac{vF_{n,B}}{2\pi N_B r_B}\cos\varphi_0$$
$$\approx \frac{4\pi^2 \mu_v N_B r_B^2 L_B}{c_B\sqrt{1-\varepsilon_B^2}} + \frac{c_B \varepsilon_B F_{n,B}}{2r_B}\sin\varphi_B \qquad 10.39$$

where φ_0 is the angle between the load direction and a velocity vector v of the journal center (Martin, 1985). The third term is important for the 'squeeze film' bearings with little relative rotation where the dynamic squeeze film effect is dominant over the hydrodynamic wedge effect. Martin (1993) illustrated that the power loss in engine bearings is due mainly to shear friction, and the contribution due to the squeeze film effect is relatively small (Fig. 10.20). For the engine system-level simplified friction calculations, the quasi-steady-state approach suggested by Taraza *et al.* (2000) can be used by ignoring the third dynamic term in equation 10.39 to assume that at each instantaneous moment within an engine cycle the load acting on the journal is a quasi-steady constant load.

In order to solve the eccentricity ε_B in equation 10.39, the lubricant force which balances the load can be approximated by using the short-bearing theory on a 180° load-carrying film (Cameron, 1981). The eccentricity ε_B in equation 10.39 can be solved by balancing the load with the hydrodynamic lubricant force in the following equation (Taraza *et al.*, 2000, 2007; Taylor, 1993a),

$$\frac{F_{n,B}}{2\pi N_B r_B \mu_v L_B}\left(\frac{c_B}{r_B}\right)^2\left(\frac{2r_B}{L_B}\right)^2 + \frac{\pi\varepsilon_B}{\left(1-\varepsilon_B^2\right)^2}\sqrt{\frac{3\pi^3\varepsilon_B^2}{100}+1} \qquad 10.40$$



10.20 Power loss – big-end bearing example of a 1300 cc engine (from Martin, 1993).

where ε_B can be solved by an iterative method when other parameters in equation 10.39 are known. Taraza *et al.* (2007) developed a fast method to solve ε_B by approximating ε_B as a function of bearing aspect ratio and the Sommerfeld number. Then, a quasi-steady-state approximation of the instantaneous friction force for a dynamically loaded bearing can be calculated using equation 10.39.

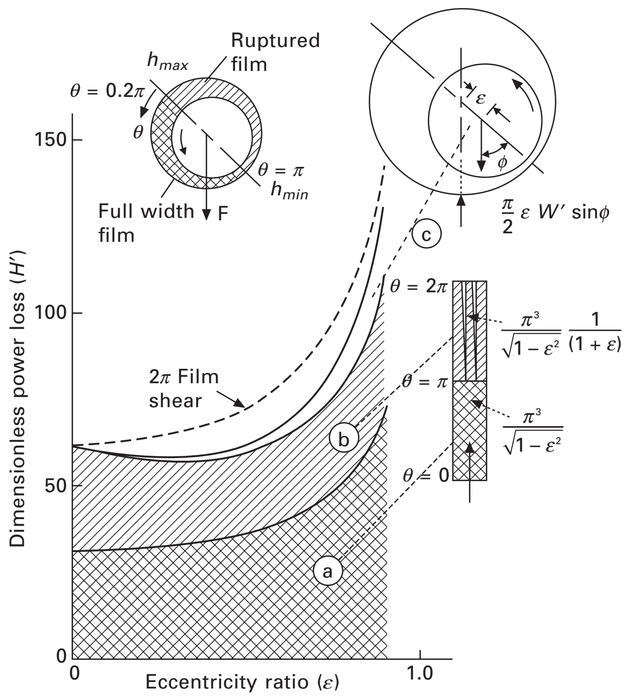
It should be noted that the above analytical equations usually overestimate the friction force because the cavitation effects are not properly included. In reality, the lubricating oil film ruptures to striated oil streams mixed with air/vapor bubbles on the bearing surface when cavitation happens. The presence of the air/vapor bubbles significantly reduces the viscous shear friction force in the cavitated region, for example, by half. A proper correction coefficient/factor to account for this cavitation effect on friction reduction can be included in the equations. A precise determination of the correction factor remains very challenging because it involves the most sophisticated cavitation modeling and because the nature of the oil film rupture varies dynamically and also changes from one component (e.g., the rotating bearing) to another (e.g., the sliding piston skirt or piston ring). The load is carried by the bearing lubricant force developed in the converging part of the clearance space (e.g., from the maximum to the minimum oil film thickness or 0–π in the so-called π film extent). In the diverging region (e.g., π–2π in a 2π film) the lubricating oil film cavitates due to the negative pressure. The oil film ruptures into many streams separated by air, gas, or vapor pockets even when the oil still flows circumferentially. The viscous shear friction in the cavitated region is lower than when it is fully filled with lubricant. The effect of cavitation or oil film rupture on friction reduction can be modeled by introducing a factor C_{cav} as

$$F_{f,B} \approx C_{cav}\left(\frac{4\pi^2\mu_v N_B r_B^2 L_B}{c_B\sqrt{1-\varepsilon_B^2}} + \frac{c_B\varepsilon_B F_{n,B}}{2r_B}\sin\varphi_B\right) \tag{10.41}$$

According to Martin (1985, Fig. 10.21), for the steady-state 2π film example, the friction force in the cavitated region (π–2π) can be assumed to be proportional to $1/(1+\varepsilon_B)$, i.e., $C_{cav,\pi-2\pi} = C_0/(1+\varepsilon_B)$, where C_0 is a constant less than $1+\varepsilon_B$. $C_{cav,0-\pi} = 1$ can be assumed for the non-cavitated region (0–π). Therefore, the overall C_{cav} for the entire 2π film in equation 10.41 can be approximately averaged as:

$$C_{cav} = 0.5\left(C_0\frac{1}{1+\varepsilon_B} + 1\right) = \frac{2+\varepsilon_B}{2+2\varepsilon_B} \quad \text{(when } C_0 = 1\text{)} \tag{10.42}$$

Note that the cavitation factor C_{cav} for the striated region can be very complex and dependent upon the severity of cavitation. C_{cav} may vary dynamically



10.21 Schematic diagram showing power loss components for steady load conditions (from Martin, 1985).

with a changing eccentricity, squeeze velocity, journal speed, and oil feed. Research on C_{cav} will help improve the quality of friction prediction in engine system design. If a more conservative prediction on the friction is wanted, C_{cav} can simply be taken as 1.

Dowson *et al.* (1996) provided a detailed discussion on a more complex calculation of the bearing friction power in Booker's mobility method. Engine bearing friction modeling is also provided by Zweiri *et al.* (2000). More sophisticated calculation formulae on engine bearing friction are summarized by Martin (1985).

10.6.3 Piston-assembly bearing friction characteristics

The connecting rod small-end and big-end bearings constitute the bearings in the piston assembly. They operate in a harsh environment under high dynamic loads and relatively low journal velocities (e.g., Zhang *et al.*, 2004). Their friction torques are important for the accuracy of piston-assembly dynamics calculation. The piston pin friction force directly affects the piston skirt tilting motion. Their friction torques need to be included in the moment balance of the piston assembly dynamics equations. Their bearing friction torques can be calculated by either a simplified approach using an equivalent boundary friction coefficient multiplied by the acting load and bearing radius, or a more sophisticated approach to include the hydrodynamic lubrication friction equations introduced in the previous section.

Suhara *et al*. (1997) measured piston pin boss bearing friction of a semi-floating press-fit piston pin in an automobile gasoline engine. They found the piston pin bearing friction force increased with the cylinder pressure in the second half of the compression stroke, the expansion stroke, and the first half of the exhaust stroke. They observed there was a spike in the friction force occurring at full load immediately after the firing TDC where the cylinder pressure was the highest. Another much smaller spike occurred at 90° crank angle after the firing TDC when the connecting rod changed direction. The friction force spikes indicated the boundary lubrication characteristics of the piston in those regions. Suhara *et al*. (1997) found the piston pin friction was in the range of 6.5% (at half load) to 16% (at full load) of the friction mean effective pressure (FMEP) of the piston skirt and rings, and was not negligible. Their findings are very similar to the analytical simulation findings when the boundary friction is assumed for the piston pin in the simulation (Xin, 1999). The piston pin and the small-end of the connecting rod only rock slightly back and forth. The piston pin friction force can be calculated by multiplying the pin load with an equivalent friction coefficient.

A very interesting Stribeck diagram was published by Suhara *et al*. (1997) for the piston pin bearing. They observed that the coefficient of friction decreases as the duty parameter decreases during the second half of the compression stroke, indicating a hydrodynamic lubrication operation. The coefficient of friction sharply increases as the duty parameter further decreases during the first half of the expansion stroke, indicating the piston pin bearing operates in the mixed lubrication regime. Suhara *et al*. (1997) believed that the rising, high friction coefficient with the duty parameter in the second half of the expansion stroke was caused by a very thin oil film which did not thicken with the increasing duty parameter. This suggested insufficient supply of lubricant oil in that region. They pointed out that design improvement to reduce friction should be emphasized for both the boundary lubrication regime in the first half of the expansion stroke and the oil starvation regime in the second half of the expansion stroke.

A small reduction in surface roughness may greatly reduce the piston pin friction, as indicated by Suhara *et al*. (1997) in their extensive experimental investigation on various design effects on piston pin friction. Improving the pin boss bearing material also has a large effect on friction reduction. Reducing bearing clearance of the piston pin boss, for example for noise reduction, may incur a friction increase in the boundary lubrication regime due to more severe asperity contacts, especially at full load under high cylinder pressures. Excessively reducing piston pin length and wall thickness, for the reason of weight reduction, may cause a large increase in unit loading and pin boss bearing deformation. This can lead to an increase in boundary friction and wear, unless other design measures are taken for counterbalance (e.g., use a better bearing material, reduce surface roughness, improve oil feed).

10.7 Valvetrain lubrication and friction

10.7.1 System design considerations for valvetrain friction

While the valvetrain can be classified as a part of the mechanical systems in the engine just like the piston and the bearings, it is also one of the four essential constituents to form the air system for a modern diesel engine. The other three constituents are the manifolds, the turbocharger, and the EGR system. The central functions of the valvetrain are to deliver the gas flow into the engine with high volumetric efficiency and to achieve certain special functions such as engine brake or advanced combustion via air flow management. The source of friction comes from its functional requirements and design arrangement. Valvetrain friction has distinct characteristics compared with its tribological counterparts such as the piston skirt, the piston rings, and the bearings. While valvetrain friction is relatively small compared to the pistons and the rings (except at low speeds), the cam–follower interface is the most challenging tribological interface in the engine due to boundary lubrication. Failure to manage valvetrain friction can cause widespread durability issues. Understanding and analyzing valvetrain friction and lubrication are of vital importance to an engine system engineer.

There are two types of valvetrains: the conventional valvetrain (i.e., the pushrod or the overhead types) and the advanced VVA system. Discussion of the friction losses of VVA is outside the scope of this book. In the conventional valvetrain, the poppet valve has proven to be the best. Other types of valving such as sleeve or rotary valves suffer from poor lubrication, bad sealing, high friction, and high oil consumption. The pushrod valvetrain has been the preferred choice for heavy-duty low-speed engines due to its many virtues such as good characteristics of lubrication (e.g., abundant oil availability, freely rotating tappet), wear, and packaging. The main disadvantage of the pushrod valvetrain is its low stiffness and excessively high vibration at high speeds. Therefore, high-speed engines tend to use the much stiffer overhead-cam (OHC) valvetrain despite the challenges in design and lubrication, especially for the pivoted follower. They operate in more arduous and hotter environments and are located in the cylinder head with less lubricant naturally available. There are two types of OHC valvetrains, namely direct acting and pivoted follower. Different configurations have their own unique characteristics of stiffness, valevtrain dynamics, spring force, lubrication and wear. For example, the direct acting OHC has to use a very large cam lobe to achieve very fast valve lift, requiring the use of large camshaft bearings and resulting in higher bearing friction. The choice of valvetrain configuration is very complex and is a system-level design topic related to performance, durability, packaging, and cost. Friction is a key attribute in valvetrain performance.

Valvetrain friction occurs mainly at the cam–follower interface. Rocker arm, camshaft bearings, seals, valve guide and follower–bore also contribute to the friction. There are two types of followers, flat-faced and roller. The roller follower has much lower friction than the flat-faced follower, and modern pushrod valvetrains usually use roller followers to reduce friction and sustain the high cam stress required (Chapter 2). The flat-faced follower has been universally used in the direct acting OHC, while the roller follower has been widely used in many pivoted follower OHC valvetrains. The rocker arm can either have a shaft (journal bearing) contact or rolling contact at the pivot. This choice affects the rocker friction significantly.

At low engine speeds, the source of valvetrain friction is the spring force, which reaches its highest level at the cranking speed and at the cam nose when the engine valve is fully open. The spring force is determined to control valve floating and valvetrain separation. At high engine speeds, the valvetrain inertia effect starts to grow to offset the spring force. The inertia force can become so high that the cam force may reach the highest level at the cam flank rather than at the cam nose. At high load conditions, the cylinder pressure is very high and the gas loading acting at the exhaust valve opening (EVO) is also high. This gas loading may produce the highest cam force at the EVO within an engine cycle. Similarly, at the high speed no-fuel motoring condition, the recompression pressure can be very high due to the high exhaust manifold pressure, and it can cause a very high cam force at the intake valve opening.

The cam–follower contact has proven to operate in the most arduous tribological conditions within the internal combustion engine. Due to the high cam force and the relatively slow rotating velocity at the interface, the cam–follower is a concentrated contact and had been believed for a long time to operate in the boundary lubrication regime. As pointed out by Taylor (1991), this traditional view may well have delayed the widespread application of the thin-film lubrication theory to this interface. Taylor (1993b) cited Müller's research work (Müller, 1966) on two cam designs dating back to the 1960s, where one cam had lower Hertzian stress and lower oil film thickness at the cam nose, and the other cam had higher Hertzian stress and higher oil film thickness. The test result showed that the first cam suffered scuffing followed by pitting while the second cam performed satisfactorily. This example shows that some elements of elastohydrodynamic lubrication occur at the cam–follower interface so that the design changes affecting the oil film thickness may play an important role. Moreover, it has been found that the majority of the cam flank region is covered by a relatively thick oil film except for the narrow cam nose region under elastohydrodynamic lubrication (Taylor, 1994). Only around the cam nose does the oil film thickness become very thin, and the additive package of a lubricant may provide effective protection via the boundary reaction film. Therefore, the overall lubrication mode should be better described as mixed lubrication, as suggested by Taylor (1991). Unfortunately, the outcome of that research from the 1960s was lost and research on valvetrain lubrication has lagged behind the bearings and the piston assemblies in general. Since the 1980s, research into applying the elastohydrodynamic lubrication principles to valvetrain problems has started to become more active.

Oil film thickness at the cam–follower interface can be changed by design. The oil film thickness at several critical locations is related to durability and wear, for example, at the cam nose and the opening ramp. Cam–follower friction can be better predicted with the mixed lubrication theory. On the durability side, cam stress is not the only dominant parameter to judge the durability of wear. Lubrication plays an equally vital role, and different oil film thickness may alter the corresponding allowable stress limit. The modern view of the valvetrain design criteria emphasizes the equal importance of using both the traditional design criteria derived from the view of the boundary lubrication (i.e., the maximum Hertzian contact stress, flash temperatures, lubricant additives, and material specifications) and the mixed/elastohydrodynamic lubrication principles to improve the minimum oil film thickness for a reliable cam–follower interface design with minimum friction and wear. In fact, the modern mixed/elastohydrodynamic lubrication theory may help improve the design criteria used in the traditional view. For example, the more accurate calculation of the friction force by using the modern mixed lubrication theory directly improves the accuracy of the prediction of flash temperature and cam surface temperature (to be detailed later).

The following are the three central tasks for the engine system engineer in valvetrain friction/lubrication work:

- minimizing valvetrain loading through air system optimization
- optimizing the cam profile/size to promote lubrication
- optimizing the tribological interfaces (e.g., selecting the roller and material, optimizing surface topography, predicting oil film thickness and friction losses).

The cam–follower interface is also subject to wear due to surface asperity contacts and high flash temperature. Wear modeling is related to valvetrain loading optimization, hence is an important part of system durability and reliability. These topics are discussed in Chapter 2 (see also Colgan and Bell, 1989). This section focuses on cam lubrication and friction.

10.7.2 Characteristics of valvetrain friction

Valvetrain friction is usually the second highest mechanical friction in the engine after the piston assembly. Among the friction losses of the various engine components, the valvetrain may account for 40% of the total friction at low speeds and 10% at high speeds. As a comparison, the piston assembly and the bearings may account for approximately 35% of the total friction at low speeds and 50% at high speeds. The crankshaft bearings may account for 5% at low speeds and 15% at high speeds. The water pump, the oil pump, and the alternator may account for 20% at low speeds and 25% at high speeds. Reducing valvetrain friction is important because a friction reduction at low speeds has a large impact on fuel economy for many vehicles running in part-load driving cycles (e.g., especially in urban city driving).

Valvetrain lubrication and friction are characterized by the following three aspects:

1. Different types of friction interfaces operating in different lubrication regimes exist for various valvetrain components. They respond differently to the changes in design and operating parameters.
2. Different friction interfaces have different characteristics of instantaneous dynamic loading and relative velocities within an engine cycle.
3. The relative proportions of each friction interface determine the overall behavior of the entire valvetrain friction.

The following friction interfaces exist in the valvetrain:

- cam–follower (mainly in the severe mixed or boundary lubrication regime, or precisely speaking, boundary lubrication dominates near the cam nose, while elastohydrodynamic lubrication dominates on the cam flanks)
- roller pin–bearing of a roller follower
- rocker–fulcrum (either a shaft or ball joint; usually in boundary lubrication)
- camshaft journal–bearing (mainly in hydrodynamic lubrication)
- camshaft seals (in boundary lubrication)
- follower–bore (in hydrodynamic lubrication)
- valve–guide (in hydrodynamic lubrication)
- valve stem seals (in boundary lubrication)
- other miscellaneous, low friction interfaces (e.g., rocker–bridge or rocker–valve stem tip, pushrod sockets).

In hydrodynamic lubrication, the friction is mainly viscous shear and it is proportional to sliding velocity and lubricant viscosity. In elastohydrodynamic lubrication, the surface asperities deform and the lubricant viscosity changes drastically with the lubricating oil film pressure. Because the contact geometry between the cam and the flat-faced follower is counter-formal (unlike the conformal contact in the engine bearings), the contact pressure is much higher and the oil film thickness is much lower than that in the engine bearings. In boundary lubrication, the friction is affected by the solid materials and lubricant additives which are related to the reaction film formed on the contact surface asperities. The properties of the bulk lubricant are of minor importance and the coefficient of friction is essentially independent of oil viscosity, load, speed, and apparent area of contact. The friction coefficient depends on the nature of the boundary lubricant, which may become less effective at higher temperatures. The boundary lubrication friction force is proportional to the load. High wear occurs in boundary lubrication. In the mixed lubrication regime, the cam–follower contact characteristics are determined by a combination of thin-film and boundary lubrication effects. In this regime, the physical properties of the bulk lubricant and the chemical properties of the boundary lubricant are all important. High oil film thickness may reduce friction. Higher engine speed promotes the elastohydrodynamic lubrication and weakens boundary lubrication at the cam–follower interface so that the overall friction coefficient reduces when the engine speed increases, as verified by Teodorescu *et al*. (2002). Their measurement showed that for a pushrod valvetrain with a flat-faced follower in a four-stroke diesel engine, the coefficient of friction at the cam–follower interface (including the follower–bore friction) decreases linearly from 0.11 at 700 rpm engine speed to 0.073 at 1700 rpm.

The valvetrain component loading and velocity exhibit strong instantaneous varying characteristics within an engine cycle. For example, at the cam–follower interface, the load at low speeds is mainly the valve spring force which peaks at the cam nose, while the load at high speeds is dominated by the dynamic inertia force which may vibrate violently from valve opening to closing. The contact sliding velocity also has its own unique kinematic characteristics. All these factors affect the lubricating oil film thickness and the lubrication regime with a strong instantaneously varying characteristic, and affect the cycle-average value of the friction force. Teodorescu *et al*. (2002) provided the experimental data for the instantaneous friction force of each major component of a pushrod valvetrain in a diesel engine.

The relative proportions of the contribution from each component to the overall valvetrain friction are dependent on designs. For example, the contribution from the cam–follower friction can be dominant for a direct acting OHC with a flat-faced follower. The contribution from its large camshaft bearing is also important. In this case, the overall valvetrain friction force at low speeds is dominated by the mixed/boundary lubrication characteristics of the cam–follower interface. The overall valvetrain friction force at high speeds may be largely influenced by the hydrodynamic lubrication characteristics of the camshaft bearings. On the other hand, if a roller follower is used for a valvetrain, the mixed/boundary lubrication characteristics of the overall valvetrain are less visible because the rolling friction contributes much less to the overall friction.

When the valvetrain friction characteristics are assessed against design or operating parameters, all the above factors need to be considered and the conclusion depends on the particular situations for the given valvetrain. In general, the following trends for the overall valvetrain friction can be concluded.

- The cam–follower friction force tends to decrease with engine speed. The hydrodynamic friction of other valvetrain components increases with engine speed. The friction of the seals basically stays constant. Because the cam–follower friction usually dominates, the overall valvetrain friction force often decreases with increasing engine speed.
- At low speeds, the proportion of cam–follower friction is greater in the overall valvetrain friction, and it is more difficult to maintain the elastohydrodynamic lubrication than at high speeds.
- For the mixed/boundary lubrication dominated valvetrains (e.g., flat-faced follower), the friction torque decreases as the engine speed increases.
- For the fully developed hydrodynamic lubrication oil film, the friction force is not sensitive to load change. For the less or marginally developed hydrodynamic lubrication oil film, a load increase may break the film and bring the regime to the mixed lubrication oil and result in a friction increase due to asperity contact. For the mixed/boundary lubrication, a load increase can increase the friction. Therefore, in the boundary–friction–dominated valvetrains or operating conditions (e.g., the flat-faced follower, low speeds, high temperatures), the valvetrain friction is more load-sensitive. In this case, a reduction in valve spring load or inertia load may have a great benefit. The effect of the load from the hydraulic lash adjuster on camshaft bearing friction is usually very small. The lash adjuster load acting on the cam base circle comes from the internal oil pressure acting on the plunger and an additional return spring load of the plunger.
- A lower oil temperature gives higher viscosity and higher viscous shear friction in the bearings and the valve guides. On the other hand, a lower oil temperature may thicken the oil film via the increased viscosity. It promotes good boundary lubricant protection for the friction modifier additives and reduces the boundary friction via less severe surface asperity contact. For the boundary/mixed friction dominated valvetrains, lower oil temperatures reduce friction. The effects of oil temperature and viscosity were investigated by Choi *et al.* (1995) for a valvetrain and certain engine bearings.
- For mechanical lash adjusters, the valvetrain lash may affect valvetrain friction to a certain extent via the duration of the valve event loading cycle.

The following design measures may reduce valvetrain friction:

- Each valvetrain configuration has its unique friction distribution and characteristics among its components. The design solution needs to be properly tailored. For example, a direct acting OHC does not have a rocker arm and is very stiff. It may use the lowest weight and spring force, but its camshaft bearings may be large due to the lack of a rocker. For this configuration, reducing its bearing friction can be effective.
- Four general guidelines are: (1) reducing the sliding motion to minimize the friction at the mixed/boundary interfaces (e.g., using a rolling contact to replace a sliding contact); (2) promoting an oil film formation at the cam–follower interface (e.g., reducing spring or inertia load, increasing oil entraining velocity, promoting lubricant availability, and preventing oil starvation via direct spray rather than oil splash); (3) using friction modifier additives in lubricant oils; and (4) reducing surface roughness of the contact.
- Using a roller follower and reducing valve spring preload or spring rate are the two most effective solutions to reduce valvetrain friction, especially at low speeds. However, it should be noted that the roller design increases the Hertzian contact stress significantly and probably requires a camshaft material change (e.g., using steel camshaft). Moreover, a softer valve spring may reduce the valvetrain separation speed unless the speed can be recovered by a lighter valvetrain mass or higher stiffness.
- Using a rolling contact may significantly reduce the friction at the rocker arm fulcrum.
- Camshaft bearing friction can be reduced by reducing the bearing diameter and length and optimizing the bearing clearance.
- Friction modified oils can significantly reduce valvetrain friction, even for the roller followers. Oil viscosity has only a minor influence on the boundary friction dominated valvetrains.

Valvetrain friction measurements and comparisons have been conducted by Armstrong and Buuck (1981) and Kovach *et al.* (1982).

10.7.3 Valvetrain lubrication and friction analysis process

In valvetrain lubrication and friction analysis, the valvetrain kinematic and dynamic analysis data and Hertzian stresses are used as input. The isothermal elastohydrodynamic lubrication theory (line or elliptic contact models) along with an assumed lubricant viscosity are used to predict the lubricating oil film thickness and the elastohydrodynamic friction force. The mixed lubrication model may be included at various levels of complexity to estimate the boundary friction caused by surface asperities. The average surface temperature can be calculated after the friction force is known. The lubricant temperature can be assumed to be equal to the average surface temperature of the cam across the contact area. Then the lubricant viscosity can be adjusted according to the lubricant temperature, and the whole calculation process may iterate until the results are convergent. Moreover, there are two special considerations to be accounted for in the friction calculations: (1) the effect of follower (tappet) rotation for the flat-faced follower; (2) the effect of roller slippage for the roller follower. Follower rotation and roller slippage are two important topics in diesel engine system design.

The valvetrain lubrication and friction analysis procedures have been described by a number of researchers. The valvetrain friction was broken down to model each component individually. In the analytical and experimental work by Staron and Willermet (1983) and Crane and Meyer (1990), the cam–follower friction force was calculated on an instantaneous basis by using the elastohydrodynamic theory, and the friction forces of other components were calculated on a cycle-average basis. Paranjpe and Gecim (1992) also took an instantaneous-friction calculation approach and compared five different valvetrain configurations (i.e., direct acting overhead cam, pushrod, end-pivoted finger follower, center-pivoted finger follower, and cam-in-head) with simulation. In their comparison, they made different valvetrains equivalent by maintaining the same valve lift, the same valvetrain no-follow speed, and the same input parameters with only a few exceptions. They found that the direct acting OHC valvetrain had the lowest friction. The two-finger follower valvetrains had the highest friction. The cam-in-head valvetrain stayed in between the previous two.

10.7.4 Cam–follower friction analysis for the flat-faced follower

The lubrication and oil film thickness conditions at the cam–follower interface are vital to wear and durability for the valvetrain. The friction at that interface is important at low speeds, but usually not at high speeds, for three reasons. First, at high speeds the cam load at the cam nose deceases because the spring force is offset by the dynamic inertia force of the valvetrain (although the cam load at the cam flank increases). Secondly, high speeds promote the hydrodynamic lubrication at the cam–follower interface to reduce friction. Thirdly, at high speeds the hydrodynamic friction forces of other valvetrain components (e.g., the camshaft bearings, the valve guides) increase and may have a larger contribution than the cam–follower interface. This may be the case when the cam–follower friction is minimized by using a roller follower. However, for flat-faced followers the friction at the cam–follower interface is usually dominant. A lot of emphasis has also been placed on the lubrication analysis of the cam–follower interface in the valvetrain friction area for the reasons of wear, scuffing, pitting, and polishing. Earlier valvetrain friction research was conducted by Naylor (1967) and Dyson (1977, 1980). Valvetrain lubrication and cam friction analysis was reviewed by Dowson *et al.* (1986), Taylor (1991, 1993b, 1994), Zhu (1993), and Teodorescu *et al.* (2003), including the analytical models and the experimental evidence to support the models.

The flat-faced follower is widely used in direct acting OHC valvetrains (a popular design option especially in light-duty applications) and some pushrod valvetrains. The kinematic and lubrication analysis of cam and flat-faced follower in a direct acting OHC valvetrain were provided by Taylor (1993b). The contact loading and Hertzian stress at the cam–follower interface can be determined by valvetrain dynamics calculations. Taylor (1993b) provided a summary and references for the prediction of the oil film thickness at the cam–follower interface by using the Reynolds equation under the assumption of rigid counter-formal surfaces. In fact, Taylor (1993b) emphasized that the elastohydrodynamic lubrication models pertaining to the surface elastic deformation and the viscoelastic effect of the lubricant must be used for the cam–follower contact to calculate the instantaneous minimum and central oil film thickness (at the center of the nominal Hertzian contact). The pressure distribution on the contact patch should satisfy both the Reynolds equation for the oil film and the elasticity equations for the deformation of the contacting bodies in the elastohydrodynamic lubrication. Dowson and Toyoda (1979) provided the original modeling. The elastohydrodynamic simplified line-contact formulae for smooth surfaces are cited here to illustrate the parametric dependence:

$$\begin{cases} H_{o,\,min} = 2.65 U_{lub}^{0.7} G_{lub}^{0.54} W_{lub}^{-0.13} \\ H_{o,central} = 3.06 U_{lub}^{0.69} G_{lub}^{0.56} W_{lub}^{-0.1} \end{cases} \qquad 10.43$$

where H_o is the dimensionless oil film thickness normalized by the radius of curvature r_{ROC} (i.e., $h_{o,min} = H_{o,min} r_{ROC}$ and $1/r_{ROC} = 1/r_{ROC1} + 1/r_{ROC2}$), U_{lub} is a dimensionless speed parameter proportional to the entraining velocity and oil viscosity ($U_{lub} = \mu_v v/(\vartheta_m r_{ROC})$, $v = 0.5(v_1 + v_2)$), G_{lub} is a dimensionless material parameter proportional to a pressure viscosity coefficient ($G_{lub} = C\vartheta_m$, where C is a pressure exponent), and W_{lub} is a dimensionless normal load parameter per unit length ($W_{lub} = F_n/(\vartheta_m r_{ROC} b_{CAM})$, where F_n is load, ϑ_m is elastic modulus of solids in contact, b_{CAM} is the cam width). Note that the oil film thickness is relatively insensitive to the load. A detailed calculation example was provided in Appendix 2 of Taylor (1993b). For the more sophisticated elliptic contact models the reader is referred to Staron and Willermet (1983) and Yang *et al.* (1996) for applications. Dowson and Higginson (1977) proposed a simplified empirical formula

$$h_{o,central} = 16(\mu_{v0} v r_{ROC})^{0.5} \qquad 10.44$$

where $h_{o,central}$ is the central oil film thickness in microns, μ_{v0} is the oil viscosity at the contact zone entry condition, v is the oil entraining velocity, r_{ROC} is the combined radius of curvature of the cam and the follower, and the speed–viscosity parameter $\mu_{v0}v$ is in N/m.

Note that these calculations are based on quasi-static (steady state) assumptions by ignoring the transient effects. The hydrodynamic lubrication 'wedge' effect is considered while the important dynamic 'squeeze film' effect is ignored. Smooth surfaces are also assumed in the calculation. More sophisticated calculations of oil film thickness may consider the effects of squeeze film (cited by Dowson *et al.*, 1986; Scales *et al.*, 1996), surface topography and the elastic deformation of the surface asperities (Lee and Patterson, 1995), and the non-Newtonian effect (Yang *et al.*, 1996). In fact, Lee and Patterson (1995) analyzed a roller follower and found that the modified oil film thickness by considering the surface roughness effect did not differ significantly from the result obtained by using the smooth-surface assumption because the elastohydrodynamic film thickness is a weak function of load.

The thickness-to-roughness ratio may be calculated subsequently to assess the severity of the lubrication and the surface asperity contact. The cam–follower interface is in the severe mixed lubrication regime. The friction force is equal to the sum of the elastohydrodynamic friction force and the boundary friction force due to surface asperity contacts,

$$F_{f,CAM} = F_{f,CAM,h} + F_{f,CAM,b} \tag{10.45}$$

Mixed lubrication models at various levels of complexity may be used to estimate the boundary friction. Staron and Willermet (1983) used the line contact elastohydrodynamic theory, and they proposed a simple mixed lubrication model as

$$\begin{cases} F_{f,CAM,b} = f_{fri,b}F_n(1-\lambda) & \text{if } \lambda < 1 \\ F_{f,CAM,b} = 0 & \text{if } \lambda \geq 1 \end{cases} \tag{10.46}$$

where $f_{fri,b}$ is the boundary friction coefficient, F_n is the cam load, and λ is the thickness-to-roughness ratio. Yang *et al.* (1996) used the more complex elliptic contact elastohydrodynamic theory. They proposed a mixed lubrication model by using the asperity contact theory of Greenwood and Tripp (1971) and assumed the area of contact ellipse is the apparent area of asperity contact. A similar mixed lubrication model was also discussed by Teodorescu *et al.* (2003) and they used the line contact EHD theory and assumed the Hertzian contact area is the apparent area of asperity contact.

Taylor (1993b) suggested using the central oil film thickness to calculate the elastohydrodynamic shear friction force at the sliding contact as follows:

$$F_{f,CAM,h} = \int_{-b_e}^{b_e} \frac{\mu_v(v_2 - v_1)}{h_{o,central}} \qquad 10.47$$

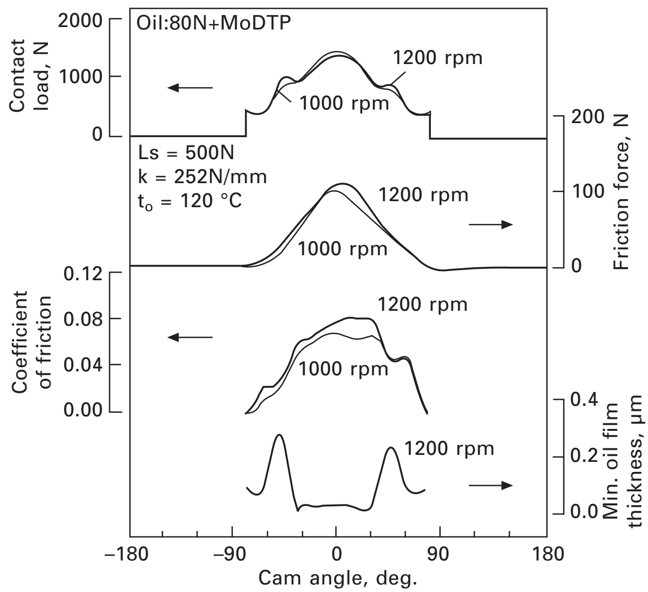
where b_e is the half width of Hertzian contact area on the cam, and $v_2 - v_1$ is the sliding velocity of the contact point between the cam and the follower. An appropriate pressure-dependent lubricant viscosity must be used in equation 10.43 because such a non-conformal contact is highly viscoelastic (i.e., the viscosity increases exponentially under high oil pressure). Because the elastohydrodynamic lubrication pressure distribution in the contact zone is almost the same as the Hertzian pressure, the pressure used to calculate viscosity can be taken from the calculated Hertzian pressure. In fact, the non-Newtonian behavior of the lubricant affecting the shear rate also plays a major role because the shear stress is lower than the Eyring strength here. Yang *et al.* (1996) and Teodorescu *et al.* (2003) provided more complex models of shear friction force by considering the non-Newtonian effect. The calculated friction force renders the coefficient of friction to be calculated, and it needs to be compared with a pre-determined limiting friction coefficient (e.g., 0.08–0.12). If the calculated friction coefficient is greater than the limiting value, the limiting value should be used. It should be noted that the EHD oil film thickness could play a secondary role on the friction force compared with the strong influences of the rheological properties of the lubricant oil and surface roughness for a very thin oil film (less than 0.1 micron).

The cam–follower friction power can be calculated by the following:

$$\dot{W}_{f,CAM} = \frac{1}{2\pi}\int_{0}^{2\pi} F_{f,CAM}\, r_C N_{CAM}\, \mathrm{d}\phi_{CAM} \qquad 10.48$$

where r_C is the perpendicular distance from the cam rotation center to the friction force vector, and ϕ_{CAM} is the cam rotational angle. The temperature rise at the cam–follower contact due to the frictional heat can be estimated based on the flash temperature concept proposed by Yang *et al.* (1996).

Figure 10.22 shows typical oil film thickness vs. cam angle (Soejima *et al.*, 1999). It should be noted that although it appears that the two spikes of high oil film thickness occupy a short duration in time or cam angle scale, they actually cover the majority of the cam flank area all the way up to near the cam nose if presented in a spatial view of the cam contour. This is due to the large variation of the velocity of the point of contact across the cam surface. Dowson *et al.* (1986) best illustrated this feature. In Fig. 10.22, it is observed that the oil film thickness is small but constant in a small region near the cam nose. It indicates a vulnerable lubrication condition in that region compared to the thicker oil film in the majority of the cam flank region. The nearly constant film thickness around the cam nose is largely dominated by the nearly constant cam radius of curvature and entraining velocity in that



10.22 Changes of contact load, friction force, friction coefficient, and minimum oil film thickness with cam angle just before and after the scuffing occurrence (from Soejima *et al.*, 1999).

region. The two most vulnerable locations subject to the boundary lubrication are the ones having theoretical zero oil film thickness, which is predicted by the quasi-static and 'wedge effect only' assumptions. The zero oil film thickness corresponds to zero mean entraining velocity due to the change in direction of the lubricant entrainment. It should be noted that in reality the oil film thickness is not zero due to some 'squeeze film' effect, as pointed out by Taylor (1993b). At the cam flank region with large oil film thickness, the thickness-to-roughness ratio is usually much larger than 1 or 3, indicating the elastohydrodynamic lubrication regime. Around the cam nose, a very thin oil film (in the order of 0.02–0.2 micron) may be maintained between the cam and the follower, and the thickness-to-roughness ratio is usually smaller than 0.5 or 1. This indicates the severe mixed lubrication regime (including the elastohydrodynamic) or even the boundary lubrication. Note that the criteria of thickness-to-roughness ratio used to judge the lubrication regimes should vary accordingly when the oil film thickness is predicted with different types of models (i.e., with or without surface topography and asperity deformation).

From equation 10.48, it is observed that the critical design parameters in cam lubrication to achieve good (large) oil film thickness are cam load and the instantaneous mean entraining velocity. The instantaneous mean entraining velocity is a function of the instantaneous radius of curvature of the contact point, lateral sliding velocity, and cam rotational speed. The cam load is affected by spring force, valvetrain mass, and engine speed. The radius of curvature is affected by the cam base circle size and the cam profile. High entraining velocity and large radius of curvature increase the oil film thickness. Taylor (1994) showed an example of a 2.0-liter engine with an end-pivot follower valvetrain. When the base circle radius was increased by 20%, the follower radius of curvature was increased by 12%, the valve spring rate was reduced by 10%, the Hertzian stress at the cam nose was reduced by 19%, the friction power was reduced by 3% and the oil film thickness at the cam nose was increased by 17%.

10.7.5 Flat-faced follower rotation

The effect of follower rotation (spin) of a flat-faced follower (tappet) on wear reduction has long been recognized. Offset design has been widely used to promote follower rotation by offsetting the follower centerline from the cam centerline (along the direction of the cam lobe width). Tapered cam profile along the lobe width has been used to make the velocities between the two mating surfaces more equal. With follower rotation, lubricant entrainment is improved, and wear is constantly distributed over a larger contact surface rather than a narrow patch. Therefore, durability is improved.

The effect of follower rotation on friction reduction was relatively new to the design community, and it inspired great interest in exploring this phenomenon by experimental and analytical means. Pieprzak *et al.* (1989), Willermet and Pieprzak (1989), Willermet *et al.* (1990), Paranjpe and Gecim (1992), Taylor (1994) and Cho *et al.* (2004) conducted experimental and analytical investigations on follower rotation and its effect on cam friction. As pointed out by Taylor (1994), in order to validate the predictive friction models, it has been common to restrain the follower rotation to make the kinematics in the experiment as precisely defined as in the theoretical model. However, in practice the flat-faced follower is free to rotate in order to enhance durability. This difference causes the discrepancy between the friction modeling result and the friction experimental test result in a real engine. Follower rotation is also related to the slip at the friction interface. Therefore, understanding, modeling, and designing the follower rotation to achieve low valvetrain friction and wear is of practical importance, despite the fact that the modeling of the dynamic friction torque driving the follower rotation is very challenging.

Follower rotation is driven by the turning torque produced by the friction force at the cam–follower interface acting at an offset distance from the follower centerline. All design/contact geometry and other forces acting on the follower affect the follower rotation, for example, the circumferential viscous shear friction force between the follower and its bore, which is a damping force to resist the rotation. The driving torque determines the follower's rotational speed. The offset distance determines the linear speed of the follower at the contact. In an extreme case, if the linear speed of the follower is equal to the linear speed of the cam lobe, the contact will have rolling friction without slip and the friction will be very low. The primary reason for the effective friction reduction due to follower rotation is that a partial rolling motion is induced between the contact surfaces, hence the friction coefficient is reduced compared to the sliding motion.

The follower rotation exhibits interesting characteristics as listed below, and some theoretical insights are provided here to explain the phenomena observed from the measurements.

- The instantaneous follower rotational speed of the follower varies as a function of cam angle. The speed gradually increases as the valve starts to open more, and reaches a maximum value near the cam nose (i.e., when the valve is fully open), then gradually decreases as the valve closes. On the cam base circle after the valve fully closes, the follower's rotational speed gradually decays before vanishing.
- At low engine speeds, the instantaneous follower's rotational speed tends to be symmetric around the cam nose. At high speeds, the pattern becomes asymmetric with the peak rotational speed located often at the cam closing side. The higher the engine speed, the more drift of the peak away from the cam nose. When the engine speed increases, the peak value of the rotational speed decreases, and it takes longer for the speed to decay to zero extended on the cam base circle. For example, Taylor (1994) showed that the peak rotational speed of the follower was 450 rpm at 549 rpm camshaft speed; and the peak follower rotational speed decreased to 300 rpm at 2,052 rpm camshaft speed. As the engine speed increases, the variation amplitude of the instantaneous rotational speed of the follower becomes smaller, or in other words, the follower rotates more evenly.
- The cycle-average rotational speed of the follower tends to increase with engine speed.
- The cycle-average rotational speed of the follower is substantially lower than the camshaft rotational speed, for example, in the order of one-tenth of the camshaft speed.
- Follower rotation is affected mainly by camshaft speed, the design of the cam–follower contact geometry (e.g., size, eccentricity) and kinematics. It is less affected by lubrication conditions.
- There was a correlation between the follower's rotational speed and the cam–follower friction torque within a cam cycle (Pieprzak *et al.*, 1989; Willermet and Pieprzak, 1989; Willermet *et al.*, 1990). Faster follower rotation results in lower friction torque. The effect is the greatest in the region of the maximum valve lift where the rotational speed is high and the contact becomes closer to rolling friction.

- High rotational speed of the follower results in a large reduction in the instantaneous friction coefficient. Follower rotation may result in an asymmetric distribution of the 'instantaneous friction coefficient vs. cam angle'.

The above characteristics can be explained by the behavior of the instantaneous cam–follower friction force, which determines the follower rotational speed and the closeness toward rolling friction. The friction force can be regarded as the product of a friction coefficient and the cam load. The friction coefficient is low at the cam flanks due to large oil film thickness, and gradually increases as it approaches the cam nose region where the boundary lubrication dominates and the oil film is very thin. At higher engine speeds, the friction coefficient tends to decrease across the cam surface due to a stronger hydrodynamic lubrication effect. Note that the measured friction coefficient at the cam–follower interface may not be symmetric with respect to the cam nose, as discovered by Willermet and Pieprzak (1989), indicating the complex nature of the lubrication regimes involved.

The cam load (force) reflects the characteristics of the valve spring load at low engine speeds with negligible effect of dynamic inertia loading. At high engine speeds, the dynamic inertia effect of the valvetrain mass may become dominant. The dynamic inertia effect increases the cam force at the cam flanks and decreases the cam force at the cam nose by offsetting the spring force. It induces several vibration peaks as well. Note that the valve spring force gradually peaks at the cam nose without vibration in a basically symmetrical fashion with respect to the cam nose. The cam force at high engine speeds looks more uniform across the entire cam event cycle. Due to the rocker shaft bearing friction, the cam force is not exactly symmetric at low engine speeds. At high speeds, the vibratory valvetrain dynamic force can easily destroy any symmetric pattern of the cam force.

The cam–follower friction force is the net complex dynamic effect of the above factors, and it may exhibit a high symmetric peak at the cam nose at low speeds, and become more asymmetric and uniform at high speeds. The follower rotation is driven by the friction force at the cam–follower interface, hence exhibits similar behavior to the dynamic friction force. In general, the greater the friction force, the faster the follower rotation. Therefore, the follower rotation is directly dependent on the specific valvetrain dynamics characteristics at different engine speeds and the cam–follower design geometry. A theoretical follower rotation model may predict the rotation very accurately at low engine speeds because the valvetrain force is relatively simple (i.e., static spring force dominated) and the friction force is easier to predict. However, at high engine speeds, the errors in valvetrain dynamics simulation can cause large discrepancies in the prediction of the friction force and the follower's rotational speed.

10.7.6 Cam–follower friction analysis for the roller follower

Roller followers have been widely used for both pushrod and overhead-cam valvetrains to significantly reduce friction. Other reasons for using the roller follower to replace the flat-faced follower include the following: (1) the design difficulty of controlling the flash temperature on the cam with the flat-faced follower; (2) the higher allowable cam stress limit of rolling contact than sliding contact; and (3) a more aggressive cam acceleration profile enabled by using a negative radius of curvature. Cam roller follower design was elaborated by Korte *et al.* (2000).

Staron and Willermet (1983) conducted a lubrication analysis for a 1.6-liter engine valvetrain to investigate the design effects on each major component. They found that using a roller follower could reduce the valvetrain friction by approximately 50%. Lee *et al.* (1994) conducted experimental work to find that the roller follower friction (including the roller pin and the cam–roller interface) accounted for 11.8% of the total valvetrain friction at 750 rpm, 12.8% at 1100 rpm, and 13.3% at 1500 rpm, compared with the 70% given by a flat-faced follower researched by Paranjpe and Gecim (1992). There are three important issues related to the friction in roller follower valvetrains: (1) roller pin bearing friction/wear; (2) the elastohydrodynamic lubrication rolling friction on the cam; and (3) roller slip (sliding or skidding) and its associated friction and wear. These issues are inter-related.

Both the roller pin friction torque and the roller inertia torque are usually quite high in magnitude, and the latter is responsible for roller slip. The roller pin bearing is believed to operate primarily in the hydrodynamic lubrication regime. Like in other journal bearings, the oil feed hole position needs to be carefully designed so that it is located in the unloaded region and does not interfere with the hydrodynamic lubrication. Colechin *et al.* (1993) presented a kinematic and lubrication analysis for a roller follower valvetrain and its roller pin bearing. Lee and Patterson (1995) measured roller pin friction and found that the roller pin operated in the hydrodynamic lubrication in the valve opening period before the cam nose (with a friction coefficient in the order of 0.001–0.002) and operated in mixed lubrication during the valve closing period after the cam nose. They found at the same Sommerfeld duty parameter the friction coefficient at the valve opening side was lower than that at the valve closing side. They suspected that the reasons could be the change of angle between the oil supply hole and the loading position, or cavitation. This again suggests the importance of oil feed hole. The roller pin friction is related to the instantaneous rotational speed of the roller, which is affected by roller slip.

The roller friction coefficient at the cam–follower interface is in the order of 0.003–0.006, while a flat-faced follower has a friction coefficient of 0.11–0.14 (Lee and Patterson, 1995). Lee and Patterson (1995) conducted a more detailed mixed lubrication analysis for the cam–roller contact by considering the surface roughness effect on the elastohydrodynamic film thickness. The cam–roller friction force is the sum of the rolling friction and the boundary asperity contact friction. When the roller rolls without slippage, the instantaneous velocity of the roller is equal to that of the cam, which is equal to the oil entraining velocity. The rolling friction can be calculated by using an elastohydrodynamic line contact model (Goksem, 1978) as follows:

$$F_{f,rolling,CAM} = \frac{4.318}{C}(G_{lub}U_{lub})^{0.658}W_{lub}^{0.0126}r_{ROC} \qquad 10.49$$

where C is a pressure viscosity coefficient, and the definition of G_{lub}, U_{lub}, W_{lub} and r_{ROC} are the same as in equation 10.43.

When roller slip occurs, the additional shear friction caused by the sliding can be calculated by using the viscous shear principle as follows:

$$F_{f,slip,CAM} = \frac{\mu_v A_c v_{slip}}{h_o} \qquad 10.50$$

where v_{slip} is the relative velocity due to the slip, h_o is the oil film thickness, and A_c is the contact area.

Roller slip is basically inevitable and it causes much higher friction than the rolling friction. Predicting roller slip is the most important topic in roller friction analysis. A pure rolling contact means the relative speed between the cam and the roller at the contact point is zero (i.e., equal 'surface' velocities). For convenience, a slip or sliding ratio is usually simply defined as a relative difference between the 'rotational' speeds of the roller and the cam. It should be noted that an equal rotational speed of the cam and the roller does not mean pure rolling. The complex kinematic relationship (e.g., radii, offset) of the cam–roller schematic determines a variable relationship between the cam speed and the roller speed at each contact point within a cam cycle in order to maintain a pure rolling. Any deviation in the cam and roller speeds from that prescribed kinematic relationship, for example caused by a force change on the roller, causes roller slip to occur.

There have been very few published studies in the area of roller slip since investigations started about 30 years ago. Among the published work, Duffy (1993), Lee *et al*. (1994), Lee and Patterson (1995) and Ji and Taylor (1998) are probably the most important. Experimental evidence revealed that roller slip occurred at the cam base circle and peaked at the two symmetric locations on the cam flanks (with respect to the cam nose) corresponding to the two large peaks of oil film thickness. There was almost no slip near the cam nose. The slip tended to increase with engine speed. In fact, the slip ratio could be either positive or negative (meaning that the roller can spin either faster or slower). The slip ratio could reach as high as 10%. Previous studies showed that roller slip became less when larger surface roughness at the cam–roller interface or lower oil viscosity was used.

Analogous to the vehicle wheel slip on a slippery road surface where the vehicle tractive force from the engine is greater than the road adhesion limit, the roller slip occurs when the 'tractive force' mismatches the total friction force at the roller follower contact point. What is different in the case of roller slip is that the 'tractive force' is the roller pin bearing friction force. The pin bearing friction force essentially comes from the rotation of the roller, which in turn originally comes from the camshaft driving torque. In fact, for this case with two friction forces involved (one at the roller pin, the other at the cam–roller contact), it is immaterial to define which one is the 'tractive force' and which one is the 'friction or resistance force'. The pin bearing force tries to slow down the roller, while the cam–roller contact friction force tries to speed up the roller. The roller's angular moment is equal to the product of the mass moment of inertia and the angular acceleration of the roller. The moment balance of the roller is given by equating the roller's angular moment to the sum of the torques coming from the cam–roller friction and the roller pin bearing friction.

As mentioned earlier, it is assumed that at a given constant camshaft rotational speed and at any moment within the cam event cycle, a unique value of the required roller rotational speed corresponding to pure rolling can be computed based on the kinematic relationship of the mechanism. That speed is denoted as the required rolling speed of the roller. The moment balance of the roller must satisfy the required rolling speed at each moment if pure rolling is to be maintained. Any deviation from that balance caused by a force change or camshaft speed change results in a violation of the pure rolling condition, i.e., roller slip will occur. The roller pin friction torque is governed by the hydrodynamic lubrication theory and affected by roller mass, pin size, and rotational speed. The cam–roller friction torque is governed by the elastohydrodynamic/mixed lubrication theory and affected by valvetrain dynamic loading, camshaft speed and surface roughness. Therefore, it is impossible to always maintain the zero slip (i.e., pure rolling) condition within the cam cycle. The angular speed of the roller calculated from its moment balance equation cannot always satisfy the pure rolling condition that is demanded by the kinematic relationship. This is the root cause of roller slip. Note that the roller does not always slip when the friction force is smaller than the tractive force.

From the above analysis, it can be observed that an increase in the roller pin bearing friction force, due to any design change or operating reason, tends to decelerate the roller to slip lagging behind the cam (i.e., slip-behind), and vice versa. An increase in the cam–roller friction force tends to accelerate the roller to slip further ahead of the cam (i.e., slip-ahead), and vice versa. Any design or operation changes affecting those two friction forces affect the roller slip. Engine speed affects roller slip in a complicated way through its impact on the valvetrain dynamic load, the elastohydrodynamic friction at the cam–roller contact and the hydrodynamic friction at the roller pin. The experimental phenomena observed from the literature can be well explained by this analysis. For example, Lee and Patterson (1995) found that slip-down usually occurred at the cam flank areas where the oil film thickness is high and the friction force is low. Slip-down also occurs on the cam base circle where the cam–roller friction is very low or zero. The slip effect was especially prominent at high speeds where the roller pin bearing friction force was high while the cam–roller friction became low. Duffy (1993) found in his experiment that low oil viscosity reduced roller slip due to an increase in the cam–roller friction force caused by the reduced elastohydrodynamic film thickness and the more severe asperity contacts. He also indicated that a slip index, defined as the angular integral of the product of the slip percentage and the follower lift over one camshaft revolution, was roughly proportional to the frictional power loss. This confirms that roller slip is very important for the friction of the roller–follower valvetain.

Roller follower slip and friction control presents an interesting and challenging task for modern engines. Coupled dynamic and lubrication modeling plays a vital role in the system optimization of roller follower kinematics and valvetrain dynamics in order to minimize roller slip, friction, and wear. Another important topic of roller motion is the skew (tilting) of the roller which affects cam stress and friction. The dynamic skew motion of roller followers was investigated by Ito (2003, 2004, 2006) and Ito and Yang (2002) based on a theoretical method developed by Ahmadi *et al.* (1983).

10.7.7 Friction analysis of valvetrain bearings and guides

Although the friction at the cam–follower interface is usually dominant in valvetrain friction (especially for the flat-faced followers), the friction from the rocker arm bearing, the camshaft bearings, and the follower/valve guides also plays an important role. Paranjpe and Gecim (1992) presented a simulation for a flat-faced-follower pushrod valvetrain at 2000 rpm engine speed, and showed that the cam–follower friction accounted for 70% of the total valvetrain friction, the rocker pivot 16%, the follower (lifter) guide 11%, and others 3%. Teodorescu *et al.* (2002) reported in a flat-faced-follower measurement that the rocker shaft bearing friction contributed 10% of the total valvetrain friction. In fact, the friction power at the cam–follower interface can be less important or lower than the total friction power of all other components, especially at high speeds. Experimental evidence on this fact was presented by Dowson *et al.* (1986).

The camshaft bearing load can be calculated with a similar approach as for the crankshaft main bearing in the multi-cylinder engine. For a four-stroke engine, the camshaft rotates at a half speed of the crankshaft. At very low engine speeds, the hydrodynamic lubrication oil film sometimes cannot be effectively developed in the camshaft bearing and it results in mixed lubrication with surface asperity contacts. Without knowing the details of the bearing minimum oil film thickness, its friction power calculation can be simplified by assuming hydrodynamic lubrication and following the method introduced in Section 10.6. For valvetrains equipped with mechanical lash adjusters, the camshaft load may have certain zero-load period (depending on the number of cylinders in the engine) which corresponds to the valve closed period. For valvetrains equipped with hydraulic lash adjusters, the camshaft load is always greater than zero because of the static load from the hydraulic pressure. During the unloaded period, the viscous shear friction force should be accounted for with a 2π film assumption without cavitation. During the statically or dynamically loaded period, the friction from both the viscous shear and the hydrodynamic pressure term should be accounted for with appropriate cavitation assumption. Staron and Willermet (1983) and Crane and Meyer (1990) used a slightly different model of the camshaft bearing to calculate the journal eccentricity, compared to the equations in Section 10.6. Staron and Willermet (1983) found that using needle bearings at the camshaft journal bearings did not reduce the friction at high speeds but could reduce the friction at very low speeds.

Staron and Willermet (1983) found that using needle bearings at the rocker arm fulcrum reduced the friction by about 10%. The rocker arm bearing friction occurs only when the valves move in each individual cylinder. The rocker bearing friction can be significant if a shaft bearing is used rather than a roller or ball bearing. The rocker shaft does not rotate constantly like the camshaft. The rocker arm is stationary when the follower is on the cam base circle, squeezing out the oil film and leading to the boundary lubrication in the rocker bearing. When the valve opens and closes, the rocker arm rocks slightly back and forth at a low velocity with heavy loading acting only on a small region of the bearing (similar to the piston pin bearing). The rocker bearing friction can be observed from the pushrod or cam load trace vs. cam angle. The load instantly drops immediately after the moment of maximum valve lift from a higher load at the valve opening side to a low load at the valve closing side. Note that the valve moving direction reverses at the maximum valve lift. This instant drop is caused by the change of direction of the friction forces. Teodorescu *et al.* (2002) measured the instantaneous rocker shaft bearing friction torque within an engine cycle. They observed that the rocker friction force varied instantaneously to follow the pattern of the pushrod force (the normal loading), indicating a boundary lubrication characteristic. They measured the friction coefficient of the rocker shaft as a constant (around 0.1–0.2), not affected by engine speed. They confirmed that the friction in the rocker bearing never reached hydrodynamic lubrication.

The rocker friction can be calculated by multiplying the bearing load by an equivalent friction coefficient (e.g., a rolling friction coefficient for a rolling contact). Note that the direct acting OHC does not have a rocker arm.

The friction losses of the valvetrain oscillating components (e.g., the valves and the followers in the guides) can be determined by assuming the clearance between the oscillating parts and the guides is constant at the concentric value by ignoring the side loading, the tilting effect, and the cavitation for simplicity. The clearance space can be assumed fully filled by the lubricant oil. The friction force of these components is the hydrodynamic viscous shear force, which is proportional to the oil dynamic viscosity, the sliding velocity, and the contact area, and inversely proportional to the guide clearance. A more sophisticated model may assume boundary lubrication at the ends of the oscillating strokes with zero sliding velocity under side loading. Paranjpe and Gecim (1992) and Wang (2007) provided the calculations of the side loading. Teodorescu *et al.* (2002) presented the measurement data of the instantaneous valve stem friction force within an engine cycle. The data showed this force was very small and only accounted for 1.5–2% of the force acting on the top of the valve stem and contributes to 2% of the total valvetrain friction loss. Their data showed this friction force did not increase with engine speed as the usual hydrodynamic friction behavior would exhibit. Instead, their valve–guide friction force stayed constant across the speed range and exhibited a mixed lubrication characteristic. The measurement data of the follower–bore friction force by Teodorescu *et al.* (2003) in a diesel pushrod valvetrain exhibited hydrodynamic lubrication characteristics (i.e., proportional to valve velocity). The friction force was in the same order of magnitude as the flat-faced cam–follower friction force.

Valve stem seals are used in modern diesel engines to drastically reduce the amount of oil passing the valve stem–guide area and the oil flow transported into the combustion chamber in order to minimize lube oil consumption and particulate emissions (Marlin *et al.*, 1994). The valve stem seals operate in the boundary lubrication regime (Netzer and Maus, 1998).

10.8 Engine friction models for system design

10.8.1 Level-1 engine friction model

The level-1 engine friction model is empirical or semi-empirical, non-instantaneous, and expressed by cycle-average FMEP. It has two types: (1) a simpler 'lumped overall' model to lump the friction of all different components together; and (2) a more complex 'breakdown overall' model to handle each component separately.

The foundation of the level-1 overall engine friction model can be traced back to about half a century ago. The effect of cylinder pressure (or engine load) on engine friction was explored by Gish *et al.* (1958). The effect of engine compression ratio on mechanical friction and thermal efficiency was studied in their work on a gasoline engine by using the indicator diagram method. They discovered a second-order correlation between the engine mechanical FMEP and the peak cylinder pressure regardless of whether the cylinder pressure was produced by the change in engine compression ratio or other measures. They concluded that the main reason for the mechanical friction to increase rapidly with the cylinder gas pressure was the high pressures acting behind the top compression ring.

The effect of peak cylinder pressure on engine friction is especially important due to the concern of possible decrease in mechanical efficiency in modern diesel engines, which usually run at very high boost and cylinder pressure levels. Peak cylinder pressure and the shape of the cylinder pressure trace are affected by intake manifold boost pressure, engine compression ratio, intake valve closing timing, and fuel injection timing.

The effect of design variables on engine friction was first systematically studied by Bishop (1964) at Ford Motor Company. He tested a wide range of different water-cooled four-stroke engines and derived certain empirical formulae for the following FMEP:

- piston FMEP as a function of cylinder bore, stroke, engine compression ratio, intake pressure, mean piston speed, and piston skirt length
- bearings FMEP as a function of cylinder bore, stroke, bearing diameter and length, and engine speed
- valvetrain FMEP as a function of cylinder bore, stroke, number of valves, valve diameter, and linearly decreasing with engine speed
- accessory FMEP, which is proportional to engine displacement and $N_E^{1.5}$ where N_E is engine speed.

Chen and Flynn (1965) at International Harvester proposed a popular formula of engine mechanical friction that has been widely used since. The engine mechanical friction they analyzed included rubbing friction and accessory loss. It was based on single cylinder diesel engine test data in the form of

$$\varpi_{FMEP} = C_0 + C_1 p_{max} + C_2 v_{mp} \qquad 10.51$$

where v_{mp} is the mean piston speed, and p_{max} is the peak cylinder pressure. They also indicated that the engine oil temperature had a significant impact on engine friction.

Millington and Hartles (1968) analyzed frictional losses in direction injection diesel engines with motoring breakdown test, and proposed a correlation on FMEP, which included both engine mechanical friction and pumping loss, as follows:

$$\varpi_{FMEP} = (\Omega - 4) + C_1 N_E + C_2 v_{mp}^2 \qquad 10.52$$

where Ω is the engine compression ratio, N_E is the engine speed, and the term

$C_2 v_{mp}^2$ is primarily for the pumping loss. They studied many design effects on engine friction, including piston rings, piston weight, skirt area, engine compression ratio, oil viscosity, bearing diameter, and accessory power.

Winterbone and Tennant (1981) proposed the following formula

$$\varpi_{FMEP} = C_0 + C_1 p_{max} + C_2 N_E \qquad 10.53$$

by assuming the effects of the independent variables, speed and load, may be linearly superimposed. They also analyzed diesel engine transient friction and concluded that the appreciably increased friction during transients compared to the steady-state levels at the same speed and load was caused by the crankshaft deflection during deceleration and acceleration and the associated changes in bearing friction.

Kouremenos *et al*. (2001) proposed a formula in the following form based on engine experimental data and the simulation of instantaneous friction force:

$$\varpi_{FMEP} = C_0 \varpi_{IMEP} + C_1 v_{mp} + C_2 p_{max} \qquad 10.54$$

Kouremenos *et al*. (2001) and Rakopoulos *et al*. (2002) believed that the influence of peak cylinder pressure is rather small (imperceptible or negligible).

An and Stodolsky (1995) simulated the effect of engine assembly mass on engine friction and vehicle fuel economy. They proposed a formula in the following form:

$$\varpi_{FMEP} = (C_0 + C_1 N_E + C_2 m_P N_E^2) + (C_3 - C_4 N_E + C_5 m_{VT} N_E^2) \qquad 10.55$$

where m_P is the piston assembly mass, m_{VT} is the valvetrain mass. The terms in the two parentheses represent the FMEP from the piston assembly and the valvetrain, respectively. The terms $m_P N_E^2$ and $m_{VT} N_E^2$ represent the contributions from the dynamic inertia loading of the piston assembly and the valvetrain, respectively. Martin (1985) commented in his bearing friction review that in general reducing mass of relevant moving parts would result in friction reduction. An and Stodolsky (1995) commented that at high engine speed (e.g., 4000 rpm) the $m_P N_E^2$ term can become dominant. Previous simulations conducted by Goenka and Meernik (1992), Keribar *et al*. (1993), Livanos and Kyrtatos (2006) as well as the experimental work by Uras and Patterson (1987) reported that the piston skirt and bearing friction forces were insensitive to load change. In fact, the effect of the inertia terms can be complex or small for two reasons. First, unlike the piston ring that is circumferentially and uniformly loaded to produce friction on one side (i.e., the ring face side), the piston skirt and bearings have different friction forces on both the loaded side and the unloaded side. When the load level (including the cylinder gas load and inertia load) changes, the load pushes the skirt or the journal to move within the clearance space eccentrically. This causes the lubricating oil film thinner on one side and thicker on the other side. Although the oil film cavitates on the unloaded side, the changes in the viscous shear friction forces on the two sides may offset each other to a great extent as long as the lubrication stays within the hydrodynamic regime. As a result, the total viscous friction force of the skirt or the bearing may possibly stay unchanged. Second, as the load increases, the friction coefficient in the hydrodynamic lubrication regime actually decreases.

Lee *et al.* (1999) at Ford Motor Company summarized their experience of using the overall engine friction models for engine system concept assessment. Their evaluations, including the critical geometry and weight of the powertrain components, were conducted based on the traditional model established by Bishop (1964) at Ford.

Ciulli (1993) reviewed the differences between various overall engine friction models and applied the models to a four-cylinder four-stroke diesel DI engine. He observed that most formulae are not directly comparable because they were obtained from different types of engines and different operating conditions.

To summarize the above findings, the level-1 lumped overall engine friction model without differentiating the contributions from each engine component can be written in the following form, as a function of certain basic engine design and operating parameters:

$$\varpi_{FMEP} = C_0 + C_1 p_{max} + C_2 v_{mp} + C_3 N_E + C_4 m_{E,mov} N_E^2 + C_5 N_E^2 \quad 10.56$$

In equation 10.56, the constant term C_0 represents the boundary lubrication friction which is independent of engine speed. The peak cylinder pressure term $C_1 p_{max}$ represents the mixed and boundary lubrication friction of the piston compression rings caused by cylinder gas pressure loading (also related to engine compression ratio). The linear term $C_2 v_{mp}$ represents the hydrodynamic lubrication friction of the sliding components (e.g., the piston skirt and the piston rings) which is proportional to the mean piston speed. The linear term $C_3 N_E$ represents the hydrodynamic lubrication friction of the rotating components (e.g., the bearings) which is proportional to the crankshaft speed. The second-order term $C_4 m_{E,mov} N_E^2$ represents a correction term related to the dynamic inertia force of the engine assembly mass, and C_4 is usually very small for well designed diesel engines. The second-order term $C_5 N_E^2$ represents the fluid pumping in the accessory work losses. Note that there is no v_{mp}^2 term in equation 10.56 for engine pumping loss since the pumping loss is not a part of the engine friction.

The level-1 overall friction model can be further expanded to include design geometries such as cylinder bore and stroke (like the formulae reviewed by Ciulli, 1993), or broken down to each component (e.g., the piston, the rings, the bearings, the seals, the valvetrain, and the accessories) to include their basic design parameters as inputs. Fujii *et al.* (1988) attempted to use an analytical approach to derive the relationship between FMEP and cylinder bore, stroke, and engine speed. Although their analysis was for a very high speed gasoline motorcycle engine, their approach is a generic methodology deserving attention. The most comprehensive level-1 breakdown overall model of engine friction was developed by Patton *et al.* (1989) for gasoline engines, later upgraded by Sandoval and Heywood (2003) at the Massachusetts Institute of Technology. This model was used by Shayler *et al.* (2005) to compare with the light-duty diesel engine test data. The basic idea in the breakdown model is to assume the coefficient of friction follows the Stribeck diagram. Therefore, the coefficient of friction is basically proportional to the Stribeck duty parameter or its square root in the hydrodynamic lubrication, varying inversely with engine speed in the mixed lubrication, or being a constant in the boundary lubrication. The friction force can be obtained by multiplying the friction coefficient by the load (a normal force). The friction terms for each component can be derived using the FMEP format shown in equation 10.6. The detailed derivations were given by Heywood (1988), Patton *et al.* (1989), and Ferguson and Kirkpatrick (2001). A modified model including the correction factors related to lubricant viscosity, piston ring tension, and cylinder bore surface roughness was summarized by Sandoval and Heywood (2003).

The level-1 breakdown overall engine friction model is summarized in the following form:

$$\varpi_{FMEP} = \left[C_1 \frac{d_B}{B_E^2 S_E n_E} + C_2 \left(\frac{\mu_v}{\mu_{v0}} \right)^n \frac{N_E^m d_B^3 L_B n_B}{B_E^2 S_E n_E} + C_3 \frac{N_E^2 d_B^2 n_B}{n_E} \right]_{Crankshaft}$$

$$+ \left[\begin{array}{l} C_4 \left(\dfrac{\mu_v}{\mu_{v0}} \right)^n \dfrac{v_{mp}^m}{B_E} + C_5 \left(\dfrac{F_{rt}}{F_{rt,0}} \right) \left(\dfrac{\sigma_{sr}}{\sigma_{sr0}} \right) \left(1 + \dfrac{C_0}{N_E} \right) \dfrac{1}{B_E^2} \\ + C_6 \left(\dfrac{\mu_v}{\mu_{v0}} \right)^n \dfrac{N_E^m d_B^3 L_B n_B}{B_E^2 S_E n_E} \end{array} \right]_{PistonAssembly}$$

$$+ \left[C_7 \left(\frac{\mu_v}{\mu_{v0}} \right)^n \frac{p_{boost}}{p_{ambient}} \cdot \Omega + C_8 \left(\frac{F_{rt}}{F_{rt,0}} \right) \frac{p_{boost}}{p_{ambient}} \cdot \Omega^{(1.33 - C_9 v_{mp})} \right]_{GasLoading}$$

$$
+\left[\begin{array}{l}
C_{10}+C_{11}\left(\dfrac{\mu_v}{\mu_{v0}}\right)^n \dfrac{N_E^m n_B}{B_E^2 S_E n_E}+C_{12}\left(1+\dfrac{C_0}{N_E}\right)\dfrac{n_{VAL}}{S_E n_E} \\
+\,C_{13}\dfrac{N_E n_{VAL}}{S_E n_E}+C_{14}\left(\dfrac{\mu_v}{\mu_{v0}}\right)^n \dfrac{l_{VAL,max}^{1.5} N_E^{0.5} n_{VAL}}{B_E S_E n_E} \\
+\,C_{15}\left(1+\dfrac{C_0}{N_E}\right)\dfrac{l_{VAL,max} n_{VAL}}{S_E n_E}
\end{array}\right]_{Valvetrain}
+\left[C_{16}+C_{17}N_E+C_{18}N_E^2\right]_{Accessories}
\tag{10.57}
$$

where the C_i are empirical constants, n is a viscosity friction exponent, and m is a speed exponent. The n and m may have different values for different components.

The terms in the first bracket of equation 10.57 represent the frictions due to the crankshaft main bearing seals, the main bearings' hydrodynamic lubrication, and the turbulent dissipations in transporting the oil through the bearings, respectively. The term d_B is the bearing diameter, L_B is the bearing length, and n_B is the number of bearings. Note that the turbulent dissipation term may be dropped and included in the oil pump work.

The terms in the second bracket represent the friction due to the hydrodynamic lubrication of the piston assembly in reciprocating motion, the piston ring friction without gas pressure loading, and the hydrodynamic lubrication of the connecting rod bearings, respectively. The term F_{rt} is the piston ring tension, μ_v is the lubricant viscosity, and σ_{sr} is the surface roughness. The subscript '0' represents a baseline value which is used to tune the model.

The terms in the third bracket represent the increase in friction due to gas pressure loading, where p_{boost} is the intake manifold boost pressure and Ω is the engine compression ratio. Note that these terms can be improved by replacing them with the more generic terms of peak cylinder pressure.

The first two terms in the fourth bracket represent the frictions due to the boundary lubrication of the camshaft bearing seals, and the hydrodynamic lubrication of the camshaft bearings, respectively. The parameter n_{VAL} is the number of valves per cylinder, and $l_{VAL,max}$ is the maximum valve lift. The third and fourth terms in the fourth bracket represent the cam–follower friction. Note that only one of them is to be used (the third for the flat-faced follower and the fourth for the roller follower). The last two terms in the fourth bracket represent the friction due to the motions of valvetrain oscillating components at the tappet–bore interface in the hydrodynamic or mixed lubrication and the valve stem–guide interface in the mixed or boundary lubrication.

The terms in the last bracket represent the friction and parasitic losses of the engine accessories such as the oil pump, the water pump, the fuel pump, and the alternator. Detailed information on engine and vehicle accessories can be found in SAE J1343 (2000), J2743 (2007), J1342 (2007), J1341 (2003), and J1340 (2003).

Note that the viscosity correction is important because the friction power at engine start-up may be twice as high as the warmed-up engine friction power (Sandoval and Heywood, 2003). The viscosity friction exponent n may vary for different components (Shayler *et al.*, 2005). Usually, $n = 0.5$. The speed exponent m on the engine speed N_E or the mean piston speed v_{mp} in equation 10.57 may use a value lower than 1 (e.g., 0.5–0.6) in order to match experimental data better at low-temperature and low-speed conditions, as attempted by Shayler *et al.* (2005). The value of m depends on the exponent used in the correlation between the coefficient of friction and a Stribeck duty parameter. Also note that the effects of component clearances and piston skirt length in equations 10.15 and 10.57 are buried in the empirical constants. If the effects of the design change on these parameters need to be evaluated, a level-2 or level-3 friction model needs to be used.

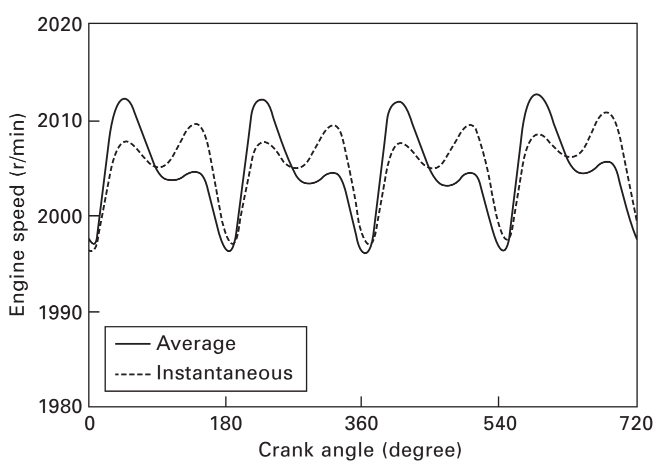
10.8.2 Level-2 engine friction model

Previous research results show that the instantaneous friction torques of the piston assembly, the crankshaft, the camshaft, and the fuel pump contain peak values much greater than their cycle-average values. There are two motivations to upgrade the level-1 engine friction model to a level-2 model: (1) to predict the instantaneous friction torque with crank angle resolution, more physics-based and less empirically, while still remaining real-time capable; and (2) to cover more engine design and operating parameters as inputs so that the effect of their design change can be evaluated, for example piston skirt length and width, piston ring tension, bearing length and diameter, clearances, valve spring preload and stiffness.

The level-2 friction model plays a very important role in engine dynamics modeling for instantaneous or transient torque prediction and cold start simulation. The level-2 model needs to analyze the friction of each component individually. The approach of the level-2 model is to compute the friction force by either using a Coulomb's type friction model, i.e., the product of friction coefficient and normal force, or using the established analytical formula for a particular component (e.g., the 'short-bearing' theory used by Taraza *et al.*, 2007). Both approaches need to be on an instantaneous crank angle basis. The normal force can be calculated by engine cycle simulation and dynamics models, which are used as engine system design tools. Some friction coefficients can be determined by the Stribeck-type diagram (e.g., piston rings), based on a lubrication parameter (as a preferred approach), or a duty parameter (traditional approach), or the film thickness-to-roughness ratio.

Unlike the level-3 model, the level-2 model does not require detailed knowledge of the spatial distributions of the lubricating oil film thickness and pressure on the component surface. The major difference between the level-2 and level-3 models is the approach to handle the coefficient of friction. The level-2 model has an advantage that some friction coefficients can come from an experimentally determined Stribeck diagram (e.g., piston rings). In the level-3 analytical models, the Reynolds equation is solved to derive the coefficient of friction. Many simplifying assumptions and boundary conditions have to be made. In fact, the engine piston assembly and bearings operate under transient thermal and dynamic conditions that are difficult to model accurately by the Reynolds equation of hydrodynamic lubrication.

Instantaneous engine friction models were developed by a group of researchers at Wayne State University, mainly for diesel engines, starting with the Rezeka–Henein model (Rezeka and Henein, 1984), followed by a series of friction research (Gardner and Henein, 1988; Taraza *et al.*, 1996; Henein *et al.*, 1997; Stanley *et al.*, 1999; Teodorescu *et al.*, 2003). Their models were reviewed by Ciulli (1993, Fig. 10.23) and summarized by Taraza *et al.* (2000, 2007). The Rezeka–Henein model has been used and modified by others (e.g., Tuccillo *et al.*, 1993), especially a group of researchers at the National Technical Univesity of Athens (Kouremenos *et al.*, 2001; Rakopoulos *et al.*, 2002, 2004a, 2004b, 2007; Rakopoulos and Giakoumis, 2009). The Rezeka–Henein model looks up the friction coefficient from a Stribeck



10.23 Trends of angular speed obtained through simulation program using formulae for average (Millington and Hartles, 1968) and for instantaneous (Rezeka and Henein, 1984) evaluation of losses (from Ciulli, 1993).

diagram to calculate the piston ring friction; uses the short-bearing analytical solution to calculate the bearing friction; and uses the elastohydrodynamic lubrication theory to calculate the cam–follower friction. The model does not handle the piston skirt friction in an effective way because a Coulomb's type friction model is used to try to calculate the viscous friction of a dynamically loaded lubricated component on an instantaneous basis.

Several important models on the instantaneous engine friction (i.e., with crank angle resolution) have been developed. These are:

- Dowson *et al*. (1996) handled the piston skirt friction modeling effectively by directly using a simplified viscous shear stress calculation and avoiding using the Stribeck diagram for friction coefficient. Their work is the most important reference for the level-2 engine friction model.
- Nakada *et al*. (1997) presented a simplified and effective model of piston friction.
- Coy (1997) used the Stribeck diagram to look up the friction coefficient and also developed an engine wear model.
- Zweiri *et al*. (2000) provided detailed derivations for a similar model to the Rezeka–Henein model.

The level-2 engine friction model is summarized in the following form:

$$\begin{aligned}
\dot{W}_{f,E}(\phi) &= \dot{W}_{f,\,OilRing} + \dot{W}_{f,\,CompressionRings} + \dot{W}_{f,\,PistonSkirt} + \dot{W}_{f,\,Bearings} \\
&\quad + \dot{W}_{f,\,Cam} + \dot{W}_{f,\,Accessories} + \dot{W}_{f,\,ValveOscillating} + \dot{W}_{f,\,Seals} \\
&= [\Sigma C_{cav} f_{fri} F_n v_{mp}]_{OilRing} + [\Sigma C_{cav} f_{fri} F_n v_{mp}]_{CompressionRings} \\
&\quad + \left[\int_0^{L_P} \left(C_{cav} \frac{\mu_v v_{mp}}{h_o} \right) \mathrm{d}A_c \right]_{PistonSkirt} + [\Sigma F_{bearings} r_B N_B]_{Bearings} \\
&\quad + [f_{fri} F_n v]_{Cam} + \left[\Sigma \left(\frac{\dot{m}p}{\eta_{pump}} + \cdots \right) \right]_{Accessories} \\
&\quad + \dot{W}_{f,\,ValveOscillating} + \dot{W}_{f,\,Seals}
\end{aligned} \tag{10.58}$$

where the friction coefficient f_{fri}, normal load F_n and relative velocity v are different for each component, and $\dot{m}$ is the pump flow rate, η_{pump} is the pump efficiency, C_{cav} is a cavitation factor accounting for the impact of cavitated or ruptured oil film on friction reduction. It should be noted that if the friction coefficient already includes the effect of cavitation, C_{cav} should be set to 1 to avoid double counting. C_{cav} reflects the effective area of the oil film contributed by the lubricant for viscous shear friction, as opposed to the air, the gas, or the vapor, in the cavitated region. C_{cav} is dependent on the severity of the cavitation and the degree of film rupture and air/

vapor filling in the gaps. Therefore, C_{cav} can be a complex function of the component's dynamic eccentricity (for piston skirt and bearings), dynamic tilting (for skirt), lateral squeeze velocity, tilting velocity (for skirt), and sliding speed, etc. Each term in equation 10.58 needs to be calculated on the instantaneous basis if possible, and its detailed formula can be found in the above mentioned references.

The piston ring friction can be calculated using one of the following methods: (1) using a pre-established Stribeck-type diagram to look up the friction coefficient by using a duty parameter or lubrication parameter that is a function of lubricant viscosity, velocity, load and contact geometry; and (2) solving the one-dimensional Reynolds equation analytically.

The piston skirt and bearing frictions should not be calculated using a Coulomb's type friction model together with a friction coefficient looked up from a Stribeck diagram because the piston skirt and engine bearings are dynamically loaded components, and therefore the zero normal load occurs with high viscous shear friction force at some crank angle locations within an engine cycle. Another reason for not using a Coulomb's type friction model for the piston skirt and the bearings is that the Coulomb's type model would bias the loaded side to over-predict the overall effect on friction when the assembly weight or the cylinder pressure increases. The viscous shear friction forces acting on the piston skirt and the bearings should be calculated using the viscous friction model shown in equation 10.58 (an example for the piston skirt).

The piston pin friction force can be calculated by using a boundary friction coefficient 0.08–0.12 multiplied by the normal acting load. The friction forces of the connecting rod big-end bearing, the crankshaft main bearing, and the camshaft bearings can be calculated by using the models described in Section 10.6. The connecting rod big-end bearing often operates in the mixed lubrication regime with high wear due to its relatively slow journal velocity. Sometimes its friction force can be approximated using the same approach as the small-end bearing (piston pin) for simplicity. The total normal force acting on a crankshaft main bearing can be simplified as a summation of the reactions produced by the two adjacent cranks/cylinders (Taraza *et al*., 2000). Other related information on journal bearing calculations can be found in Schnurbein (1970), Kozhevnikov (1974), and Mourelatos *et al*. (1987). Taraza *et al*. (2000, 2007) provided detailed models for the friction calculations of the bearings, the cam, and the engine accessories.

The friction torque during engine transients is believed to be different from its steady-state value at the same engine speed and fueling conditions (Winterbone and Tennant, 1981). Rakopoulos *et al*. (2004b, 2007) provide detailed discussions on this topic.

10.8.3 Level-3 engine friction model

The level-3 engine friction model refers to solving the Reynolds lubrication equation at different levels of complexity, from the simplest rigid-body fully-flooded hydrodynamic lubrication without surface roughness to the most complex thermo-elastohydrodynamic lubrication including the effects of surface roughness and advanced algorithms of lubricant cavitation. Usually, the level-3 model is not real-time capable if a two-dimensional Reynolds equation needs to be solved by numerical computation at each time step in the coupled lubrication–dynamics simulation. The motivations to research the level-3 model include the following:

1. Solve the one-dimensional Reynolds equation for the piston ring to deduce the friction coefficient to facilitate the piston ring modeling in the level-2 engine friction model. Note that it is possible to solve the one-dimensional Reynolds equation analytically with fast computing speed.
2. Understand the potential of component design changes in friction reduction with acceptable sufficient oil film thickness.
3. A large amount of non-real-time engine friction data can be produced by using the level-3 model and then converted to a real-time friction model by using surface fit or neural network techniques for engine controls development and transient system design. The work in this area was introduced by Wilhelm *et al.* (2007) and Meng *et al.* (2007a).

Detailed modeling information on the level-3 models is elaborated in the earlier sections. It is worth noting that the work conducted by Goenka *et al.* (1992) and Paranjpe and Cusenza (1992) at General Motors provides good examples of the level-3 modeling. Livanos and Kyrtatos (2006) summarized the solutions of the Reynolds equations for the piston rings, the piston skirt, and the bearings.

10.9 References and bibliography

Ahmadi N, Keer L M and Mura T (1983), 'Non-Hertzian contact stress analysis for an elastic half space – normal and sliding contact', *International Journal of Solids and Structures*, 19 (4), 357–373.

Ai X, Cheng H S, Hua D, Moteki K and Aoyama S (1998), 'A finite element analysis of dynamically loaded journal bearings in mixed lubrication', *Tribology Transactions*, 41(2), 273–281.

An F and Stodolsky F (1995), 'Modeling the effect of engine assembly mass on engine friction and vehicle fuel economy', SAE paper 950988.

Andersson P, Tamminen J and Sandstrom C E (2002), 'Piston ring tribology – a literature survey', *VVT Tiedotteita – Valtion Teknillinen Tutkimuskeskus*, 2178, 8–105.

Andersson S, Soderberg A and Bjorklund S (2007), 'Friction models for sliding dry, boundary and mixed lubricated contacts', *Tribology International*, 40, 580–587.

Arcoumanis C, Duszynski M, Flora H and Ostovar P (1995), 'Development of a piston-ring lubrication test-rig and investigation of boundary conditions for modelling lubricant film properties', SAE paper 952468.

Arcoumanis C, Ostovar P and Mortier R (1997), 'Mixed lubrication modelling of a Newtonian and shear thinning liquids in a piston-ring configuration', SAE paper 972924.

Armstrong W B and Buuck B A (1981), 'Valve gear energy consumption: effect of design and operation parameters', SAE paper 810787.

Audette W E and Wong V W (1999), 'A model for estimating oil vaporization from the cylinder liner as a contributing mechanism to engine oil consumption', SAE paper 1999-01-1520.

Baba Y and Hoshi M (1986), 'Analysis on driving forces of oil pumps for internal combustion engines', SAE paper 860230.

Beardmore J M (1982), 'Piston stick slip noise generation mechanism', SAE paper 820753.

Bishop I N (1964), 'Effect of design variables on friction and economy', SAE paper 640807.

Blair W L, Hoult D P and Wong V W (1990), 'The role of piston distortion on lubrication in a reciprocating engine', *Transactions of the ASME*, *Journal of Engineering for Gas Turbines and Power*, 112, 287–300.

Booker J F (1965), 'Dynamically loaded journal bearings: mobility method of solution', *Transactions of the ASME*, *Journal of Basic Engineering*, 187, 537–546.

Cameron A (1966), *The Principles of Lubrication*, Wiley, New York.

Cameron A (1981), *Basic Lubrication Theory*, 3rd edn, Ellis Horwood, Chichester.

Campbell J, Love P P, Martin F A and Rafique S O (1967), 'Bearings for reciprocating machinery: a review of the present state of theoretical, experimental and service knowledge', *Proc. IMechE*, Part 3A, 182, 51–74.

Cann P, Ioannides, E, Jacobson B and Lubrecht A A (1994), 'The lambda ratio – a critical re-examination', *Wear*, 175(1–2), 177–188.

Carden P, Bell D, Priest M and Barrell D (2006), 'Piston assembly friction losses: comparison of measured and predicted data', SAE paper 2006-01-0426.

Chen J and Richardson D E (2000), 'Predicted and measured ring pack performance of a diesel engine', SAE paper 2000-01-0918.

Chen S K and Flynn P F (1965), 'Development of a single cylinder compression ignition research engine', SAE paper 650733.

Chittenden R J and Priest M (1993), 'Analysis of the piston assembly, bore distortion and future developments', in Taylor C M (ed.), *Engine Tribology*, Elsevier Science, Amsterdam, pp. 241–270.

Cho M-R, Kim H-J, Moon T-S and Han D-C (2004), 'Theoretical and experimental evaluation of tappet rotation for a direct acting valve train system', *International Journal of Vehicle Design*, 34(1), 35–51.

Choi J K, Min B S and Oh D Y (1995), 'A study on the friction characteristics of engine bearing and cam/tappet contacts from the measurement of temperature and oil film thickness', SAE paper 952472.

Ciulli E (1992), 'A review of internal combustion engine losses, part 1: specific studies on the motion of pistons, valves and bearings', *Proc. IMechE*, *Part D: Journal of Automobile Engineering*, 206, 223–236.

Ciulli E (1993), 'A review of internal combustion engine losses, part 2: studies for global evaluations', *Proc. IMechE*, *Part D: Journal of Automobile Engineering*, 207, 229–240.

Colechin M, Stone C R and Leonard H J (1993), 'Analysis of roller-follower valve gear', SAE paper 930692.

Colgan T and Bell J C (1989), 'A predictive model for wear in automotive valve train systems', SAE paper 892145.

Comfort A (2003), 'An introduction to heavy-duty diesel engine frictional losses and lubricant properties affecting fuel economy – Part I', SAE paper 2003-01-3225.

Coy R C (1997), 'Practical applications of lubrication models in engines', *IMechE Proceedings of the First World Tribology Congress on New Directions in Tribology*, September, pp. 197–209.

Crane M E and Meyer R C (1990), 'A process to predict friction in an automotive valve train', SAE paper 901728.

Curtis J M (1981), 'Piston ring dynamics and its influence on the power cylinder performance', SAE paper 810935.

Das P K (1976), 'Analysis of piston ring lubrication', SAE paper 760008.

De Petris C, Giglio V and Police G (1996), 'Some insights on mechanisms of oil consumption', SAE paper 961216.

Dowson D (1993), 'Piston assemblies; background and lubrication analysis', in Taylor C M (ed.), *Engine Tribology*, Elsevier Science, Amsterdam, pp. 213–240.

Dowson D (1998), *History of Tribology*, 2nd edn, Professional Engineering Publishing, London.

Dowson D and Higginson G R (1977), *Elasto-Hydrodynamic Lubrication*, SI edn, Pergamon Press, New York.

Dowson D and Toyoda S (1979), 'A central film thickness formula for elastohydrodynamic line contacts', *Elastohydrodynamics and Related Topics*, *Proceedings of the 5th Leeds-Lyon Symposium on Tribology*, Mechanical Engineering Publications (IMechE), pp. 60–65.

Dowson D, Economou P N, Ruddy B L, Strachan P J and Baker A J S (1979), 'Piston ring lubrication – part II: theoretical analysis of a single ring and a complete ring pack', in Rohde S M, Wilcock D F and Cheng H S (eds), *Energy Conservation Through Fluid Film Lubrication Technology*: *Frontiers in Research and Design*, ASME, New York, pp. 23–52.

Dowson D, Harrison P and Taylor C M (1986), 'The lubrication of automotive cams and followers', *Twelfth Leeds-Lyon Symposium on Tribology – Mechanisms and Surface Distress*, Butterworths, Oxford, pp. 305–322.

Dowson D, Taylor C M and Yang L (1996), 'Friction modelling for internal combustion engines', *Tribology Series*, 31, 301–318.

Duffy P E (1993), 'An experimental investigation of sliding at cam to roller tappet contacts', SAE paper 930691.

Dursunkaya Z and Keribar R (1992), 'Simulation of secondary dynamics of articulated and conventional piston assemblies', SAE paper 920484.

Dursunkaya Z, Keribar R and Richardson D E (1993), 'Experimental and numerical investigation of inter-ring gas pressures and blowby in a diesel engine', SAE paper 930792.

Dursunkaya Z, Keribar R and Ganapathy V (1994), 'A model of piston secondary motion and elastohydrodynamic skirt lubrication', *Transactions of the ASME*, *Journal of Tribology*, 116, 777–785.

Dyson A (1977), 'Elastohydrodynamic lubrication and wear of cams bearing against cylindrical tappets', SAE paper 770018.

Dyson A (1980), 'Kinematics and wear patterns of cam and finger follower automotive valve gear', *Tribology International*, 13(3), 121–132.

Economou P N, Dowson D and Baker A J S (1979), 'Piston ring lubrication – part I: the historical development of piston ring technology', in Rohde S M, Wilcock D F and Cheng H S (eds), *Energy Conservation Through Fluid Film Lubrication Technology: Frontiers in Research and Design*, ASME, New York, pp. 1–21.

Essig G, Kamp H and Wacker E (1990), 'Diesel engine emissions reduction – the benefits of low oil consumption design', SAE paper 900591.

Ferguson C R and Kirkpatrick A T (2001), *Internal Combustion Engines – Applied Thermosciences*, 2nd edn, John Wiley & Sons, New York.

Feuga L and Bury C (1984), 'Piston and ring mechanical losses', SAE paper 841296.

Fujii I, Yagi S, Sono H and Kamiya H (1988), 'Total engine friction in four stroke SI motorcycle engine', SAE paper 880268.

Furuhama S (1987), 'Tribology on reciprocating internal combustion engines', *JSME International Journal*, 30, 1189–1199.

Furuhama S and Sasaki S (1983), 'New device for the measurement of piston frictional forces in small engines', SAE paper 831284.

Furuhama S, Hiruma M and Tsuzita M (1979), 'Piston ring motion and its influence on engine tribology', SAE paper 790860.

Furuhama S, Takiguchi M and Tomizawa K (1981), 'Effect of piston and piston ring designs on the piston friction forces in diesel engines', SAE paper 810977.

Gardner T P and Henein N A (1988), 'Diesel starting: a mathematical model', SAE paper 880426.

George S, Balla S, Gautam V and Gautam M (2007), 'Effect of diesel soot on lubricant oil viscosity', *Tribology International*, 40, 809–818.

Gish R E, McCullough J D, Retzloff J B and Mueller H T (1958), 'Determination of true engine friction', *SAE Transactions*, 66, 649–661.

Glidewell J and Korcek S (1998), 'Piston ring/cylinder bore friction under flooded and starved lubrication using fresh and aged engine oils', SAE paper 982659.

Goenka P K and Meernik P R (1992), 'Lubrication analysis of piston skirts', SAE paper 920490.

Goenka P K and Paranjpe R S (1992), 'A review of engine bearing analysis methods at General Motors', SAE paper 920489.

Goenka P K, Paranjpe R S and Jeng Y R (1992), 'FLARE: an integrated package for friction and lubrication analysis of automotive engines, part I: overview and application', SAE paper 920487.

Goksem P G (1978), 'The effect of viscous shear heating on both film thickness and rolling traction in an EHL line contact', *Transactions of the ASME*, *Journal of Lubrication Technology*, 100, 346–352.

Greenwood J A and Tripp J H (1971), 'The contact of two nominally flat rough surfaces', *Proc. IMechE*, 185, 625–633.

Guertler R W (1986), 'Excessive cylinder wear and bore polishing in heavy duty diesel egnines: causes and proposed remedies', SAE paper 860165.

Gulwadi S D (2000), 'Analysis of tribological performance of a piston ring pack', *Tribology Transactions*, 43(2), 151–162.

Guzzomi A L, Hesterman D C and Stone B J (2010), 'Some effects of piston friction and crank or gudgeon pin offset on crankshaft torsional vibration', *Journal of Ship Research*, 54 (1), 41–52.

Haddad S D and Tjan K-T (1995), 'An analytical study of offset piston and crankshaft designs and the effect of oil film on piston slap excitation in a diesel engine', *Mechanism and Machine Theory*, 30(2), 271–284.

Hairer E and Wanner G (1996), *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, Springer, Heidelberg.

Henein N A, Huang S and Bryzik W (1997), 'A new approach to determine lubrication regimes of piston-ring assemblies', *Transactions of the ASME*, *Journal of Tribology*, 119, 808–816.

Herbst H M and Priebsch H H (2000), 'Simulation of piston ring dynamics and their effect on oil consumption', SAE paper 2000-01-0919.

Heywood J B (1988), *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.

Hitosugi H, Nagoshi K, Komada M and Furuhama S (1996), 'Study on mechanism of lubricating oil consumption caused by cylinder bore deformation', SAE paper 960305.

Hu Y, Cheng H S, Arai T, Kobayashi Y and Aoyama S (1994), 'Numerical simulation of piston ring in mixed lubrication – a nonaxisymmetrical analysis', *Transactions of the ASME*, *Journal of Tribology*, 116, 470–478.

Ito A (2003), 'An analysis of skew motion of roller tappets in an OHV type DI diesel engine, 2nd report: comparison between measurement and calculation', *JSAE Annual Congress Proceedings*, No. 1-03, pp. 17–20.

Ito A (2004), 'An analysis of skew motion of roller tappets in an OHV type DI diesel engine, 3rd report: the estimation of the roller skew motion', *JSAE Annual Congress Proceedings*, No. 27-04, pp. 15–18.

Ito A (2006), 'A study on the measurement and estimation method of skew motion of roller-tappets in an OHV type DI diesel engine', SAE paper 2006-01-3348.

Ito A and Yang L (2002), 'An analysis of skew motion of roller-tappets in an OHV type DI diesel engine, 1st report: measurement and calculation', in Dowson D, Priest M, Dalmaz G and Lubrecht A A (eds), Tribology Series, Vol. 40, *Boundary and Mixed Lubrication: Science and Applications*, pp. 535–538.

Ito A, Shirakawa H, Nakamura M, Yoshida K and Akiyama H (2005), 'A study on the mechanism of lubricating oil consumption of diesel engines – 1st report: the effect of the design of piston skirt on lubricating oil consumption', SAE paper 2005-01-2169.

Jacobson B (1996), 'How lubricants behave in EHL contacts', *Tribology Series*, 31, 555–558.

Jacobson B (1997), 'At the boundary between lubrication and wear', in Hutchings I (ed.), *New Directions in Tribology*, Mechanical Engineering Publications, London, pp. 291–298.

Jeng Y-R (1992a), 'Friction and lubrication analysis of a piston-ring pack', SAE paper 920492.

Jeng Y-R (1992b), 'Theoretical analysis of piston-ring lubrication, part I – fully flooded lubrication', *Tribology Transactions*, 35(4), 696–706.

Jeng Y-R (1992c), 'Theoretical analysis of piston-ring lubrication, part II – starved lubrication and its application to a complete ring pack', *Tribology Transactions*, 35(4), 707–714.

Ji F and Taylor C M (1998), 'A tribological study of roller follower valve trains, part 1: a theoretical study with a numerical lubrication model considering possible sliding', *Tribology Series*, 34, 489–497.

Kapadia R G, Bovington C H and Lee-Prudhoe I (2007), 'A tribological investigation of lubricant effects on bearing friction and fuel economy for passenger car engines and heavy duty diesel engines', *Lubrication Science*, 19, 11–23.

Keribar R and Dursunkaya Z (1992a), 'A comprehensive model of piston skirt lubrication', SAE paper 920483.

Keribar R and Dursunkaya Z (1992b), 'Modeling of piston secondary dynamics and tribology', *Proceedings of Society of Automotive Engineers*, P-256, 657–677.

Keribar R, Dursunkaya Z and Flemming M F (1991), 'An integrated model of ring pack performace', *Transactions of the ASME*, *Journal of Engineering for Gas Turbines and Power*, 113, 382–389.

Keribar R, Dursunkaya Z and Ganapathy V (1993), 'An integrated design analysis methodology to address piston tribological issues', SAE paper 930793.

Kimura T, Takahashi K and Sugiyama S (1999), 'Development of a piston secondary motion analysis program with elastically deformable piston skirt', SAE paper 1999-01-3303.

Klingebiel F and Kahlstorf U (2000), 'Simulating engine lubrication systems with 1-D fluid flow models', SAE paper 2000-01-0284.

Kluck C E, Olsen P W and Skriba S W (1986), 'Lubrication system design considerations for heavy-duty diesel engines', SAE paper 861224.

Knoll G D and Peeken H J (1982), 'Hydrodynamic lubrication of piston skirts', *Transactions of the ASME*, *Journal of Lubrication Technology,* 104, 504–509.

Knoll G, Peeken H, Lechtape-Gruter R and Lang J (1996), 'Computer-aided simulation of piston and piston ring dynamics', *Transactions of the ASME*, *Journal of Engineering for Gas Turbines and Power*, 118, 880–886.

Korte V, Glas T, Lettmann M, Krepulat W and Steinmetz C (2000), 'Cam roller follower design for heavy duty diesel engines', SAE paper 2000-01-0525.

Kouremenos D A, Rakopoulos C D, Hountalas D T and Zannis T K (2001), 'Development of a detailed friction model to predict mechanical losses at elevated maximum combustion pressures', SAE paper 2001-01-0333.

Kovach J T, Tsakiris E A and Wong L T (1982), 'Engine friction reduction for improved fuel economy', SAE paper 820085.

Kozhevnikov S N (1974), 'Structure and dynamics of statically indeterminate mechanisms', ASME paper 74-DET-104.

Kuo T-W, Sellnau M C, Theobald M A and Jones J D (1989), 'Calculation of flow in the piston-cylinder-ring crevices of a homogeneous-charge engine and comparison with experiment', SAE paper 890838.

Lee J and Patterson D J (1995), 'Analysis of cam/roller follower friction and slippage in valve train systems', SAE paper 951039.

Lee J, Patterson D J, Morrison K M and Schwartz G B (1994), 'Friction measurement in the valve train with a roller follower', SAE paper 940589.

Lee S, Shannon B A, Mikulec A and Vrsek G (1999), 'Applications of friction algorithms for rapid engine concept assessments', SAE paper 1999-01-0558.

Li C H (1982), 'Piston thermal deformation and friction considerations', SAE paper 820086.

Li D F and Ezzat H A (1983), 'An automotive piston lubrication model', *ASLE Transactions*, 26(2), 151–160.

Ling F F, Klaus E E and Fein R S (1969), *Boundary Lubrication – An Appraisal of World Literature*, ASME, New York.

Livanos G and Kyrtatos N P (2006), 'A model of the friction losses in diesel engines', SAE paper 2006-01-0888.

Ludema K C (1984), 'A review of scuffing and running-in of lubricated surfaces, with asperities and oxides in perspective', *Wear*, 100, 315–331.

Ma J, Ryan T W, Winter J and Dixon R (1996), 'The piston ring shape and its effects on engine performance', SAE paper 960052.

Ma M-T and Smith E H (1996), 'Implementation of an algorithm to model the starved lubrication of a piston ring in distorted bores: prediction of oil flow and onset of gas blow-by', *Proc. IMechE*, *Part J: Journal of Engineering Tribology*, 210, 29–44.

Ma M-T, Smith E H and Sherrington I (1995a), 'A three-dimensional analysis of piston ring lubrication. Part 1: modelling', *Proc. IMechE*, *Part J: Journal of Engineering Tribology*, 209, 1–14.

Ma M-T, Smith E H and Sherrington I (1995b), 'A three-dimensional analysis of piston ring lubrication. Part 2: sensitivity analysis', *Proc. IMechE*, *Part J: Journal of Engineering Tribology*, 209, 15–27.

Ma M-T, Sherrington I, Smith E H and Grice N (1997), 'Development of a detailed model for piston-ring lubrication in IC engines with circular and non-circular cylinder bores', *Tribology International*, 30(11), 779–788.

Maekawa K, Mitsutake S and Morohoshi S (1986), 'A study on engine lubricating oil consumption by computer simulation', SAE paper 860546.

Mansouri S H and Wong V W (2004), 'Effects of piston design parameters on piston secondary motion and skirt-liner friction', SAE paper 2004-01-2911.

Marlin P J, Fiedler D F and Bruns I J (1994), 'Development of valve seals for turbocharged compression ignition engines', SAE paper 940588.

Martin F A (1983), 'Developments in engine bearing design', *Tribology International*, 16(3), 147–164.

Martin F A (1985), 'Friction in internal combustion engine bearings', in *Combustion Engines – Reduction of Friction and Wear*, Paper C67/85, *IMechE Conference Publications*, Institution of Mechanical Engineers, pp. 1–17.

Martin F A (1993), 'Engine bearing design: design studies, wider aspects and future developments', in Taylor C M (ed.), *Engine Tribology*, Elsevier Science, Amsterdam, pp. 113–157.

McGeehan J A (1978), 'A literature review of the effects of piston and ring friction and lubricating oil viscosity on fuel economy', SAE paper 780673.

McGeehan J A (1979), 'A survey of the mechanical design factors affecting engine oil consumption', SAE paper 790864.

McGeehan J A, *et al.* (2002), 'API CI-4: The first oil category for diesel engines using cooled exhaust gas recirculation', SAE paper 2002-01-1673.

McGeehan J A, *et al.* (2006), 'API CJ-4: Diesel oil category for both legacy engines and low emission engines using diesel particulate filters', SAE paper 2006-01-3439.

McGeehan J A, *et al.* (2007), 'API CJ-4: Diesel oil category for pre-2007 engines and new low emission engines using cooled exhaust gas recirculation and diesel particulate filters', SAE paper 2007-01-1966.

Meng F-M, Zhang Y Y, Hu Y Z and Wang H (2007a), 'Numerical study of influences of hard particles in lubricant on tribological performances of the piston ring', *Proc. IMechE, Part C: Journal of Mechanical Engineering Science*, 221, 361–372.

Meng F-M, Hu Y Z, Wang H and Zhang Y Y (2007b), 'Analysis of the dynamic performances of a piston-crankshaft system considering oil-film forces reconstructed by a neural network', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 221, 171–180.

Meng F-M, Wang J-X and Xiao K (2010), 'A study of the influences of particles in the gas flow passage of a piston ring pack on the tribological performances of the piston ring', *Proc. IMechE, Part C: Journal of Mechanical Engineering Science*, 224, 201–215.

Mian M A (1997), 'Design and analysis of engine lubrication systems', SAE paper 970637.

Michail S K and Barber G C (1995), 'The effects of roughness on piston ring lubrication, part I: model development', *Tribology Transactions*, 38(1), 19–26.

Millington B W and Hartles E R (1968), 'Frictional losses in diesel engines', SAE paper 680590.

Monaghan M L (1988), 'Engine friction – a change in emphasis: a new approach which may result in significant fuel consumption gains', *Industrial Lubrication and Tribology*, 40(2), 4–11.

Mourelatos Z P, Parsons M G and Karni Z H (1987), 'Steady-state equilibrium of a flexible shaft in a hydrodynamic journal bearing', *Transactions of the ASME*, *Journal of Tribology*, 109, 374–375.

Muller R (1966), 'The effect of lubrication on cam and tappet performance', *Motortechnische Z.*, 27, 58.

Munro R and Parker A (1975), 'Transverse movement analysis and its influence on diesel piston design', SAE paper 750800.

Nakada T, Yamomoto A and Abe T (1997), 'A numerical approach for piston secondary motion analysis and its application to the piston related noise', SAE paper 972043.

Nakamura M, Hayashi H and Ito A (2005), 'A study on the mechanism of lubricating oil consumption of diesel engines – 2nd report: mechanism of oil film generation on piston skirt', SAE paper 2005-01-2167.

Nakamura M, Hayashi H and Ito A (2006), 'A study on the mechanism of lubricating oil consumption of diesel engines – 3rd report: effect of piston motion on piston skirt oil film behavior', SAE paper 2006-01-3349.

Nakayama K, Yasutake Y, Takiguti M and Furuhama S (1997), 'Effect of piston motion on piston skirt friction of a gasoline engine', SAE paper 970839.

Nakayama K, Tamaki S, Miki H and Takiguchi M (2000), 'The effect of crankshaft offset on piston friction force in a gasoline engine', SAE paper 2000-01-0922.

Naylor H (1967), 'Cam and friction drives', *Proc. IMechE*, 182(3A), 237–247.

Netzer J and Maus K-H (1998), 'Improvements of valve stem seals to meet future emission requirements', SAE paper 980581.

Oetting H, Pundt D and Ebbinghaus W (1984), 'Friction in the piston group and new ideas for piston design', SAE paper 841299.

Offner G, Herbst H M and Priebsch H H (2001), 'A methodology to simulate piston secondary movement under lubricated contact conditions', SAE paper 2001-01-0565.

Oh K P, Li C H and Goenka P K (1987), 'Elastohydrodynamic lubrication of piston skirts', *Transactions of the ASME*, *Journal of Tribology*, 109, 356–362.

Paranjpe R S and Cusenza A (1992), 'FLARE: an integrated package for friction and lubrication analysis of automotive engines, part II: experimental validation', SAE paper 920488.

Paranjpe R and Gecim B (1992), 'Comparative friction assessment of different valve-train types using the FLARE (friction and lubrication analysis) code', SAE paper 920491.

Paranjpe R S, Tseregounis S I and Viola M B (2000), 'Comparison between theoretical calculations and oil film thickness measurements using the total capacitance method for crankshaft bearings in a firing engine', *Tribology Transactions*, 43(3), 345–356.

Parker D A (1990), 'The tribology of automobile components: development of verified predictive design techniques', *Proc. IMechE*, *Part D: Journal of Automobile Engineering*, 204, 1–19.

Parker D A and Adams D R (1982), 'Friction losses in the reciprocating internal combustion engine', IMechE paper C5/82.

Patir N and Cheng H S (1979), 'Application of average flow model to lubrication between rough sliding surfaces', *Transactions of the ASME, Journal of Lubrication Technology*, 101(2), 220–230.

Patton K J, Nitschke R G and Heywood J B (1989), 'Development and evaluation of a friction model for spark ignition engines', SAE paper 890836.

Pawlus P (1996), 'A study of the dependence of the functional properties of the cylinder liner surface layer on the operating conditions', *Proc. IMechE*, *Part J: Journal of Engineering Tribology*, 210, 17–27.

Pawlus P (1997), 'Change of cylinder surface topography in the initial stage of engine life', *Wear*, 209, 69–83.

Pieprzak J M, Willermet P A and Klassen D (1989), 'A tappet rotation monitor based on light reflectance – development and functional testing', SAE paper 890722.

Priest M, Taylor R I, Dowson D and Taylor C M (1996), 'Boundary conditions for Reynolds equation with particular reference to piston ring lubrication', *Tribology Series*, 31, 441–452.

Rakopoulos C D and Giakoumis E G (2009), *Diesel Engine Transient Operation: Principles of Operation and Simulation Analysis*, Springer-Verlag London Limited, London.

Rakopoulos C D, Hountalas D T, Koutroubousis A P and Zannis T C (2002), 'Application and evaluation of a detailed friction model on a DI diesel engine with extremely high peak combustion pressures', SAE paper 2002-01-0068.

Rakopoulos C D, Giakoumis E G, Hellas S A L, Rakopoulos D C (2004a), 'The effect of friction modelling on the prediction of turbocharged diesel engine transient operation', SAE paper 2004-01-0925.

Rakopoulos C D, Giakoumis E G, Hountalas D T and Rakopoulos D C (2004b), 'The effect of various dynamic, thermodynamic and design parameters on the performance of a turbocharged diesel engine operating under transient load conditions', SAE paper 2004-01-0926.

Rakopoulos C D, Giakoumis E G and Dimaratos A M (2007), 'Evaluation of various dynamic issues during transient operation of turbocharged diesel engine with special reference to friction development', SAE paper 2007-01-0136.

Rao V D N, Boyer B A and Cikanek H A (1999), 'Effects of honing and surface texture of cylinder bore on engine mechanical friction', ASME paper 99-ICE-216.

Rezeka S and Henein N A (1984), 'A new approach to evaluate instantaneous friction and its components in internal combustion engines', SAE paper 840179.

Richardson D E (2000), 'Review of power cylinder friction for diesel engines', *Transactions of the ASME*, *Journal of Engineering for Gas Turbines and Power*, 122, 506–519.

Richardson D E and Borman G L (1992), 'Theoretical and experimental investigations of oil films for application to piston ring lubrication', SAE paper 922341.

Rosenberg R (1982), 'General friction considerations for engine design', SAE paper 821576.

Ruddy B L, Dowson D, Economou P N and Baker A J S (1979), 'Piston ring lubrication – part III: the influence of ring dynamics and ring twist', in Rohde S M, Wilcock D F and Cheng H S (eds), *Energy Conservation Through Fluid Film Lubrication Technology*: *Frontiers in Research and Design*, ASME, New York, pp. 191–215.

SAE Surface Vehicle Information Report J357 (2006), 'Physical and chemical properties of engine oils'.

SAE Surface Vehicle Information Report J2227 (2006), 'International tests and specifications for automotive engine oils'.

SAE Surface Vehicle Recommended Practice J1343 (2000), '(R) Information relating to duty cycles and average power requirements of truck and bus engine accessories'.

SAE Surface Vehicle Recommended Practice J1340 (2003), 'Test method for measuring power consumption of air conditioning and brake compressors for trucks and buses'.

SAE Surface Vehicle Recommended Practice J1341 (2003), 'Test method for measuring power consumption of hydraulic pumps for trucks and buses'.

SAE Surface Vehicle Recommended Practice J1423 (2003), '(R) Classification of energy conserving engine oil for passenger cars, vans, sport utility vehicles, and light-duty trucks'.

SAE Surface Vehicle Recommended Practice J1342 (2007), 'Test method for determining power consumption of engine cooling fan drive systems'.

SAE Surface Vehicle Recommended Practice J2743 (2007), 'Air compressor size rating recommended practice – truck and bus'.

SAE Surface Vehicle Standard J2612 (2002), 'Internal combustion engines – piston vocabulary'.

SAE Surface Vehicle Standard J300 (2004), 'Engine oil viscosity classification'.

SAE Surface Vehicle Standard J183 (2006), '(R) Engine oil performance and engine service classification (other than 'energy conserving')'.

Sanda S, Murakami M, Noda T and Konomi T (1997), 'Analysis of lubrication of a piston ring package (effect of oil starvation on oil film thickness', *JSME International Journal*, Series B, 40(3), 478–486.

Sandoval D and Heywood J B (2003), 'An improved friction model for spark-ignition engines', SAE paper 2003-01-0725.

Sawicki J T and Yu B (2000), 'Analytical solution of piston ring lubrication using mass conserving cavitation algorithm', *Tribology Transactions*, 43(3), 419–426.

Scales L E, Rycroft J E, Horswill N R and Williamson B P (1996), 'Simulation and observation of transient effects in elastohydrodynamic lubrication', SAE paper 961143.

Schipper D J and De Gee A W J (1995), 'Lubrication modes and the IRG diagram', *Lubrication Science*, 8(1), 27–35.

Schipper D J, Vroegop P H and De Gee A W J (1991), 'Prediction of lubrication regimes of concentrated contacts', *Lubrication Science*, 3(2), 131–139.

Schnurbein E V (1970), 'A new method of calculating plain bearings of statically indeterminate crankshafts', SAE paper 700716.

Scholz B and Bargende M (2000), 'Three-dimensional simulation of the piston group', SAE paper 2000-01-1239.

Shah P, Mourelatos Z P and Patel P (2007), 'Piston secondary dynamics considering elastohydrodynamic lubrication', SAE paper 2007-01-1251.

Shayler P J, Leong D K W and Murphy M (2005), 'Contributions to engine friction during cold, low speed running and the dependence on oil viscosity', SAE paper 2005-01-1654.

Shiao Y J and Moskwa J J (1993), 'An investigation of load force and dynamic error magnitude using the lumped mass connecting rod model', SAE paper 930617.

Shin S, Cusenza A and Shi F (2004), 'Offset crankshaft effects on SI engine combustion and friction performance', SAE paper 2004-01-0606.

Soejima M, Ejima Y, Wakuri Y and Kitahara T (1999), 'Improvement of lubrication for cam and follower', *Tribology Transactions*, 42(4), 755–762.
Spearot J A, Murphy C K and Deysarkar A K (1989), 'Interpreting experimental bearing oil film thickness data', SAE paper 892151.
Spikes H A (1995), 'Boundary lubricating films', *Proc. Int. Trib. Conf.*, Yokohama, 1995, *Satellite Forum on Tribochemistry*, Japan Soc. Trib., Tokyo, October.
Spikes H A (1997), 'Mixed lubrication – an overview', *Lubrication Science*, 9(3), 221–253.
Stanley R, Taraza D, Henein N and Bryzik W (1999), 'A simplified friction model of the piston ring assembly', SAE paper 1999-01-0974.
Staron J T and Willermet P A (1983), 'An analysis of valve train friction in terms of lubrication principles', SAE paper 830165.
Suhara T, Ato S, Takiguchi M and Furuhama S (1997), 'Friction and lubrication characteristics of piston pin boss bearings of an automotive engine', SAE paper 970840.
Sui P C and Ariga S (1993), 'Piston ring pack friction and lubrication analysis of an automotive engine using a mixed lubrication model', SAE paper 931937.
Tamminen J, Sandstrom C-E and Andersson P (2006), 'Influence of load on the tribological conditions in piston ring and cylinder liner contacts in a medium-speed diesel engine', *Tribology International*, 39, 1643–1652.
Tanaka M (1999), 'Recent analyses and designs of thick-film bearings', *Journal of Japanese Society of Tribologists*, 44(5), 308-319.
Taraza D, Henein N A and Bryzik W (1996), 'Experimental determination of the instantaneous frictional torque in multicylinder engines', SAE paper 962006.
Taraza D, Henein N and Bryzik W (2000), 'Friction losses in multi-cylinder diesel engines', SAE paper 2000-01-0921.
Taraza D, Henein N A, Ceausu R and Bryzik W (2007), 'Engine friction model for transient operation of turbocharged, common rail diesel engines', SAE paper 2007-01-1460.
Taylor C M (1991), 'Valve train lubrication analysis', *The Seventeenth Leeds-Lyon Symposium on Tribology – Vehicle Tribology, Volume 18*, pp. 119–130.
Taylor C M (ed.) (1993a), *Engine Tribology*, Elsevier Science, Amsterdam.
Taylor C M (1993b), 'Valve train – cam and follower: background and lubrication analysis', in Taylor C M (ed.), *Engine Tribology*, Elsevier Science, Amsterdam, pp. 159–181.
Taylor C M (1994), 'Fluid film lubrication in automobile valve trains', *Proc. IMechE, Part J: Journal of Engineering Tribology*, 208, 221–234.
Taylor C M (1998), 'Automobile engine tribology – design considerations for efficiency and durability', *Wear*, 221, 1–8.
Taylor R I (1997), 'Engine friction: the influence of lubricant rheology', *Proc. IMechE, Part J: Journal of Engineering Tribology*, 211, 235–246.
Taylor R I (2000), 'Heavy duty diesel engine fuel economy: lubricant sensitivities', SAE paper 2000-01-2056.
Teodorescu M, Taraza D, Henein N and Bryzik W (2002), 'Experimental analysis of dynamics and friction in valve train systems', SAE paper 2002-01-0484.
Teodorescu M, Taraza D, Henein N and Bryzik W (2003), 'Simplified elastohydrodynamic friction model of the cam-tappet contact', SAE paper 2003-01-0985.
Teraguchi S, Suzuki W, Takiguchi M and Sato D (2001), 'Effects of lubricating oil supply on reductions of piston slap vibration and piston friction', SAE paper 2001-01-0566.

Thirouard B, Tian T and Hart D P (1998), 'Investigation of oil transport mechanisms in the piston ring pack of a single cylinder diesel engine, using two dimensional laser induced fluorescence', SAE paper 982658.

Tian T, Wong V W and Heywood J B (1996a), 'A piston ring-pack film thickness and friction model for multigrade oils and rough surfaces', SAE paper 962032.

Tian T, Noordzij B, Wong V W and Heywood J B (1996b), 'Modeling piston-ring dynamics, blowby, and ring-twist effects', *ASME 1996 Fall Technical Conference*, ICE-Vol. 27-2, Volume 2, pp. 67–80.

Tian T, Rabute R, Wong V W and Heywood J B (1997), 'Effects of piston-ring dynamics on ring/groove wear and oil consumption in a diesel engine', SAE paper 970835.

Tian T, Wong V W and Heywood J B (1998), 'Modeling the dynamics and lubrication of three piece oil control rings in internal combustion engines', SAE paper 982657.

Ting L L (1985), 'A review of present information on piston ring tribology', SAE paper 882355.

Ting L L (1993a), 'Development of a reciprocating test rig for tribological studies of piston engine moving components – part I: rig design and piston ring friction coefficients measuring method', SAE paper 930685.

Ting L L (1993b), 'Development of a reciprocating test rig for tribological studies of piston engine moving components – part II: measurement of piston ring friction coefficients and rig test confirmation', SAE paper 930686.

Ting L L and Mayer J E (1974a), 'Piston ring lubrication and cylinder bore wear analysis, part I – theory', *Transactions of the ASME*, *Journal of Lubrication Technology*, 96, 258–266.

Ting L L and Mayer J E (1974b), 'Piston ring lubrication and cylinder bore wear analysis, part II – theory verification', *Transactions of the ASME*, *Journal of Lubrication Technology*, 96, 305–314.

Tuccillo R, Arnone L, Bozza F, Nocera R and Senatore A (1993), 'Experimental correlations for heat release and mechanical losses in turbocharged diesel engines', SAE paper 932459.

Uras H M and Patterson D J (1983), 'Measurement of piston and ring assembly friction instantaneous IMEP method', SAE paper 830416.

Uras H M and Patterson D J (1984), 'Effect of some lubricant and engine variables on instantaneous piston and ring assembly friction', SAE paper 840178.

Uras H M and Patterson D J (1985), 'Oil and ring effects on piston-ring assembly friction by the instantaneous IMEP method', SAE paper 850440.

Uras H M and Patterson D J (1987), 'Effect of some piston variables on piston and ring assembly friction', SAE paper 870088.

Urabe M, Tomomatsu T, Ishiki K, Takiguchi M and Someya T (1998), 'Variation of piston friction force and ring lubrication condition in a diesel engine with EGR', SAE paper 982660.

Wacker E A, Sander W and Schmidt K (1978), 'Heavy-duty diesel engine piston design for low blowby and oil consumption', ASME paper 78-DGP-16.

Wakabayashi R, Takiguchi M, Shimada T, Mizuno Y and Yamauchi T (2003), 'The effects of crank ratio and crankshaft offset on piston friction losses', SAE paper 2003-01-0983.

Wakuri Y, Ono S, Soejima M and Masuda K (1981), 'Oil-film behaviour of reciprocating slider with circular profile (optical measurement of oil-film separation boundary)', *Bulletin of the JSME*, 24(194), 1462–1469.

Wakuri Y, Soejima M, Ejima Y, Hamatake T and Kitahara T (1995), 'Studies on friction characteristics of reciprocating engines', SAE paper 952471.

Wang Y (2007), *Introduction to Engine Valvetrains*, SAE International, Warrendale, PA.

Werner L (1987), 'Control of piston stick slip noise in a 4-cylinder gasoline engine', SAE paper 870991.

Wilhelm C, Winsel T, Ayeb M, Theuerkauf H J, Brandt S, Busche E, Longo C and Knoll G D (2007), 'A combined physical/neural approach for real-time models of losses in combustion engines', SAE paper 2007-01-1345.

Willermet P A and Pieprzak J (1989), 'Some effects of lubricant composition and tappet rotation on cam/tappet friction', *Transactions of the ASME*, *Journal of Tribology*, 111(3), 683–691.

Willermet P A, Pieprzak J M and Dailey D P (1990), 'Tappet rotation and friction reduction in a center pivot rocker arm contact', *Transactions of the ASME*, *Journal of Tribology*, 112(4), 655–661.

Winship J W (1967), 'Designing an automotive engine piston', SAE paper 670020.

Winterbone D E and Tennant D W H (1981), 'The variation of friction and combustion rates during diesel engine transients', SAE paper 810339.

Wong V W, Tian T, Lang H, Ryan J P, Sekiya Y, Kobayashi Y and Aoyama S (1994), 'A numerical model of piston secondary motion and piston slap in partially flooded elastohydrodynamic skirt lubrication', SAE paper 940696.

Xin Q (1999), *Analytical and Numerical Study of Lubrication Dynamics for Internal Combustion Engine Piston Assemblies and Journal Bearings*, D.Sc. (Doctor of Science) degree dissertation, Washington University, St. Louis, Missouri, USA.

Yagi T and Yamagata I (1982), 'Experimental method of determining piston profile by use of composite materials', SAE paper 820769.

Yang L, Ito A and Negishi H (1996), 'A valve train friction and lubrication analysis model and its application in a cam/tappet wear study', SAE paper 962030.

Yang Q and Keith T G (1995), 'An elastohydrodynamic cavitation algorithm for piston ring lubrication', *Tribology Transactions*, 38(1), 97–107.

Yang Q and Keith T G (1996a), 'Two-dimensional piston ring lubrication – part I: rigid ring and liner solution', *Tribology Transactions*, 39(4), 757–768.

Yang Q and Keith T G (1996b), 'Two-dimensional piston ring lubrication – part II: elastic ring consideration', *Tribology Transactions*, 39(4), 870–880.

Zhang C, Cheng H S and Wang Q J (2004), 'Scuffing behavior of piston-pin/bore bearing in mixed lubrication – part II: scuffing mechanism and failure criterion', *Tribology Transactions*, 47, 149–156.

Zhu G (1993), 'Valve trains – design studies, wider aspects and future developments', in Taylor C M (ed.), *Engine Tribology*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, pp. 183–211.

Zhu D, Cheng H S, Arai T and Hamai K (1992), 'A numerical analysis for piston skirts in mixed lubrication – part I: basic modeling', *Transactions of the ASME*, *Journal of Tribology*, 114, 553–562.

Zhu D, Hu Y-Z, Cheng H S, Arai T and Hamai K (1993), 'A numerical analysis for piston skirts in mixed lubrication – part II: deformation considerations', *Transactions of the ASME*, *Journal of Tribology*, 115, 125–133.

Zweiri Y H, Whidborne J F and Senvirante L D (2000), 'Instantaneous friction components model for transient engine operation', *Proc. IMechE*, *Part D: Journal of Automobile Engineering*, 214, 809–824.

11
Noise, vibration, and harshness (NVH) in diesel engine system design

Abstract: This chapter addresses NVH characteristics in diesel engine system design. By focusing on engine noise – a critical performance attribute for engine competitiveness, this chapter provides a comprehensive coverage of the NVH issues that a system engineer can evaluate by using engine system design/analysis tools. The chapter starts by introducing the fundamental principles of powertrain and diesel engine NVH, and establishes a three-level system modeling approach to engine noise. It summarizes the noise characteristics and noise-reduction design measures for both the overall engine and individual subsystems such as the noises from combustion, piston slap, valvetrain, geartrain, cranktrain, auxiliary, and aerodynamic sources.

Key words: noise, vibration, and harshness (NVH), noise identification, transient noise, engine NVH variation, combustion noise, piston slap, valvetrain, geartrain, cranktrain, intake and exhaust noise.

11.1 Overview of noise, vibration, and harshness (NVH) fundamentals

11.1.1 Basic concepts of NVH

NVH refers to an engine performance attribute represented by noise, vibration, and harshness. Sound is generated by a travelling wave of the fluctuation in the ambient air pressure. The speed of a sound wave is given by

$$v_{sound} = \sqrt{\kappa \cdot R_{gas} \cdot T} \qquad 11.1$$

where κ is the ratio of specific heat capacities ($\kappa = 1.41$ for air), R_{gas} is the gas constant of air ($R_{gas} = 287$ J/kg.K), and T is the ambient air temperature in K. The speed of sound is approximately 340 m/s at normal ambient temperatures. A sound pressure is generated when the sound wave propagates.

Most engine noises are originated by a source of energy (e.g., combustion of the fuel), generated by the vibratory motion of the solid surfaces of various engine components with high oscillatory accelerations, radiated to the surrounding air, and perceived by the human ear. Some other engine noises are produced aerodynamically by inducting the intake air into the engine and discharging the exhaust gas out of the engine tailpipe.

Vibration is an oscillatory mechanical motion. Engine vibration is excited by the cylinder pressure gas force and the inertia force acting on the cranktrain components, and transmitted through the engine mount to the drivertain and the entire vehicle body. Excessive vibration causes discomfort and even durability failures of the component.

Harshness usually refers to an unpleasant and subjective characteristic of noise or vibration, such as a short duration of noise spikes or tactile shock spikes, and the after-shake oscillatory vibrations.

Noise is perceived by the human ear for both loudness and sharpness after the air pressure fluctuations are converted to the mechanical motion of the cochlea which responds at different locations depending upon the excitation frequency of the sound. Sound characteristics can be described by sound pressure level and frequency with equal loudness contours. The sound pressure level has a unit of decibel (i.e., one tenth of a Bel), which in fact is a unit of power. It should be noted that the sound pressure level is not a measure of the loudness of a sound because the human ear is not equally sensitive to all frequencies. If the sound pressure of one noise is concentrated in a higher frequency range where the ear is more sensitive, it may be perceived louder than another noise with equal sound pressure level but concentrated in a lower frequency region that is less sensitive to the ear. In order to obtain the levels which bear a closer relationship to loudness judgment than the sound pressure levels, three different networks of frequency weighting (A, B, and C) were incorporated in sound level meters. The A-weighting most closely mimics the human ear, and it is most widely used in noise control work. The A-weighted sound level is expressed in the unit of dB(A) or dBA. A human whispering may be around 1 kHz and 30 dB(A). A normal conversation is at 0.5 kHz and 65 dB(A). A machine operator can be exposed to sounds around 0.5–5 kHz and 90–100 dB(A). The noise level in residential areas is around 50 dB(A). A vehicle noise can be 85 dB(A). Generally, a sound pressure level above 75 dB(A) is regarded as noisy, and a noise level above 90 dB(A) is considered very high. The threshold of pain is around 120 dB(A). In order to reduce or prevent hearing damage, standards organizations (OSHA) and national labor laws stipulate the limits of noise and impose hearing protection measures.

The audible frequency range (20 Hz–20 kHz) can be divided into bands with one octave wide. An octave is a frequency interval between two sounds whose frequency ratio is two (Table 11.1(a–b)). In the frequency analysis of noise signals, the narrow band spectrum data (e.g., obtained by Fast Fourier Transfer) provide the highest resolution but are more qualitative in nature. Usually, 1/3 (one-third) octave, sometimes octave and 1/12 octave, band spectrum with center frequencies is used to present more quantitative information in noise analysis. Another type of frequency band, the 'critical bands' of noises, was defined to divide the frequency range into 24 bands in the unit of Bark (Zwicker definition, Table 11.1(c)) with each band

Table 11.1 Noise frequency band definitions

(a) Octave band center frequency (unit: Hz)

31.5	63	125	250	500	1000	2000	4000	8000	16000

(b) One-third octave band center frequency (unit: Hz)

31.5	40	50	63	80	100	125	160	200	250
315	400	500	630	800	1000	1250	1600	2000	2500
3150	4000	5000	6300	8000	10000	12500	16000	20000	25000

(c) Zwicker's critical band definition

Critical band (Bark)	1	2	3	4	5	6	7	8
Center frequency (Hz)	50	150	250	350	450	570	700	840
Bandwidth (Hz)	100	100	100	100	110	120	140	150
Critical band (Bark)	9	10	11	12	13	14	15	16
Center frequency (Hz)	1000	1170	1370	1600	1850	2150	2500	2900
Bandwidth (Hz)	160	190	210	240	280	320	380	450
Critical band (Bark)	17	18	19	20	21	22	23	24
Center frequency (Hz)	3400	4000	4800	5800	7000	8500	10500	13500
Bandwidth (Hz)	550	700	900	1100	1300	1800	2500	3500

having a center frequency from 50 Hz in the lowest band to 13,500 Hz in the highest band; and it has a bandwidth from 100 Hz in each of the several low-frequency bands to 3500 Hz in the highest band.

Ideally engine noises are measured at different positions around the engine in a free field environment (i.e., anechoic without sound reflections or echoes) with electrostatic microphones to measure the sound pressure (a scalar 'point' property) or using intensity probes to measure the sound intensity (a vector 'field' property). Sound power is a measure of the acoustic power radiated from the source. Sound intensity is the sound power per unit area, given by

$$I_s = p_s v_{air} \tag{11.2}$$

where the acoustic velocity v_{air} is the velocity of the local oscillating motion of the air particles. The sound power $\dot{W}_s$, sound intensity I_s, and sound pressure p_s are related for an engine by the following:

$$\dot{W}_s = \int I_s \cdot \mathrm{d}A = I_s A = \frac{p_s^2}{\rho v_{sound}} A \tag{11.3}$$

where ρ is the air density and A is the area.

Sound magnitude is characterized by sound pressure or sound power. Because the perception of the human ear to loudness is proportional to the logarithmic scale of the sound pressure rather than a linear scale, a logarithmic scale is usually used to define the sound pressure level in the unit of decibel (dB) as follows:

$$p_{SPL} = 10 \cdot \log_{10}\left(\frac{p_s}{p_{s,ref}}\right)^2 = 20 \cdot \log_{10}\frac{p_s}{p_{s,ref}} \tag{11.4}$$

where $p_{s,ref}$ is a reference sound pressure, usually equal to 20 micro-pascals (μPa), which is the minimum sound level the human ear can hear. A sound pressure of 20 μPa corresponds to a sound pressure level of 0 dB. A doubling of any value of the sound pressure corresponds to an increase in sound pressure level of 6.02 dB. A multiplication of the sound pressure by a factor of ten corresponds to an increase in sound pressure level of 20 dB. Sound pressure is inversely proportional to the distance from the source. The sound pressure level decreases by 6 dB for each doubling of the distance from the source.

The sound power level is given by the following in the unit of dB:

$$\dot{W}_{SWL} = 10 \cdot \log_{10}\frac{\dot{W}_s}{\dot{W}_{s,ref}} \tag{11.5}$$

where $\dot{W}_{s,ref}$ is a reference sound power, usually equal to 1×10^{-12} watt.

The sound power level and the sound pressure level are related by the following:

$$p_{SPL} = \dot{W}_{SWL} - 20 \cdot \log_{10} l - 10.9 + C_s \qquad 11.6$$

where l is the distance from the source of the noise in meters, and C_s is a correction term given by Harris (1979).

The sound intensity level is given by the following in the unit of dB:

$$I_{SIL} = 10 \cdot \log_{10} \frac{I_s}{I_{s,ref}} \qquad 11.7$$

where $I_{s,ref}$ is a reference sound intensity, usually equal to 1×10^{-12} W/m^2.

Sound intensity measurement can estimate sound power. Intensity measurement techniques provide information about the radiation characteristics of sound sources and their spatial distribution.

Engine vibration can be measured by using the low-mass accelerometers rigidly mounted to the test structure. The vibration signals, such as displacement, velocity and acceleration can also be converted to vibration displacement level, velocity level and acceleration level, respectively, in the unit of dB as follows:

$$l_{VDL} = 20 \cdot \log_{10} \frac{l}{l_{ref}} \qquad 11.8$$

$$v_{VVL} = 20 \cdot \log_{10} \frac{v}{v_{ref}} \qquad 11.9$$

$$a_{VACL} = 20 \cdot \log_{10} \frac{a}{a_{ref}} \qquad 11.10$$

where l_{ref} is a reference displacement, for example 1×10^{-12} m; v_{ref} is a reference velocity, e.g., 1×10^{-9} m/s; and a_{ref} is a reference acceleration, e.g., 1×10^{-6} m/s^2. Moreover, mobility refers to the ratio of the velocity response at the excitation point on the structure to the applied force at the point.

Noise and vibration data are characterized by an overall level and the levels of detailed signals. The overall level usually refers to either an average (root-mean-square or RMS) or peak signal, often averaged from four microphones installed at different locations one meter away from the engine surface. The detailed signals include a time history or frequency breakdown of the sound pressure level or vibration data. Often A-weighted overall sound pressure levels and one-third octave spectra are used in noise testing. For more information about noise fundamentals, the reader is referred to Harris (1979) and Hickling (2005).

11.1.2 Objective and subjective assessment of noise

Noise and vibration signals can be measured and evaluated objectively. However, because the human body has its own frequency response (e.g., the ear's response to different frequencies), NVH problems are also largely tied to subjective impressions or sound quality issues, which are the subject of the field of psychoacoustics. The human ear has a logarithmic response to the magnitude of sound pressure waves and is most sensitive to frequencies of about 1 kHz. The human ear is sensitive to broad-band impulsive noises. Poor diesel sound quality may cause subjective annoyance even if the sound pressure level is low. For more in-depth information about human hearing and sound quality evaluation methods, the reader is referred to Sasaki and Nakashima (2007).

Human hearing reacts nonlinearly to sound pressure. The unit of decibel was conceived partly for that reason. Although the simple frequency weightings (e.g., A-weighting) try to mimic human hearing response, they do not provide information about sound quality. Many psychoacoustic parameters are based on a modified frequency scale, the tonality scale, from 0 to 24 Bark. Psychoacoustic or sound quality parameters include the following:

- loudness (in the unit of sone)
- sound volume (phon)
- severity or sharpness (a weighting that emphasizes the high frequencies in the unit of acum)
- fluctuation strength or amplitude (characterizing the extremely low frequency such as the less than 20 Hz unpleasant sound signals)
- roughness (assessing the rough modulations in the 20–300 Hz frequency range)
- tonality (classifying the pure tonal content in the noise)
- repetition frequency of periodic sound as perceived periodicity in Hertz
- impulsiveness or knocking measured by kurtosis
- harmony or sound pressure distribution
- spatial selectivity.

The models and evaluation methods of these psychoacoustic parameters have been developed to facilitate the design decisions related to sound quality. The direct injection diesel engine is the most important source of interior noise for a diesel vehicle. The subjective noise character of the engine can be evaluated using AVL's Annoyance Index. The index links the perceived sound quality to measurable psychoacoustic parameters so that the engine noises may be described by the index as a function of loudness, perceived periodicity, sharpness, and impulsiveness.

In engine development the major objectives of noise control are to reduce the noise level in dB(A) to meet legislative requirements and to reduce the annoyance of the noise to satisfy customer requirements on the subjectively perceived sound quality. The sound pressure level and annoyance are different attributes. Two different sound levels apart by several dB(A) may be observed subjectively to have equal annoyance. Measures to reduce the engine noise level in dB(A) do not necessarily lead to an improved sound quality. Often it may even result in deterioration in the subjective noise characteristics (Schiffbanker *et al*., 1991). More information about the subjective characteristics of diesel engine noises, noise measurement methods, and diesel engine noise data examples can be found in Rust *et al*. (1989) and Corcione *et al*. (1989).

11.2 Vehicle and powertrain noise, vibration, and harshness (NVH)

11.2.1 Noise regulations

As the regulations on pollutant emissions of diesel engines become more stringent, the limits in vehicle noise regulations have also been drastically reduced over the last 30 years. NVH development has been subject to vehicle-level NVH regulations since the first pass-by noise regulation in 1970 in the EU and in 1971 in Japan and later in 1978 in the US Vehicle noises are usually tested as the following: (1) the pass-by noise in front of a by-stander; (2) the noise received by the operator under various load conditions; or (3) the noise received by the passenger in the cabin. The US federal noise regulations for trucks require 80 dB(A) at 15 m (EPA CFR Part 205, SAE J366) for pass-by noise (during mid-gear full-pedal acceleration), and 85 dB(A) at 15 m (EPA CFR Part 325, SAE J1096) for stationary noise (during no-load acceleration and deceleration, i.e., from idle to maximum governed engine speed and then to idle).

Pass-by noise measurements are required for new vehicles to fulfill the legal regulations on vehicle exterior noise. The measurement is conducted to obtain sound pressure levels in dB(A) radiated by an accelerating vehicle. In order to effectively reduce the vehicle's exterior noise through design, engineers need to know the dominant noise sources and their transfer paths.

Correlating the vehicle pass-by noise to a design target of engine noise is an important engineering practice. The diesel engine is a major NVH excitation source in the vehicle. When the engine is running, a small portion of the energy involved is lost through engine vibration to the surrounding air as sound. To meet the vehicle noise target of 80 dB(A) during acceleration pass-by, an engineering target lower than 80 dB(A) needs to be established, for example 78 dB(A) to ensure production audit by reserving a certain margin to cover tolerances and variations. The contribution of engine noise to total truck pass-by noise level also needs to be determined, for example 75 dB(A). Note that every 3 dB reduction corresponds to having the sonic energy decreased by half. Other contributions to the vehicle pass-by noise come from the tire and the vehicle body. It is important to recognize the relative importance of the contributions from the engine and the tire to the total vehicle noise level. If the total noise is low and the tire noise is high, the engine's contribution becomes less important.

Engine sound level measurement is normally conducted according to the SAE J1074 procedure in an engine test cell. A noise target (for example 93 dB(A)) in engine noise control can be set at one meter away from the engine surface with the engine operating at the steady-state rated power condition. Such a target may approximately represent the noise level required by the engine in the vehicle in order to meet the vehicle pass-by noise regulation level at 80 dB(A). The target cascading from the vehicle level noise on the road during acceleration to the engine level noise in the test cell running at steady-state power is largely empirical, and can be very complex. However, this is an important area where a system engineer can play a major role to contribute to make this translation more analytically elegant and accurate.

Moreover, it should be noted that the global noise regulations are not directly comparable because they are tied to different measurement procedures. For example, the test is conducted at 7.5 m away from the microphone position under the European noise regulation ISO R362. On the other hand, the same test is conducted at 15 m away from the microphone position under the US SAE measurement standards. Previous practice has demonstrated that the European noise regulation seems more stringent (Wodtke and Bathelt, 2004). More information on vehicle noise regulations can be found in Reinhart (1991) and Cherne (1993). The SAE procedures listed in the section of references at the end of this chapter also provide good fundamental knowledge on vehicle and engine noises and their measurement methods.

11.2.2 Classification of powertrain and drivetrain NVH problems

NVH is one of the most important vehicle attributes, along with drivability, durability, ride, and handling. Vehicle NVH can be classified into three types of issues from a receiver's perspective, namely acoustic, visual, and tactile. Examples of those issues include the noise at the driver's and the passenger's ears, the noise perceived by a by-stander, the vibration of the rearview mirror, the vibration of the steering wheel, and seat shake.

Vehicle NVH can also be classified into two categories, interior and exterior. The interior NVH issues deal with the noise sound level, sound quality, and vibration experienced by the occupants inside the vehicle cabin. The design criterion of the interior cabin noise can be as simple as a sound pressure level received by the driver or all occupants at certain vehicle speed and load, or as complicated as full sound quality metrics (e.g., the Zwicker sound quality metrics). The exterior noise radiated by the vehicle has a direct impact on the quality of life for the people who hear the noise. The exterior noise is subject to the pass-by noise regulation. The diesel engine contributes to a large extent to both the interior and exterior noises in sound pressure level and sound quality.

Moreover, vehicle NVH problems can be classified according to the dynamic excitation sources. The external excitation sources include road surface, wind and other environmental effects. The internal excitation sources include powertrain (engine and transmission) and drivetrain forces, such as engine combustion, engine reciprocating imbalance, engine vibration and torque cyclicity, engine-driven accessory disturbances, engine mount vibration, intake and exhaust noises, gear meshing variation, torque converter imbalance, driveshaft and half-shaft imbalances, tire and wheel imbalance, and brake-induced forces. Diesel trucks usually have three major noise sources: the engine (including the exhaust system), the cooling fan, and the tires. Note that the intake and exhaust aerodynamic noises are generally considered as a part of the engine noise because its origin is the gas wave dynamics inside the engine, while the intake and exhaust structure-borne noises (e.g., intake system mount, exhaust hanger, and heat shield) can be considered as a part of the vehicle noise.

Engine noise is no doubt the most important noise in powertrain development. However, transmission noise (mainly gear rattle) sometimes may become prominent, for example the gear rattle at idle in a manual transmission. The rattle noise is related mainly to transmission errors and variations in gear rotation. Gear rattle noise increases if the powertrain oscillates at its resonance frequency. Note that unique NVH issues exist in hybrid powertrains, for example the torque blending mismatch between the engine and the electric motor, the driveline vibrations caused by low-speed electric motor torque ripple, and the motor gear rattle noise. The powertrain and drivetrain NVH issues are discussed in more detail by Steyer *et al.* (2005), Juang *et al.* (2006), Wellmann *et al.* (2007), Tousignant *et al.* (2009), and Govindswamy *et al.* (2009).

Depending on the vehicle speed the major noise sources can be simply traced to two sources, the engine noise and the road–tire noise. The engine noise dominates at low vehicle speeds and the road–tire noise dominates at high vehicle speeds. Wind noise is usually relatively low.

Vehicle noises are classified into two categories, structure-borne and airborne. Their definitions were given by Anderton and Zheng (1993):

> The sound entering the air via an acoustic path is the sound radiated by an internal sound source, e.g., valve impact which propagates through the

acoustic medium, and is then transmitted through the wall of the (valve) cover. This type of radiated noise is referred to as airborne sound. The sound entering the air via a structural path is the sound which has its source in the vibrational energy that has been transmitted through the structure (either flexural or compressional waves) from an engine's and its component's (including valve mechanism itself) vibrational force. This type of radiated sound is referred to as structure-borne sound.

The structure-borne noise transmits via the engine mountings, drive shafts and other connecting elements into the vehicle body. The airborne noise transmits directly from the engine compartment to pass through the body shell wall to enter into the vehicle to contribute to the total interior noise. Automotive structure-borne noise is usually below 500–1000 Hz in a low frequency range. The airborne noise is usually in a wide range of 300 Hz to 7 kHz, generally in a high frequency range. Detailed classifications of structure-borne and airborne noises and the source of excitations are detailed by Riding and Weeks (1991).

According to frequency range the noise can be classified as low frequency noise and high frequency noise. For instance, the structure-borne noise which is transferred to the vehicle body via the engine mounts occurs in the frequency range below 800 Hz and can be regarded as a low frequency noise. On the other hand, the structure-borne noise radiated from the engine surface to the surrounding air occurs in the frequency range above 800 Hz and can be called high frequency noise.

11.2.3 Vehicle powertrain and drivetrain NVH process

Vehicle powertrain and drivetrain NVH development processes are described in Castillo (2001), Alt *et al*. (2003), Laux *et al*. (2005), Mori *et al*. (2005), March *et al*. (2005a, 2005b), and Afaneh *et al*. (2007). Understanding how the engine NVH target is cascaded down from the vehicle level is very important for diesel engine system design.

11.3 Diesel engine noise, vibration, and harshness (NVH)

11.3.1 Classification of engine NVH

The area of engine vibration is very broad, including engine balance, engine internal excitation forces, and engine mount vibration. Noise is a more direct performance attribute than vibration for customer satisfaction although they are related. The engine noise sources originate from combustion processes, mechanical movements, and air charging/discharging processes. The excited vibrations are absorbed and transmitted to the structure, and the noise is generated from the vibrating surfaces. The engine noise includes that excited by high-frequency impulsive forces (e.g., combustion pressures, piston slap impacts) and that excited by low-frequency inertia loads acting on the components. Engine noise includes the noises from combustion, piston slap, valvetrain, fuel injector, cranktrain, engine block (including covers and heat shields), geartrain, accessory, turbocharger, and intake and exhaust flows. They can be classified into three types: combustion noise, mechanical noise, and aerodynamic noise. The combustion and mechanical noises are mostly structure-borne noises.

The mechanical noise is caused mainly by the mechanical impacts between the moving components due to the cylinder pressure loading and inertia forces. The impacts happen at the engine bearings, the piston–liner interface, the valve seat, and the gear drives. The mechanical noise also includes the noise due to the hydraulic pressure fluctuations in the auxiliaries such as the fuel injection system and the oil pump, and the noise due to the electromagnetic mechanisms in the alternator.

The combustion noise is affected mainly by the rate of rise of the cylinder pressure and the lasting duration of the high rate of rise. It is the noise above the engine firing frequency and its harmonics in the frequency domain. It is directly due to the combustion excitation in the cylinder. It is induced by the in-cylinder gas pressure wave oscillations, and is the radiated noise from the vibratory surfaces of the cylinder head, the piston, the connecting rod, the crankshaft, and the engine block.

The aerodynamic noise is mainly an airborne noise induced by the gas flow pressure waves in the intake and exhaust systems, the turbocharger, and the cooling fan. Often the fan noise is considered as a part of the vehicle noise.

The experimental and analytical investigations on engine noise and its controls are normally conducted in the following areas:

- overall sound pressure levels from motoring, no load to full load across the entire engine speed range
- detailed sound pressure levels in the frequency/order spectrum and/or the time/crank angle domain at different engine speeds and/or loads (e.g., the contour map of sound pressure levels in the domain of 'noise frequency vs. engine speed' – the Cambell diagram; the contour map of sound power levels in the domain of 'noise frequency vs. crank angle' – the Wigner–Ville analysis)
- sound intensity measurement
- sound quality measurement and analysis (e.g., binaural head recording, objective and subjective assessment)
- noise source identification for the contributions from the structure-borne and airborne noises in the frequency spectrum, noise radiation ranking

- Noise source identification for the contributions from combustion noise, mechanical noise, and aerodynamic noise (from intake, exhaust, turbocharger, and cooling fan) in the frequency spectrum
- breakdown of the mechanical noise at different engine speeds and loads in the frequency spectrum (e.g., piston assembly, valvetrain, cranktrain, block, geartrain, belt, chain, fuel system, pumps, and accessories)
- differences between the steady state and the transient noises
- noise transfer path and structural attenuation in the frequency spectrum
- sound radiation from the engine surfaces (e.g., engine block, mounts, crankcase, crank pulley, flywheel cover, front cover, valve cover, oil pan, heat shields, manifolds, engine mounts, cylinder head, air filter, exhaust pipe, turbocharger, pumps, alternator, and starter).

11.3.2 Differences between diesel and gasoline engine noises

It is well recognized that the noise level of the diesel engine is higher than that of the gasoline engine. The sound quality of the diesel engine also tends to be worse due to its internal impulsive excitations which induce high-frequency noises. The heterogeneous combustion in the diesel engine is generally noisier than the homogeneous combustion in the gasoline engine due to the high rate of rise of the in-cylinder pressure after the ignition delay. The so-called diesel knocking refers mainly to the noise in the frequency range of 500–6000 Hz. It has been recognized that diesel knocking is the major contributor to poor sound quality perceived by the customer, particularly at the low-idle condition.

The peak cylinder pressure in the gasoline engine is much lower than that in the diesel engine. The moving components of the diesel engine are designed relatively heavier and stronger than those in the gasoline engine in order to meet the durability requirements at higher cylinder pressures. Therefore, the diesel engine has higher mechanical impact noise than the gasoline engine if compared at the same engine speed. However, if compared at their respective rated speeds, their noise levels are often similar because the gasoline engine has a higher rated speed. The acoustic similarity of the combustion noise and the piston related noise in the diesel engine often make their differentiation and a subjective rating of the piston noise difficult (Künzel *et al.*, 2000, 2001). Moreover, the diesel engine usually has more auxiliary components such as the fuel injection pump and the turbocharger. They are potential sources of additional noise and vibration.

On the other hand, the gasoline engine exhibits some noises that the diesel engine does not have, for example the piston pin ticking noise at the low-speed no-load conditions (Werkmann *et al.*, 2005; Moshrefi *et al.*, 2007), the cold start piston clatter noise due to a tertiary motion (Pollack *et al.*, 2005), and the prominent stick slip piston noise emitted from the excited crankshaft (Beardmore, 1982; Werner, 1987). These noises are due to the unique characteristics of the gas loading and the mass of the gasoline engine (Künzel *et al.*, 2000).

Stucklschwaiger *et al.* (1999) compared the differences in NVH between the diesel and gasoline engines. The design configurations related to NVH features at the vehicle and powertrain levels were discussed. Engine mounting, engine balance, bank angles, balancing shaft design, and other NVH design issues were elaborated for light-duty trucks in their study.

11.3.3 Diesel engine noise characteristics

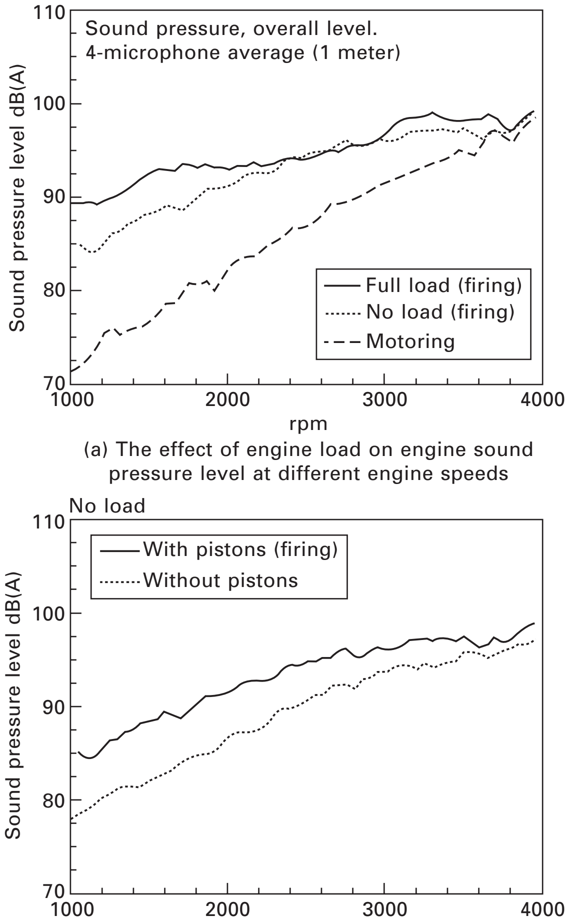
Understanding the contributions to the total engine noise from combustion, mechanical impacts and aerodynamic effects in the engine speed–load domain and knowing their behavior in the frequency spectrum are important for identifying and prioritizing corrective measures to control the noise. For example, for combustion noise dominated engines design improvement should be focused on combustion and fuel system design, emissions calibration, and structural attenuation.

Engine full-load noise characteristics are especially important. Reduction of the engine noise particularly under full-load operation is often required in order to comply with vehicle noise regulations. The mechanical noise usually increases with engine speed. Figure 11.1 shows that there is a large difference between the firing and motoring noise levels, especially at low engine speeds. At higher engine speeds the difference becomes smaller due to the increasing mechanical noise. Figure 11.1 also shows the effect of the piston on the noise.

Informative measurement results revealing the fundamental characteristics of diesel engine noises are provided by Badawi *et al.* (2007). They plotted the typical NVH events in the time–frequency domain, including valvetrain operation, fuel injection, combustion, and piston slap. Govindswamy *et al.* (2007) illustrate the diesel impulsive noises including diesel knocking, injector ticking, and gear rattling. Diesel engine brake noise is introduced in Chapter 6.

11.3.4 Engine noise identification

In order to design a low noise engine one must know the sources and paths of the noise generation and transmission. Several methods have been used to separate the noise sources. An engine is fired and motored respectively to separate the mechanical sources from the combustion sources. Lead covering techniques are used to identify the noise radiating characteristics



(a) The effect of engine load on engine sound pressure level at different engine speeds

(b) The effect of removing the piston on engine sound pressure level at no load

11.1 Effect of diesel engine operating conditions on noise level.

of any individual surface or component of the engine to be evaluated. The entire engine is wrapped with a lead sheet of absorbing material (acoustically masked), and then each component (e.g., the oil pan, the manifold) is exposed one at a time to determine its contribution to the engine noise. The breakdown and individual contributions of the mechanical noise can also be obtained by removing the subject component one at a time during the test.

It is usually difficult to clearly separate the combustion noise from the mechanical noise because the latter is affected by the gas loading coming from the in-cylinder pressure. In order to separate these noises, the engine noise can be measured at different levels of combustion excitation, and then the overall and spectral contributions from the combustion and mechanical sources can be estimated. For example, at a given frequency, if the data of 'measured sound pressure level of the engine vs. the decibel level of the cylinder pressure' form a line, the slope of the line would indicate the structural attenuation for the cylinder pressure, and the interception would approximately reflect the sound pressure level of the mechanical noise. The structural attenuation essentially reflects a type of transfer function of the system, which is measured by the ratio of the output of the system (response) at a certain location to the input of the system (i.e., the driving force) at another location. Different transfer functions are obtained for different input and output locations. Extensive research has been conducted in the area of separating the combustion noise from the mechanical noise. This area is important for accurately estimating the noise contribution from the combustion in order to reduce it.

11.3.5 Transient engine noise

Most engine noise research has been conducted typically based on steady-state noise measurements because the highly dynamic acoustic test cells are not commonly available. However, the vehicle noise legislation requires the pass-by test to be conducted during transient acceleration operation. Diesel engine transient noise is significantly different from that at the corresponding steady-state conditions (i.e., if compared at the same engine speed and same fueling rate). For example, the transient noise could exceed the steady-state full-load noise by up to 6 dBA (Dhaenens *et al*., 2001). If the vehicle accelerates after a period of idling or light-load running, the noise during the acceleration transient will be higher than that at the steady-state conditions due to longer ignition delay and higher cylinder pressure gradient (rise rate). The longer ignition delay can be caused by colder intake air temperature, lower boost pressure due to turbocharger lag, lower combustion chamber wall temperature, and more advanced dynamic injection timing. In fact, the transient engine noise depends on the running history before the transient event occurs. Different running history may produce different conclusions on the difference between the transient noise and the steady-state noise (Shu *et al*., 2005b, 2006). The difference in noise is influenced largely by the difference in ignition delay. Diesel engine noise under transient conditions was studied by Head and Wake (1980), Rust and Thien (1987), Dhaenens *et al*. (2001), and Shu *et al*. (2005b, 2006). Modeling the noise difference between the transient (especially the acceleration transient as in the drive-by noise test) and the steady-state conditions is very important.

11.3.6 Diesel engine NVH variation

Diesel engine NVH variation is a very important issue to consider in NVH development, measurement and overall system design/planning. It includes test variability and engine-to-engine variability. The statistical variability may mask the small difference caused by the design changes. The theory of engine NVH variability and the application in diesel engine noise reduction were introduced by Reinhart *et al*. (2003). They illustrated that NVH design decisions are often made based on a small measured difference that is statistically insignificant. It is important to design a test plan in order to achieve statistically significant results. They used an example of the SAE J1074 noise test to emphasize that if the data from a single comparison test are used to evaluate the effect of a design change, the difference must be greater than 0.8 dB to be more than 90% confident that there is a real difference. In other words, a measured 0.8 dB difference can indicate a true difference anywhere between 0 dB (no change) and 1.6 dB (a substantial change)! In NVH system design, using advanced simulation tools to accurately quantify the benefit of design changes (especially the small changes), applying the theory of design-for-variability in robust design, and considering the NVH variability are three key elements to make the design successful.

11.3.7 Engine NVH development process

Good NVH characteristics are a major competitive advantage for an engine product. Due to their complex nature, NVH failures often cause costly late design changes. It is important to take NVH into account in diesel engine system design from the concept stage and to develop practical and advanced engineering rules that can be used in system design.

The engine NVH development process generally consists of two stages: (1) the concept design and analysis stage; and (2) the product design and testing validation stage. In the concept stage, NVH criteria are defined in target setting/cascading with a top-down approach from the system level cascaded to the component level. Internal engine excitation forces or noise sources (e.g., combustion, mechanical, and aerodynamic noises) are simulated. Usually, NVH problems are analyzed in a chain of input–transfer–response. The input or excitation sources of the NVH problems can be handled by multi-body system dynamics in engine system design to identify the force loading. Since the best approach to control diesel engine NVH is to control it at the source of excitation, engine system design plays a critical role in the NVH development process because all the major engine design parameters and the engine loads are optimized and determined at the system design level. The transfer path of the NVH problems is usually handled by finite element analysis (FEA) or statistical energy analysis (SEA). The acoustic response can be simulated by the boundary element method (BEM). Once the prototype design is available, FEA is conducted to calculate the modal and vibration responses of engine components and assemblies, and BEM is used to predict the radiated noise levels.

In the product design and testing stage, the engine sound pressure level (using SAE J1074), sound intensity, and sound power are measured for the entire engine and certain individual sources (e.g., intake and exhaust noises) with a bottom-up approach from the component level integrated up to the system level. Sound quality metrics are also evaluated. Frequency response function and transfer path analyses are conducted. Engine mount vibration and torsional vibration are measured. The truck pass-by noise (SAE J366, Ruffinen *et al*., 1995) and the stationary noise (SAE J1096) are measured, and the vehicle interior noise is evaluated.

In the detailed design and FEA analysis phase, the structural vibration response is predicted in terms of natural frequency, mode shape, and damping. The simulation accuracy of FEA depends on how accurately the following are specified in the model: physical properties of the structural material, boundary conditions of the restrained structure, the type and the location of the dynamic forcing function, the number of modes, and the frequency range of interest. The FEA forms mass and stiffness matrices that are used to determine the eigenvalues and eigenvectors (or natural frequency and mode shapes) of the system. The damping, along with the natural frequency, determines the dynamic behavior of the system in terms of amplitude, overshoot, and settling time. FEA can play a major role in the system integration phase by predicting the overall engine noise after all the component designs are proposed. The process of low-noise engine design is discussed in detail by Beidl *et al*. (2001).

11.3.8 Design measures to reduce engine NVH

Design measures to reduce engine NVH generally fall into four major categories: (1) reducing the strength of the excitation at the source by engine design or operation measures; (2) muffling or silencing (for aerodynamic noise); (3) reducing or isolating sound transmission path by structural attenuation; and (4) noise insulation and encapsulation. Engine system design focuses on the first two categories, and they will be detailed in later sections. The other two categories are usually component design work. However, in order to predict the overall engine noise accurately, the design characteristics in the latter two categories need to be understood by the system engineer.

In structural attenuation, there are basically four methods to reduce the noise radiated from the engine surfaces: (1) increasing the stiffness and resonant frequency of the structure (e.g., adding ribs/fins or increasing wall thickness); (2) reducing the surface area; (3) increasing the noise transmission loss (e.g., using absorption or barrier materials for the airborne noise); and (4) interrupting the noise transfer path (e.g., structural isolation, mounting, damping by the use of mass dampers for the structure-borne noise). When the structure stiffness is increased the structural resonances are shifted toward higher frequencies where the amplitudes of the dynamic excitation forces become smaller (Wodtke and Bathelt, 2004). The mechanism of noise damping is to convert the sound energy to another type of energy, normally heat. Typical examples of noise damping are using the parts made of foam or porous materials. Light-weight engine designs have been used to increase the engine power-to-weight ratio, but lowering the engine structure mass will generally weaken structural attenuation and result in higher noise. Engine design needs to reach the best compromise between engine performance and vibro-acoustic comfort.

The mechanism of noise insulation is to block and reflect the sound waves. In addition to the engine block surface, a large amount of noise is radiated from the accessories or shields that are attached to the basic structure, such as the oil pan and the turbocharger. These attachments usually have light structures and their noises are easier to control than for the basic engine block. The attached components should be fastened at the locations having low vibration or high stiffness in order to avoid high noise. Local sound covers (e.g., valve cover, oil pan cover, geartrain cover) can reduce their radiated noises effectively. Damping materials may be applied on the surface of the covers or shields to further dissipate the sound. Lastly, total engine encapsulation is an effective measure to reduce the engine noise. However, total engine encapsulation is expensive and increases the engine weight and size. It also requires special ventilation design in order to prevent the engine surface from overheating.

Heavy-duty diesel engine NVH is summarized by Walker (1999). Light-duty diesel engine NVH is briefly introduced by Schulte (1999) and Wolf *et al.* (2003). Reviews on the diesel engine noise are provided by Austen and Priede (1959), Grover and Lalor (1973), Priede (1980), Hickling and Kamal (1982), Challen and Croker (1982), Russell (1982), Yawata and Crocker (1983), Haddad (1984), Cuschieri and Richards (1985), Farnell and Riding (1999), and Gaikwad *et al.* (2007). The diesel engine structural design guidelines for noise control are summarized by Priede *et al.* (1969), Jenkins and Kuehner (1973), Anderton and Priede (1982), and Challan (1982). Overall design strategies of low-noise diesel engines are elaborated by Anderton (1984), Brandl *et al.* (1987) and Boesch (1987). Diesel engine idle noise characteristics has been researched by Miura and Kojima (2003). A good design example of a modern low-NVH diesel engine is provided by Kwak *et al.* (2007). Diesel engine acoustic encapsulation technique is presented by Nathak *et al.* (2007).

11.3.9 The approach of diesel engine system design to NVH

The term NVH has been widely used to address the noise and vibration problems together. Indeed, noise and vibration are closely related. They are both performance attributes of the engine. In fact, vibration is also an attribute of structural durability because excessive vibration may cause structural failures. The central theme of diesel engine system design is to conduct a good performance-based design subject to durability constraints with the focus at the system level rather than at the component level. Therefore, many vibration problems at the component durability level are outside the scope of engine system design (e.g., dynamic stress evaluation of the crankshaft torsional damper, bending block, rocker cover, and oil pan). However, the overall planning of designing a low-noise and low-vibration engine does belong to the scope of engine system design.

The NVH attribute often conflicts with other attributes such as packaging, durability, cost, and other performance attributes. System-level optimization is required to balance these attributes. Due to the extreme complexity of NVH issues, experimental work is often needed on the actual parts in order to make final judgments. However, analytical methods can be effective in narrowing down the specification candidates and in shortening the confirmation cycle.

A 'top-down' system design is not equal to the 'bottom-up' summation of all the component designs in NVH. At most the latter can be called 'system synthesis' in a late validation stage of the design cycle. Defining the roles and distinguishing between the system work and the component work in the NVH area are important in order to avoid losing focus and unnecessary repeating between the two. The system design work is characterized by the following: (1) ownership of planning the system-level NVH loading and its associated hardware and software; (2) predicting the system-level NVH operating characteristics; and (3) system optimization. A system engineer cares less about the local design solutions of a component unless that design affects the entire system.

In general there are five advanced simulation tools that are used in the NVH area for virtual powertrain analysis, namely (1) gas wave dynamics; (2) engine cycle simulation; (3) multi-body dynamics; (4) finite element analysis (FEA); and (5) the boundary element method (BEM). The first three tools are particularly suitable for engine system NVH analysis. They are used by a system engineer in his/her daily work to analyze the system behavior of the gas flows and system vibrations, and he/she has the advantage to take these tools a step further to apply them in the NVH area. The FEA and BEM tools are usually very time-consuming and normally more suitable for the component-level design analysis. In particular, the multi-body dynamics in engine dynamic simulations includes valvetrain dynamics, piston-assembly dynamics, cranktrain dynamics, engine balance, and geartrain dynamics.

Three levels of NVH models can be considered in diesel engine system design to predict the excitation sources, vibrations, and even noises from various levels of complexity. These models are as follows:

- Level-1 system noise model, which does not have crank-angle resolution.
- Level-2 system NVH model, which has instantaneous crank-angle resolution for all the excitation sources. It does not have to be real-time capable but needs to be computationally fast for system optimization. This model is the primary approach in diesel engine system design for NVH. A real-time capable level-2 model can be used for real-time engine controls related to noise issues such as fault diagnostics and active noise control measures.
- Level-3 system NVH model, which involves FEA or BEM type of simulation, and is computationally expensive and cannot be real-time capable. It predicts structural attenuation (or transfer function for the noise) and noise radiation. The output of the model includes the detailed velocity distribution on engine surfaces and the noise emitted. The level-3 model can be used in the system synthesis analysis stage with a detailed simulation.

A combination of analytical and empirical approaches is required in the modeling. These models will be detailed in Section 11.12.

11.4 Combustion noise

Combustion noise is an important concern in combustion system design and performance calibration. Combustion is the primary source of noise in most naturally aspirated direct injection diesel engines. Although the combustion noise in turbocharged diesel engines is not dominant at high-speed and high-load steady-state conditions, it can become dominant at idle, light-load, or acceleration conditions.

Combustion noise is transferred to the top deck of the cylinder head and the cylinder wall. It is also transferred to the piston, the connecting rod, the crankshaft, the main bearings, and the crankcase. Combustion noise is related to the rate of rise of the in-cylinder pressure ($\mathrm{d}p/\mathrm{d}\phi$) in the premixed combustion phase and also related to the lasting duration of the high rate of rise, which is directly governed by the duration of ignition delay. It is believed that the sharply rising cylinder pressure induces highly transient oscillatory shock waves and excites the cylinder block structure to vibrate and emit noise. Moreover, the gas waves during the heterogeneous combustion process are reflected by the cylinder wall to form highly oscillatory gas waves to produce noise. Because the natural frequencies of the cylinder structural components are mostly at the medium or high frequency, the structure is excited by the combustion to generate noise in the medium- to high-frequency range. The noise is perceived by the ear as impulsive and unpleasant.

Combustion noise is sensitive to the rate of rise of the cylinder pressure. A small change in the cylinder pressure trace may not affect the output power significantly, but it may affect the combustion noise appreciably. A frequency-domain analysis of the in-cylinder pressure trace is necessary in order to analyze the sensitivity and identify appropriate solutions to control the combustion noise. The combustion noise is affected mainly by the cylinder pressure level (in dB) in the medium-frequency range (e.g., 100–1000 Hz) that is related to $dp/d\phi$, and also in the high-frequency range (e.g., higher than 1000 Hz) that is related to $d^2p/d\phi^2$. In the low-frequency range the pressure levels exhibit an oscillating and decaying pattern that reflects the firing frequency of the engine and its multiple harmonics, just like a typical periodic forcing function would act. The continuous monotonic pattern in the medium- to high-frequency range indicates a rising slope of the cylinder pressure in the time domain. A larger $dp/d\phi$ gives a flatter slope of the pressure level (in dB) in the frequency domain, resulting in higher combustion noise.

Tung and Crocker (1982) discovered that for turbocharged diesel engines the frequency content of the combustion gas pressure up to about 300 Hz is related to the peak cylinder pressure. The frequencies between 300 Hz and 2000 Hz are related to the maximum rate of rise of the cylinder pressure. The frequencies above 2000 Hz are related to both the amplitude and the duration of the second-order pressure derivative. They also found that the frequencies of the cylinder pressure fluctuations are closely related to the cavity resonance frequencies in the combustion chamber.

Because the resonant frequencies of many components in the diesel engine are in the medium- and high-frequency range (e.g., above 1000 Hz), it is relatively easier to excite the structure vibration and noise in the medium- to high-frequency range than in the low-frequency range. Therefore, the large amplitude of the cylinder pressure in the low-frequency range usually does not translate to high vibration and noise. The term 'structural attenuation' is defined as the difference between the cylinder pressure in dB and the sound pressure level emitted from the engine surface. Inherently, the engine has high (strong) structural attenuation in the low-frequency range and low attenuation in the high-frequency range. When the structure stiffness of the cylinder bore is very high the resonant frequency of the structure becomes very high. If the resonant frequency is higher than the frequency of the combustion excitation, it can effectively attenuate the high-frequency combustion noise.

Diesel knocking or clatter noise is an impulsive noise. It is most significantly observed at low engine speeds/loads. Diesel knocking can be reduced by combustion calibration and/or by reducing the structural transfer function for the combustion noise.

Advanced diesel engines use multiple combustion modes over the operating range in order to meet stringent emissions standards. In addition, diesel aftertreatment systems may require step (sudden) changes in parameters such as air–fuel ratio, EGR rate, and fuel injection timing. Sudden changes in the combustion modes or the operating parameters may cause sudden changes in the noise level and the sound quality perceived by the customer. This poses new NVH challenges. A smooth gradual transition between different operating modes must be achieved by engine calibration and control algorithms in order to eliminate the NVH issues to make the transitions undetectable by the driver.

Combustion noise can be reduced by reducing the cylinder pressure level (in dB) or increasing the structural attenuation (especially in the medium- and high-frequency range). Note that it is the shape, rather than the maximum level, of the cylinder pressure trace that affects the frequency spectrum of the pressure. Basically, when the ignition delay is reduced, the rise rate of the cylinder pressure and the high-frequency contents of the pressure are reduced, thus the combustion noise can be reduced. In general, the following factors can reduce the combustion noise: the combustion chamber design, which can reduce the ignition delay and the rise rate of the cylinder pressure; indirect injection combustion chamber; higher compression pressure and temperature at the TDC; higher compression ratio; increased intake manifold boost pressure; higher cylinder wall temperature; retarded fuel injection timing; higher EGR rate; lower engine speed; lower load or lower fueling rate; increasing structural attenuation of the cylinder bore (e.g., increasing the structural stiffness by using larger stroke-to-bore ratio or more cylinders, using a stiffer liner and block); and suppressing the vibration of the top deck by design changes and raising the powertrain mode frequency (e.g., adding ribs in the cylinder head top deck as reinforcement).

The timing and quantity of pilot fuel injection and the timing and pressure of main injection affect the combustion noise. Modern fuel injection systems offer flexibility for NVH tuning, but this often has to be compromised by emissions and performance requirements during engine combustion development and calibration. If less opportunity could be offered from the combustion calibration for noise reduction, the transfer characteristics of the engine structure and the noise transfer paths in the vehicle would become more important for the combustion noise control in both sound pressure level and sound quality. This requires a system approach to optimize.

Reviews on engine combustion noise are provided by Challen (1975), Anderton (1979), and Wolschendorf *et al.* (1991). The methodology of separating the combustion noise and the mechanical noise was proposed by Brandl *et al*. (2007). Separation of the combustion noise and the piston slap noise was researched by Badaoui *et al*. (2005). The transfer function of engine structural attenuation was researched experimentally by Shu *et al*. (2005a). Combustion noise optimization by structural attenuation and an application of using the combustion noise meter was presented by Wang *et al*. (2007). Torregrosa *et al*. (2007) developed an important innovative approach for combustion noise assessment by using cylinder pressure decomposition, and they showed that the result was more accurate than the classical 'block attenuation curve' approach and the method may be used as a promising alternative to calculate the combustion noise. Regarding the effect of fuels, the impact of diesel cetane number on engine noise was measured by Machado and De Melo (2005). The effects of fuel system design, fuel injection strategies and emissions calibration optimization on engine noise were studied extensively by Russell *et al*. (1990), Kohketsu *et al*. (1994), Tabuchi *et al*. (1995), Badami *et al*. (2002), Mallamo *et al*. (2002), Roy and Tsunemoto (2002), and Mendez and Thirouard (2008). Diesel cold start noise was analyzed by Alt *et al*. (2005). Detailed in-cylinder CFD modeling of combustion noise origins and sensitivities was conducted by Blunsdon *et al*. (1995) and Luckhchoura *et al*. (2008). Diesel combustion noise variations (cycle-to-cycle and cylinder-to-cylinder fluctuations) were studied by Gazon and Blaisot (2006). An overall combustion noise optimization process in powertrain product development was described by Alt *et al*. (2001).

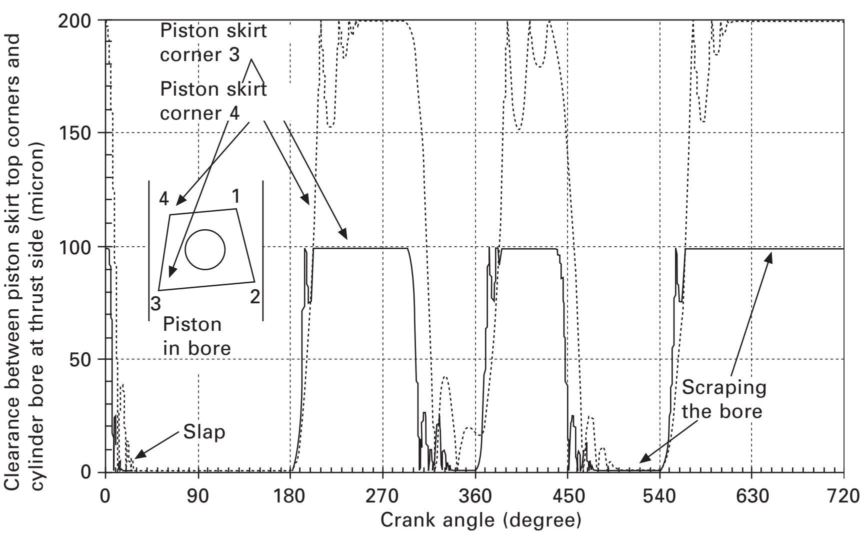
11.5 Piston slap noise and piston-assembly dynamics

11.5.1 Piston slap noise

There are three possible types of piston noise, namely piston rattling noise (i.e., the top land contacts the cylinder bore), piston pin ticking noise (i.e., pin-bearing impact), and piston slap noise (i.e., the skirt contacts the bore). The first two types of noise can be avoided or eliminated by proper design. Piston slap cannot be eliminated because it is caused by the piston secondary motions within the skirt-to-bore clearance that is originated from the crank–slider mechanism by nature. Piston slap is usually the largest contributor to mechanical noise, especially in the diesel engine. For a clearance-free piston, the side thrust is a low-frequency forcing function which is related to engine speed (see Chapter 10). With the clearance in the real engine, the time history of the side force acting on the liner is changed by the additional sharp impact forces when the piston moves within the clearance. These impulsive impact forces are the high-frequency forcing functions driving the cylinder liner and the engine block to vibrate and radiate an impulsive type of noise. The piston slap noise is also transferred from the piston to the connecting rod and the crankshaft and finally to the engine block. Moreover, piston slap causes liner cavitation erosion in heavy-duty diesel engines with the induced excessive liner vibration (Yonezawa and Kanda, 1985). Some design parameters that can reduce piston skirt friction unfortunately affect piston slap adversely.

In engine system design, a good understanding of the following topics is required: (1) the characteristics of piston slap; (2) the modeling approaches; and (3) an optimized overall planning for the piston assembly to balance the trade-offs between fuel economy and noise. Although the piston skirt mass, the skirt flexibility, and the cylinder pressure load are quite different between the diesel engine and the gasoline engine, they share many similar characteristics. Some references mentioned in this section are from the gasoline engine. Introduction to piston slap excitation, noise, and its related design features are provided by Ross and Ungar (1965), Munro and Parker (1975), Whitacre (1990), Slack and Lyon (1982), De Luca and Gerges (1996), Chien (1995), Künzel *et al.* (2001), and Fabi *et al.* (2007).

Piston slap between the skirt and the cylinder bore is caused by the secondary motions (transverse or lateral, and tilting) driven by the alternating piston side thrust within the skirt-to-bore clearance. Not only moving transversely, the piston also tilts around the piston pin and this usually results in the upper (or lower) portion of the skirt slapping against the bore. There are several piston slap events within one engine cycle due to those side thrust reversals (Fig. 11.2). The most significant one is usually the slap right after the firing TDC (0° crank angle). For this impact, the piston moves across the firing



11.2 Simulation of cold piston slap motion without effective lubrication.

TDC to transition from a sliding motion against the anti-thrust side of the piston in the late compression stroke to a slapping event on the thrust side just after the TDC. The gas load acting on the piston may create a moment around the piston pin to turn the piston, thus affecting the slap noise. Efforts should be directed to reduce the most severe piston slap near the firing TDC for piston slap noise control.

Piston slap noise is affected by all the factors involved in this mechanism, i.e.,

- *Piston side thrust force*: lower reciprocating mass, lower engine speed, lower cylinder pressure, and larger ratio of the connecting rod length to the crank radius can reduce the side thrust, thus reducing the piston slap noise (e.g., Oetting *et al.*, 1984).
- *The moment about the piston pin*: lower moment of inertia, proper piston pin offset, the crankshaft offset, the cylinder pressure acting point, the moments from the cylinder pressure force, the gravity force, the lubricant normal forces, the piston ring lateral friction force, and the piston pin friction force may reduce or alter the moment about the piston pin to reduce the piston slap noise. The friction force between the ring and its groove bottom is in the boundary lubrication regime (with a coefficient of friction equal to 0.1–0.2) and has a large influence on piston slap timing. When the ring floats, the resistance to piston slap from the ring disappears and this usually aggravates piston slap. Munro and Parker (1975) reported that a ring lateral friction force with a coefficient of ring–groove friction equal to 0.1 can halve the impact velocity and reduce the kinetic energy by a factor of 4.
- *The allowable travel distance of the piston before hitting the wall*: smaller skirt-to-bore clearance can reduce the piston slap noise. For example, lower liner temperature decreases the clearance by liner contraction. Higher piston temperature may reduce the clearance. Using smaller skirt-to-bore clearance can reduce the severity of piston slap but at the expense of increased viscous shear friction. There is a trade-off between low engine noise and high mechanical efficiency when the piston clearance is designed.
- *The damping force to resist the piston secondary motions*: adequate oil supply on the skirt can reduce piston slap significantly. Lower tension in the piston rings (especially the oil-control ring) can increase the oil film thickness on the cylinder liner. It would allow a larger damping effect of the oil film to cushion the piston slap thus reducing the impact velocity and the noise. The lubricating oil film thickness and lubricant force, which are also affected by the skirt length, the lubricant viscosity, and the surface waviness or roughness, may reduce the piston impact velocity and thus the slap noise. Ryan *et al.* (1994) showed that there

appeared to be an optimum oil viscosity to minimize the piston slap noise; either a higher or lower viscosity increased the slap intensity. Using a longer skirt length as a better guiding and damping surface may reduce the piston slap noise. Increasing the contact area of the piston slap by bore or skirt design changes (e.g., the piston ovality) to enhance oil film damping may reduce the piston slap noise.

- *Piston impact velocity*: all the above factors eventually affect the piston impact lateral velocity, which partly characterizes the severity of the slap.
- *Piston mass*: the piston mass contributes to the impact impulse or kinetic energy. Larger piston mass and higher impact velocity make the piston slap noise louder.
- *The contact area during the slap*: the contact area affects the transient elastic collision process and the impact force. If the piston slap occurs over a larger contact area, the impact energy can be better absorbed to reduce the slap noise. Both vertical shape (skirt profile) and circumferential shape (ovality) affect the contact area and hence the piston slap noise.
- *The stiffness and the damping of the parts in contact*: the stiffness and the damping affect the impact force during the elastic impact process or the coefficient of restitution. If the softer portion of the piston skirt (e.g., the lower portion of the skirt) slaps the bore, the noise will be lower due to larger deformation. The elastic distribution of the skirt stiffness should be uniform. It is important to increase the top land clearance to avoid the contact between the very stiff top land and the bore. The top land is essentially a solid disc of metal of high stiffness. Its contact with the cylinder wall produces a sharp rattling noise. For the reasons of noise and scuffing, the top land should not touch the cylinder wall.
- *The structural sound attenuation characteristics of the cylinder liner/block.*

Among the design factors, offsetting the piston pin relative to the lateral location of the center of gravity of the piston is the most commonly used technique to control the piston slap noise. As explained in the piston slap mechanism above, it is the moment about the piston pin that controls the piston tilting. The moment is affected by both the lateral pin offset and the vertical pin position relative to the center of gravity of the piston. The cylinder gas force largely influences the titling moment. The lateral forces (e.g., the lubricant force) play an equally important role to control the titling moment. Therefore, the effectiveness of the lateral pin offset is dependent on the vertical pin position. With the pin offset toward the thrust side, the gas force will rotate the piston about the piston pin toward the anti-thrust side. This rotation assures the bottom of the skirt reverses to contact the thrust side before the top of the skirt crosses over, thus reducing the force that the top-side reversal would otherwise generate. The bottom portion of the skirt is usually less rigid than the top portion, thus the piston slap can become less noisy. On the other hand, offsetting the pin to the anti-thrust side causes large slap noise because the reverse moment on the piston causes the upper stiff portion of the skirt to contact the bore near the firing TDC. However, offsetting to the anti-thrust side may yield a small (often marginal) decrease in skirt friction. It should be noted that a large piston pin offset may cause excessive piston tilting around the TDC and cause increased blow-by, oil consumption, and friction. Sometimes there are trade-offs between the piston slap noise and the piston tilting. An optimized skirt profile design may relieve this trade-off by altering the lubricant moment acting around the piston pin. As observed from the above-mentioned factors, piston slap control is a complex task, but there are many opportunities to optimize it.

The importance of piston slap noise depends on engine applications. For example, piston slap predominates in marine diesels which have relatively large piston clearances, while it is less prominent in small gasoline engines. The piston slap noise is especially prominent when the engine is cold and the piston clearance is large without effective lubrication (e.g., at cold start). The noise increases with engine speed and peak cylinder pressure. The piston slap noise is most apparent at cold start and idle conditions, as well as at low-speed high-load where other noises are relatively less apparent. Künzel *et al.* (2001) discovered that the piston slap noise was most prominent (audible) at low engine speeds (e.g., 1000–2000 rpm) from low loads to high loads for passenger car diesel engines. Another important scenario is that the piston slap noise occurs prominently after a cold start, when the piston-to-bore clearance is at a maximum but the metal is cold without effective lubrication. For example, Richmond and Parker (1987) found that at mid-speed and low-load (e.g., 1600 rpm, one-third load, accelerating to 30 mph after cold start) the piston slap noise may become the most intrusive. The primary design measure to minimize the piston slap noise is to optimize the piston secondary motions at all operating conditions so that with a change in the skirt–bore contact pattern only a minimum amount of impact energy is transmitted into the engine structure. The two most commonly used techniques to control the piston slap noise are reducing the skirt-to-bore clearance and offsetting the piston pin. Piston skirt profile also plays an important role in noise control.

Cylinder liner or block vibration (or acceleration) has been proven to be a good indicator of piston slap noise. It is found that the liner vibration correlates with the piston impact kinetic energy very well. Kamiya *et al.* (2007) used small thin-film pressure sensors to directly measure the oil film pressure at the piston slap locations to try to understand the excitation force at the slap location. They discovered that a clear correlation exists between the oil film pressure near the top of the skirt (located at the anti-thrust side) and the cylinder liner acceleration measured near the liner top. This verifies that when the piston slap occurs, a large squeeze film reaction happens in hydrodynamic lubrication to produce the oil pressure. It indicates that the piston slap velocity on the lubricated surface can also be used as an indicator for the slap noise.

Piston slap noise and liner/block vibration measurements have been conducted extensively over the past 30 years (DeJong and Parsons, 1982; Furuhama and Hirukawa, 1983; Kaiser *et al*., 1988; Richmond and Parker, 1987; Vora and Ghosh, 1991; Kamp and Spermann, 1995; Ryan *et al*., 1994; Nakada *et al*., 1997; Teraguchi *et al*., 2001). The measurement results facilitate good understanding of the parametric dependency of piston slap and provide support for analytical model development.

11.5.2 Piston-assembly dynamics modeling for piston slap

Piston slap is a very complex dynamic process involving the following key topics:

- multi-body dynamics of the entire piston assembly and the crankshaft
- multi-phase dynamics including the transitions between non-scraping and scraping motions
- rigid-body impact and rebound dynamics
- boundary lubrication friction for surface asperity load carrying
- lubrication and squeeze film effect
- elastic impact and rebound
- deformation
- cylinder liner vibration response.

Numerical modeling can provide valuable tools to reveal the physical mechanisms that are difficult to measure, and to predict the effects of design changes. The ultimate goal of piston slap modeling is to predict the slap noise. However, this can be extremely complex and time-consuming in computation. Even in the experimental work, direct measurement of piston slap noise is difficult because it is not easy to separate out the piston slap noise from other noises. Instead, alternative and easier available parameters such as the liner vibration (acceleration) signal are used to judge the severity of the piston slap. With proper calibration the liner acceleration data can be converted through a transfer function to an estimated piston slap noise. A similar approach can be taken in simulation by evaluating the kinetic energy loss or impact impulse. Piston slap can be accessed by using the piston impact and rebound velocities as key parameters. Experimental work showed that the piston slap noise is almost linearly proportional to the total kinetic energy due to piston impact excitation. A time history and frequency spectrum of the impact force are also important.

Piston slap is essentially driven by the side thrust. After a series of impact and rebound events within a short period of time, when the impact velocity is damped to negligibly small, the piston enters a scraping phase of motion. During the scraping phase the skirt-to-bore contact portion is supported by the surface asperity contact under the boundary friction, and the rest of the skirt may undergo hydrodynamic lubrication. The lubrication dynamics formulation of the piston assembly is introduced in Chapter 10. This section will address the key issues in the detailed level-2 and level-3 NVH modeling of piston slap and the link between piston slap and piston friction. Both the level-2 and level-3 models have instantaneous crank-angle resolution. When the model is complex enough to include structural attenuation or noise calculation with FEA and the BEM, the model is at level-3. The simplified level-1 modeling of piston slap will be introduced in the system-level summary in Section 11.12.

The research on piston slap modeling started in the 1970s. The numerical modeling initially focused on using the multi-body multi-phase dynamics to calculate the piston slap velocity for a rigid-body cylindrical piston, and to calculate the piston rebound velocity based on an assumed coefficient of restitution. The piston-to-bore friction was calculated by using a simple model of Coulomb friction without the hydrodynamic lubrication model. The model predicts the trends reasonably well compared to the experimental measurement data of the piston secondary motions, and is able to reveal many fundamental mechanisms of the piston slap. The model is especially suitable for the cold engine conditions where the skirt lubrication is not effectively established. Typical work in this category includes Wilson and Fawcett (1974), Haddad and Howard (1980), Haddad and Tjan (1995), and Haddad (1995). Detailed derivations of the multi-body dynamics equations of piston assembly including impact dynamics are given by Wilson and Fawcett (1974), Haddad (1995), and Xin (1999). When the piston dynamics model is simplified from multi-body to single-body (i.e., only the piston skirt), errors are introduced in the calculations of the piston side thrust and the tilting moment. Some researchers use this method for simplicity. However, the piston slap timing and the impact velocity cannot be accurately modeled with the simplified method. Piston slap and scraping motion measurements were conducted by Rohrle (1975) and Sander *et al.* (1979).

The energy transferred from the piston slap to excite the block vibration and radiate the noise comes from the piston impact against the bore. The impact is not a one-time event. Instead, it is a consecutive series of impact events with gradually decreasing impact velocity due to the dissipation of impact energy into the structural material. The impact and rebound events are modeled by using the concept of coefficient of restitution, which reflects the attenuation in the rebound velocity relative to the impact velocity. When the impact velocity becomes negligibly small and there is still a normal force to push the skirt against the bore, the piston motion can be regarded as a scraping motion sliding on the bore with the boundary lubrication. It should be noted that the rebound motion is important in modeling because it is possible that after the impact events the piston will rebound to leave the bore without recurring impacts and continue with another non-scraping motion without encountering the boundary lubrication friction. In the scraping phase of the motion, the load carried by asperity contacts is calculated based on the force and the moment balances of the piston. The friction can be calculated by either assuming a friction coefficient or using more sophisticated asperity contact models.

The piston impact is usually assumed to occur over an infinitesimally small time interval with negligible changes in the mechanism's kinematic positions between right before and just after the impact. When the impact occurs, the instantaneously released kinetic energy is calculated using the theory of conservation of momentum from both the lateral and rotational piston velocities. The kinetic energy dissipated at the impact depends on the coefficient of restitution used. The topics of the coefficient of restitution, the impact and the rebound on lubricated and non-lubricated surfaces, and the impact with friction are important subjects in the areas of mechanical impact dynamics and tribology research. The impact dynamics is the bridge between the non-scraping phase of motion and the scraping phase of motion. The assumed coefficient of restitution provides the initial condition of the piston lateral velocity after the rebound in order to carry on the numerical solution of the ordinary differential equations of the dynamic motion. Note that the focus of piston slap analysis is on the series of impact events, while the focus of the boundary lubrication friction analysis is on the scraping motion after the impact events.

The existence of effective lubrication on the piston skirt depends on particular engine operating conditions. Furuhama and Hirukawa (1983) found that in their experiment under the low-temperature cold idle condition, almost all the spaces in the large piston-to-bore clearance were filled with gas, and severe oil starvation happened with little squeeze film action from the oil. Under this circumstance, a model without lubrication is justified.

Although many useful simulation results of the piston impact and the scraping motions can be obtained by ignoring the skirt lubrication model, experimental results showed that adequate oil supply and oil attainment in the skirt area can reduce the piston slap noise. A lubrication model may affect the predicted piston lateral, tilting, and scraping motions and the impact velocity in the following aspects (either more accurate or less):

- A skirt lubrication model may significantly change the moment balance around the piston pin due to the oil film distribution on the skirt. This is particularly true when a non-cylindrical barrel-shape piston skirt is modeled where strong lubricant forces are developed simultaneously at both the thrust and anti-thrust sides and the lubricant moments play a key role.

- A skirt lubrication model provides an oil film damping effect mainly through the dynamic squeeze film effect to cushion/reduce the piston impact velocity.
- The existence of an oil film affects the equivalent stiffness and damping of the contact area during the piston slap process, thus affecting the kinetic energy dissipation, the coefficient of restitution (the piston rebound), and the piston slap noise.
- A skirt lubrication model affects the force and moment balances of the skirt in the scraping motion so that the load carried by the asperity contacts is affected in the calculation. When the piston scrapes on the bore, the hydrodynamic lubricating oil film pressure may support a large portion of the load, and the rest of the load has to be carried by the surface asperity contacts. This may affect the accuracy of the boundary lubrication friction during the scraping.

The piston skirt lubrication model can be as simple as a damper or as complex as a full Reynolds equation. Theoretically, according to the Reynolds equation of the squeeze film lubrication, the damping factor of a squeeze film damper is inversely proportional to the oil film thickness to the power of 3. Haddad and Tjan (1995) provided a simplified oil film thickness model and a semi-empirical equation of the oil pressure as a function of skirt clearance. Gerges *et al.* (2005) used a simplified dynamic model consisting of a spring (representing the air bubbles in the oil film generated during the impact event) and a damping dashpot (representing the oil film) connected in series to simulate the behavior of the oil film mixed with the air bubbles, and they found that the oil aeration could greatly reduce the load carrying capacity of the oil film. Kobayashi *et al.* (2007) used an elastohydrodynamic lubrication model including the aeration effect of the air bubbles inside the oil film. Nakada *et al.* (1997) used a stiffness element and a damper to model the skirt, a viscous damper to model the oil film, and a stiffness element and a damper to model the liner. These elements were connected in series. In order to speed up the calculation, they simplified the Reynolds equation from two-dimensional to one-dimensional by dropping the tilting-motion term and only considered the lateral motion. They obtained a closed-form analytical solution of the oil pressure and the lubricant force. Even such a highly simplified lubrication model could predict piston slap motion quite accurately compared to their measurement data. Wong *et al.* (1994) introduced the two-dimensional Reynolds equation to simulate the lubricated piston slap with elastohydrodynamic and deformation effects for a barrel-shape skirt.

Deformation modeling for piston slap covers two aspects: the component deformation before the impact, and the elastic deformation during the short period of the impact process. Deformation modeling plays an important role in piston slap analysis and may enhance the understanding of the real elastic impact process through modeling the details of the impact force. First, the deformation of the components affects the clearance between the skirt and the liner, thus affecting the location and the severity of the piston slap via the piston impact velocity and the oil film damping force. Secondly, the elastic impact between the piston skirt and the liner is a highly complex transient collision process. During the impact, a part of the impact kinetic energy is transformed into elastic deformation, and the remaining kinetic energy becomes lower for the second impact. As an alternative approach, elastic impact modeling may be more accurate than the right-body impact modeling in simulating the time history of the impact force. The impact force is basically equal to the side thrust plus the impulsive slap force. In the rigid-body impact model, the coefficient of restitution is used to assume the impact and the rebound processes occur within a very short time interval, and a large average impulsive impact force during the process is obtained by using the ratio of the impact impulse to the assumed short time duration. Although the impulsive force is much larger than the side thrust, the sudden application of the side thrust acting on the impact surface after the piston travels across the clearance is also important for the noise, as Lalor *et al.* (1980) pointed out. In this approach, the piston impact velocity is the only key parameter because the impact duration can be uncertain. In an elastic impact model, the impact force is a more important parameter, which can be equal to the side thrust superpositioned by a large oscillating/vibratory impulsive force of the piston–bore impact (i.e., the side thrust is amplified by the vibration force). The oscillating force consists of the elements such as the stiffness, the damping, and the piston impact velocity (all in the lateral direction that is perpendicular to the liner wall) of the piston–film–liner elastic system. The oscillating force governs the piston impact and the rebound behavior, and may provide more details during the impact process than the rigid-body assumption. Compared with the rigid-body impact model, one advantage of the elastic impact model is that the impact forcing function can be more accurately transformed from the time domain to the frequency domain in order to (1) directly analyze the frequency characteristics of the piston slap noise, and (2) correlate its harmonic contents with other frequency signals commonly used in the NVH area (e.g., the liner acceleration). The piston side thrust has low frequencies (e.g., less than 500 Hz), while the addition of an impulsive piston slap force can result in dominant force amplitudes at high frequencies (e.g., 1–3 kHz). The piston impact force and the liner vibration response can be used to estimate the piston slap noise and the cylinder liner cavitation.

In the elastic impact modeling area, Lalor *et al.* (1980) presented a simplified piston slap model using a mass–spring system to represent the piston, the liner, and the engine block. Paranjpe (1998) used a similar simple spring–damper system to model the elastic impact process and the piston–liner contact forcing function in the frequency domain. Ohta *et al.* (1987) provided an important modeling approach to simulate the transient impact force during the piston slap process and the dynamic vibration response of the coupled piston–liner system by using the stiffness and dashpot elements. They calculated the time history and the frequency spectrum of the impact forcing function. Their calculated time history and frequency spectrum of the liner acceleration substantially agreed with measured results. The work by Ohta *et al.* (1987) is one of the most important achievements in the piston slap area.

In the component deformation modeling, early research on the piston slap simulation with the skirt/bore deformation was conducted by Sander *et al.* (1979) without a lubrication model. Kageyama *et al.* (1994) used a simplified contact deformation (stiffness) model for a barrel skirt without lubrication. Patel *et al.* (2007) used a complementarity method to model the dry-contact piston slap for a non-cylindrical piston with the effects of bore distortion and skirt deformation to calculate the kinetic energy loss due to piston slap and to estimate the friction force.

Directly predicting the cylinder liner vibration accurately can be important for a more direct prediction of the piston slap noise. Ohta *et al.* (1987) provided a practical approach to the dynamic modeling of the piston–liner vibration system by using a set of ordinary differential equations, which is suitable for the engine system-level design analysis. Kobayashi *et al.* (2007) presented a highly complicated method to use finite element models of the cylinder block and the piston to simulate the piston slap, the deformations and the liner vibration. Their computation is time-consuming. Although acceptable for the component-level design, this approach is not suitable for an early concept-level engine system design.

In summary, from the perspective of diesel engine system design, a viable approach for a system engineer to analyze and quantify the piston slap noise in the NVH area is to adopt a computational approach based on the ordinary differential equations of the multi-body multi-phase dynamics of the piston assembly and the cylinder liner. With certain fidelity of the impact forcing details and the recurring impact and rebound events, the time history and the frequency spectrum of the piston impact forcing function and the cylinder liner vibration response can be predicted. This requires instantaneous crank-angle-resolution modeling. The calculated piston impact kinetic energy loss summed over one engine cycle is the indicator of the sound pressure level.

The engine noise level (dB) due to piston slap can be estimated by

$$p_{SPL} = \dot{W}_{SWL} - 10\log_{10}(4\pi l^2) \qquad 11.11$$

where p_{SPL} is the sound pressure level, $\dot{W}_{SWL}$ is the sound power level, l is the distance from the engine center to the microphone. The sound power of piston slap is estimated by

$$\dot{W}_s = Z_{sm} Z_{sr} \left(C \cdot \sum_i E_{k,\,slap,\,i} \right) \tag{11.12}$$

where Z_{sm} is the mechanical impedance, Z_{sr} is the radiation impedance, $\sum_i E_{k,\,slap,\,i}$ is the total kinetic energy loss of all the piston impact events within one engine cycle, and C is a model calibration constant to offset the errors in the kinetic energy loss. Z_{sm} and Z_{sr} are measured or simulated known inputs.

11.5.3 The effects of piston design and engine operation on piston slap

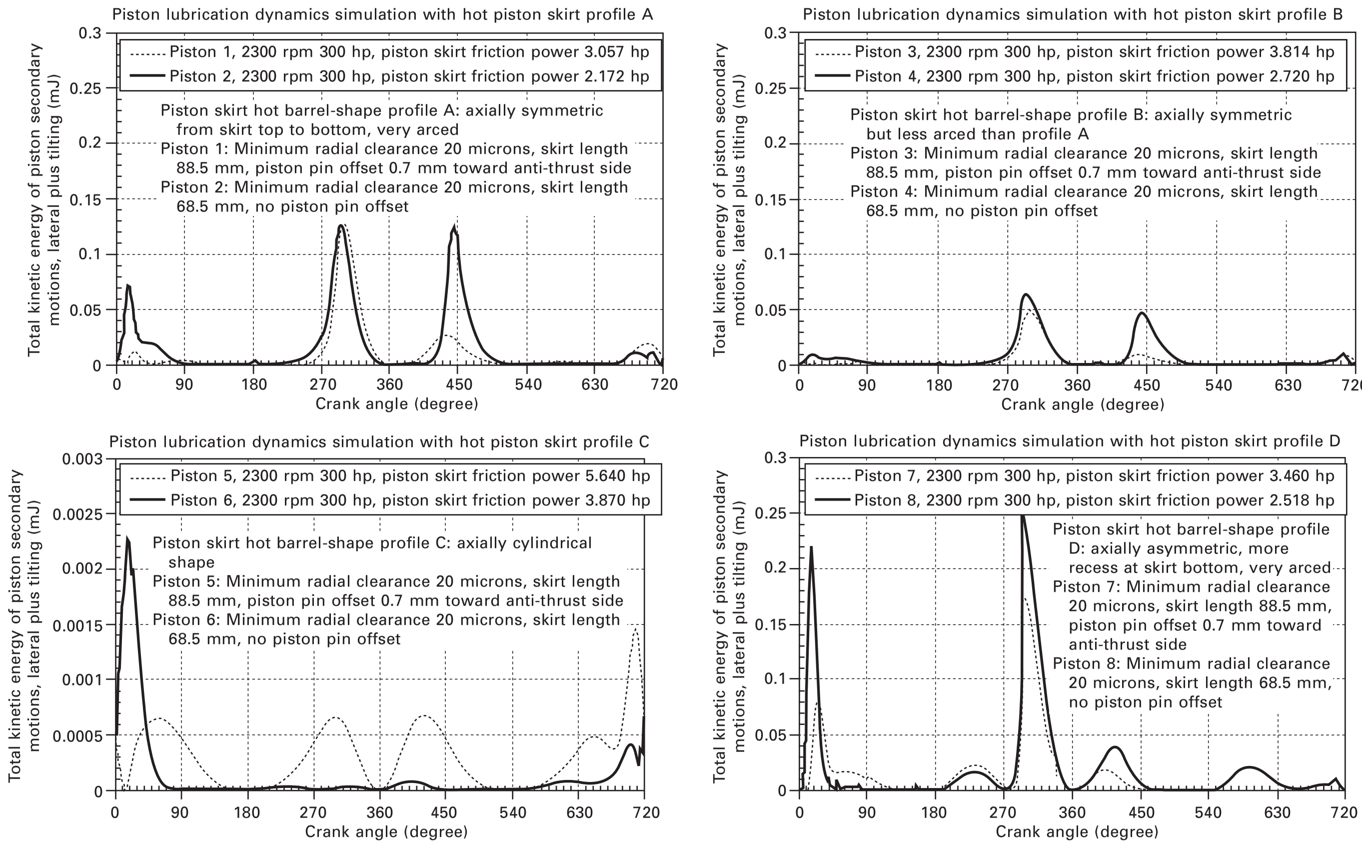
Within an entire engine cycle, piston slap usually becomes the most severe and noisiest during the expansion stroke due to the high in-cylinder firing pressure in the expansion stroke (0–180°). Offsetting the piston pin to the thrust side usually reduces the severity of the piston slap at the more rigid upper portion of the piston skirt during the expansion stroke. But the piston slap at the less stiff lower portion of the skirt in other strokes might increase. Figure 11.2 shows the simulated piston slap and rebound motions of a cold piston with simplified skirt profile without effective skirt lubrication during one engine cycle (at 2300 rpm engine speed). Figure 11.3 illustrates the instantaneous kinetic energy calculated for different piston designs under lubricated conditions. These simulations provide insights into the complex physical processes that are difficult to measure.

11.6 Valvetrain noise

Usually the contribution of the valvetrain noise to the engine sound pressure level is small. However, the valvetrain noise can be a sound quality issue due to its high-frequency nature (e.g., above 3 kHz). Valvetrain NVH is an important topic to consider in overall engine NVH planning and valvetrain design.

The excitation mechanisms of the valvetrain noise include three sources:

- *Cam acceleration excitation*: The opening and closing cam positive acceleration peaks (or more fundamentally, the cam jerks – the derivatives of cam accelerations) excite the valvetrain vibration due to the inertia force at high engine speeds with high frequencies. Any discontinuity or non-smoothness in the cam acceleration will produce a 'hammering' shock loading effect to excite the valvetrain to vibrate violently and produce noise.



11.3 Piston lubrication dynamics simulation on kinetic energy of different piston designs.

- *Valvetrain impact excitation*: These include the impulsive impact loading at the valve opening between the cam and the follower (for mechanical lash adjusters), at the valve closing between the valve and the valve seat, and during the valve event when valvetrain separation and jump/bounce occurs at high speeds. The most common valvetrain impact is valve seating. Often the primary source of the noise in the pushrod valvetrain is the valve seating impact.
- *Frictional vibration*: This occurs at the severe boundary lubrication conditions where the asperity metal contact occurs between the cam and the follower, typically at the two locations near the cam nose where the theoretical lubricant entraining velocity becomes zero. This noise can be prominent at low speeds.

The experimental work of valvetrain noise identification by Hanaoka and Fukumura (1973), Dent and Chen (1989), Kalser *et al*. (1991), and Suh and Lyon (1999) showed evidence of these valvetrain noises. These noises can be reflected in a measured valve acceleration–cam angle diagram (Hanaoka and Fukumura, 1973) or a time–frequency diagram (Suh and Lyon, 1999). There are usually three groups of vibration (highly oscillatory) signals of the valve acceleration located at the cam opening flank, the cam closing flank, and the valve closing. They are high frequency signals, especially at valve seating. Cylinder head vibrates at the similar crank angle timing as the valve acceleration excitation peaks at the valve opening and closing sides. The valvetrain noise generally increases with engine speed.

The valvetrain noise is transmitted via both airborne and structure-borne paths. Anderton and Zheng (1993) found that the airborne valvetrain noise contributed greatly to the total valvetrain noise, particularly at high engine speeds above 2000–3300 rpm. The airborne noise inside the valve cover might be dominated by the valve seating impact. Suh and Lyon (1999) discovered that there were two transmission paths of the vibration through the structure for the excitation due to the valve seating impact: (1) directly transmitted to the cylinder head via the valve seat (being a more dominant path); and (2) transmitted via the valvetrain components and the camshaft to the cylinder head. The vibration was transmitted from the cam to the cylinder head surface for the excitation due to cam acceleration.

Suh and Lyon (1999) found that the contributions to the cylinder head vibration response from the excitation sources of cam acceleration and valve seating impact were comparable in magnitude. They concluded that both sources could be dominant to the noise radiation from the structure surface in the frequency range up to 6 kHz. They found that valve seating excited the vibration in a very wide band frequency spectrum and it was the solely dominant source for the frequency range above 10–20 kHz. Savage and Matterazzo (1993) conducted an experiment on a 3.3 L gasoline engine to investigate the possible influential factors for the valvetrain noise, such as cam jerk level, valve spring load, tappet-to-bore clearance, valve stem clearance and finish, rocker arm bearing clearance, valve overlap, and cylinder head mass and damping. They discovered that lowering the cam jerk level had a considerably larger effect on reducing the valvetrain noise (sound pressure level) than eliminating the valve seating. When the valve seating was eliminated by using a shimmed tappet (i.e., no seating), the valvetrain noise did not reduce until the engine speed was higher than 3000 rpm. They concluded that cam acceleration or jerk level was the most important factor affecting the valvetrain noise over the entire engine speed range, and valve seating velocity control only played a significant role in noise reduction at high speeds (e.g., above 3500 rpm).

The valvetrain noise can be controlled by reducing the excitation source or impeding the noise transmission path. The following design measures provide a few examples:

- Use very smooth cam acceleration profile and reduce cam acceleration level. Match the cam acceleration with the valvetrain stiffness. This controls the valvetrain inertia excitation related to the valve opening velocity, separation, and bounce. Smooth acceleration can often be achieved by advanced cam design techniques and optimization. Reducing cam acceleration or jerk level often results in smaller valve flow area and deterioration in engine volumetric efficiency. Overall valvetrain system optimization can minimize the compromise between the valvetrain noise and the engine fuel economy.
- Use high precision to machine the cam profile to make the cam acceleration and jerk profiles of the manufactured cam become smooth. Experimental data showed that manufacturing precision is very important for valvetrain noise (Hanaoka and Fukumura, 1973).
- Reduce valvetrain vibration by using higher valvetrain stiffness and lower mass.
- Use hydraulic lash adjuster to eliminate valvetrain clearances and the impact at valve opening.
- Reduce valve seating velocity by cam acceleration and vibration control. An appropriate kinematic cam ramp velocity is also important. Lowering the valve seating velocity by lowering the cam ramp velocity has been the traditional design approach to reduce the valvetrain noise. In fact, the valve seating velocity can be reduced more effectively by using smooth cam acceleration, and this may reduce the valvetrain inertia excitation as well.
- Use smaller tappet-to-bore clearance to reduce the impact.
- Reduce the excitation levels of the cam bearing force.
- Reduce or eliminate the valvetrain friction-induced vibration noise by designing to achieve greater oil film thickness.

- Control the noise transmission path. Use special high-frequency-isolation materials to reduce the structure-borne noise. Increase the cylinder head mass, the stiffness, and the material damping to reduce the surface vibration level of the cylinder head.
- Reduce the noise radiation efficiency. Valve cover can provide attenuation of the airborne noise generated under the valve cover. The valve cover needs to provide a low vibration response to structure-borne noise by high damping and stiffness or isolation.

Valvetrain noise analysis can be conducted by a system engineer as follows in a level-2 approach (i.e., with instantaneous crank-angle resolution). At a given engine speed–load mode, the valvetrain vibration cam force and the valve seating velocity are calculated with a valvetrain dynamics model. The time-history of the force and the impulse data are converted to the frequency spectrum of the excitation sources by using the Fast Fourier Transfer. The spectrum of the vibration response of the cylinder head surface can be calculated by multiplying the excitation spectrum with a known (measured) transfer function characterizing the transmission path. The total response spectral sum is the superposition of the two excitation mechanisms, the cam force, and the valve seating. Finally the frequency spectrum of the radiated noise can be calculated after applying the A-weighting to obtain the A-weighted spectrum. A practical approach of the valvetrain noise calculation was provided by Kalser *et al*. (1991).

11.7 Geartrain noise

The gear rattle noise is a primary NVH issue for automotive drivertains such as in the transmissions. It is also important in the timing gears and the gears used to transmit torque to drive the accessories such as the fuel pump in the diesel engine. The engine transmits non-uniform torques from the cranktrain to the drivetrain or the geartrain and that causes the gear rattle noise.

Clearances are inevitably required between the idle surfaces of the engaging gear teeth in geared/toothed connections (e.g., automotive gear boxes, manual transmissions, clutches, gear couplings) in order to allow for thermal expansion and manufacturing tolerances. When the gears are lightly loaded with strongly oscillating torque and/or rotate at low speeds, there is a high possibility of meshing teeth separation and the resulting vibration impact between the teeth surfaces. Such a vibratory energy excitation of the gear teeth within the working clearance generates the annoying rattle noise. The gear rattle noises in the powertrain and the drivetrain are related to torsional vibrations. Another annoying gear noise is the whine noise which is excited by the transmission error at the gear mesh due to manufacturing errors and tooth deflection under load. The gear rattle noise is affected by the transmission input speed fluctuation, kinematic transmission error, the backlash in the tooth meshes, and the friction force acting on the gears.

The ever increasing cylinder pressure and fuel injection pressure dramatically increase the torsional excitation of the geartrain in the diesel engine. The gear impact noise may become a significant noise source at many operating conditions especially at full load. The geartrain noise level is a strong function of the size of the geartrain, the number of gear meshes, the location of the geartrain (front or rear), and the magnitude of torsional inputs from the crankshaft and the fuel system (Zhao and Reinhart, 1999). Optimizing gear profile and minimizing the transmission error by reducing manufacturing and assembly errors may also reduce the noise.

Engine geartrain noise was investigated by Spessert and Ponsa (1990) and Zhao and Reinhart (1999). Geartrain dynamics and gear motion/rattle modeling are highly complex. Good references are provided by Croker *et al*. (1995), Padmanabhan *et al*. (1995), Meisner and Campbell (1995), Lahey *et al*. (2001), and Wang *et al*. (2001).

11.8 Cranktrain and engine block noises

The crankshaft torsional vibration, the thin sections in the engine block and the covers (e.g., the valve cover, the oil pan) bolted to the crankcase and the cylinder block are important sources of NVH. Sophisticated commercial software is available (e.g., Ricardo's ENGDYN) to predict the time-domain response of the coupled cranktrain and engine block system with nonlinear oil film lubrication models (e.g., Offner *et al*., 2004). The radiated noise can be calculated from the surface normal velocity using the Rayleigh equation. The radiated sound power and the noise radiation efficiency can be calculated for each vibrating surface that is approximated as a flat plate. The sound intensity can also be calculated across the surface. Critical modes of vibration of the powertrain can be identified. Cranktrain dynamics and its interaction with the cylinder block can be reliably studied in simulation.

Details of crankshaft, crankcase, and engine block vibration and noise are provided by Russell (1972), Ochiai and Yokota (1982), and Maetani *et al*. (1993). Engine mounts were reviewed by Shangguan (2009). Cranktrain dynamics and engine balance have been extensively covered in the literature (e.g., Thomson, 1978; Lee *et al*., 2000).

11.9 Auxiliary noise

The noise contribution from the auxiliary devices is significant in modern low-noise diesel engines. The oil pump (tooth pump) noise is excited mainly by oil pressure peaks. The generator noise comes mainly from the aerodynamic noise of the fan and the rotor, and the mechanical noise of its bearings. Fuel system torsional vibration due to high cyclic torque affects the geartrain noise. Some fuel systems may have very low cyclic torque, thus contributing little to the gear noise. Haller *et al*. (1993) reviewed the noise excitation by auxiliary devices in heavy-duty diesel engines including oil pumps, generators, reciprocating air compressors, and hydraulic oil pumps. The hydraulic fluid borne noise was elaborated by Skaistis (1975).

11.10 Aerodynamic noises

11.10.1 Intake noise

The excitation source of the engine intake air noise is the pressure fluctuation produced by the reciprocating piston motion and the valve timing. The intake noise is produced by the air pressure pulsation in the inlet duct and the high air flow velocity as the turbulent air flow passes through the intake valve opening area. The noise due to the mechanical vibration of the intake pipes can also be associated as a part of the intake noise, or called secondary airborne noise.

The intake noise is usually dominated by low-frequency contents although the intake noise associated with the varying valve flow area contains a wide range of frequencies. The strongest low-frequency content in the intake noise is directly related to engine speed and given by the following:

$$f_{q,intake} = i\frac{N_E n_E}{60\, n_s} \tag{11.13}$$

where i is the harmonic number, N_E is the engine crankshaft speed in rpm, n_E is the number of cylinders, $n_s = 2$ for four-stroke engines and $n_s = 1$ for two-stroke engines.

The intake noise increases as the engine speed increases mainly because the inlet air flow rate, the air pressure pulsation amplitude and frequency increase with engine speed. Intake system design has a major influence on the intake noise through gas pressure wave dynamics. Engine displacement, intake valve size, valve opening timing, intake runner and manifold all affect the intake noise. Moreover, the shape of the air inlet opening (e.g., circular, or flattened in the aspect ratio) has an effect on the intake air noise, and the noise may decrease as the circular cross-sectional area of the opening is decreased or throttled. Excessively flattened inlet design may produce extra whistling noise.

There are generally three methods to control the intake noise: (1) using silencers; (2) reducing the pulsation magnitude of the intake air pressure; and (3) reducing the turbulence level of the intake valve air flow. Silencers are the most effective design measures to reduce the intake and exhaust noises. The intake air noise can be reduced by two types of silencers, the absorptive (dissipative) type and the reactive type.

The absorptive silencers use a duct or parallel ducts lined with sound-absorptive materials to dissipate the acoustic energy to heat. They are effective in the medium- and high-frequency ranges. They tend to have poor attenuation at low frequencies and quick erosion in the absorptive lining due to the high mean flow velocities. The use of adsorptive silencers in internal combustion engines is generally limited (Onorati, 1999).

The reactive silencers include expansion chambers (e.g., large-volume air cleaners) and cancellation resonators (e.g., the Helmholtz or quarter wavelength type). The resonator is tuned to certain frequency or frequencies with a specific design of volume, hole opening area, effective length and internal structure (e.g., number of perforates or internal orifices along the passage). When the induction system produces an air pressure wave the resonator reflects a waveform with the same frequency and amplitude but opposite phase thereby cancelling the induction noise at that frequency. The resonator is only effective within the narrow operating frequency range it is tuned for, usually in the low- and medium-frequency ranges. Wide band resonators can be used to reduce specific noises from the induction system or the turbocharger.

The silencer performance is characterized by insertion loss and transmission loss. The insertion loss refers to the difference in the sound pressure level at a certain location before and after using the silencer. The transmission loss refers to the difference in the sound power level between the inlet and outlet of the silencer (Tao and Seybert, 2003).

It should be noted that all the silencing techniques usually conflict with the tight packaging requirement in the engine compartment. Silencers should be designed to suit the frequency spectrum of the noise source with minimal pressure losses and good packaging size. Number of stages, volume, pressure drop, and expansion ratio are the critical design parameters in silencer selection. The intake silencer is usually designed as a part of the air cleaner (filter) and it is a standard component in almost all engines.

A review of piston engine intake and exhaust acoustic design is provided by Davies (1996). Gas wave dynamics simulation of engine intake acoustics was conducted by Silvestri *et al.* (1994).

11.10.2 Exhaust noise

The excitation source of the engine exhaust noise is the pressure pulses in the exhaust duct that are generated by the periodic charging and discharging processes of the engine. The high-velocity exhaust flow in the ducts generates significant turbulence and vortex shedding and the associated noises. The noise due to the mechanical vibration of the exhaust pipe can also be associated as a part of the exhaust noise, called shell noise. Like the intake noise, the exhaust noise is usually dominated by low-frequency contents although the exhaust noise associated with the varying valve flow area contains a wide range of frequencies. The frequency of the strongest exhaust noise contents in the low-frequency range can be estimated in a similar way to the intake noise by using equation 11.13.

Exhaust noise and its resonant amplification are largely affected by the flow resistance in the aftertreatment system. Sharp resonances can be reduced by flow resistance. The flow resistance is caused by the viscous force acting between the exhaust gas and the tube walls or among the turbulent gas flows of vortices. The flow resistance (i.e., exhaust restriction) results in a loss in the static pressure and negatively affects engine performance.

The exhaust noise level increases as the engine speed or load increases mainly because the exhaust gas flow rate, the pressure pulsation amplitude and frequency increase with engine speed or load. Exhaust system design has a major influence on the exhaust noise through gas pressure wave dynamics. Engine displacement, exhaust valve size, valve opening timing, in-cylinder pressure level at exhaust valve opening, and exhaust runner and manifold designs all affect the exhaust noise. Long and thin exhaust tailpipe may help reduce the exhaust noise. However, excessively thin tailpipe increases the exhaust restriction and thus hurts the engine breathing performance. Overall automotive exhaust acoustic system design is discussed by Pang *et al*. (2003).

Similar to the intake noise control, there are generally three methods to control exhaust noise: (1) using silencers; (2) reducing the pulsation magnitude of the exhaust gas pressure; and (3) reducing the turbulence level of the exhaust valve gas flow. Exhaust noise control is reviewed by Munjal (1981).

Exhaust noise silencing can be achieved by using reactive (i.e., reflective, sound wave reflection, or cancellation) type mufflers, reactive-dissipative perforate type mufflers, or absorptive (i.e., dissipative) silencers, or a combination of them. The constriction type of silencers relies on orifice restrictions and causes excessively high flow restriction; therefore it is not desirable. The absorptive and cancellation types of silencers are usually the preferred choice. The absorptive silencer reduces the exhaust noise by dissipating the sound energy in porous medium, while the cancellation type silencer cancels the sound pressure waves by out-of-phase waves to reduce the noise.

The design of the silencer is usually a compromise between acoustic attenuation and exhaust (or intake) flow restriction (pressure drop). The goal is to achieve good engine breathing performance at acceptable noise levels. The overall radiated tailpipe noise is the sum of both the 'gas pulse' noise and the aerodynamic 'gas flow' noise (Onorati, 1999). Silencer design needs to avoid the generation of excessive gas flow noise inside the silencer. The mechanism and the theory of the exhaust pulse noise and the flow noise were explained by Kunz (1999) and Pang *et al*. (2005). The pulse noise can be simulated by using nonlinear gas wave dynamics models in engine cycle simulations. The flow noise usually has to be estimated based on semi-empirical formulae. Analytical silencer (muffler) design was presented by Yadav *et al*. (2007).

Engine cycle and gas wave dynamics simulations play a key role in exhaust noise control. Engine exhaust system acoustics modeling was reviewed by Jones (1984) and Onorati (1999). The basic theory of exhaust acoustic modeling was presented by Ghafouri and Ricci (1993). Engine exhaust noise modeling was conducted by Onorati (1995), Isshiki *et al*. (1996), Morel *et al*. (1999), and Siano *et al*. (2005). The details of silencer simulation with the one-dimensional nonlinear gas wave dynamics models were provided by Onorati (1999). The simulation of the unsteady flows in the coupled engine–muffler system can predict the sound pressure level spectra of the radiated tailpipe noise and the influence of muffler design on engine performance. The nonlinear gas dynamics model is more advanced and reliable than the over-simplified linear acoustic model which is based on the hypothesis of small pressure perturbations in the pipes. Onorati (1999) stated in his review article that the silencer simulation technique has reached a stage where it is capable of giving good predictions of attenuation curves compared with the measured data over a broad frequency band for a wide range of muffling pipe systems.

Diesel particulate filter (DPF) has been widely used in modern diesel engines to control particulate matter. Although they can be separated, the DPF and the muffler are often combined into one unit in order to save cost and packaging space. In this case, the design of the combined unit needs to balance the requirements on particulate filtration and exhaust noise attenuation. Understanding the noise attenuation characteristics of the DPF as a muffler is a new challenging and interesting area to the diesel engine industry. Integrated noise and particulates control is the future direction in acoustic modeling and design. DPF acoustics was researched by Katari *et al*. (2004), Allam and Åbom (2005, 2006), and Feng *et al*. (2008).

In summary, exhaust noise prediction needs to be conducted in the following three areas:

- nonlinear simulation of gas wave dynamics to predict the gas pulse noise
- using and developing empirical formula to predict the gas flow noise
- DPF acoustic simulation.

There are two approaches to simulate the exhaust noise: (1) using a detailed silencer or DPF model (e.g., obtained from the supplier) in the nonlinear gas dynamics model; and (2) using a simplified flow restriction model and available attenuation spectra curves of the silencer (or the DPF) or empirical correlations to estimate the tailpipe noise.

11.10.3 Turbocharger noise

Turbocharging changes the intake/exhaust noises and the engine NVH characteristics. The compressor and the turbine reduce the sound propagation through the intake and the exhaust systems and the intake/exhaust noise is generally reduced because the compressor housing behaves like a small reactive silencer element and the turbine usually dissipates the exhaust flow pulsations. The turbine wheel damps the exhaust pulsations so that the exhaust noise at the turbine outlet is much lower than the exhaust noise in naturally aspirated engines.

However, the turbocharger may generate sound quality issues such as the whooshing noise (Evans and Ward, 2006) due to the high flow and the whining noise related to the number of blades in the compressor and the turbine. The turbocharger whining noise in different frequency ranges can be reduced by proper rotor balance specification or by increasing the gap between the wheel and the flap of the turbine (Lu and Jen, 2007).

The heat shield of the turbocharger is another source of noise. Lu and Jen (2007) showed that the turbocharger had 2–4 dB noise reduction after the heat shield was removed. They pointed out that the heat shield required for the turbocharger has a negative effect on noise due to its large radiating surface and weak rigidity. Therefore, the shape and the stiffness of the heat shield should be optimized to reduce the radiation noise. Lu and Jen (2007) found that the noise refinement by a double-layer heat shield design was around 1–3 dB compared with a single-layer shield.

Other turbocharger noises include the following: (1) some resonances in certain frequency range at the exhaust tailpipe; (2) some local resonances with the turbocharger supporting bracket when the component stiffness is not high enough; and (3) some high-frequency noise associated with the high rotating speed, and certain sub-harmonic and imbalance vibrations caused by rotor dynamics and the blade vibration which is induced by the exhaust gas pulsations.

Silencers can be used on compressors to alleviate their turbulent intake flow noise problems. The silencers are usually hole-chamber resonators, interference resonators (e.g., Herschel–Quincke tube) or absorptive silencers.

Turbocharger noise is reviewed by Rämmal and Åbom (2007). Turbocharger noise control was researched by Evans and Ward (2005), Trochon (2001), and Lee *et al*. (2009a). Turbocharger structural vibration is analyzed by Inagaki *et al*. (1993). The acoustics of charge air cooler, which is related to turbocharging, is analyzed by Knutsson and Åbom (2007).

11.10.4 Cooling fan noise

The cooling fan is used for engine coolant temperature and heat rejection controls. The cooling fan noise is a significant part of vehicle NVH and important for automotive cabin quietness and comfort. As engine power or heat rejection increases in low-NO_x engines, it is increasingly challenging to design a cooling fan that has a small size, low noise, low power, and low cost. The cooling fan operates in a wide range of cooling air mass flows, and its peak efficiency and minimum noise operating points do not necessarily coincide. Optimizing both aerodynamic performance and noise characteristics is important in selecting the best fan design for a given application.

The cooling fan noise is related to fan size (specific diameter), fan speed (specific speed), blade design, and airway structures. Design measures such as decreasing tip clearance and optimizing pitch angle may increase the fan efficiency and reduce the fan noise. Fan blade design can also minimize the fan vibration and the fan noise. For example, the fan blade passage noise is one important element of the total fan noise, and is caused by the non-uniformity of the upstream and downstream flows of the fan rotor. The discrete noise at the blade passing frequency can cause serious noise problems. Appropriate blade spacing design may reduce the noise. Optimizing the airway structures (flow passages) can improve the uniformity of the inflow air distribution and reduce the lift fluctuations of the fan blades thus reducing the noise level of the fan. Moreover, psychoacoustic characteristics (i.e., subjective assessment of the sound comfort and annoyance) of the cooling fan are also important, in addition to the requirements of sound pressure level. The engine cooling fan noise and performance was elaborated by Mellin (1980).

11.11 Engine brake noise

Engine brake performance and design are introduced in Chapter 6. The compression-release engine brake has unique noise characteristics. Details on engine brake noise are discussed in Section 6.4.8. The noise of the compression-release engine brake is an exhaust-flow induced noise related to the gas wave dynamics in the exhaust manifold and the tailpipe. The noise is caused by a sudden blow-down event of the in-cylinder charge near the braking TDC. The noise can be controlled by using better braking mechanisms that control the engine braking valve flow characteristics, or it can be controlled by advanced mufflers. Engine cycle simulation coupled with an acoustic simulation model of the exhaust system can be used to analyze the engine brake noise. More information on engine brake noise can be found in Reinhart and Wahl (1997) and Wahl and Reinhart (1997).

11.12 Diesel engine system design models of noise, vibration, and harshness (NVH)

11.12.1 Level-1 system noise model

The level-1 system noise model used in diesel engine system design is a non-instantaneous (non-crank-angle-resolution) semi-empirical model. The coefficients or constants in such a model are limited to certain particular engines from which the formula is derived, and it cannot be broadly applied to pursue accurate absolute values. This type of model is useful to understand the basic major parametric dependency of the noise. It can be used to roughly estimate the relative effect or trend.

The overall engine noise level can be expressed as:

$$p_{SPL,E} = \sum_i p_{SPL,i} \qquad 11.14$$

where $p_{SPL,E}$ is the overall engine sound pressure level at one meter from the engine surface. $p_{SPL,i}$ represents the contribution from combustion, piston slap, valvetrain, gears, fuel injection and accessories.

Jenkins (1975) pointed out that the engine noises induced by different excitation sources can be expressed by the following functions:

$$\begin{cases} p_{SPL,Combustion} = f(N_E^{n_c}, B_E, F_E) \\ p_{SPL,PistonSlap} = f(N_E^2, B_E, p, f_P) \\ p_{SPL,Valvetrain} = f(N_E^{5.8}, f_{VT}, c_{VT}) \\ p_{SPL,FuelInjection} = f(N_E^{4.3}, \dot{m}_{fuel}, v_{inj}) \end{cases} \qquad 11.15$$

where N_E is the engine speed, B_E is the cylinder bore diameter, F_E is the engine load, n_c is a function of the combustion system, p is the cylinder pressure, f_P is a piston design factor, f_{VT} is a valvetrain design factor, c_{VT} is the valvetrain lash, $\dot{m}_{fuel}$ is the fuel rate, and v_{inj} is the velocity of the injector at the start of the injection.

The most basic form of the overall engine noise model was deduced based on the noise measurements on a large number of engines for combustion-induced noise (Anderton *et al.*, 1970; Challen, 1975; Corcione *et al.*, 2003) as follows:

$$p_{SPL,E} = C_1 \log_{10} N_E + C_2 \log_{10} B_E + C_3 \qquad 11.16$$

where $p_{SPL,E}$ is the overall engine sound pressure level in dB(A) at one meter from the engine surface, N_E is the engine speed (rpm), B_E is the cylinder bore diameter, and C_1, C_2 and C_3 are empirical constants obtained by fitting the measurement data. C_1 is related to the cylinder pressure level spectra (i.e., the combustion system), and C_3 reflects the engine structural characteristics. Note that the two most influential parameters reflected in equation 11.16 are engine speed and cylinder bore size. The latter actually reflects the combustion chamber size. The load dependency of the engine noise was found to be very weak. For four-stroke turbocharged diesel engines, $C_1 = 40$, $C_2 = 50$, $C_3 = -66.5$ if B_E is in inches and $C_3 = -136.7$ if B_E is in mm (Anderton *et al.*, 1970). The implication of equation 11.16 is that it is important to choose a low rated speed and a small bore size (or using more cylinders for a given engine displacement) in engine system design for low noise characteristics. Note that equation 11.16 was obtained from diesel engines having low BMEP and low fuel injection pressure in the old era.

For a given cylinder bore size and engine speed, there is a large variation in the engine noise due to the difference in the combustion system. In order to analyze the effect of the cylinder pressure rise on the combustion noise and characterize different combustion systems, Hawksley and Anderton (1978) proposed

$$n_c = 4.3 - 0.21\, f_{rpr} \qquad 11.17$$

where n_c is a combustion index, and f_{rpr} is the rate of pressure rise in bar per degree crank angle. f_{rpr} varies greatly with engine load. Based on the following relationship (Anderton *et al.*, 1970) about noise intensity,

$$I_s \propto N_E^{n_c} B_E^5 \qquad 11.18$$

they proposed that the engine noise level for turbocharged diesel engines followed

$$p_{SPL,E} = 30\log_{10} N_E + 50\log_{10} B_E + (2.1 f_{rpr} - 13)\log_{10}\frac{5455}{N_E} - 103 \qquad 11.19$$

Note that this formula includes the effect of engine load via the parameter f_{pr}.

Anderton (1979) summarized the relationship between the overall engine noise level and five key parameters (i.e., engine speed, bore diameter, cylinder pressure spectrum at a given frequency, a combustion index, and a shape factor) based on the noise measurements on a large number of engines as:

$$p_{SPL,E} = 10\, n_c \cdot \log_{10}\frac{N_E}{1000} + 50\log_{10} B_E + C_{pr(1.0)} - C_2 \qquad 11.20$$

where $p_{SPL,E}$ is the overall engine noise level in dB(A), N_E is the engine speed (rpm), B_E is the cylinder bore diameter (mm), $C_{pr(1.0)}$ is the value (in dB) of the simplified and reduced cylinder pressure spectrum at 1.0 Hz/rev/min, n_c is the combustion index, which is equal to one tenth of the slope of the simplified and reduced cylinder pressure spectrum in dB/decade, and C_2 is an A-weighted shape factor as a function of the combustion index n_c. The average constants in equation 11.20 of different types of engines were given by Anderton (1979). For example, $C_{pr(1.0)}$ = 144 dB and n_c = 5.1 for turbocharged two- and four-stroke diesel engines.

Tung and Crocker (1982) used a simpler formula in the following form:

$$p_{SPL,E} = C_1 \log_{10} N_E + C_2 \log_{10} \dot{W}_E + C_3 \tag{11.21}$$

where $\dot{W}_E$ is the engine brake power.

The research focus on mechanical noise has been concentrated on the piston slap noise because it is the most important mechanical noise in diesel engines. Usami *et al.* (1975) provided a formula by using measured data of an indirect injection diesel engine at 3700 rpm as follows:

$$\begin{cases} E_{k,\,slap} \propto c_P^{1.3} \\ p_{SPL,E} = 10 \log_{10}(c_{P1}^{1.3} + 1.74\, c_{P2}^{1.3}) - 80.4 \end{cases} \tag{11.22}$$

where $E_{k,slap}$ is the impact energy of piston slap, c_{P1} and c_{P2} are the piston-to-bore clearances at the skirt top and bottom, respectively. Note that this equation is independent of engine speed and only includes basically one parameter, the piston-to-bore clearance c_P.

Lalor *et al.* (1980) gave the noise intensity due to the kinetic energy of piston slap as a function of engine speed N_E and piston-to-bore clearance c_P,

$$I_s \propto N_E^{0.667} c_P^{1.333} \tag{11.23}$$

Priede (1982) developed a formula for the sound intensity of piston slap as

$$I_s \propto \left(\frac{\mathrm{d}F_P}{\mathrm{d}t} m_P \right)^{1/3} c_P^{1.333} \cdot K_{s,\,liner} \cdot B_E \tag{11.24}$$

where I_s is the sound intensity, c_P is the piston-to-bore clearance, m_P is the piston mass, $K_{s,liner}$ is the cylinder liner stiffness, $\mathrm{d}F_P/\mathrm{d}t$ is the variation rate of the piston lateral force.

Other basic design calculation guidelines for low-noise diesel engines were provided by Jenkins *et al.* (1973).

Stout *et al.* (2005) proposed the following empirical formula for three V6 gasoline engines:

$$\begin{cases} p_{SPL,E} = 50\log_{10} N_E + 40\log_{10} S_E + 30\log_{10} B_E - 223.5 \\ p_{SPL,mech} = 0.9 \times 50\log_{10} N_E + 1.6 \times 40\log_{10} S_E \\ \qquad -0.7 \times 30\log_{10} B_E - 158 \end{cases} \quad 11.25$$

where $p_{SPL,E}$ is the overall engine noise level, $p_{SPL,mech}$ is the mechanical noise level, and S_E is the engine stroke.

Zhao and Reinhart (1999) proposed the following empirical formulae including the influence of the geartrain noise based on their measurements on 14 heavy-duty diesel engines of 1994 EPA or Euro II emissions (from 3 liters to 16 liters engine displacement):

$$\begin{cases} p_{SPL,E,Rated} = 93.9 + 2.6(n_G - 1) + 4.8 i_F \\ p_{SPL,E,PeakTorque} = 91.5 + 2.5(n_G - 1) + i_F \\ p_{SPL,E,LowIdle} = 85.9 + 0.8(n_G - 1) \\ p_{SPL,E,HighIdle} = 99 + 1.2(n_G - 1) \end{cases} \quad 11.26$$

where $p_{SPL,E,Rated}$ is the overall engine noise level at rated power measured according to SAE J1074 at 1 m from the engine surface, n_G is the number of gear meshes between the crankshaft and the fuel system, and $i_F = 1$ for front geartrains, while $i_F = 0$ for rear geartrains. This study suggests that the engine noise level is more sensitive to the number of gear meshes at full load than at part load.

It is observed that all the above equations actually provide limited values to diesel engine system design due to their limited range of parametric dependency and simplified relationships. More physics-based models need to be developed for engine system design, and they are the level-2 models.

11.12.2 Level-2 system NVH model

The level-2 NVH model used in diesel engine system design possesses the following features on an instantaneous basis with crank-angle resolution:

- Model the excitation source at various complexity levels.
- Use a 'virtual' combustion noise meter to estimate the combustion noise by processing the cylinder pressure data generated from cycle simulations.
- Use a piston-assembly-liner dynamics model to estimate the piston slap noise.
- Use a valvetrain dynamics model to estimate the valvetrain vibration and noise.
- Use a gas wave dynamics cycle simulation model coupled with an

acoustic model of the silencer or the DPF to predict the intake and exhaust noises.
- The model can be real-time capable, if simplified, for advanced active noise control in engine controls and for fault diagnostics.

Note that all the modeling tools involved in these topics are the tools used by a system design engineer in his/her daily work. The system engineer has the expertise and a clear incentive to conduct this NVH design/analysis work at the system level.

Combustion noise meters (e.g., the Lucas CAV and AVL combustion noise meters) have been widely used in the NVH and engine calibration areas to conveniently measure the effects of combustion (or essentially the cylinder pressure changes) on engine noise. The combustion noise meter was developed based on the following principle about the relationship between the frequency spectrum of the cylinder pressure and the measured frequency spectrum of the total engine noise level at one meter from the engine surface: by subtracting the spectra of the in-cylinder pressure from the engine noise spectra, the structural attenuation can be determined. Such a structural attenuation spectrum can be obtained from many engines through measurement and then an average attenuation spectrum can be calculated and entered into the combustion noise meter. Then, the meter can be used to predict the combustion noise once the cylinder pressure data are known without measuring the overall engine noise level. The noise change due to the cylinder pressure change can be regarded as the contribution of the combustion noise change. This technique is acceptable for engines having similar spectra of structural attenuation.

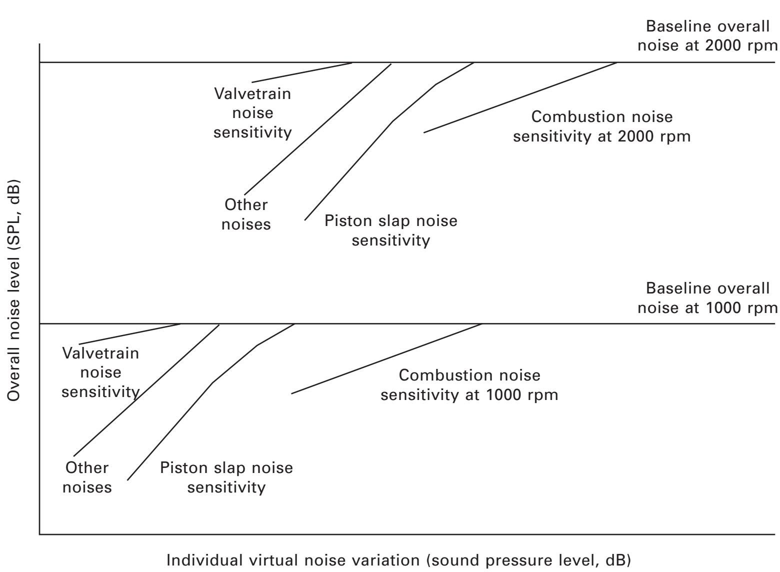
The combustion noise is equal to the cylinder pressure minus the structural attenuation (calculated in frequency spectra in the unit of dB). The combustion chamber pressure signals are commonly measured in engine development in crank-angle resolution. The combustion meter uses the crank-angle-resolution signals from a cylinder pressure transducer to calculate the combustion noise level, which would be the noise level measured 1 m from the engine if only the 'combustion related' noise were included. A combustion meter uses a frequency analysis (Fast Fourier Transfer) of the cylinder pressure signals to generate frequency spectra (e.g., one-third octave). The total noise level is the logarithmic mean value of the spectrum. Then, it subtracts the engine block structural attenuation function via an attenuation filter (usually a mean free-field response filter representing an average engine mass/structure response based on the experiments conducted on many engines) from the frequency spectrum of the cylinder pressure to obtain the combustion noise spectrum. The attenuation filter may also include a function to filter out the combustion chamber resonance with selectable low-pass filters (e.g., 10 kHz low-pass). The combustion noise signals are then processed by an A-weighting filter to produce the noise sensitive to human hearing in the unit of dB(A). Finally, a single root-mean-square value of the combustion noise is produced. The combustion noise can be measured in this way by using cylinder pressure instead of using a sound meter. Moreover, the mechanical noises are separated out from the measured combustion noise. The meter can also be used for recording the combustion noise during transient conditions.

Cylinder pressure derivative with respect to time is an output of the combustion noise meter. Usually, there is a direct correlation between the cylinder pressure derivative and the combustion noise. The correlation between the peak cylinder pressure and the combustion noise is much less clear although increasing the cylinder pressure sometimes leads to increased combustion noise for a given turbocharged engine. Note that peak cylinder pressure affects the mechanical noises such as the piston slap noise and the gear rattle noise. Details of the combustion noise meters are provided by Russell (1984), Russell and Young (1985), Reinhart (1987), and Wang *et al.* (2007). More accurate accounting for the structural attenuation characteristics is discussed by Lee M *et al.* (2009). The limitations of the classical 'block attenuation curve' method and a more advanced alternative approach were presented by Torregrosa *et al.* (2007).

A system design engineer uses engine cycle simulations to analyze the cylinder pressure signals in a virtual world. Although the cylinder pressure data obtained from the engine cycle simulation may usually contain less details related to the pressure derivatives (e.g., $dp/d\phi$ or dp/dt) than the real test data and cannot reflect the design changes of the combustion system, the simulation data are useful to correctly reflect certain operating parameter changes such as the effects of engine speed, fueling rate, air–fuel ratio, EGR rate, fuel injection timing, ignition delay, and cetane number. The combustion noise is sensitive to these parameters (Reinhart, 1987) and they are important system design parameters. Therefore, a virtual combustion noise meter can be developed to estimate the combustion noise changes in the entire speed–load domain and during acceleration transients for major system design parameters. The structural attenuation function can be measured and developed in the model for a particular engine. Note that the structural attenuation of a modern diesel engine can be overall much higher than that used in a combustion noise meter which is based on the old engine structure data. Even if the structural attenuation response is not very accurate, this approach is still effective to predict the relative difference in the combustion noise instead of the absolute values. With further development of advanced zero- or one-dimensional combustion models the prediction capability for combustion and cylinder pressure details in the cycle simulation software will continue to improve. Moving to the virtual world for combustion noise analysis is the right direction, and diesel engine system design can play a key role in this process by effectively integrating combustion noise analysis into the system design process. Related references about combustion noise prediction can be found in Corcione *et al.* (2003) and Scarpati *et al.* (2007). Note that a cylinder pressure decomposition technique (i.e., decomposing the total in-cylinder pressure signal to the compression-expansion, combustion, and resonance pressures) may also be considered to predict the combustion noise (Siano and Bozza, 2009).

The methodologies and the models of the level-2 piston noise and valvetrain noise models are described in Sections 11.5 and 11.6. It should be noted that, unlike the combustion noise, it is difficult to separate/measure the piston slap noise and the valvetrain noise in engine experiments due to the difficulty of measuring the impact events. Simulation in the virtual world plays a key role in these NVH areas.

Overall engine noise synthesis is an important part of system NVH analysis. The synthesis refers to the following (Fig. 11.4): (1) the sound-pressure-level synthesis by combining all the noises from different sources to predict the overall noise level; (2) the frequency-domain synthesis analysis for sound quality; (3) the sensitivity comparison of any design/operation changes between different subsystems (i.e., different sources of noise) to balance/coordinate at a system level to address the most important issues; and (4) the synthesis of transient noises in the time domain. In the synthesis analysis of the overall engine noise from the noises of each source (combustion, piston slap, valvetrain, etc.), the results depend on which noise is dominant. For



11.4 Concept of diesel engine NVH system synthesis analysis.

example, the combustion noise is dominant in naturally aspirated diesel engines, and sometimes the combustion noise is only 1–2 dB lower than the overall noise. However, at some conditions the combustion noise level can be well below the overall noise in a turbocharged diesel engine, bearing little effect on the overall noise. In the synthesis analysis, not only the difference between the overall noise level and each component noise level (i.e., the contribution) is important, the slope (larger or less than 1) and the shape (linear or nonlinear) of the sensitivity curve of each individual noise source are also important and informative for system design.

Useful references on system-level powertrain NVH analysis about the cranktrain, the valvetrain, and the engine block are provided by Stout (1997, 1999 and 2001), Stout *et al*. (2007), and Morel *et al*. (2003).

11.12.3 Level-3 system NVH model

The level-3 system NVH model predicts the structural attenuation, the transfer function, and the noise radiation with FEA and/or BEM. It does not have real-time capability. It tends to be used in a system synthesis stage to assemble the critical component models together, or used at a detailed component design level. The primary focus of engine system design is to produce system specifications in the early stage of the design. Therefore, the sophisticated level-3 model is not the primary focus in system design.

References for the level-3 models of powertrain dynamics and structural noise radiation include Moulin (2003), Payer and Platnick (1998), Priebsch and Krasser (1998), Schneider *et al*. (2002), Hayes and Quantz (1982), Loibnegger *et al*. (1997), Sung *et al*. (1997), Seybert *et al*. (1997), and Richardson and Riding (1997).

11.13 References and bibliography

11.13.1 References and bibliography on general NVH

Afaneh A H, Abdelhamid M K and Qatu M S (2007), 'Engineering challenges with vehicle noise and vibration in product development', SAE paper 2007-01-2434.

Allam S and Åbom M (2005), 'Acoustic modelling and testing of diesel particulate filters', *Journal of Sound and Vibration*, 288(1–2), 255–273.

Allam S and Åbom M (2006), 'Sound propagation in an array of narrow porous channels with application to diesel particulate filters', *Journal of Sound and Vibration*, 291(3–5), 882–901.

Alt N W, Nehl J, Heuer S and Schlitzer M W (2003), 'Prediction of combustion process induced vehicle interior noise', SAE paper 2003-01-1435.

Anderton D (1979), 'Relation between combustion system and engine noise', SAE paper 790270.

Anderton D (1984), 'Design options for low noise in IC engines', in Haddad S D and Watson N (eds), *Design and Applications in Diesel Engineering*, Ellis Horwood Limited, Chichester.

Anderton D and Priede T (1982), 'Applications of engine noise control research', in Hickling R and Kamal M M (eds), *Engine Noise – Excitation, Vibration, and Radiation*, Plenum Press, New York.

Anderton D and Zheng J H (1993), 'A new measurement method for separating airborne and structure borne sound from an IC engine's valve train mechanism', SAE paper 931335.

Anderton D, Grover E C, Lalor N and Priede T (1970), 'Origins of reciprocating engine noise – its characteristics, prediction, and control', ASME paper 70-WA/DGP-3.

Austen A E W and Priede T (1959), 'Origins of diesel engine noise', SAE paper 590127.

Badawi B A, Kholosy M, Omer A A and Shahin M A (2007), 'Identification of diesel engine events from acoustic signals using independent component analysis and time-frequency analysis', SAE paper 2007-01-2278.

Beardmore J M (1982), 'Piston stick slip noise generation mechanism', SAE paper 820753.

Beidl C, Rust A and Rasser M (2001), 'Key steps and methods in the development of low noise engines', *Sound and Vibration*, 35(4), 48–54.

Boesch N J W (1987), 'The development of low-noise DI diesel engines', SAE paper 870951.

Brandl F K, Affenzeller J and Thien G E (1987), 'Some strategies to meet future noise regulations for truck engines', SAE paper 870950.

Castillo M S (2001), 'Noise control for heavy duty trucks, a systems approach', SAE paper 2001-01-1531.

Challen B J (1975), 'The effect of combustion system on engine noise', SAE paper 750798.

Challen B J (1982), 'Techniques for engine noise control', in Hickling R and Kamal M M (eds), *Engine Noise – Excitation, Vibration, and Radiation*, Plenum Press, New York.

Challen B J and Croker D M (1982), 'A review of recent progress in diesel engine noise reduction', SAE paper 820517.

Cherne K D (1993), 'Motor vehicle noise regulations', SAE paper 931281.

Corcione F E, Mattia M G and Paciucci R (1989), 'Acoustic intensity measurements of noise emission from a light duty TCDI diesel engine', SAE paper 891130.

Corcione F E, Siano D, Vaglieco B M, Corcione G E, Lavorgna M, Viscardi M, Ladevaia M and Lecce L (2003), 'Analysis and control of noise emissions of a small single cylinder DI diesel engine', SAE paper 2003-01-1459.

Croker M D, Amphlett S A and Barnard A I (1995), 'Heavy duty diesel engine gear train modelling to reduce radiated noise', SAE paper 951315.

Cuschieri J M and Richards E J (1985), 'On the prediction of impact noise, part VIII: diesel engine noise', *Journal of Sound and Vibration*, 102(1), 21–56.

Davies P O A L (1996), 'Piston engine intake and exhaust system design', *Journal of Sound and Vibration*, 190(4), 677–712.

Dent W T and Chen C R (1989), 'A new approach for the analysis of OHC engine cam shape', SAE paper 891768.

Dhaenens M, Linden G, Nehl J and Thiele R (2001), 'Analysis of transient noise behavior of a truck diesel engine', SAE paper 2001-01-1566.

Evans D and Ward A (2005), 'Minimising turbocharger whoosh noise for diesel powertrains', SAE paper 2005-01-2485.

Evans D and Ward A (2006), 'The reduction of turbocharger whoosh noise for diesel powertrains', *8th International Conference on Turbochargers and Turbocharging,* Woodhead Publishing Limited Cambridge, pp. 29–42.

Farnell R and Riding D (1999), 'Engine noise', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 487–507.

Feng L, Gao W and Niu W (2008), 'An investigation on acoustic model and test of diesel particulate filter', SAE paper 2008-01-1757.

Gaikwad A, Mahale P and Raju S (2007), 'Noise assessment of diesel engines', SAE paper 2007-26-038.

Ghafouri M M and Ricci G (1993), 'A numerical method for the prediction of exhaust noise in internal combustion engine exhaust systems', SAE paper 931348.

Govindswamy K, Wolff K and Heuer S (2007), 'Aspects of powertrain noise with special emphasis on impulsive noise', SAE paper 2007-01-2411.

Govindswamy K, Wellmann T and Eisele G (2009), 'Aspects of NVH integration in hybrid vehicles', SAE paper 2009-01-2085.

Grover E C and Lalor N (1973), 'A review of low noise diesel engine design at ISVR', *Journal of Sound and Vibration*, 28(3), 403–431.

Haddad S D (1984), 'Mechanically-induced noise and vibration in diesel engines', in Haddad S D and Watson N (eds), *Design and Applications in Diesel Engineering*, Ellis Horwood Limited, Chichester.

Haller H, Beisenbusch K, Spessert B, Cordes H and Querengasser J (1993), 'Noise excitation by auxiliary units of internal combustion engines', SAE paper 931293.

Hanaoka M and Fukumura S (1973), 'A study of valve train noises and a method of cam design to reduce the noises', SAE paper 730247.

Harris C M (ed.) (1979), *Handbook of Noise Control*, 2nd edn, McGraw-Hill, New York.

Hayes P A and Quantz C A (1982), 'Determining vibration, radiation efficiency, and noise characteristics of structural designs using analytical techniques', SAE paper 820440.

Head H E and Wake J D (1980), 'Noise of diesel engines under transient conditions', SAE paper 800404.

Hickling R (2005), 'A plea for linear units as an alternative to decibels and octaves', SAE paper 2005-01-2549.

Hickling R and Kamal M M (eds) (1982), *Engine Noise – Excitation, Vibration, and Radiation*, Plenum Press, New York.

Inagaki M, Yoshikawa K, Wakabayashi M and Takama K (1993), 'Structural vibration analysis in turbocharger-exhaust systems', SAE paper 931318.

Isshiki Y, Shimamoto Y and Wakisaka T (1996), 'Simultaneous prediction of pressure losses and acoustic characteristics in silencers by numerical simulation', SAE 960637.

Jenkins S H (1975), 'Analysis and treatment of diesel-engine noise', *Journal of Sound and Vibration*, 43(2), 293–304.

Jenkins S H and Kuehner H K (1973), 'Diesel engine noise reduction hardware for vehicle noise control', SAE paper 730681.

Jenkins S H, Lalor N and Grover E C (1973), 'Design aspects of low-noise diesel engines', SAE paper 730246.

Jones A D (1984), 'Modeling the exhaust noise radiated from reciprocating internal combustion engines – a literature review', *Noise Control Engineering Journal*, 23(1), 12–31.

Juang T B, Harvey K E, Burzlaff M W and Rockett P J (2006), 'Analytical and experimental development to improve powertrain system induced vehicle NVH issues', SAE paper 2006-01-1533.

Kalser H J, Deges R, Schwarz D and Meyer J (1991), 'Investigations on valve train noise in multi-valve engines', SAE paper 911062.

Katari A, Syed M, Sickels M, Wahl T and Rajadurai S (2004), 'Effect of aspect ratio on pressure drop and acoustics in diesel particulate filters', SAE paper 2004-01-0695.

Knutsson M and Åbom M (2007), 'Acoustic analysis of charge air coolers', SAE paper 2007-01-2208.

Kunz F (1999), 'Semi-empirical model for flow noise prediction on intake and exhaust systems', SAE 1999-01-1654.

Kwak W-H, Kim H-S, Jung I-S, Yoo D-K and Kang K-T (2007), 'NVH optimization of HMC 3.0L V6 diesel engine', SAE paper 2007-01-2371.

Lahey H-P, Steffens C and Schultz C (2001), 'Simulation method for geartrain NVH assessment and optimization', SAE paper 2001-01-1593.

Laux P C, Blauwkamp K J, Cho P E and Unglenieks R J (2005), 'Robust process for the design and development of automotive noise control systems (a.k.a. acoustic integration)', SAE paper 2005-01-2395.

Lee S-D, Shannon B A and Mikulec A (2000), 'Cranktrain system geometry optimization and dynamic assembly of parametric solid models', *International Journal of Vehicle Design*, 23(1), 16–29.

Lee I J, Selamet A, Kim H, Kim T C and Kim J (2009a), 'Design of a multi-chamber silencer for turbocharger noise', SAE paper 2009-01-2048.

Lee I, Selamet A, Kim T-C and Kim J (2009b), 'A multi-chamber silencer for turbocharger noise', SAE paper 2009-01-2048.

Loibnegger B, Rainer G P, Bernard L, Micelli D and Turino G (1997), 'An integrated numerical tool for engine noise and vibration simulation', SAE paper 971992.

Lu M H and Jen M U (2007), 'Gain and loss of NVH characteristics from naturally aspirated to turbocharged engine', SAE paper 2007-01-2282.

Maetani Y, Niikura T, Suzuki S, Arai S and Okamura H (1993), 'Analysis and reduction of engine front noise induced by the vibration of the crankshaft system', SAE paper 931336.

March J, Strong G, Gregory S and Rediers B (2005a), 'Achieving diesel vehicle appeal, part 1: vehicle NVH perspective', SAE paper 2005-01-2484.

March J, Ward A, Bennett C and Towalski C (2005b), 'Achieving diesel vehicle appeal, part 2: powertrain NVH perspective', SAE paper 2005-01-2489.

Meisner S and Campbell B (1995), 'Development of gear rattle analytical simulation methodology', SAE paper 951317.

Mellin R C (1980), 'Noise and performance of automotive cooling fans', SAE paper 800031.

Miura Y and Kojima N (2003), 'Noise generating mechanism at idling for a four-cylinder in-line diesel engine', SAE paper 2003-01-1720.

Morel T, Silvestri J, Goerg K-A and Jebasinski R (1999), 'Modeling of engine exhaust acoustics', SAE paper 1999-01-1665.

Morel T, Keribar R and Leonard A (2003), 'Virtual engine/powertrain/vehicle simulation tool solves complex interacting system issues', SAE paper 2003-01-0372.

Mori T, Takaoka A and Maunder M (2005), 'Achieving a vehicle level sound quality target by a cascade to system level noise and vibration targets', SAE paper 2005-01-2394.

Moshrefi N, Mazzella G, Yeager D and Homco S (2007), 'Gasoline engine piston pin tick noise', SAE paper 2007-01-2290.

Moulin L (2003), 'Simulation study of the internal mechanical excitations to reduce the noise emitted by a heavy duty engine', SAE paper 2003-01-1727.

Munjal M L (1981), 'Evaluation and control of the exhaust noise of reciprocating IC engines', *Shock and Vibration Digest*, 13(1), 5–14.
Nathak S S, Rao M D and Derk J R (2007), 'Development and validation of an acoustic encapsulation to reduce diesel engine noise', SAE paper 2007-01-2375.
Ochiai K and Yokota K (1982), 'Light-weight, quiet automotive DI diesel engine oriented design method', SAE paper 820434.
Offner G, Priebsch H H, Ma M T, Karlsson U, Wikstrom A and Loibnegger B (2004), 'Quality and validation of cranktrain vibration predictions – effect of hydrodynamic journal bearing models', *Multi-Body Dynamics: Monitoring and Simulation Techniques-III*, pp. 255–271.
Onorati A (1995), 'Numerical simulation of exhaust flows and tailpipe noise of a small single cylinder diesel engine', SAE 951755.
Onorati A (1999), 'Numerical simulation of unsteady flows in internal combustion engine silencers and the prediction of tailpipe noise', in Winterbone D E and Pearson R J, *Design Techniques for Engine Manifolds*, SAE International, Warrendale, PA, pp. 267–340.
Padmanabhan C, Rook T E and Singh R (1995), 'Modeling of automotive gear rattle phenomenon: state of the art', SAE paper 951316.
Pang J, Kurrle P, Qatu M, Rebandt R and Malkowski R (2003), 'Attribute analysis and criteria for automotive exhaust systems', SAE paper 2003-01-0221.
Pang J, Rebandt R, Knapp G, Qatu M, Demmith D and Sheng G (2005), 'Flow excited noise analysis of exhaust', SAE paper 2005-01-2352.
Payer E and Platnick J L (1998), 'The use of CAE for the optimization of engine NVH behaviour: state-of-the-art and future trends', SAE paper 980274.
Pollack M, Govindswamy K and Hartwig M (2005), 'Cold start engine clatter noise evaluations', SAE paper 2005-01-2455.
Priebsch H H and Krasser J (1998), 'Simulation of vibration and structure borne noise of engines – a combined technique of FEM and multi body dynamics', *The 16th CAD-FEM Users' Meeting*, Bad Neuenahr, Germany, October 7–9, 1998.
Priede T (1980), 'In search of origins of engine noise – an historical review', SAE paper 800534.
Priede T (1982), 'Road vehicle noise', in White R G (ed.), *Noise and Vibration*, Ellis Horwood Limited, Chichester, pp. 389–437.
Priede T, Grover E C and Lalor N (1969), 'Relation between noise and basic structural vibration of diesel engines', SAE paper 690450.
Rämmal H and Åbom M (2007), 'Acoustics of turbochargers', SAE paper 2007-01-2205.
Reinhart T E (1991), 'US vehicle noise regulations and the effects on vehicle condition', SAE paper 912709.
Reinhart T E and Wahl T J (1997), 'Reducing compression brake noise', SAE paper 971870.
Reinhart T E, Sampath A, Bagga K S and Leistensnider G W (2003), 'NVH variations in diesel engine populations', SAE paper 2003-01-1723.
Richardson S H and Riding D H (1997), 'Predictive design support in the achievement of refined power for the Jaguar XK8', SAE paper 972041.
Riding D and Weeks R (1991), 'The application of noise simulation techniques to conceptual automotive powertrain design', SAE paper 911077.
Ruffinen M A Z, Hamilton D A, Hayes P A and Reinhart T E (1995), 'J366 driveby variability', SAE paper 951357.

Russell M F (1972), 'Reduction of noise emissions from diesel engine surfaces', SAE paper 720135.

Russell M F (1982), 'Diesel engine noise: control at source', SAE paper 820238.

Rust A and Thien G E (1987), 'Effect of transient conditions on combustion noise of NA-DI diesel engines', SAE paper 870989.

Rust A, Schiffbanker H and Brandl F K (1989), 'Complete NVH optimisation of a passenger vehicle with a DI diesel engine to meet subjective market demands and future legislative requirements', SAE paper 890125.

SAE Surface Vehicle Information Report J1242 (1991), '(R) Acoustic emission test methods'.

SAE Surface Vehicle Recommended Practice J986 (1998), 'Sound level for passenger cars and light trucks'.

SAE Surface Vehicle Recommended Practice J57 (2000), 'Sound level of highway truck tires'.

SAE Surface Vehicle Recommended Practice J1074 (2000), 'Engine sound level measurement procedure'.

SAE Surface Vehicle Recommended Practice J1207 (2000), 'Measurement procedure for determination of silencer effectiveness in reducing engine intake or exhaust sound level'.

SAE Surface Vehicle Recommended Practice J1477 (2000), 'Measurement of interior sound levels of light vehicles'.

SAE Surface Vehicle Recommended Practice J336 (2001), 'Sound level for truck cab interior'.

SAE Surface Vehicle Recommended Practice J1492 (2008), '(R) Measurement of light vehicle stationary exhaust system sound level engine speed sweep method'.

SAE Surface Vehicle Recommended Practice J1096 (2000), 'Measurement of exterior sound levels for heavy trucks under stationary conditions'.

SAE Surface Vehicle Standard J1470 (1998), '(R) Measurement of noise emitted by accelerating highway vehicles'.

SAE Surface Vehicle Standard J2555 (2000), 'Vehicle idle gear rattle evaluation procedure'.

SAE Surface Vehicle Standard J366 (2001), 'Exterior sound level for heavy trucks and buses'.

SAE Surface Vehicle Standard J88 (2006), '(R) Sound measurement – off-road work machines – exterior'.

SAE Surface Vehicle Standard J1169 (2007), 'Measurement of light vehicle exhaust sound level under stationary conditions'.

SAE Surface Vehicle Standard J2805 (2008), 'Measurement of noise emitted by accelerating road vehicles'.

Sasaki M and Nakashima K (2007), 'Human auditory models and sound quality evaluation method for diesel noise', SAE paper 2007-01-2219.

Savage J and Matterazzo J (1993), 'Application of design of experiments to determine the leading contributors to engine valvetrain noise', SAE paper 930884.

Schiffbanker H, Brandl F K and Thien G E (1991), 'Development and application of an evaluation technique to assess the subjective character of engine noise', SAE paper 911081.

Schneider M, Lahey H-P, Steffens C and Sonntag H-D (2002), 'CAE process to eliminate powertrain noise and vibration', SAE paper 2002-01-0459.

Schulte H S-H (1999), 'Passenger car engines', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 524–551.

Seybert A F, Hamilton D A and Hayes P A (1997), 'Prediction of radiated noise from engine components using the BEM and the Rayleigh integral', SAE paper 971954.

Shangguan W-B (2009), 'Engine mounts and powertrain mounting systems: a review', *International Journal of Vehicle Design*, 49(4), 237–258.

Siano D, Corcione F E, Bozza F, Gimelli A and Manelli S (2005), 'Characterization of the noise emitted by a single cylinder diesel engine: experimental activities and 1D simulation', SAE paper 2005-01-2483.

Siano D, Valentino G, Corcione F , Bozza F, Arnone L and Manelli S (2007), 'Experimental and numerical analyses of performances and noise emission of a common rail light duty D.I. diesel engine', SAE paper 2007-24-0017.

Silvestri J, Morel T and Costello M (1994), 'Study of intake system wave dynamics and acoustics by simulation and experiment', SAE paper 940206.

Skaistis S J (1975), 'Hydraulic fluidborne noise mechanics', SAE paper 751179.

Spessert B and Ponsa R (1990), 'Investigation in the noise from main running gear, timing gears and injection pump of DI diesel engines', SAE paper 900012.

Steyer G, Voight M and Sun Z (2005), 'Balancing competing design imperatives to achieve overall driveline NVH performance objectives', SAE paper 2005-01-2308.

Stout J L (1997), 'Valvetrain unbalance and its effects on powertrain NVH', SAE paper 971993.

Stout J L (1999), 'Concept level powertrain radiated noise analysis', SAE paper 1999-01-1746.

Stout J L (2001), 'Engine excitation decomposition methods and V engine results', SAE paper 2001-01-1595.

Stout J L, Rahimo G and Chien M-H (2005), 'Engine radiated noise prediction modeling using noise source decomposition and regression analysis', SAE paper 2005-01-2383.

Stout J L, Chien M-H, Rahimo G, Hoffman D, Williams R, Festag G, Holmes E, Mallela R and Mancini M (2007), 'NVH design and development of the Duratec35 engine from Ford Motor Company', SAE paper 2007-01-2414.

Stucklschwaiger W, Hargreaves N M and Beidl C V (1999), 'NVH considerations to make modern diesel engines an attractive alternative for SUV powertrains', SAE paper 1999-01-1772.

Suh I-S and Lyon R H (1999), 'An investigation of valve train noise for the sound quality of IC engines', SAE paper 1999-01-1711.

Sung S H, Nefske D J, Chen F H K and Fannin M P (1997), 'Development of an engine system model for predicting structural vibration and radiated noise of the running engine', SAE paper 972039.

Tao Z and Seybert A F (2003), 'A review of current techniques for measuring muffler transmission loss', SAE paper 2003-01-1653.

Thomson W (1978), *Fundamentals of Automotive Engine Balance*, Mechanical Engineering Publications, London.

Tousignant T, Wellmann T, Govindswamy K, Heuer S and Workings M (2009), 'Application of combustion sound level (CSL) analysis for powertrain NVH development and benchmarking', SAE paper 2009-01-2168.

Trochon E P (2001), 'A new type of silencers for turbocharger noise control', SAE paper 2001-01-1436.

Wahl T J and Reinhart T E (1997), 'Developing a test procedure for compression brake noise', SAE paper 972038.

Walker S (1999), 'Larger engine noise and vibration control', in Challen B and Baranescu

R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 511–519.
Wang M Y, Manoj R and Zhao W (2001), 'Gear rattle modeling and analysis for automotive manual transmissions', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 215, 241–258.
Wellmann T, Govindswamy K, Braun E and Wolff K (2007), 'Aspects of driveline integration for optimized vehicle NVH characteristics', SAE paper 2007-01-2246.
Werkmann M, Tunsch M and Kuenzel R (2005), 'Piston pin related noise – quantification procedure and influence of pin bore geometry', SAE paper 2005-01-3967.
Werner L (1987), 'Control of piston stick slip noise in a 4-cylinder gasoline engine', SAE paper 870991.
Wodtke H-W and Bathelt H (2004), 'Noise emissions', in Basshuysen R and Schafer F (eds), *Internal Combustion Engine Handbook*, SAE International, Warrendale, PA, pp. 753–768.
Wolf A, Gaertner U and Droege T (2003), 'Noise reduction and sound design for diesel engines – an achievable development target for US passenger cars?', SAE paper 2003-01-1719.
Yadav P S, Muthukumar A, Kiran V V P, Tandon V and Raju S (2007), 'Optimized design of silencer – an integrated approach', SAE paper 2007-26-037.
Yawata Y and Crocker M J (1983), 'Identification of internal noise sources in diesel engines', SAE paper 831330.
Zhao H and Reinhart T E (1999), 'The influence of diesel engine architecture on noise levels', SAE paper 1999-01-1747.

11.13.2 References and bibliography on combustion noise

Alt N W, Wiehagen N, Steffens C and Heuer S (2001), 'Comprehensive combustion noise optimization', SAE paper 2001-01-1510.
Alt N, Sonntag H-D, Heuer S and Thiele R (2005), 'Diesel engine cold start noise improvement', SAE paper 2005-01-2490.
Anderton D (1979), 'Relation between combustion system and engine noise', SAE paper 790270.
Badami M, Mallamo F, Millo F and Rossi E E (2002), 'Influence of multiple injection strategies on emissions, combustion noise and BSFC of a DI common rail diesel engine', SAE paper 2002-01-0503.
Badaoui M E, Daniere J, Guillet F and Serviere C (2005), 'Separation of combustion noise and piston-slap in diesel engine – part I: separation of combustion noise and piston-slap in diesel engine by cyclic Wiener filtering', *Mechanical Systems and Signal Processing*, 19, 1209–1217.
Blunsdon C A, Dent J C and Das S (1995), 'Modelling the origins of combustion noise in the indirect injection diesel engine', SAE paper 952432.
Brandl F, Veit J, Arndt R and Priebsch H-H (2007), 'The application of a new software tool for separating engine combustion and mechanical noise excitation', SAE paper 2007-01-2376.
Carlucci P, Ficarella A, Chiara F, Giuffrida A and Lanzafame R (2004), 'Preliminary studies on the effects of injection rate modulation on the combustion noise of a common rail diesel engine', SAE paper 2004-01-1848.
Challen B J (1975), 'The effect of combustion system on engine noise', SAE paper 750798.

Costa M, Siano D, Allocca L, Montanaro A and Bozza F (2009a), 'Light duty diesel engine: optimization of performances, noxious emission and radiated noise', SAE paper 2009-32-0105.

Costa M, Siano D, Valentino G, Corcione F E and Bozza F (2009b), 'Prediction and optimization of the performances, noxious emissions and radiated noise of a light duty common-rail diesel engine', SAE paper 2009-24-0011.

Dhaenens M, Linden G, Nehl J and Thiele R (2001), 'Analysis of transient noise behavior of a truck diesel engine', SAE paper 2001-01-1566.

Gazon M and Blaisot J-B (2006), 'Cycle-to-cycle fluctuations of combustion noise in a diesel engine at low speed', SAE paper 2006-01-3410.

Hawksley G J and Anderton D (1978), 'Studies into combustion and noise in turbocharged engines', IMechE paper C72/78.

Kohketsu S, Mori K, Kato T and Sakai K (1994), 'Technology for low emission, combustion noise and fuel consumption on diesel engine', SAE paper 940672.

Lee M, Bolton J S and Suh S (2009), 'Estimation of the combustion-related noise transfer matrix of a multi-cylinder diesel engine', *Measurement Science and Technology*, 20(1), 015106.

Luckhchoura V, Won H W, Sharma A, Paczko G and Peters N (2008), 'Investigation of combustion noise development with variation in start of injection using 3-dimensional simulations by applying representative interactive flamelet (RIF) model', SAE paper 2008-01-0950.

Machado G B and De Melo T C C (2005), 'Diesel cetane number versus noise emission', SAE paper 2005-01-2150.

Mallamo F, Badami M and Millo F (2002), 'Analysis of multiple injection strategies for the reduction of emissions, noise and BSFC of a DI CR small displacement non-road diesel engine', SAE paper 2002-01-2672.

Mendez S and Thirouard B (2008), 'Using multiple injection strategies in diesel combustion: potential to improve emissions, noise and fuel economy trade-off in low CR engines', SAE paper 2008-01-1329.

Reinhart T E (1987), 'An evaluation of the Lucas combustion noise meter on Cummins B series engines', SAE paper 870952.

Roy M M and Tsunemoto H (2002), 'Effect of injection pressure and split injection on exhaust odor and engine noise in DI diesel engines', SAE paper 2002-01-2874.

Russell M F (1984), 'Combustion noise in automotive diesel engines', in Haddad S D and Watson N (eds), *Design and Applications in Diesel Engineering*, Ellis Horwood Limited, Chichester, pp. 203–252.

Russell M F and Young C D (1985), 'Measurement of diesel combustion noise', *Proc. IMechE Conference Autotech '85*, Birmingham, October.

Russell M F, Young C D and Nicol S W (1990), 'Modulation of injection rate to improve direct injection diesel engine noise', SAE paper 900349.

Scarpati J, Wikström A, Jönsson O, Glav R, Händel P and Hjalmarsson H (2007), 'Prediction of engine noise using parameterized combustion pressure curves', SAE paper 2007-01-2373.

Shu G-Q, Wei H-Q and Han R (2005a), 'The transfer function of combustion noise in DI-diesel engine', SAE paper 2005-01-2486.

Shu G-Q, Wei H-Q and Liang X-Y (2005b), 'Experimental study on combustion noise in transient conditions of DI-diesel engine', SAE paper 2005-01-2487.

Shu G-Q, Wei H-Q, Wang Y-J, Yang Z-Y and Wei J-S (2006), 'Secondary influence factors of combustion noise mechanism under transient conditions of DI-diesel engine', SAE paper 2006-01-1530.

Siano D and Bozza F (2009), 'Combustion noise prediction in a small diesel engine finalized to the optimization of the fuel injection strategy', SAE paper 2009-01-2077.

Tabuchi T, Fujitani N and Makino N (1995), 'Recent technology to improve engine combustion noise and exhaust emission by optimizing fuel injection system', SAE paper 951791.

Torregrosa A J, Broatch A, Martin J and Monelletta L (2007), 'Combustion noise level assessment in direct injection diesel engines by means of in-cylinder pressure components', *Measurement Science and Technology*, 18, 2131–2142.

Tung V T C and Crocker M J (1982), 'Diesel engine noise and relationship to cylinder pressure', SAE paper 820237.

Wang S, Chalu C and Gautier F (2007), 'Optimization of combustion noise of modern diesel engines for passenger cars', SAE paper 2007-01-2379.

Wolschendorf J, Dürnholz M and Schmillen K (1991), 'The IDI diesel engine and its combustion noise variations', SAE paper 910228.

11.13.3 References and bibliography on piston slap noise

Chien M H (1995), 'Engine impact measurement and quantification', SAE paper 951236.

DeJong R G and Parsons N E (1982), 'Piston slap noise reduction in a Vee-block diesel engine', SAE paper 820240.

De Luca J C and Gerges S N Y (1996), 'Piston slap excitation: literature review', SAE paper 962395.

Fabi P, Flekiewicz M, Madej H, Wojnar G and Flekiewicz B (2007), 'Influence of piston slap on engine block vibration', SAE paper 2007-01-2163.

Furuhama S and Hirukawa K (1983), 'Piston slap motion and engine noise under low temperature idling operation of diesel engines', SAE paper 830066.

Gerges S N Y, De Luca J C and Lalor N (2005), 'Effect of cylinder lubrication on piston slap', SAE paper 2005-01-2165.

Haddad S D (1995), 'Theoretical treatment of piston motion in IC piston engine for the prediction of piston slap excitation', *Mechanism and Machine Theory*, 30(2), 253–269.

Haddad S D and Howard D A (1980), 'Analysis of piston slap-induced noise and assessment of some methods of control in diesel engines', SAE paper 800517.

Haddad S D and Tjan K-T (1995), 'An analytical study of offset piston and crankshaft designs and the effect of oil film on piston slap excitation in a diesel engine', *Mechanism and Machine Theory*, 30(2), 271–284.

Kageyama H, Hara S and Kawabata Y (1994), 'Study of the simulation of piston skirt contact', JSAE paper 9430095.

Kaiser H J, Schmillen K and Spessert B (1988), 'Acoustical optimization of the piston slap by combination of computing and experiments', SAE paper 880100.

Kamiya M, Kobayashi T, Mihara Y and Someya T (2007), 'Measurement of piston skirt oil-film pressure under piston slap', SAE paper 2007-01-2215.

Kamp H and Spermann J (1995), 'New methods of evaluating and improving piston related noise in internal combustion engines', SAE paper 951238.

Kobayashi T, Takahashi Y and Bell D J (2007), 'How to predict the piston slap-noise using 3D piston motion simulation', SAE paper 2007-01-1245.

Künzel R, Tunsch M and Werkmann M (2000), 'Piston related noise with spark ignition

engines – characterization, quantification and mechanisms of excitation', SAE paper 2000-01-3311.
Künzel R, Werkmann M and Tunsch M (2001), 'Piston related noise with diesel engines parameters of influence and optimization', SAE paper 2001-01-3335.
Lalor N, Grover E C and Priede T (1980), 'Engine noise due to mechanical impacts at pistons and bearings', SAE paper 800402.
Munro R and Parker A (1975), 'Transverse movement analysis and its influence on diesel piston design', SAE paper 750800.
Nakada T, Yamomoto A and Abe T (1997), 'A numerical approach for piston secondary motion analysis and its application to the piston related noise', SAE paper 972043.
Oetting H, Pundt D and Ebbinghaus W (1984), 'Friction in the piston group and new ideas for piston design', SAE paper 841299.
Ohta K, Irie Y, Yamamoto K and Ishikawa H (1987), 'Piston slap induced noise and vibration of internal combustion engines (1st report, theoretical analysis and simulation)', SAE paper 870990.
Paranjpe R S (1998), 'Development of a math-based piston noise model', SAE paper 980564.
Patel P, Mourelatos Z P and Shah P (2007), 'A comprehensive method for piston secondary dynamics and piston-bore contact', SAE paper 2007-01-1249.
Richmond J W and Parker D A (1987), 'The quantification and reduction of piston slap noise', *Proc. IMechE*, *Part D: Transport Engineering*, 201(4), 235–244.
Rohrle M D (1975), 'Affecting diesel engine noise by the piston', SAE paper 750799.
Ross D and Ungar E E (1965), 'On piston slap as a source of engine noise', ASME paper 65-OGP-10.
Ryan J P, Wong V W, Lyon R H, Hoult D P, Sekiya Y, Kobayashi Y and Aoyama S (1994), 'Engine experiments on the effects of design and operational parameters on piston secondary motion and piston slap', SAE paper 940695.
Sander W, Steidle W and Wacker E (1979), 'Piston movement and its influence on noise of automotive engines', SAE paper 790272.
Slack J W and Lyon R H (1982), 'Piston slap', in Hickling R and Kamal M M (eds), *Engine Noise – Excitation, Vibration, and Radiation*, Plenum Press, New York.
Teraguchi S, Suzuki W, Takiguchi M and Sato D (2001), 'Effects of lubricating oil supply on reductions of piston slap vibration and piston friction', SAE paper 2001-01-0566.
Usami T, Wada S and Sonoda S (1975), 'Piston slap noise of indirect combustion diesel engine', SAE paper 750801.
Vora K C and Ghosh B (1991), 'Vibration due to piston slap and combustion in gasoline and diesel engines', SAE paper 911060.
Whitacre J P (1990), 'Automotive gasoline engine piston noise, sources and solutions', SAE paper 901491.
Wilson R and Fawcett J N (1974), 'Dynamics of the slider-crank mechanism with clearance in the sliding bearing', *Mechanism and Machine Theory*, 9, 61–80.
Wong V W, Tian T, Lang H, Ryan J P, Sekiya Y, Kobayashi Y and Aoyama S (1994), 'A numerical model of piston secondary motion and piston slap in partially flooded elastohydrodynamic skirt lubrication', SAE paper 940696.
Xin Q (1999), *Analytical and Numerical Study of Lubrication Dynamics for Internal Combustion Engine Piston Assemblies and Journal Bearings*, D.Sc. (Doctor of Science) degree dissertation, Washington University, St Louis, Missouri, USA.
Yonezawa T and Kanda H (1985), 'Analysis of cavitation erosion on cylinder liner and cylinder block', SAE paper 850401.

12
Diesel engine heat rejection and cooling

Abstract: This chapter develops the analysis methods of heat rejection based on the engine energy balance of the first law of thermodynamics for diesel engine system design. A theoretical analysis of engine miscellaneous heat losses forms the foundation of the analysis approach. The concept of base engine coolant heat rejection is proposed and its characteristics are illustrated. Cooling system design calculations are then discussed to address cooler performance, cooling capability, and coolant temperatures.

Key words: heat rejection, cooling, first law of thermodynamics, energy balance, miscellaneous heat losses, base engine coolant heat rejection, low-heat-rejection engine, cooler, sink temperature, waste heat recovery.

12.1 Engine energy balance analysis

12.1.1 Challenges from thermal load to heat rejection control

Turbocharged diesel engines have a higher thermal load than non-turbocharged diesel engines. Engine heat transfer and thermal loading have been extensively researched in the past several decades. Engine heat transfer fundamentals are introduced by Heywood (1988). Thermal loading and engine cooling are discussed in detail by Sitkei (1974) and Challen and Baranescu (1999). The thermal load refers to the gas temperature and the heat flux a component is exposed to. A higher thermal load in an engine is typically characterized by higher gas temperatures in the intake manifold (e.g., caused by the hot air from the compressor outlet), the cylinder, and the exhaust manifold. The higher thermal load causes many durability problems, such as cracks in the cylinder head and exhaust manifold, and damage in the turbine rotor. Managing the thermal load and reducing the exhaust manifold gas temperature have been important design challenges for turbocharged engines.

The successful measures in thermal load control for non-EGR engines used to include the following: (1) reducing the intake manifold temperature by using air charge cooling; (2) reducing the compressor outlet air temperature by increasing the compressor efficiency; (3) increasing fresh air scavenging from the intake port to the exhaust port by increasing the pressure differential between the intake manifold and the exhaust manifold through turbocharger matching or manifold design; (4) increasing fresh air scavenging by increasing the valve overlap through cam design; (5) increasing the air–fuel ratio to reduce the exhaust temperature; (6) enhancing the cooling capacity of the cylinder head; and (7) adjusting the fuel injection timing and the shape of heat release rate to reduce the exhaust temperature.

However, many of the above measures cannot be applied to modern EGR engines. For example, enhancing fresh air scavenging with a negative engine delta P (i.e., intake manifold pressure higher than exhaust manifold pressure) is impossible due to the need to drive EGR flow. On the other hand, for EGR engines adding cooled EGR can reduce the in-cylinder gas temperature and the NO_x concentration due to the higher specific heat of the EGR gases. Meanwhile, the exhaust gas temperature and the thermal load can be reduced accordingly. However, cooling the EGR increases the total engine heat rejection (heat transfer rate) regardless of whether it is cooled by engine coolant, cooling air or low-temperature radiator coolant. The heat rejection increases as more EGR is used. Essentially, in modern diesel engines, a large portion of the thermal load is moved by EGR from the hot gas to the coolant because the engine needs to run cooler in order to meet the NO_x emissions requirement. The design challenge is then shifted to how to manage the heat rejection due to the EGR gas. In a modern powertrain development program, one of the top priorities is to control the engine heat rejection and design a packageable cooling system to control the radiator inlet coolant temperature.

The coolant heat rejection in modern diesel engines increases mainly for three reasons: high EGR rate, low intake manifold gas temperature, and high power rating. In order to meet stringent emissions regulations, high EGR rate and very cold EGR are required to achieve low oxygen concentration and low mixture gas temperature in the intake manifold. The increasing customer demand on engine power rating results in a proportional increase in the fueling rate and the coolant heat rejection. Owing to the difficulties in heat rejection control and cooling system design (Bowman *et al.*, 2005), all the heat rejection data need to be scrutinized for accuracy. Any inaccurate estimation method becomes obsolete. A critical task is to predict heat rejections (including the important cylinder heat rejection) accurately at various extreme ambient conditions.

12.1.2 System methodology of heat rejection and cooling analysis

Experimental errors in coolant flow rate and temperature cause variations in the measurement of engine coolant heat rejection. The heat rejection errors can range from 3 to 20 kW and become unacceptably large for precise design. In order to measure the engine cylinder coolant heat rejection directly, experimental errors need to be minimized. Alternatively, an indirect approach with sufficient accuracy can be used to calculate the coolant heat rejection.

A numerical simulation of the cylinder heat transfer by using in-cylinder CFD is usually inaccurate or impractical due to the extremely complicated conditions of the in-cylinder turbulent flows. A practical approach needs to be developed for engine system design to handle the prediction of coolant heat rejection and cooler design parameters. This chapter presents a methodology for the calculations. The methodology includes the following features:

1. The foundation is built upon estimating engine miscellaneous heat losses.
2. Two energy balance methods of the first law of thermodynamics with different boundaries are used to calculate and compare the cylinder heat rejection.
3. The percentage of base engine coolant heat rejection lost to fuel energy is selected as a characteristic parameter for the basis of heat rejection analysis.
4. The Woschni correlation of in-cylinder heat transfer is selected in an engine cycle simulation model and is calibrated on the basis of item (3) above.
5. The cooling medium temperatures of the coolers and the radiator are estimated.
6. The cooling capability of the coolers and the radiator is defined.

12.1.3 Energy balance analysis methods of the first law of thermodynamics

As a part of the methodology of heat rejection analysis in engine system design, engine energy balance equations are developed to use the gas flow rates and gas temperatures to calculate heat rejections. Two methods of calculating the energy balance are presented below, depending on where the control volume boundary is drawn for the system.

Method 1 sets the control volume boundary for gas flows at the compressor inlet and the turbine outlet, and the boundary for the fuel flow is set at the fuel tank. If low-pressure-loop EGR is used, the boundary is set after the EGR pickup. For a coolant-cooled turbocharged engine, as an example, the energy balance equation is given as follows:

$$\begin{aligned} &\dot{m}_{comp} h_{CompIn} + \dot{m}_{fuel} h_{FuelIn} + \dot{m}_{fuel} q_{LHV} \\ &= \dot{W}_E + \dot{W}_{Eacc} + \dot{m}_{exh} h_{TurbOut} + \dot{Q}_{base-coolant} + \dot{Q}_{EGRcooler} \\ &+ \dot{Q}_{CAC} + \dot{Q}_{ISC} + \dot{Q}_{FuelCooler} + \dot{Q}_{miscellaneous,1} \end{aligned} \qquad 12.1$$

In equation 12.1, $\dot{m}_{comp}$ is the compressor air or gas flow rate, h_{CompIn} is the specific enthalpy of the compressor inlet air or gas flow, $\dot{m}_{fuel}$ is the fuel flow rate, and h_{FuelIn} is the specific enthalpy of the fuel flow at the fuel tank temperature. Note that $\dot{m}_{fuel}h_{FuelIn}$ is usually very small compared to the enthalpy of the air flows. The q_{LHV} is the lower heating value of the fuel which is related to fuel formulation (see ASTM American National Standard D4868-00 and SAE J1498. Note that the enthalpy of vaporization of diesel fuels is usually small compared to their lower heating value. $\dot{W}_E$ is the engine firing brake power. $\dot{W}_{Eacc}$ is a certain engine accessory power including the alternator, the air compressor, and the cooling fan. The $\dot{m}_{exh}$ is the exhaust gas flow rate, and $h_{TurbOut}$ is the specific enthalpy of the turbine outlet exhaust gas flow. $\dot{Q}_{base-coolant}$ is defined as the base engine coolant heat rejection (to be detailed later). $\dot{Q}_{EGRcooler}$ is the EGR cooler heat rejection. $\dot{Q}_{CAC}$ is the charge air cooler heat rejection. $\dot{Q}_{ISC}$ is the compressor inter-stage cooler heat rejection (if any). $\dot{Q}_{FuelCooler}$ is the fuel cooler heat rejection (if any). $\dot{Q}_{miscellaneous,1}$ is the miscellaneous heat losses such as the convection and radiation heat transfer from the exhaust manifold, the EGR tubing (excluding the EGR cooler), the engine block and the turbocharger surface, and the thermal energy of the unburned or incompletely burned fuel (up to 1–2% of the fuel energy).

Method 2 sets the control volume boundary for gas flows at the intake manifold and the turbine inlet, and the boundary for the fuel flow is set at the fuel tank. If high-pressure-loop EGR is used, the boundary is set before the EGR pickup. As an example, for a coolant-cooled turbocharged engine the energy balance is given as follows:

$$\begin{aligned}&\dot{m}_{air}h_{IMT,air} + \dot{m}_{EGR}h_{IMT,EGR} + \dot{m}_{fuel}h_{FuelIn} + \dot{m}_{fuel}q_{LHV}\\ &= \dot{W}_E + \dot{W}_{Eacc} + \dot{W}_{EGRpump} + \dot{W}_{supercharger} + \dot{m}_{ex}h_{TurbIn}\\ &\quad + \dot{Q}_{base-coolant} + \dot{Q}_{FuelCooler} + \dot{Q}_{miscellaneous,2}\end{aligned} \tag{12.2}$$

In equation 12.2, $h_{IMT,air}$ is the specific enthalpy of the air flow entering the intake manifold, $h_{IMT,EGR}$ is the specific enthalpy of the EGR flow entering the intake manifold, h_{TurbIn} is the specific enthalpy of the turbine inlet exhaust gas flow, $\dot{W}_{EGRpump}$ is the EGR pump power (if any), and $\dot{W}_{supercharger}$ is the mechanical supercharger power (if any). It should be noted that $\dot{Q}_{miscellaneous,2}$ is lower than $\dot{Q}_{miscellaneous,1}$. Moreover, the heat rejections of the transmission cooler and other small auxiliary coolers are not explicitly included in the above equations. For example, the heat generated inside the transmission comes from the gear or torque converter frictional power losses, which are already a part of the engine crankshaft output power. Similarly, vehicle accessory resistance power, such as the power from the power steering pump and the air conditioning compressor, needs to be overcome by a part of the engine crankshaft power $\dot{W}_E$.

Given the turbocharger power balance between the compressor and the turbine, equations 12.1 and 12.2 can be derived as equivalent (i.e., if equation 12.1 is subtracted from 12.2, all the terms will cancel out to zero on both sides of the equality sign). The gas flow enthalpy is calculated by using

$$h = \int_0^T c_p \cdot \mathrm{d}T$$

where the specific heat c_p must be used as a function of temperature and gas species composition for acceptable accuracy. The cooler heat rejection can be calculated by the following:

$$\dot{Q}_{cooler} = \dot{m}_{gas} c_p (T_{CoolerInletGas} - T_{CoolerOutletGas}) \qquad 12.3$$

In order to characterize a base engine heat rejection, it is important to distinguish two types of working fluids: (1) the in-cylinder flows (including air, EGR, and fuel); and (2) the outside-cylinder cooling media (including coolant and oil flows). The pressures and temperatures of the gas and fuel flows at the entry of the cylinder are affected by their respective pumping power (e.g., a compressor or supercharger for the fresh air flow, an EGR pump for the EGR gas flow, a fuel pump for the fuel flow) and cooler heat rejection (e.g., charge air cooler, EGR cooler, fuel cooler). The in-cylinder gas flows participate in the engine cycle processes and their energy leaves the cylinder as shaft work, exhaust enthalpy, and heat rejection. On the other hand, the water pump power and the oil pump power for pumping the outside-cylinder cooling media eventually only dissipate as heat rejection. Therefore, it is usually convenient to define the 'base engine coolant heat rejection' to include the heat rejections from the following:

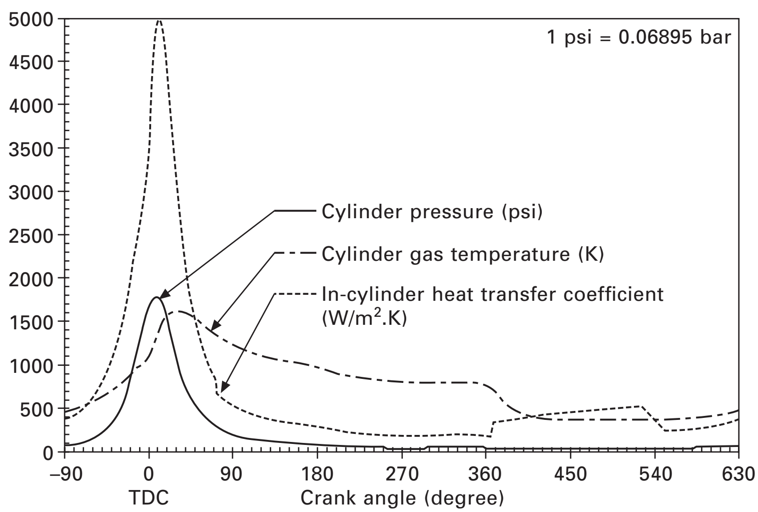
- the cylinders (the piston, the valves, the liner, and the cylinder head, including the exhaust ports)
- the oil cooler (including piston cooling and mechanical rubbing friction)
- the water pump power and the oil pump power.

The base engine heat rejection is a critical parameter which characterizes the fundamental design and operating features of the base engine (i.e., excluding the effect of the coolers, except for the oil cooler). Different engines can be compared on the common basis of the percentage of fuel energy lost to the base engine coolant heat rejection in order to evaluate their performance competitiveness. It should be noted that if the cylinder heat transfer is the focus of the study, engine friction and water/oil pump power should be excluded from the definition of base engine coolant heat rejection.

In equations 12.1 and 12.2, the only unknown is $\dot{Q}_{base-coolant}$. If there were no experimental errors in gas temperature and flow rate measurements, the same result of base engine coolant heat rejection would have been obtained by using either method 1 or 2. Therefore, the above energy balance methods can serve the following five purposes:

1. Calculate coolant heat rejection accurately for cooling system design, including the cylinder heat rejection. Note that the vehicle radiator coolant heat rejection is equal to the sum of the heat rejections from the base engine, EGR cooler, and other coolant-cooled coolers.
2. Compare the calculated coolant heat rejections from methods 1 and 2 with available heat rejection data coming from the direct coolant flow measurements in order to verify the data accuracy for each method.
3. Check the accuracy of engine performance test data of gas temperatures and flow rates.
4. Reveal the engine energy balance distribution based on the first law of thermodynamics, and reveal the competitiveness of the base engine design in term of heat losses.
5. Tune the cylinder heat transfer model (e.g., a multiplier for the Woschni in-cylinder gas-side heat transfer coefficient) in the engine cycle simulation to match with the engine test data on the percentage of fuel energy lost to the base engine coolant heat rejection.

Figure 12.1 shows a tuned in-cylinder heat transfer coefficient. Note that the heat transfer coefficient is at the gas side from the in-cylinder gas to the cylinder inner wall. It is much more dominant in terms of thermal resistance than the coolant-side heat transfer coefficient if compared in the overall heat transfer coefficient from the gas to the coolant. Also note that when



12.1 Engine in-cylinder cycle simulation of gas pressure, temperature, and heat transfer.

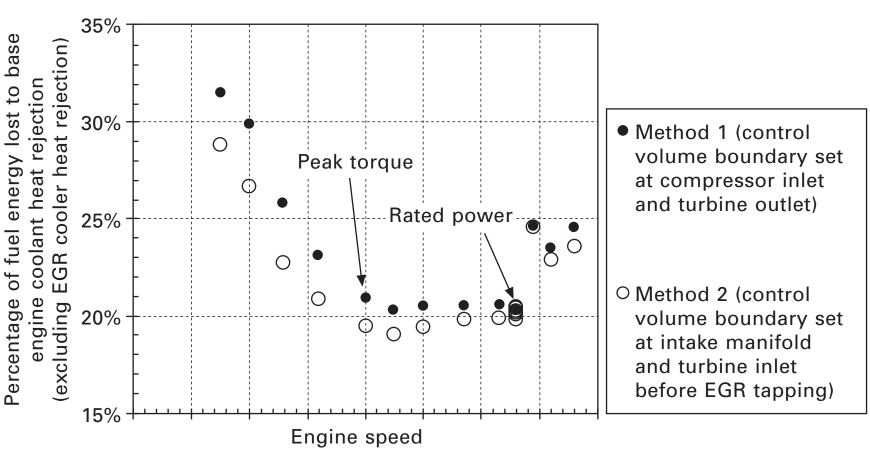
the fuel injection timing is advanced, the shape of the in-cylinder pressure, temperature, and heat transfer coefficient traces change or shift accordingly, and the percentage of base engine coolant heat rejection in fuel energy increases. Figure 12.2 shows an example of the energy balance calculations using engine test data at full load. It is observed that the results given by the two methods match well. Figure 12.3 shows typical energy balance distributions of diesel engines. Another example of diesel engine energy balance was provided by Wallace and Kremer (2009).

12.2 Engine miscellaneous energy losses

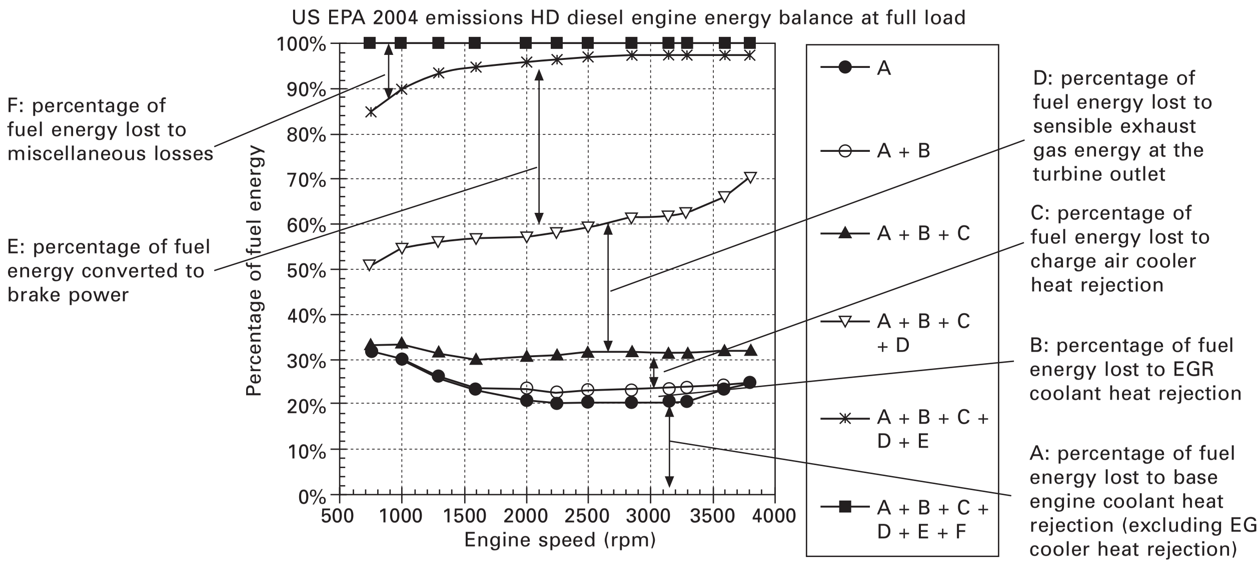
12.2.1 Overview of miscellaneous energy losses

In the energy balance equations 12.1 and 12.2, the miscellaneous loss is treated as a known quantity in order to calculate the base engine coolant heat rejection. The miscellaneous loss is not negligible for the precise control of heat rejections in modern high-EGR engines (Fig. 12.3). In fact, the analysis of the miscellaneous loss is very complex. Its heat transfer part is related to engine speed (i.e., the time scale for heat transfer) and the exhaust manifold gas temperature that depends on engine load. Only after the miscellaneous heat losses are calculated, can the coolant heat rejection be estimated in engine system design by using method 1 or 2. Therefore, the analysis of the miscellaneous heat losses is of paramount importance in the heat transfer theory in diesel engine system design. The complete theory on miscellaneous heat losses was developed by Xin and Zheng (2009).

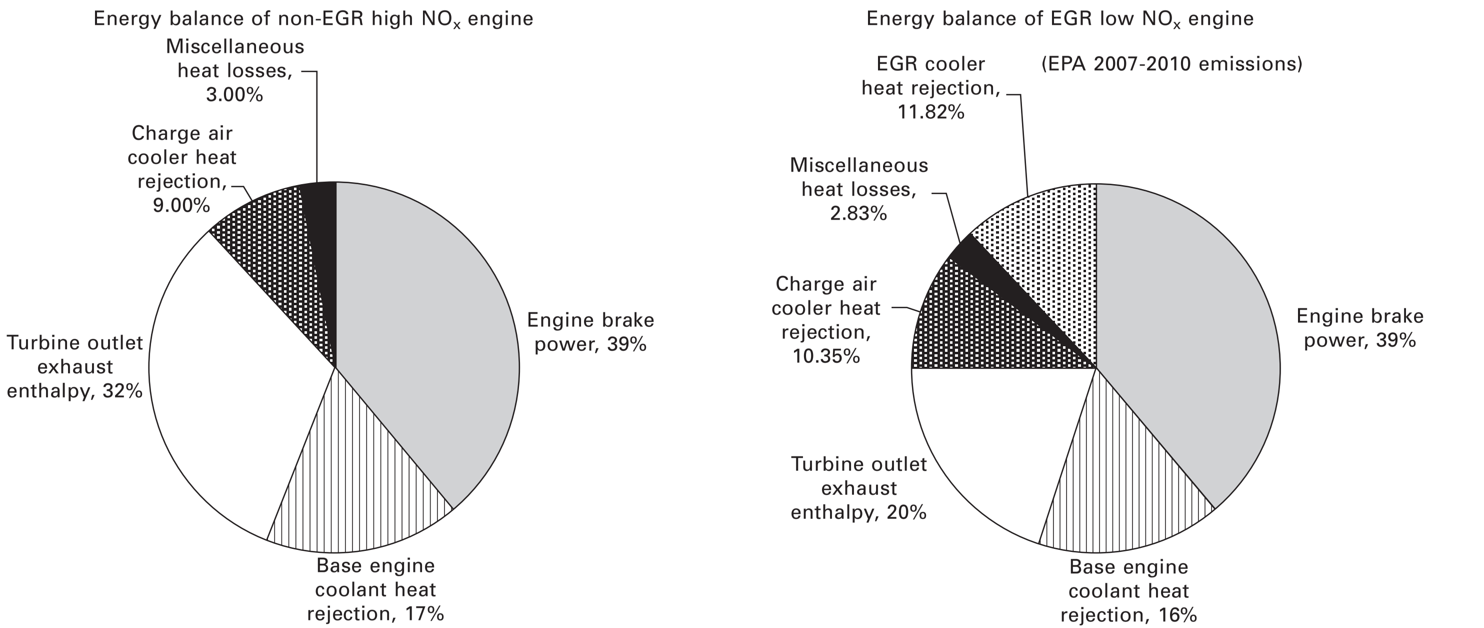
Xin and Zheng (2009) found that the percentage of fuel energy lost to the



12.2 Energy balance calculation on full-load lug curve.



12.3 Engine energy balance comparison between non-EGR and EGR engines.



12.3 Continued

miscellaneous heat losses is a function of certain characteristic gas temperature (e.g., exhaust manifold gas temperature), cooling medium temperature, engine speed and load. Figure 12.4 shows that the percentage of the miscellaneous losses increases when engine speed or load decreases. They also evaluated the miscellaneous heat losses for a water-cooled diesel engine using energy balance method 1. If the losses due to incomplete combustion are negligible and natural convection heat transfer occurs around the engine in a test cell or an engine compartment, 3% of fuel energy lost to the miscellaneous heat is a reasonable estimate with method 1 at full load from peak torque to rated power. The calculated percentage difference between methods 1 and 2 is about 1.8% (Fig. 12.5).

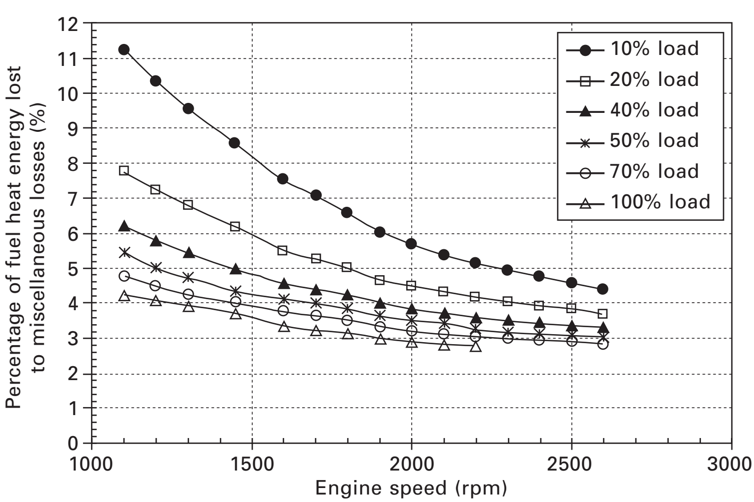
12.2.2 Parametric dependency of miscellaneous heat losses

Denote the percentage of miscellaneous heat losses in fuel energy as

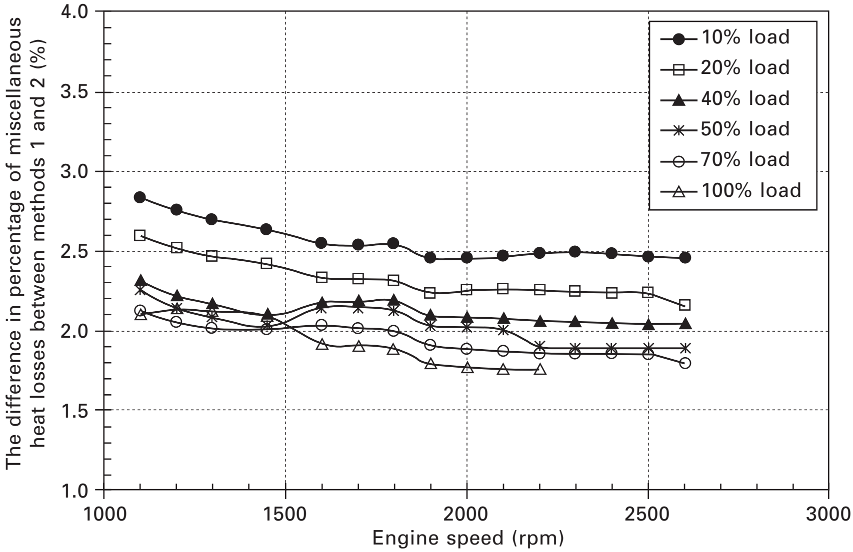
$$G = \frac{\dot{Q}_{miscellaneous}}{\dot{Q}_{fuel}} \qquad 12.4$$

where $\dot{Q}_{fuel}$ is the fuel energy rate. The miscellaneous heat losses of energy balance method 1 can be expressed in the following form (Xin and Zheng, 2009):

$$\dot{Q}_{miscellaneous,1} = C_1 + C_2 T_{ch} + C_3 T_{ch}^4 \qquad 12.5$$



12.4 Miscellaneous heat losses as functions of engine speed and load (calculated with method 1).



12.5 Calculated difference in the percentage of miscellaneous heat losses between the energy balance methods 1 and 2.

where C_1, C_2 and C_3 are coefficients. T_{ch} is a characteristic gas temperature (e.g., exhaust manifold gas temperature). The term C_2T_{ch} represents the convective heat transfer from the engine surface to the cooling medium. The term $C_3T_{ch}^4$ represents the radiation heat transfer. Note that the complex nonlinear effect of the temperature in radiation heat transfer is lumped into the coefficients. Substituting equation 12.5 into 12.4, the following is obtained:

$$G_1 = \frac{C_1 + C_2T_{ch} + C_3T_{ch}^4}{C_4 + C_5N_E + C_6J_E + C_7N_EJ_E} \qquad 12.6$$

where the fuel energy rate is modeled as a function of engine speed N_E and engine brake torque J_E according to the trend observed from engine test data, and the subscript 1 of G represents energy balance method 1.

The characteristic temperature (exhaust manifold gas temperature here) is affected primarily by engine speed, load, air–fuel ratio, EGR rate, and fuel injection timing. The measured exhaust manifold gas temperature is a function of engine speed and load. It is observed that the temperature is predominantly linearly proportional to the engine torque within a wide range with a weak dependency on engine speed.

When the engine torque J_E is kept constant, the numerator of equation 12.6 can be simplified as a constant. Therefore, the percentage of miscellaneous heat losses of energy balance method 1 becomes

$$G_{1,J} = \frac{C_8}{C_9 + C_{10} N_E} \tag{12.7}$$

where C_8, C_9 and C_{10} are coefficients. The parametric dependency of the percentage of miscellaneous heat losses upon engine speed shown in Fig. 12.4 is theoretically explained by equation 12.7.

If the engine speed is constant, the percentage of miscellaneous heat losses is

$$G_{1,N} = \frac{C_1 + C_2 T_{ch} + C_3 T_{ch}^4}{C_{11} + C_{12} J_E} \approx \frac{C_1 + C_2 T_{ch}}{C_{11} + C_{12} J_E} \tag{12.8}$$

where C_{11} and C_{12} are coefficients. Generally, the radiation term $C_3 T_{ch}^4$ in equation 12.8 is small. For simplicity, $C_3 = 0$. From engine test data it was observed that the engine brake torque increases much faster than the exhaust temperature. For example, when the engine brake torque increases sixfold from 100 ft.lb to 600 ft.lb, the exhaust manifold gas temperature becomes only approximately doubled from 450°F to 950°F. Therefore, as engine brake torque increases, the percentage of miscellaneous heat losses decreases.

12.2.3 Engine scaling for miscellaneous heat losses

The miscellaneous heat losses can be calculated by heat transfer equations once the input data are complete. When the engine block emissivity, the block area, the convection heat transfer coefficient, and the average block temperature are not available, a simplified analysis can be performed assuming that the ratio of the miscellaneous heat losses $(\dot{Q}_{miscellaneous})$ between two similar engines is proportional to $(V_{E1}/V_{E2})^{2/3}$, where V_{E1} and V_{E2} are the displacements of the two engines. In fact, $V_E^{2/3}$ is proportional to the characteristic heat transfer area associated with the miscellaneous heat losses. The similarity between the two engines refers to their respective configurations such as inline or Vee, EGR piping, and characteristic temperature in heat transfer losses.

The percentage of the miscellaneous heat losses can be approximated by

$$G_1 \propto \frac{K_{h1} T_{ch1} V_{E1}^{2/3}}{\left(\dfrac{J_{E1} N_{E1}}{\eta_{th1}}\right)} \propto \frac{K_{h1} T_{ch1} V_{E1}^{2/3} \eta_{th1}}{\varpi_{BMEP1} V_{E1} N_{E1}} = \frac{K_{h1} T_{ch1} \eta_{th1}}{\varpi_{BMEP1} V_{E1}^{1/3} N_{E1}} \tag{12.9}$$

where the subscript 1 represents engine 1, $K_{h1} T_1 V_{E1}^{2/3}$ is an approximate estimate of the miscellaneous heat losses, and $J_{E1} N_{E1}/\eta_{th1}$ is the fuel energy rate. The η_{th} is the engine brake thermal efficiency. Similarly, the percentage of miscellaneous heat losses of engine 2 can be approximated by

$$G_2 \propto \frac{K_{h2} T_{ch2} \eta_{th2}}{\varpi_{BMEP2} V_{E2}^{1/3} N_{E2}} \tag{12.10}$$

Therefore, the ratio of the percentages between engine 1 and engine 2 is

$$\frac{G_1}{G_2} = \frac{K_{h1}}{K_{h2}} \cdot \frac{T_{ch1}}{T_{ch2}} \cdot \frac{\eta_{th1}}{\eta_{th2}} \cdot \frac{\varpi_{BMEP2}}{\varpi_{BMEP1}} \cdot \frac{N_{E2}}{N_{E1}} \cdot \left(\frac{V_{E2}}{V_{E1}}\right)^{1/3} \qquad 12.11$$

When the two engines possess similarity, each of the first several terms at the right-hand side of equation 12.11 becomes 1. The following approximation is then obtained as a scaling rule:

$$\frac{G_1}{G_2} = \left(\frac{V_{E2}}{V_{E1}}\right)^{1/3} \qquad 12.12$$

12.3 Characteristics of base engine coolant heat rejection

12.3.1 Definition of base engine coolant heat rejection

Heavy-duty direct injection diesel engines usually have an oil cooler that uses the engine coolant as cooling medium. Consequently, the engine coolant heat rejection includes the heat removed by the oil cooler. Oil cooler heat rejection consists mainly of two parts: the piston cooling and the engine rubbing friction power loss. For system simulation, the total rubbing friction can be obtained by subtracting the calculated pumping loss and accessory power from the measured engine motoring power. The engine accessories here include the water pump, the oil pump, the fuel pump, the alternator, and the air compressor for the service brakes. A practical method for estimating the motoring power is the Willan's line (Heywood, 1988), where a plot of fuel flow rate vs. BMEP is used to extrapolate to zero fuel flow rate in order to determine the motoring mean effective pressure (MEP) or power.

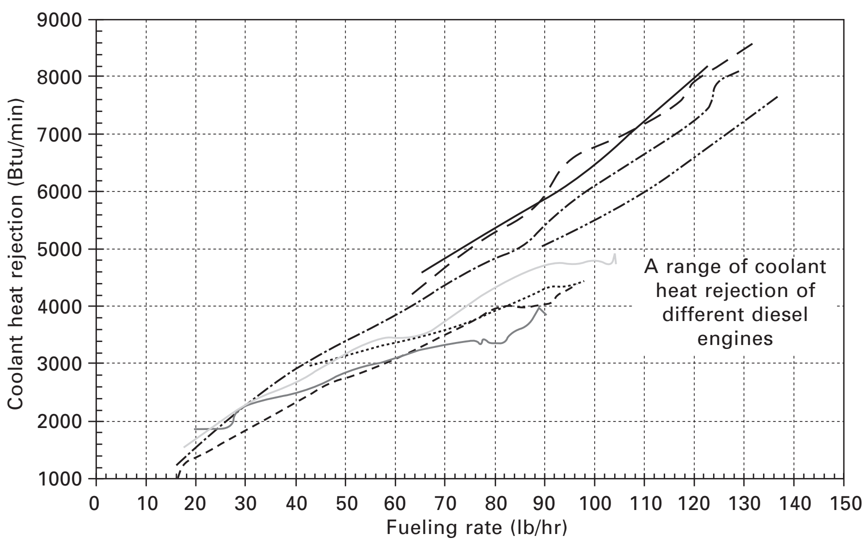
A more detailed global friction model (Taraza *et al*., 2000) of individual friction components (e.g., piston assembly, bearings, valvetrain) and auxiliaries is desirable in engine system simulation to enhance accuracy. The model is constructed with relatively simple physics governing the friction. The most important design parameters of the components are included in the global model. However, a more detailed friction model is needed for subsystem or component design analysis. A very complex tribological model of piston-assembly lubrication dynamics is shown in Chapter 10 as an example of an advanced analytical design tool used in mechanical subsystem design.

Exhaust port heat rejection is a part of the base engine coolant heat rejection. Its calculation is presented by Hires and Pochmara (1976), Rush (1976), and Malchow *et al*. (1979). Cylinder head cooling studies are reported by Norris *et al*. (1993). Piston cooling and heat rejection is reported by French (1972), Pimenta and Filho (1993), Mian (1997), Varghese *et al*. (2005), and Agarwal and Goyal (2007). Oil cooler design and testing information is provided by SAE J1244 (2008), J1468 (2006), J2414 (2005), Adams (1999) and Adams *et al*. (1999). Engine coolant heat rejection is discussed in Alkidas (1993), Imabeppu *et al*. (1993), Shayler *et al*. (1996) Franco and Martorano (1999), Campbell *et al*. (2000), and Luján *et al*. (2003). Heavy-duty diesel engine lubrication system design is introduced by Kluck *et al*. (1986).

Coolant heat rejection can be specified in the following three forms: (1) a heat transfer rate in kW or Btu/min (Fig. 12.6); (2) brake specific heat rejection in, for example, kW/hp; and (3) a percentage of the fuel energy that contributes to the coolant heat rejection. The percentage is usually more suitable to compare different engines as a competitive benchmarking parameter.

12.3.2 Low-heat-rejection engines

Coolant heat rejection is dominated by the large gas-to-wall thermal resistance in the thermal boundary layer of the combustion chamber, rather than by the smaller wall-to-coolant thermal resistance on the coolant side. Despite the fact that ceramics have a much lower thermal conductivity than metals (by one or two orders of magnitude), 'adiabatic' (or so-called low heat rejection or LHR) engines using ceramic materials as the cylinder wall insulation do not reduce heat rejection by an order of magnitude. The reason is that the thermal resistance of the gas side plays a major role in the overall thermal



12.6 Competitive engine data of base engine coolant heat rejection as a function of fuel flow rate.

resistance. Improving engine thermal efficiency through in-cylinder heat rejection reduction has been the key objective of LHR engines. It was found that there was a reduction in the ignition delay and the premixed combustion and an increase in the diffusion combustion duration in the LHR engines when compared to the conventionally cooled engines. An adverse aspect of LHR engines is that the much hotter cylinder wall temperature heats up the induction air and results in a reduction in volumetric efficiency and an increase in NO_x emissions. Overall, the findings and conclusions on LHR engines remain inconsistent and inconclusive. An improvement in LHR engines' fuel consumption has been reported in the range of 4–10% (Jaichandar and Tamilporai, 2003). The in-cylinder heat transfer characteristics of LHR engines and their tribological impact are very complex and still not fully understood by researchers.

Low heat rejection diesel engines have been extensively researched by Morel *et al*. (1986), Assanis (1989), Kobori *et al*. (1992), Sun *et al*. (1994), Yonushonis (1997). Kamo *et al*. (1987, 1996, 2000a, 2000b, 2003), Kamo (2000), Sharma and Gaur (1990), Reddy *et al*. (1990), Woods *et al*. (1992), Kimura *et al*. (1992), Schwarz *et al*. (1993), Bryzik *et al*. (1993), Jaichandar and Tamilporai (2003, 2004), Tamilporai *et al*. (2003), Hergart *et al*. (2005), Sutor *et al*. (2005), Saad *et al*. (2007), Rakopoulos and Giakoumis (2007), and Yasar (2008).

12.3.3 Sensitivity of base engine coolant heat rejection

The most critical parameter for engine cooling system design is heat rejection because it affects engine outlet coolant temperature, durability, and vehicle front-end cooling packaging. Once the heat rejection is known, the cooling medium flow rates of the radiator and the charge air cooler, the water pump power, and the cooling fan power can be calculated relatively easily.

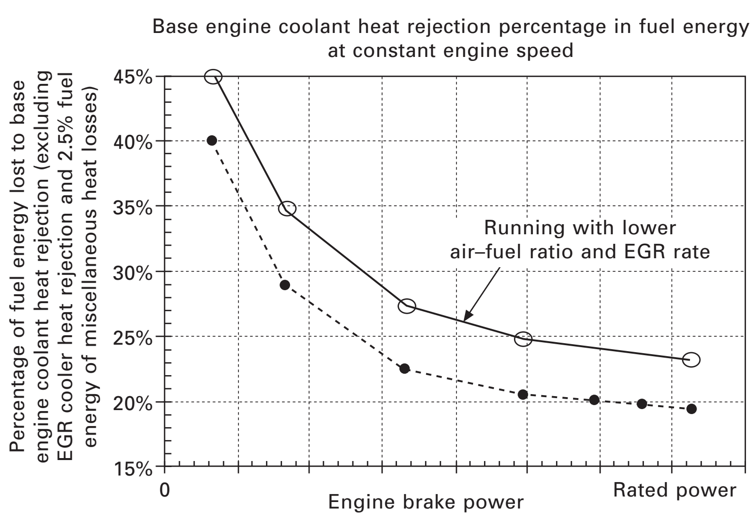
In addition to being characterized by the brake specific heat rejection, the base engine coolant heat rejection can be characterized as a 'percentage' of fuel energy rate. The base engine heat rejection 'percentage' is affected by cylinder heat transfer area, instantaneous heat transfer coefficient, and engine friction. The relevant design or operating parameters include the following: (1) cylinder bore diameter and stroke; (2) volume-to-surface ratio of the combustion chamber, and engine compression ratio; (3) cylinder liner, piston, and exhaust port design, especially the metal surface area exposed to heat transfer; (4) swirl ratio and in-cylinder turbulence; (5) the ratio of in-cylinder charge mass (air plus EGR) to fuel mass; (6) intake manifold gas temperature; (7) fuel injection timing; (8) engine speed and load; (9) mean piston speed; (10) water pump and oil pump power.

Note that the larger the in-cylinder charge-to-fuel ratio, the lower the base engine heat rejection percentage. Moreover, the more retarded fuel injection timing, the lower the percentage but the higher the BSFC (Fig. 12.1). The coolant temperature, the coolant-side convection heat transfer coefficient, and the cylinder head material (cast iron vs. aluminum alloy) have relatively small impacts on coolant heat rejection. Good engine performance and low heat rejection design require a small percentage of fuel energy lost to the base engine coolant heat rejection. Figure 12.7 was obtained by engine cycle simulations and it shows that the base engine heat rejection percentage increases when the engine load decreases. Figure 12.8 shows that directionally in-cylinder swirl reduction results in lower gas-side heat transfer coefficient and hence lower base engine coolant heat rejection and lower BSFC. Figure 12.9 shows that the exhaust port length has an important impact on the port heat rejection. Figure 12.10 shows the sensitivity of the base engine coolant heat rejection with respect to engine power, EGR rate, air–fuel ratio, and turbocharger efficiency, .

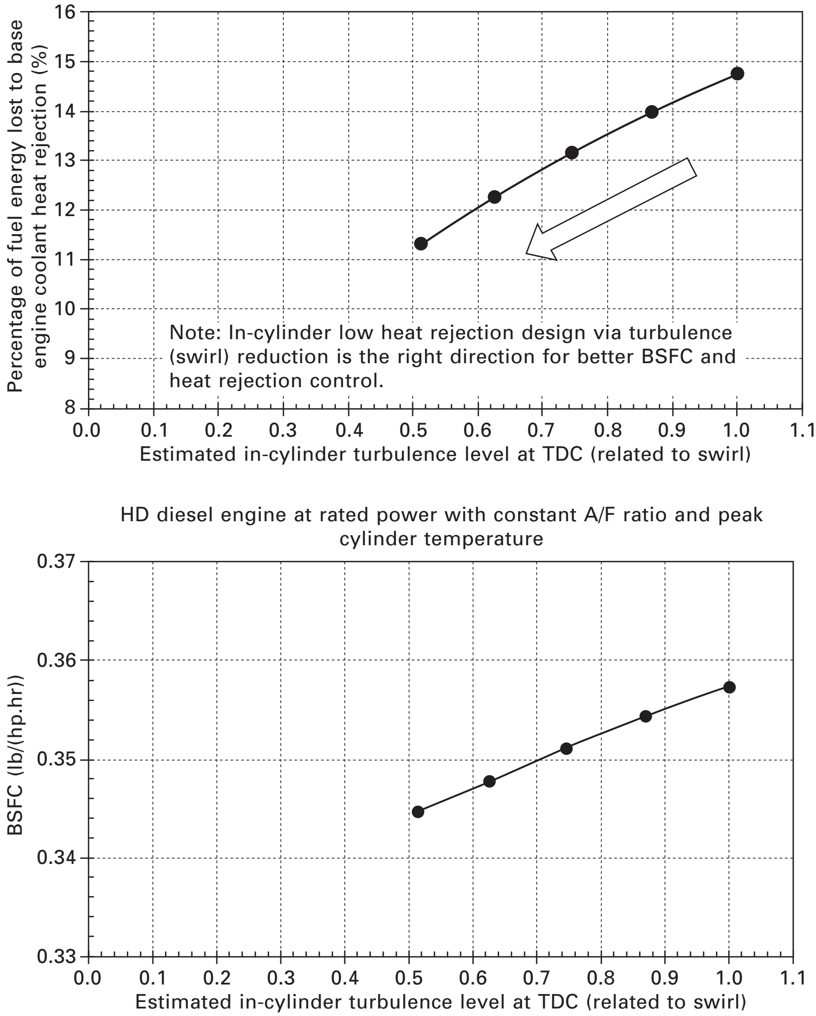
12.4 Cooling system design calculations

12.4.1 Cooling system design considerations

Cooling system performance is important for engine performance. The purpose of cooling is to maintain the engine component metal temperature and the temperature gradient at appropriate levels (e.g., about 380°C for cast iron and a lower temperature for aluminum alloys, 180–200°C for the top piston-ring groove to prevent the deterioration of the lubricating oil film). Undercooled engines may suffer from the following problems: loss of material

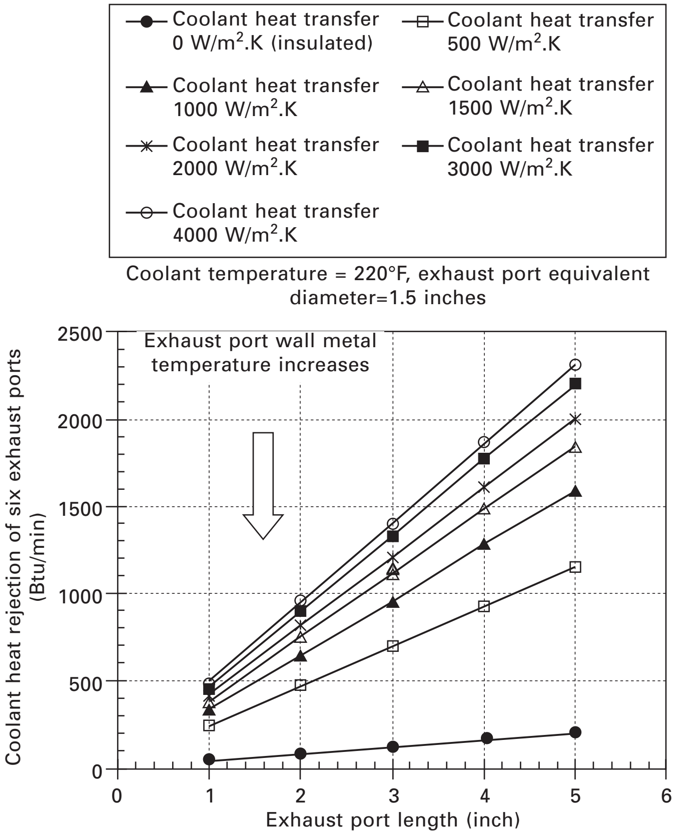


12.7 Characteristics of base engine coolant heat rejection.



12.8 Impact of in-cylinder turbulence on base engine coolant heat rejection.

strength, high thermal strain, lubricant oil degradation, overheating and scuffing of the power cylinder components, heating the intake air, and lower volumetric efficiency. On the other hand, overcooled engines may exhibit poor combustion and fuel economy, excessive heat rejection, high wear at the piston rings and the liner, and high friction and noise. An excessively low cylinder liner temperature, even below the condensation temperature of the burnt gases, may cause liner corrosion. The maximum cooling capacity of heavy-duty diesel engines is usually designed based on an in-vehicle condition running at the rated power in summer at high altitude with air-

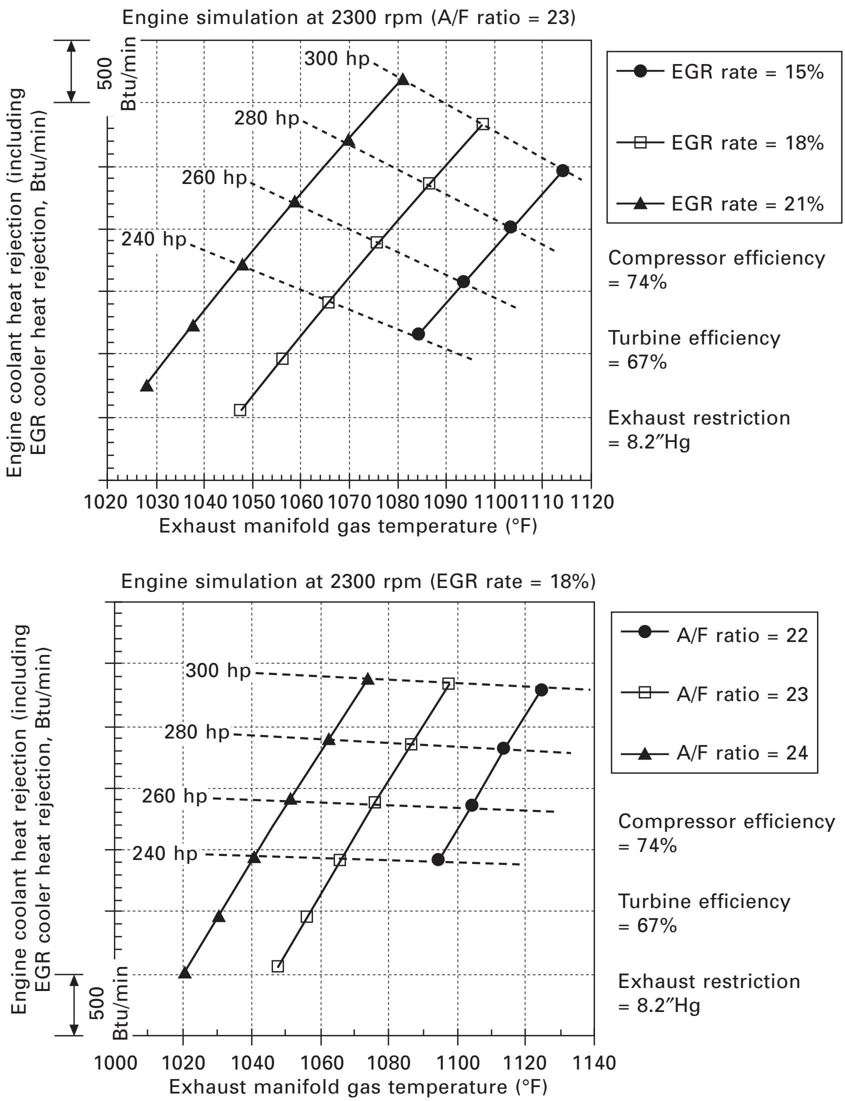


12.9 Effect of exhaust port length on engine coolant heat rejection at rated power.

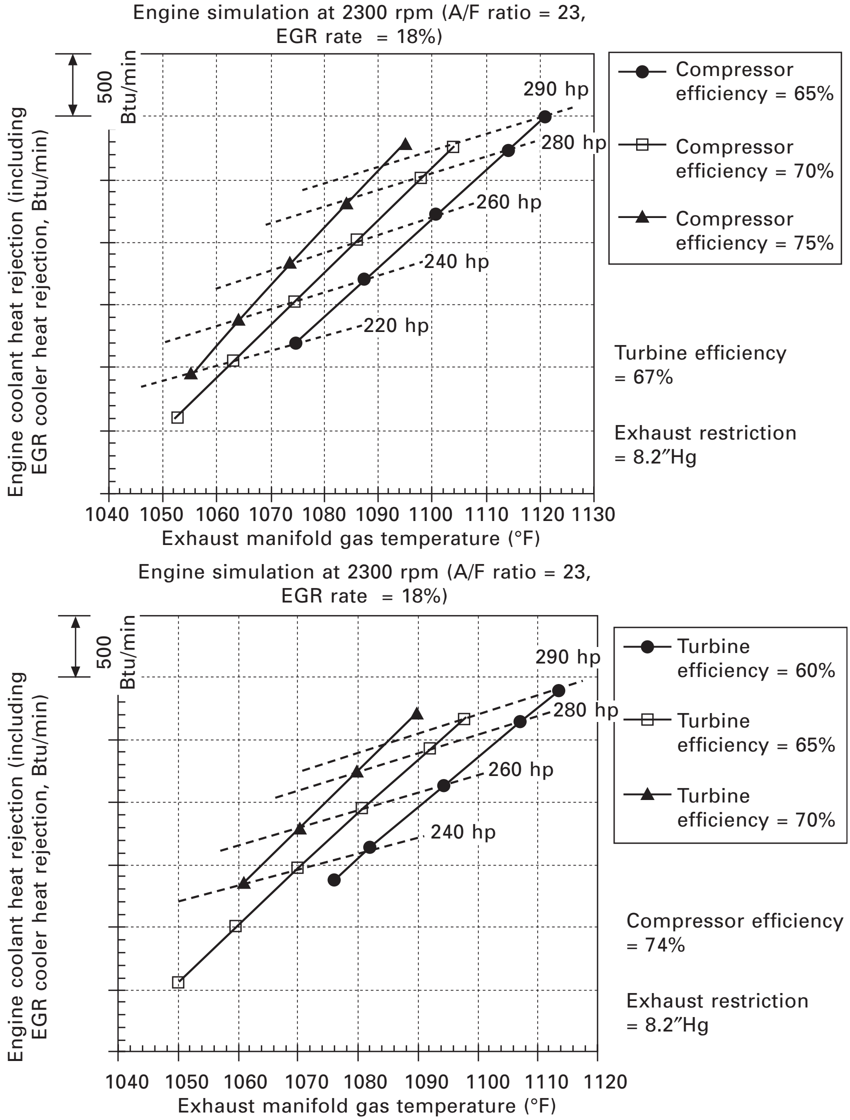
conditioning switched on. At part load or in a cold climate, the engine coolant can be excessively cold and can adversely affect engine combustion and performance. Moreover, overcooling in cooler design for certain operating conditions requires additional measures to fix the problem. For example, if the charge air cooler outlet air temperature is lower than the dew temperature, water is condensed in the cooler and may require a device to separate the condensate. Another example is the fouling and corrosion problems of the hydrocarbons and the acidic vapor condensate in the EGR cooler when the EGR cooler outlet gas temperature is too low at part load.

To reach a target intake manifold gas temperature for emissions control, there is a balance between the charge air cooler size (or effectiveness) and the EGR cooler size within their allowable packaging constraints. Overcooling the intake manifold mixture sometimes may be detrimental to NO_x control. The increase in the ignition delay with cold charge is accompanied by an increase in the premixed portion of the fuel injection and NO_x emissions.

EGR cooler heat transfer performance and flow restriction deteriorate significantly when soot and sulfate particulates deposit on the cooling tube surface. In the design specification of cooler sizing, a sufficient margin of effectiveness and pressure drop should be reserved to account for EGR cooler fouling. The explanation of EGR cooler fouling and the related references are



12.10 Parametric simulation of design effects on engine coolant heat rejection.



12.10 Continued

provided in the durability discussions in Section 2.12. EGR cooler design and cooling performance are elaborated by Ap (2000), Stolz *et al.* (2001), Chalgren *et al.* (2002), Melgar *et al.* (2004), Honma *et al.* (2004), and Mosburger *et al.* (2008). Air-cooled charge air cooler is widely used in diesel vehicles although sometimes water-cooled charge air coolers are used (Kern and Wallner, 1986; Chalgren *et al.*, 2004). Other vehicle cooling system design issues or components (e.g., underhood thermal management, radiator, cooling fan) are introduced by Habchi *et al.* (1994), Avequin *et al.* (2001), Muto *et al.* (2001), Chang *et al.* (2003), Shah (2003), and Scott and McDonald (2005).

In coolant selection, forced convection or sub-cooled boiling should be maintained, and saturated boiling should be avoided. With saturated boiling, there is a risk of vapor blanket and film boiling (see Section 2.12.2) which can cause insulation overheating and thermal fatigue of the cooler component. More information on engine coolant is provided by SAE J814 (2007), Hannigan (1993), and Brosel (1999).

In cooling system design, various configurations with different cooling media and cooler arrangements may be proposed in order to control the heat rejection under the vehicle front-end packaging constraints while achieving the same target of intake manifold gas temperature. The hottest engine outlet coolant temperature usually occurs at an intermediate speed between the peak torque and the rated power due to the nature of the matching between the engine heat rejection and the operating characteristics of the radiator and the water pump. Engine system analysis needs to be conducted under the harshest environmental conditions in order to cover the worst application scenarios of the highest engine-outlet coolant temperature. The extreme conditions may include sea-level altitude at 38°C or 100°F hot ambient, 1676 m or 5500 feet altitude at 38°C, or 10,000 feet on a hot day. The elevated compressor inlet air temperature due to in-vehicle underhood heating should be considered. Moreover, the effect of using air conditioning and hot air recirculation around the radiator and the charge air cooler should not be ignored when analyzing the cooling medium (sink) temperature.

SAE J631 (2007), J1004 (2004), J1148 (2004), J1393 (2004), J2082 (1992), J1994 (2008), and J1339 (2009) provide fundamental information on engine cooling systems.

Fundamentals of engine cooling system design are introduced by Bosch and Real (1990) and Kanefsky *et al.* (1999). Engine cooling system performance sensitivity is analyzed by Rahman and Sun (2003), Rahman *et al.* (2001), and Savage *et al.* (2007). An analysis of radiator top tank temperature has been conducted by Tang *et al.* (2008). Engine cooling system design and optimization are reported by Valaszkai and Jouannet (2000), Koch and Haubner (2000), Pantow *et al.* (2001), Hughes and Wiseman (2001), Chalgren and Allen (2005), and Burke *et al.* (2010). Cooling system simulations have been carried out by Sakai *et al.* (1994), Mohan *et al.* (1997), Ap (1999), and Hughes *et al.* (2001). A simulation methodology of EGR and engine cooling systems and their controls is given by Chalgren *et al.* (2002).

12.4.2 Cooler performance calculations

The cooler performance design specification includes effectiveness, heat rejection, and pressure losses at both the gas side and the cooling medium side. Cooler thermal performance calculations are basically based on two formulas: (1) the definition of cooler effectiveness and (2) the equation for the heat transfer rate. Taking EGR cooler as an example, its heat transfer calculation is illustrated below. The cooler effectiveness is defined as:

$$\begin{aligned}\varepsilon_{cooler} &= \frac{\text{Actual heat transfer rate}}{\text{Max possible heat transfer rate}}\\ &= \frac{\dot{m}_{gas}c_{p,gas}(T_{gas,in} - T_{gas,\,out})}{(\dot{m}c_p)_{min}(T_{gas,in} - T_{coolant,in})}\\ &= \frac{(\dot{m}_{coolant}\, c_{p,coolant}(T_{coolant,out} - T_{coolant,in})}{(\dot{m}c_p)_{min}(T_{gas,in} - T_{coolant,in})}\\ &= \frac{T_{gas,in} - T_{gas,\,out}}{T_{gas,in} - T_{coolant,in}}\end{aligned} \qquad 12.13$$

The cooler heat rejection is expressed as

$$\dot{Q}_{cooler} = K_h \cdot A_h \cdot \Delta T_{mean} \qquad 12.14$$

where K_h is the overall heat transfer coefficient, A_h is the heat transfer area, and the mean temperature difference of the heat exchanger is defined as

$$\Delta T_{mean} = \frac{(T_{gas,in} - T_{coolant,in}) - (T_{gas,out} - T_{coolant,out})}{\ln\left(\dfrac{T_{gas,in} - T_{coolant,in}}{T_{gas,out} - T_{coolant,out}}\right)}$$

The cooler heat transfer rate can also be expressed as

$$\begin{aligned}\dot{Q}_{cooler} &= \dot{m}_{gas}\, c_{p,gas}(T_{gas,in} - T_{gas,out})\\ &= \dot{m}_{coolant}\, c_{p,coolant}\,(T_{coolant,out} - T_{coolant,in})\end{aligned} \qquad 12.15$$

For parallel-flow cooling, the cooler effectiveness can be calculated by

$$\varepsilon_{cooler} = \frac{1 - \exp\left[\dfrac{-K_h A_h}{\dot{m}_{gas}c_{p,gas}}\left(1 + \dfrac{\dot{m}_{gas}c_{p,gas}}{\dot{m}_{coolant}\, c_{p,coolant}}\right)\right]}{1 + \dfrac{\dot{m}_{gas}c_{p,gas}}{\dot{m}_{coolant}\, c_{p,coolant}}} = \frac{1 - \mathrm{e}^{-f_{NTU}(1+\tau_c)}}{1 + \tau_c} \qquad 12.16$$

For counter-flow cooling, the cooler effectiveness can be calculated by

$$\varepsilon_{cooler} = \frac{1 - \mathrm{e}^{-f_{NTU}(1-\tau_c)}}{1 - \tau_c \cdot \mathrm{e}^{-f_{NTU}(1-\tau_c)}} \qquad 12.17$$

where the number of transfer units (NTU) is defined as

$$f_{NTU} = \frac{K_h A_h}{\dot{m}_{gas} c_{p,gas}} \qquad 12.18$$

and

$$\tau_c = \frac{\dot{m}_{gas} c_{p,gas}}{\dot{m}_{coolant} c_{p,coolant}} \qquad 12.19$$

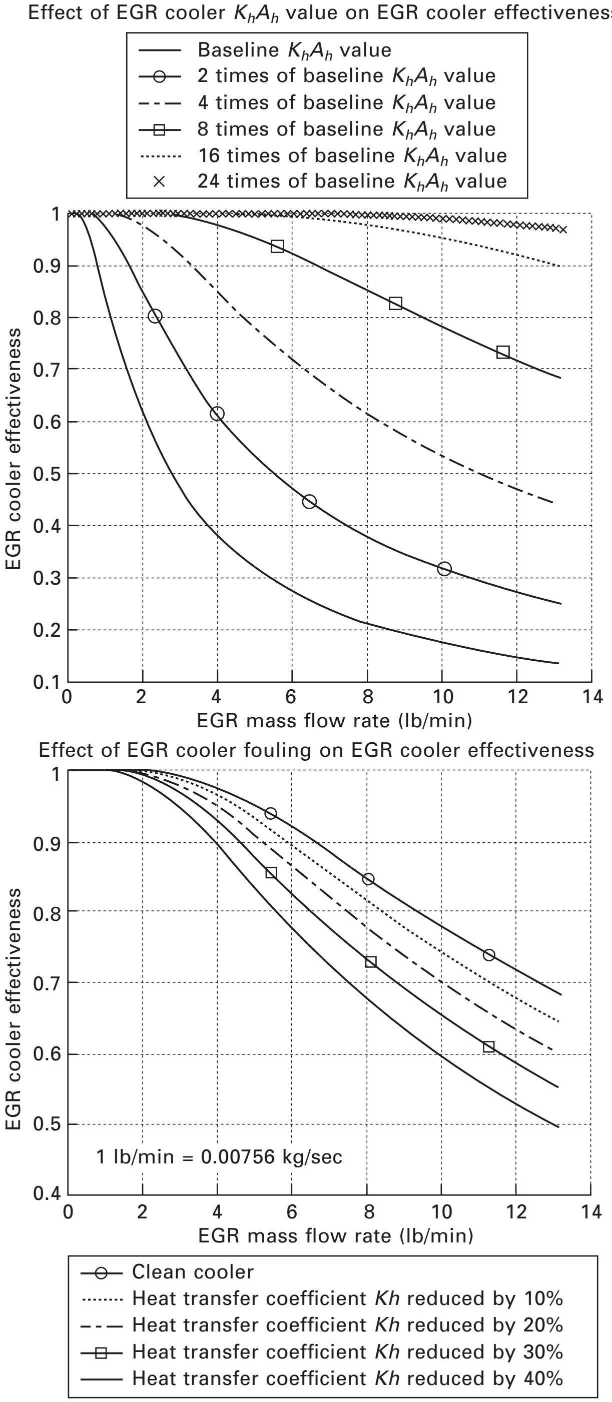
For the EGR cooler, the value of τ_c is very small (around 0.02). From equation 12.16 or 12.17, it is observed that the EGR cooler effectiveness is a function of EGR gas mass flow rate and cooler heat transfer capacity $K_h A_h$. Note that cooler effectiveness does not depend on the cooler cooling medium temperature. Figure 12.11 shows the calculated characteristics of EGR cooler effectiveness. The steady-state performance of the charge air cooler and vehicle radiator can be calculated in a similar way. In engine cycle simulations, the engine coolant temperature and the cooler cooling medium temperatures can be calculated by the above equations to couple with the computation of engine coolant heat rejection.

As to the transient performance of the coolers, Pearson *et al.* (2000) provided a theoretical analysis on the wave-action model of a charge air cooler's boundary, and a methodology to predict the instantaneous effectiveness and heat transfer coefficient of the cooler as a function of air mass flow rate.

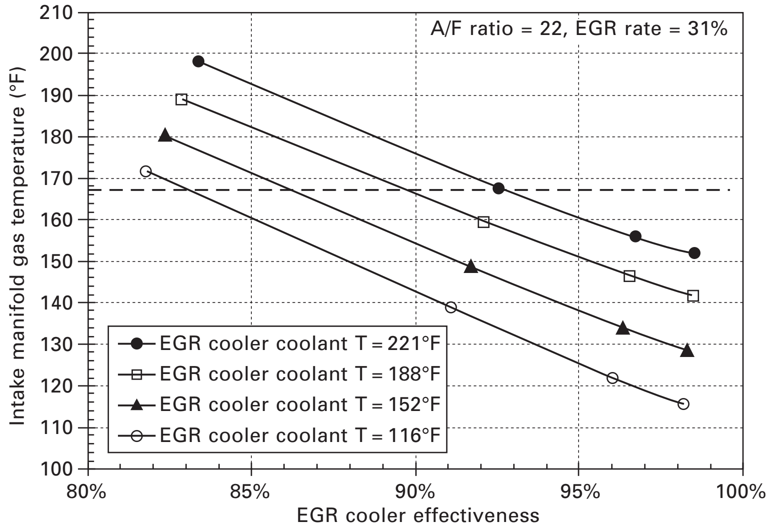
12.4.3 Cooler sink temperature and cooling capability

The calculation of cooler sink temperature is important for engine system design because it affects cooling system design and the predictions of heat rejection and engine air or gas temperatures at the cooler outlet. The sink temperature refers to the cooling medium inlet temperature. For the charge air cooler or the radiator, the sink temperature is the cooling air temperature in front of the cooler or the radiator. For the EGR cooler or the liquid-cooled inter-stage cooler, the sink temperature is the coolant inlet temperature of the cooler. For the engine cylinder, the sink temperature is the radiator outlet coolant temperature. It is obvious that the sink temperature is dependent on the vehicle cooling system configuration (e.g., whether the charge air cooler is placed in front of or behind the radiator). Figure 12.12 shows the effect of EGR cooler sink temperature on EGR cooler sizing. It is observed from this example that to reach the same intake manifold temperature of 167°F, if the EGR cooler sink temperature increases by 10°F, the EGR cooler effectiveness has to increase by 1% in order to maintain the same intake manifold gas temperature.

One important parameter in engine system design is the radiator inlet coolant temperature (i.e., the top tank temperature). Applying equations



12.11 Cooler effectiveness characteristics.



12.12 Relationship between EGR cooler heat transfer capacity (effectiveness) and cooler coolant temperature (simulation at rated power 2300 rpm 300 hp).

12.13 and 12.15 to the radiator, the following equation can be derived to calculate this temperature:

$$T_{RAD,InletCoolant} = \frac{\dot{Q}_{RAD,hrj}}{\varepsilon_{RAD}(\dot{m}c_p)_{min}} + T_{sink,RAD} \qquad 12.20$$

If the radiator is placed behind a charge air cooler (as is usually the case), $T_{sink,RAD}$ is equal to the hot cooling air temperature at the outlet of the charge air cooler, and it can be calculated by

$$T_{CAC,CoolingAirOut} = \frac{\dot{Q}_{CAC,hrj}}{\dot{m}_{CoolingAir}c_{p,air}} + T_{CAC,CoolingAirIn} \qquad 12.21$$

where $T_{CAC,CoolingAirIn}$ is the sink temperature of the charge air cooler and

$$T_{CAC,CoolingAirIn} = T_{AMB} + \Delta T_{recir} \qquad 12.22$$

ΔT_{recir} is a temperature rise due to the recirculation effect in front of the charge air cooler. Note that $\dot{m}_{CoolingAir}$ is dependent upon vehicle speed, fan speed, and altitude level (due to ambient air density change). The radiator outlet coolant temperature can be calculated after the radiator inlet coolant temperature is obtained, provided the radiator heat rejection and the coolant flow rate are known.

Denote $(\dot{m}c_p)_{min} = (\dot{m}_{cal}c_p)_{min} f_{cm}$, where $\dot{m}_{cal}$ is the cooling or cooled (whichever is minimum) flow rate at an engine speed used for model calibration/tuning, and f_{cm} is a cooling capability multiplier for the flow rate (to be detailed later). Also denote $\varepsilon_{RAD} = \varepsilon_{RAD,cal} f_{c\varepsilon}$, where $\varepsilon_{RAD,cal}$ is the radiator effectiveness at the model calibration condition, and $f_{c\varepsilon}$ is a cooling capability multiplier for the effectiveness. Moreover, define the cooling capability of the radiator as

$$\Phi_{RAD} = \varepsilon_{RAD} (\dot{m}c_p)_{min} = \varepsilon_{RAD,cal}\, f_{c\varepsilon,RAD} \cdot (\dot{m}_{cal} c_p)_{min} f_{cm,RAD} \qquad 12.23$$

Substituting 12.23 into 12.20, the following are obtained:

$$\Phi_{RAD} = \frac{\dot{Q}_{RAD,hrj}}{T_{RAD,InletCoolant} - T_{sink,RAD}} \qquad 12.24$$

$$T_{RAD,InletCoolant} = \frac{\dot{Q}_{RAD,hrj}}{\Phi_{RAD}} + T_{sink,RAD} \qquad 12.25$$

$$T_{sink,RAD} = \frac{\dot{Q}_{CAC,hrj}}{(\dot{m}_{cal} c_p)_{min}\, f_{cm,CAC}} + T_{AMB} + \Delta T_{recir} \qquad 12.26$$

Equation 12.24 is used to calibrate the model for cooling capability, for example at one operating condition with a cooling air flow rate $\dot{m}_{cal}$ and effectiveness ε_{cal}. At other operating conditions, the cooling air flow rate changes to $\dot{m}_{cal} f_{cm}$, and the effectiveness changes to $\varepsilon_{cal} f_{c\varepsilon}$. The cooling capability Φ_{RAD} changes accordingly to $\Phi_{RAD,cal} f_{c\varepsilon} f_{cm}$. The multiplier $f_{c\varepsilon} f_{cm}$ converts the cooling capability from that of the model calibration condition to the capability at other engine speed or load mode, or other ambient conditions such as high altitude. Therefore, if $f_{c\varepsilon}$ and f_{cm} are known, the radiator coolant temperature at any other operating condition can be calculated using equations 12.25 and 12.26.

Define $f_{c\varepsilon} \triangleq f_{c\varepsilon 1}\, f_{c\varepsilon 2}\, f_{c\varepsilon 3}$ and $f_{cm} \triangleq f_{cm1}\, f_{cm2}\, f_{cm3}\, f_{cm4}$. $f_{c\varepsilon 1}$ is an altitude multiplier for cooler effectiveness. At sea-level altitude, $f_{c\varepsilon 1} \triangleq 1$. At high altitude,

$$f_{c\varepsilon 1} = 1 + \frac{\rho_{air,altitude} - \rho_{air,SeaLevel}}{\rho_{air,SeaLevel}} \cdot x\% \qquad 12.27$$

The air density is calculated by

$$\rho_{air} = \frac{p_{AMB}}{287 \cdot T_{air}} \qquad 12.28$$

where ρ_{air} is in kg/m^3, p_{AMB} is in Pa, and T_{air} is in K. The model 12.27 is based on an assumption that every 1% of the radiator effectiveness reduction corresponds to $x\%$ of the reduction in cooling air mass flow rate or air density. For example, if 35% of effectiveness reduction at high altitude corresponds to 75% reduction in radiator's cooling air mass flow rate ratio, this means on average every 1% of effectiveness reduction corresponds to 2.14% reduction in cooling air flow rate or density, i.e., $x = 2.14$. This example gives roughly 3.6°F increase in the top tank temperature for every 1000 feet altitude increase.

The parameter $f_{c\varepsilon 2}$ is a multiplier for radiator or CAC cooling air flow uniformity and mal-distribution. At the model calibration condition, $f_{c\varepsilon 2} \triangleq 1$. The parameter $f_{c\varepsilon 3}$ is a multiplier to account for the change in effectiveness due to the variation in $(\dot{m}c_p)_{min}$. As shown in Fig. 12.13, $f_{c\varepsilon 3}$ can be modeled as

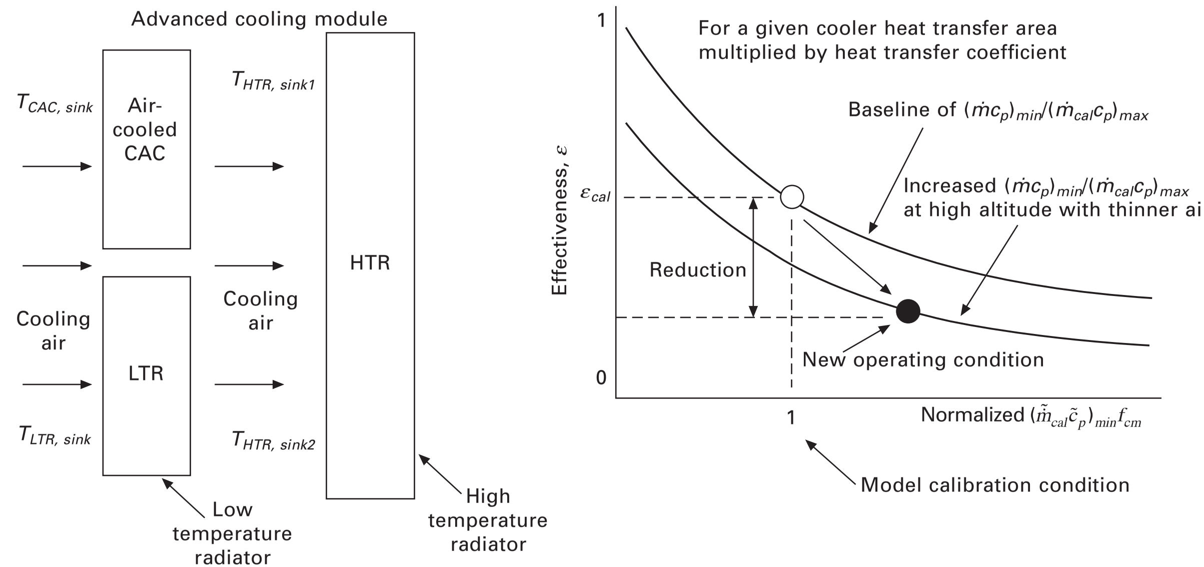
$$f_{c\varepsilon 3} = \frac{1 - (1 - \varepsilon_{cal})[(\tilde{\dot{m}}_{cal}\tilde{c}_p)_{min} f_{cm}]^n}{\varepsilon_{cal}} \qquad 12.29$$

where $\tilde{\dot{m}}_{cal}$ is the normalized cooling or cooled (whichever is minimum) mass flow rate at the model calibration condition. The parameter ε_{cal} is a known effectiveness term at the model calibration condition. The exponent n in equation 12.29 can be adjusted to change the shape of the effectiveness characteristic curve (e.g., $n = 0.5$).

The multiplier f_{cm1} is a cooling fan speed multiplier for the cooling air flow rate and is proportional to engine speed. The parameter f_{cm2} is a vehicle ram air speed multiplier and is proportional to vehicle speed. The parameter f_{cm3} is a water pump speed multiplier and is proportional to engine speed. The parameter f_{cm4} is a multiplier for vehicle underhood cooling air flow restriction and any model calibration uncertainty. At the model calibration condition, $f_{cm1} = f_{cm2} = f_{cm3} = f_{cm4} \triangleq 1$. If $(\dot{m}_{cal}c_p)_{min}$ refers to the coolant flow (for the radiator) or the charge air flow (for the air-cooled CAC), $f_{cm1} = f_{cm2} = f_{cm4} = 1$ because they do not apply in that case. For air-cooled CAC or if $(\dot{m}_{cal}c_p)_{min}$ refers to the cooling air flow rate for the radiator, $f_{cm3} = 1$ because it does not apply. It should be noted that when $(\dot{m}c_p)_{min}$ changes between cooling air flow and liquid coolant flow at different engine speeds or operating conditions, the equations 12.25 and 12.26 involving the term $(\dot{m}_{cal}c_p)_{min}\, f_{cm,RAD}$ cannot be used. In that case, the actual $(\dot{m}c_p)_{min}$ at each speed or operating condition must be used as in equation 12.20.

Equation 12.25 shows that the radiator coolant temperature depends directly on the cooling capability. In engine system design, a normal practice is to estimate the required change in the normalized radiator cooling capability (i.e., essentially effectiveness multiplied by cooling air flow rate) until the engine outlet coolant temperature reaches the design target under the predicted heat rejection. For example, a normalized cooling capability of 1.58 means 58% larger than the baseline cooling capability.

Engine simulation shows that although the coolant heat rejection reaches the highest value at rated power, the radiator inlet coolant temperature usually



12.13 Concepts of cooler capability and cooling sink temperature.

reaches the highest (worst) value at peak torque or at medium speed and full load. The coolant temperature is higher at hotter ambient temperatures, and higher at high altitude. Figure 12.14 shows a simulation example at the full-load lug curve.

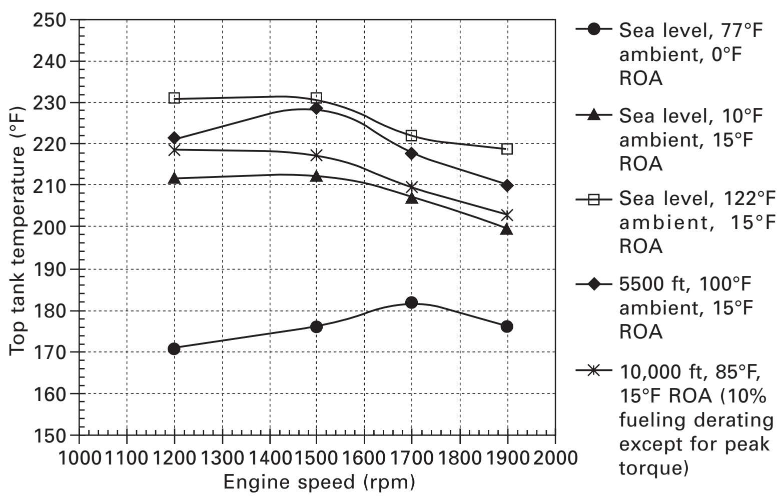
The methodology introduced in this section on cooler sink temperature and cooling capability is not only important for predicting the radiator inlet temperature in cooling system design and heat rejection control, but also necessary to close the loop in the prediction of engine coolant heat rejection by iterating the cooler sink temperature in the engine cycle simulation model until convergence.

12.5 Engine warm-up analysis

Engine warm-up is a typical transient process that affects emissions, noise, and friction. Although warm-up is not a typical condition used for diesel engine system design, understanding its transient heat rejection is helpful for engine modeling. For completeness some references on warm-up investigation are included here (Shayler *et al*., 1993, 1997; Jarrier *et al*., 2000, 2002; Samhaber *et al*., 2001).

12.6 Waste heat recovery and availability analysis

Waste heat recovery (WHR) is an important area for modern diesel engines. EGR heat recovery and turbocompounding are the two primary WHR technologies. In high-EGR engines, although the coolant heat rejection is



12.14 HD diesel engine coolant temperature simulation at full load.

high, the availability of the engine coolant is rather low due to its relatively low temperature. A large amount of exhaust gas flows to the EGR circuit instead of going through the turbine. Therefore, the exhaust availability at the turbine outlet in high-EGR engines is lower than that in the non-EGR engines, thus the importance of turbocompounding is not so critical. Instead, the highest potential for WHR is the EGR heat. Organic Rankine cycle has been attempted to recover the EGR heat to convert part of it to mechanical shaft power for the engine in order to both reduce coolant heat rejection and improve fuel economy. More detailed information about WHR can be found in DiBella *et al*. (1983), Diehl *et al*. (2001), Crane *et al*. (2001), Kapich (2002), Li and Figliola (2004), Chammas and Clodic (2005), Arias *et al*. (2006), Stobart and Weerasinghe (2006), Hountalas *et al*. (2007), and Teng *et al*. (2006, 2007a, 2007b).

12.7 References and bibliography

Adams M (1999), 'New developments in oil/coolant heat exchangers for heavy duty engine and transmission applications', SAE paper 1999-01-2823.

Adams M, Banzhaf M and Weiss E G (1999), 'First-time use of aluminum for engine oil coolers in heavy commercial vehicles', SAE paper 1999-01-0235.

Agarwal A K and Goyal S K (2007), 'Experimental and numerical investigations of jet impingement cooling of piston of heavy-duty diesel engine for controlling the non-tail pipe emissions', SAE paper 2007-01-0763.

Alkidas A C (1993), 'Effects of operating parameters on structural temperatures and coolant heat rejection of a SI engine', SAE paper 931124.

Ap N S (1999), 'A simple engine cooling system simulation model', SAE paper 1999-01-0237.

Ap N S (2000), 'Exhaust heat exchange coefficient in a pipe of an internal combustion engine: EGR cooler and passenger compartment heating applications', SAE paper 2000-01-0966.

Arias D A, Shedd T A and Jester R K (2006), 'Theoretical analysis of waste heat recovery from an internal combustion engine in a hybrid vehicle', SAE paper 2006-01-1605.

Assanis D N (1989), 'Effect of combustion chamber insulation on the performance of a low heat rejection diesel engine with exhaust heat recovery', *Heat Recovery Systems and CHP*, 9(5), 475–484.

ASTM American National Standard (reapproved 2005), 'Standard test method for estimation of net and gross heat of combustion of burner and diesel fuels', D4868-00.

Avequin S, Potier M, Mahe C, Gille G and Martins C (2001), 'Engine cooling multi-exchanger', SAE paper 2001-01-1749.

Bosch D J and Real J D (1990), 'Heavy truck cooling systems', SAE paper 900001.

Bowman J P, Krishnan S S and Nalim R (2005), 'Cooling challenges of modern truck diesel engines', *Proceedings of ICES2005, ASME Internal Combustion Engine Division 2005 Spring Technical Conference*, April 5–7, Chicago, IL, pp. 439–445.

Brosel M B (1999), 'Overview of engine coolant testing in Europe with particular regard to its development in Germany', *ASTM Special Technical Publication*, Issue 1335, pp. 392–408.

Bryzik W, Schwarz E, Kamo R and Woods M (1993), 'Low heat rejection from high output ceramic coated diesel engine and its impact on future design', SAE paper 931021.

Burke J and Haws J (2001), 'Vehicle thermal systems modeling using FLOWMASTER2', SAE paper 2001-01-1696.

Burke R D, Brace C J, Hawley J G and Pegg I (2010), 'Review of the systems analysis of interactions between the thermal, lubricant, and combustion processes of diesel engines', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 224, 681–704.

Campbell N A F, Hawley J G, Robinson K and Leathard M J (2000), 'Review of predictive analysis applied to IC engine coolant heat transfer', *Journal of the Institute of Energy*, 73(495), 78–86.

Chalgren R D and Allen D J (2005), 'Light duty diesel advanced thermal management', SAE paper 2005-01-2020.

Chalgren R D, Parker G G, Arici O and Johnson J (2002), 'A controlled EGR cooling system for heavy duty diesel applications using the vehicle engine cooling system simulation', SAE paper 2002-01-0076.

Chalgren R D, Barron L and Bjork D R (2004), 'A controllable water cooled charge air cooler (WCCAC) for diesel trucks', SAE paper 2004-01-2614.

Challen B and Baranescu R (eds) (1999), *Diesel Engine Reference Book*, second edition, SAE International, Warrendale, PA.

Chammas R E and Clodic D (2005), 'Combined cycle for hybrid vehicles', SAE paper 2005-01-1171.

Chang F-C, Malipeddi S, Uppuluri S and Shapiro S (2003), 'Underhood thermal management of off-highway machines using 1D-network simulations', SAE paper 2003-01-3405.

Crane D, Jackson G and Holloway D (2001), 'Towards optimization of automotive waste heat recovery using thermoelectrics', SAE paper 2001-01-1021.

Cui Y, Deng K and Wu J (2001), 'A direct injection diesel combustion model for use in transient condition analysis', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 215, 995–1004.

DiBella F A, DiNanno L R and Koplow M D (1983), 'Laboratory and on-highway testing of diesel organic Rankine compound long-haul vehicle engine', SAE paper 830122.

Diehl P, Haubner F, Klopstein S and Koch F (2001), 'Exhaust heat recovery system for modern cars', SAE paper 2001-01-1020.

Franco A and Martorano L (1999), 'Methods to evaluate in-cylinder heat transfer and thermal load in the small internal combustion engines', SAE paper 1999-01-1252.

French C C J (1972), 'Piston cooling', SAE paper 720024.

Habchi S D, Ho S Y, Elder J and Singh S (1994), 'Airflow and thermal analysis of underhood engine enclosures', SAE paper 940316.

Hannigan H J (1993), 'Review of automotive engine coolant technology', *ASTM Special Technical Publication*, Issue 1192, pp. 6–10.

Hergart C-A, Louki A and Peters N (2005), 'On the potential of low heat rejection DI diesel engines to reduce tail-pipe emissions', SAE paper 2005-01-0920.

Heywood J B (1988), *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.

Hires S D and Pochmara G L (1976), 'An analytical study of exhaust gas heat loss in a piston engine exhaust port', SAE paper 760767.

Honma J, Murao T, Yamashita Y, Tsujita M and Sugihara H (2004), 'Development of a highly efficient and reliable multi-tube EGR cooler', SAE paper 2004-01-1446.

Hountalas D T, Katsanos C O and Lamaris V T (2007), 'Recovering energy from the diesel engine exhaust using mechanical and electrical turbocompounding', SAE paper 2007-01-1563.

Hughes C A V and Wiseman M W (2001), 'Feasibility of intelligent control strategies to reduce cooling system size', SAE paper 2001-01-1759.

Hughes C, Mitts C, Jacquelin F and Wiseman M (2001), 'Heavy duty truck cooling system design using co-simulation', SAE paper 2001-01-1707.

Imabeppu S, Shimonosono H, Hirano Y, Fujigaya K and Inoue K (1993), 'Development of a method for predicting heat rejection to the engine coolant', SAE paper 931114.

Jaichandar S and Tamilporai P (2003), 'Low heat rejection engines – an overview', SAE paper 2003-01-0405.

Jaichandar S and Tamilporai P (2004), 'The status of experimental investigations on low heat rejection engines', SAE paper 2004-01-1453.

Jarrier L, Champoussin J C, Yu R and Gentile D (2000), 'Warm-up of a DI diesel engine: experiment and modeling', SAE paper 2000-01-0299.

Jarrier L, Gyan P and Champoussin J C (2002), 'Thermo-hydraulic oil loop modeling for ICE warm-up investigation', SAE paper 2002-01-2197.

Kamo L S (2000), 'Preliminary investigation of developmental low heat rejection engine piston tribology', *ASME Internal Combustion Engine Division (publication) ICE*, 33, 155–164.

Kamo L S, Kleyman A S, Bryzik W and Mekari M (1996), 'High temperature tribology for piston ring and cylinder liner in advanced low heat rejection engines', *ASME Internal Combustion Engine Division (publication) ICE*, 27, 89–96.

Kamo L, Woods M, Bryzik W and Mekari M (2000a), 'Thermal barrier coatings for monolithic ceramic low heat rejection diesel engine components', SAE paper 2000-01-1236.

Kamo L, Kamo R, Bryzik W, Mekari M, Jin Y S and Li S H (2000b), 'Advancements in high temperature cylinder liner and piston ring tribology', SAE paper 2000-01-1237.

Kamo L, Saad P, Mnatsakanov R, Bryzik W and Mekari M (2003), 'Development of high temperature diesel engine piston ring and cylinder liner tribology', SAE paper 2003-01-1104.

Kamo R, Bryzik W and Glance P (1987), 'Adiabatic engine trends – worldwide', SAE paper 870018.

Kanefsky P, Nelson V A and Ranger M (1999), *A Systems Engineering Approach to Engine Cooling Design*, the 44th L. Ray Buckendale Lecture, SP-1541, SAE paper 1999-01-3780, SAE International, Warrendale, PA.

Kapich D (2002), 'Turbo-hydraulic engine exhaust power recovery system', SAE paper 2002-01-2731.

Kern J and Wallner R (1986), 'Water-cooled charge air coolers for heavy diesel engines', IMechE paper C116/86.

Kimura S, Matsui Y and Itoh T (1992), 'Effects of combustion chamber insulation on the heat rejection and thermal efficiency of diesel engines', SAE paper 920543.

Kluck C E, Olsen P W and Skriba S W (1986), 'Lubrication system design considerations for heavy-duty diesel engines', SAE paper 861224.

Kobori S, Kamimoto T and Toralde M (1992), 'Combustion in low-heat-rejection diesel engines', *JSME International Journal*, *Series* 2: *Fluids Engineering, Heat Transfer, Power, Combustion, Thermophysical Properties*, 35(1), 1–9.

Koch F W and Haubner F G (2000), 'Cooling system development and optimization for DI engines', SAE paper 2000-01-0283.

Li H and Figliola R S (2004), ‘Optimization of an automotive cooling system based on exergy analysis’, SAE paper 2004-01-3541.

Luján J M, Serrano J R, Arnau F and Dolz V (2003), ‘Heat transfer model to calculate turbocharged HSDI diesel engines performance’, SAE paper 2003-01-1066.

Malchow G L, Sorenson S C and Buckius R O (1979), ‘Heat transfer in the straight section of an exhaust port of a spark ignition engine’, SAE paper 790309.

Melgar A, Horrillo A, Cabaco G, Castaño C and Grande J A (2004), ‘Theoretical and experimental methodology for the improvement of EGR-coolers design’, SAE paper 2004-01-0052.

Mian M A (1997), ‘Design and analysis of engine lubrication systems’, SAE paper 970637.

Mohan K V, Arici O, Yang S L and Johnson J H (1997), ‘A computer simulation of the turbocharged diesel engine as an enhancement of the vehicle engine cooling system simulation’, SAE paper 971804.

Morel T, Keribar R, Blumberg P N and Fort E F (1986), ‘Examination of key issues in low heat rejection engines’, SAE paper 860316.

Mosburger M, Fuschetto J, Assanis D, Filipi Z and McKee H (2008), ‘Impact of high sulfur military JP-8 fuel on heavy duty diesel engine EGR cooler condensate’, SAE paper 2008-01-1081.

Muto S, Sugimoto T, Utikawa A and Yamamoto M (2001), ‘Development of a cooling module containing a radiator and a condenser, part 1: product design’, SAE paper 2001-01-1018.

Norris P M, Wepfer W, Hoag K L and Courtine-White D (1993), ‘Experimental and analytical studies of cylinder head cooling’, SAE paper 931122.

Pantow E, Kern J, Banzhaf M, Lutz R and Tillmann A (2001), ‘Impact of US02 and Euro4 emission legislation on power train cooling challenges and solutions for heavy duty trucks’, SAE paper 2001-01-1716.

Pearson R J, Bassett M D, Winterbone D E and Bromnick P A (2000), ‘Comprehensive charge-cooler model for simulating gas dynamics in engine manifolds’, SAE paper 2000-01-1264.

Pimenta M M and Filho R (1993), ‘Cooling of automotive pistons: study of liquid-cooling jets’, SAE paper 931622.

Rahman S and Sun R (2003), ‘Robust engineering of engine cooling system’, SAE paper 2003-01-0149.

Rahman S, Kayupov M, Li J and Mourelatos Z P (2007), ‘Balance between reliability and robustness in engine cooling system optimal design’, SAE paper 2007-01-0594.

Rakopoulos C D and Giakoumis E G (2007), ‘Study of the transient operation of low heat rejection turbocharged diesel engine including wall temperature oscillations’, SAE paper 2007-01-1091.

Reddy C S, Domingo N and Graves R L (1990), ‘Low heat rejection engine research status: where do we go from here?’, SAE paper 900620.

Rush J H (1976), ‘Exhaust port heat rejection in a piston engine, a preliminary report’, SAE paper 760766.

Saad D, Saad P, Kamo L, Mekari M, Bryzik W, Schwarz E and Tasdemir J (2007), ‘Thermal barrier coatings for high output turbocharged diesel engine’, SAE paper 2007-01-1442.

SAE Surface Vehicle Standard J631 (2007), ‘Radiator Nomenclature’.

SAE Surface Vehicle Information Report J814 (2007), ‘Engine coolants’.

SAE Surface Vehicle Information Report J1004 (2004), ‘Glossary of engine cooling system terms’.

SAE Surface Vehicle Recommended Practice J1148 (2004), '(R) Engine charge air cooler (CAC) nomenclature'.
SAE Surface Vehicle Recommended Practice J1244 (2008), '(R) Oil cooler nomenclature and glossary'.
SAE Surface Vehicle Recommended Practice J1339 (2009), 'Test method for measuring performance of engine cooling fans'.
SAE Surface Vehicle Recommended Practice J1393 (2004), '(R) Heavy duty vehicle cooling test code R'.
SAE Surface Vehicle Recommended Practice J1468 (2006), 'Application testing of oil-to-air oil coolers for heat transfer performance'.
SAE Surface Vehicle Information Report J1498 (2005), 'Heating value of fuels'.
SAE Surface Vehicle Recommended Practice J1994 (2008), 'Laboratory testing of vehicle and industrial heat exchangers for heat transfer and pressure drop performance'.
SAE Surface Vehicle Information Report J2082 (1992), 'Cooling flow measurement techniques'.
SAE Surface Vehicle Standard J2414 (2005), 'Application testing of oil-to-water oil coolers for heat transfer performance'.
Sakai T, Ishiguro S, Sudoh Y, Raab G and Hager J (1994), 'The optimum design of engine cooling system by computer simulation', SAE paper 942270.
Samhaber C, Wimmer A and Loibner E (2001), 'Modeling of engine warm-up with integration of vehicle and engine cycle simulation', SAE paper 2001-01-1697.
Savage J, Tang Y, Xie Z and Rahman S S (2007), 'Sensitivity analysis of powertrain cooling system performance', SAE paper 2007-01-0598.
Schwarz E, Reid M, Bryzik W and Danielson E (1993), 'Combustion and performance characteristics of a low heat rejection engine', SAE paper 930988.
Scott T C, McDonald L K (2005), 'Cooling fan modeling to support robust AC/cooling system simulation', SAE paper 2005-01-1905.
Shah R K (2003), 'Advances in automotive heat exchanger technology', SAE paper 2003-01-0533.
Sharma J K and Gaur R R (1990), 'Some analytical investigations on a four stroke diesel engine for high specific output and low heat rejection', SAE paper 900618.
Shayler P J, Christian S J and Ma T (1993), 'A model for the investigation of temperature, heat flow and friction characteristics during engine warm-up', SAE paper 931153.
Shayler P J, Chick J P and Ma T (1996), 'Effect of coolant mixture composition on engine heat rejection rate', SAE paper 960275.
Shayler P J, Chick J P and Ma T (1997), 'Correlation of engine heat transfer for heat rejection and warm-up modeling', SAE paper 971851.
Sitkei G (1974), *Heat Transfer and Thermal Loading in Internal Combustion Engines*, Akademiai Kaido, Budapest.
Stobart R and Weerasinghe R (2006), 'Heat recovery and bottoming cycles for SI and CI engines – a perspective', SAE paper 2006-01-0662.
Stolz A, Fleischer K, Knecht W, Nies J and Strähle R (2001), 'Development of EGR coolers for truck and passenger car application', SAE paper 2001-01-1748.
Sun X, Wang W G, Bata R M and Gao X (1994), 'Performance evaluation of low heat rejection engines', *Transactions of the ASME*, *Journal of Engineering for Gas Turbines and Power*, 116(4), 758–764.
Sutor P, Schwarz E E and Pangilinan H (2005), 'Major advances in tribology for low heat rejection engines – a ten-year overview', *Proceedings of the World Tribology Congress III – WTC 2005*, pp. 575–576.

Tamilporai P, Baluswamy N, Jawahar P M, Subramaniyam S, Chandrasekaran S, Vijayan K, Jaichandar S, Rani J J and Arunachalam K (2003), 'Simulation and analysis of combustion and heat transfer in low heat rejection diesel engine using two zone combustion model and different heat transfer models', SAE paper 2003-01-1067.

Tang Y, Grall N and Savage J (2008), 'Various impacting factors on a radiator top tank temperature', SAE paper 2008-01-1438.

Taraza D, Henein N and Bryzik W (2000), 'Friction losses in multi-cylinder diesel engines', SAE paper 2000-01-0921.

Taylor S, Clark N, Gautam M and Wayne W S (2004), 'Diesel emissions prediction from dissimilar cycle scaling', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 218, 341–352.

Teng H, Regner G and Cowland C (2006), 'Achieving high engine efficiency for heavy-duty diesel engines by waste heat recovery using supercritical organic-fluid Rankine cycle', SAE paper 2006-01-3522.

Teng H, Regner G and Cowland C (2007a), 'Waste heat recovery of heavy-duty diesel engines by organic Rankine cycle, part I: hybrid energy system of diesel and Rankine engines', SAE paper 2007-01-0537.

Teng H, Regner G and Cowland C (2007b), 'Waste heat recovery of heavy-duty diesel engines by organic Rankine cycle, part II: working fluids for WHR-ORC', SAE paper 2007-01-0543.

Valaszkai L and Jouannet B (2000), 'Cooling system optimization for Euro4 – EPA/02 heavy duty trucks', SAE paper 2000-01-0964.

Varghese M B, Goyal S K and Agarwal A K (2005), 'Numerical and experimental investigation of oil jet cooled piston', SAE paper 2005-01-1382.

Wallace S J and Kremer G G (2009), 'Diesel engine energy balance study operating on diesel and biodiesel fuels', *Proceedings of ASME International Mechanical Engineering Congress and Exposition*, 17, 337–343.

Woods M, Bryzik W and Schwarz E (1992), 'Heat rejection from high output adiabatic diesel engine', SAE paper 920541.

Xin Q and Zheng J (2009), 'Theoretical analysis of internal combustion engine miscellaneous heat losses', SAE paper 2009-01-2881.

Yasar H (2008), 'First and second law analysis of low heat rejection diesel engine', *Journal of the Energy Institute*, 81(1), 48–53.

Yonushonis T M (1997), 'Overview of thermal barrier coatings in diesel engines', *Journal of Thermal Spray Technology*, 6(1), 50–56.

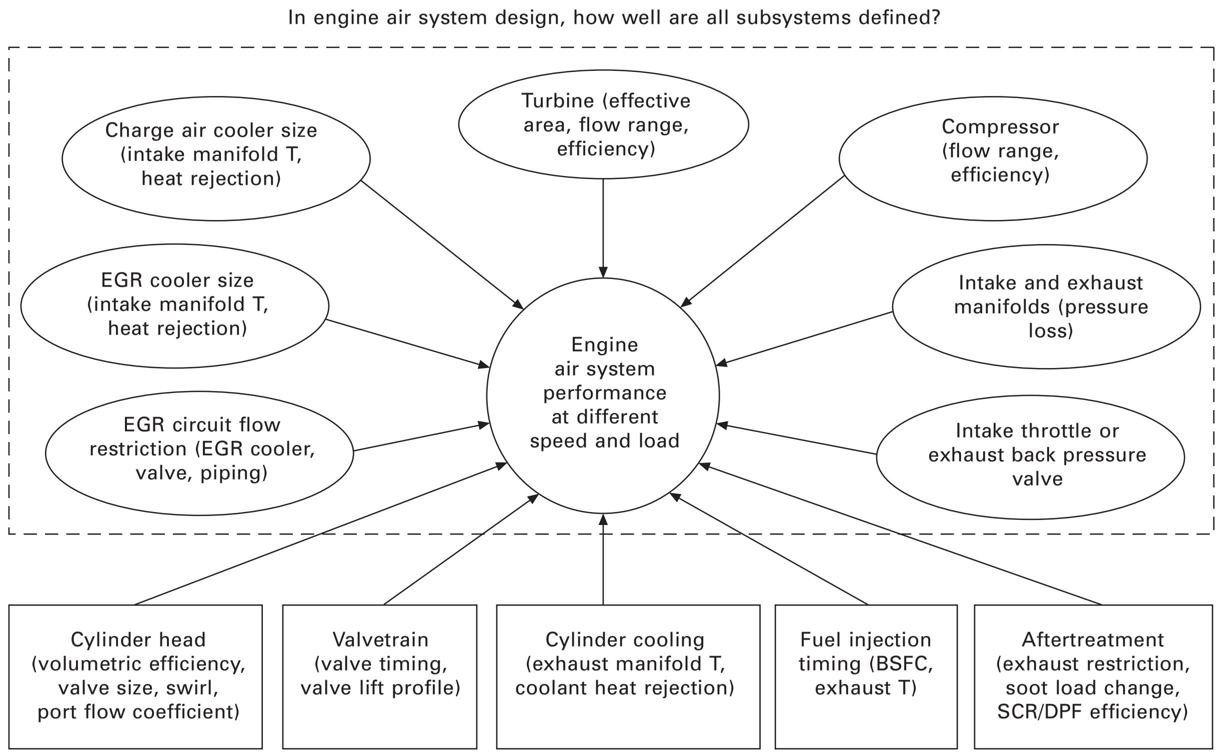
13
Diesel engine air system design

Abstract: This chapter develops the design theory of diesel engine air system (including the turbocharger, the manifolds, and the exhaust gas recirculation (EGR) system) with engine cycle simulations. It addresses the applications of the theory of pumping loss control and subsystem interactions that is developed in Chapter 4. The chapter begins with an overview of the air system requirements determined by the emissions recipe (i.e., air–fuel ratio, EGR rate, intake manifold gas temperature, etc.), followed by a discussion of different EGR system configurations, before elaborating turbocharger matching and exhaust manifold design. The chapter concludes with a comprehensive system analysis of the second law of thermodynamics for modern turbocharged EGR diesel engines.

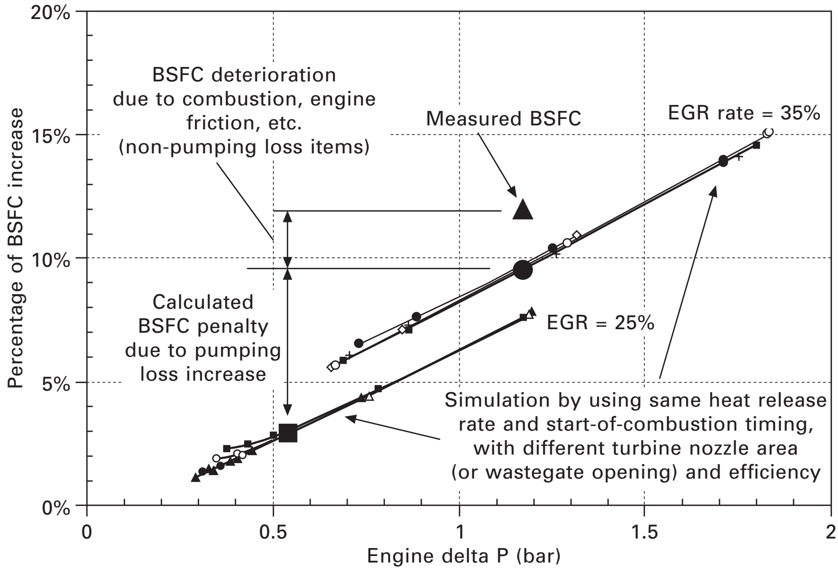
Key words: air system, turbocharger matching, manifold, EGR system, engine cycle simulation, pumping loss, subsystem interaction, second law of thermodynamics.

13.1 Objectives of engine air system design

The diesel engine air system consists of intake and exhaust manifolds, turbocharger, charge air cooler (CAC), EGR circuit, intake throttle (IT) and exhaust back-pressure valve (EBP valve), other air/gas control valves, and valvetrain. Its relationship with other systems is shown in Fig. 13.1. The goal of air system design is to achieve desirable air–fuel ratio, EGR rate, exhaust gas temperature, and oxygen concentration to meet the requirements of both engine-out emissions and aftertreatment operation with minimum pumping loss and low BSFC without violating the design constraints in the desirable operating region of engine speed and load under all driving and ambient conditions. Low pumping loss is achieved via high volumetric efficiency or low engine delta P. The design constraints include the maximum limits of cylinder pressure, exhaust manifold gas pressure and temperature, compressor outlet air temperature, turbocharger speed, and coolant heat rejection. The design constraints also include certain minimum limits such as acceptable intake manifold gas temperature and compressor surge margin. Note that the requirements resulting from aftertreatment devices are new challenges for the air system. The core of air system design is pumping loss control. Chapter 4 discusses the pumping loss theory and the mathematical formulations behind air system design principles. This chapter focuses on their engine applications. Figure 13.2 illustrates the direct correlation between engine delta P and



13.1 The boundary of diesel engine air system design.

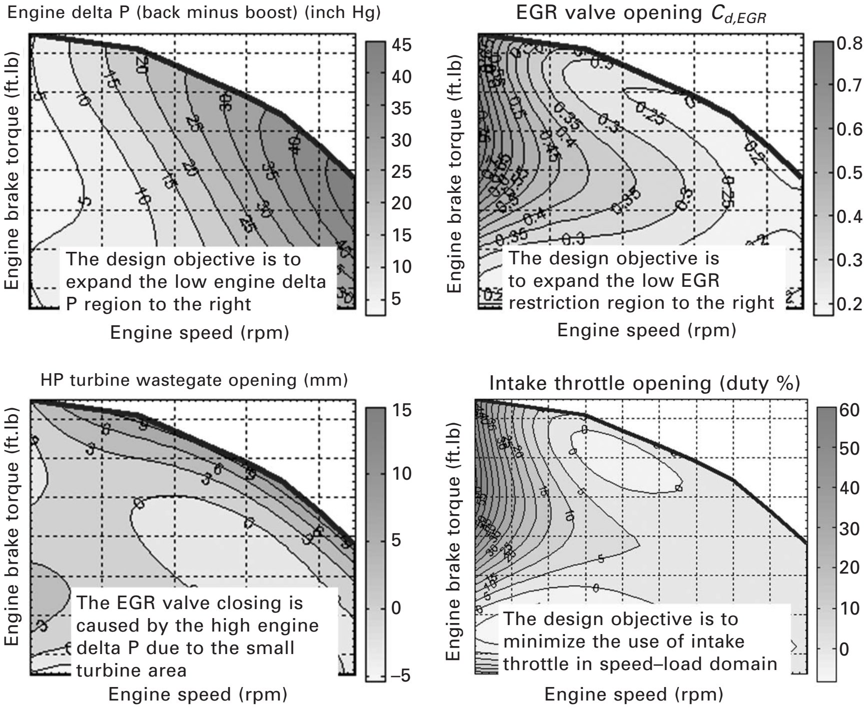


13.2 Effect of engine delta P on BSFC at rated power.

BSFC. Figure 13.3 shows the typical behavior of air system control factors and their design directions in the engine speed–load domain.

13.2 Overview of low-emissions design and air system requirements

In the base engine design with a given displacement, there are five factors directly related to air system performance in terms of volumetric efficiency, heat rejection, pumping loss, and mechanical friction. The factors are: (1) number of cylinders; (2) stroke-to-bore (S/B) ratio; (3) valve flow area related to number of valves and valve size; (4) valve overlap height; and (5) the 'K-factor' used to measure the proportion of air available for combustion. The K-factor refers to the ratio of piston bowl volume to total combustion chamber volume at the TDC. Fewer cylinders result in fewer parts and lower material cost, but they increase the reciprocating mass of each cylinder. The reciprocating inertia force is high at high speeds. Using more cylinders is an option to reduce the inertia force. Increasing the number of cylinders also increases the surface-to-volume ratio and heat losses. A larger S/B ratio reduces piston ring circumference and the potential air leakage during cold start and may also reduce the surface-to-volume ratio for the TDC volume and heat losses. A smaller S/B ratio allows bigger intake or exhaust valve size to achieve higher volumetric efficiency and lower pumping loss. Engine friction and inertia force increase as the mean piston speed increases, which



13.3 Engine air system control parameters in speed–load domain.

is proportional to the S/B ratio. The K-factor affects air system design because it is related not only to the tolerances of the piston assembly, the cylinder head, and the gasket, but also to the size of the valve recession or the piston cutout and hence the valve overlap height at the TDC. The greater the K-factor, the lower the dead volume in the combustion chamber. Valve overlap is a key design parameter for engine breathing performance and valvetrain dynamics. Reducing the clearance volume can increase the K-factor, reduce soot emissions, and improve the smoke limit. Competitive engine data analysis indicates that there seems to be a trend for the K-factor to be a function of the S/B ratio. As the S/B ratio increases from 1.0 to 1.2, the K-factor range increases from 0.7 to 0.8 approximately. The optimum system design concept should be based on the best trade-off among the S/B ratio, the K-factor, and the valve overlap height.

The prevailing cylinder head design in modern diesel engines uses a four-valve head. Compared with the two-valve head design, it has the following benefits: (1) higher volumetric efficiency; (2) enabling a centralized vertical fuel injector; (3) the feasibility of shutting off one intake port to regulate the swirl; and (4) a lower valvetrain weight and better valvetrain dynamics. In intake port design, the appropriate swirl ratio needs to be determined based on the following considerations: (1) matching with the fuel injection parameters for good penetration and mixing; (2) matching with the combustion chamber shape for proper air motion and utilization; (3) matching with the cylinder bore size to control wall impingement of the fuel spray and evaporation of the fuel film on the cylinder wall; (4) an appropriate compromise with port flow discharge coefficient or volumetric efficiency; (5) a proper balance between the swirl levels at low speeds and high speeds; and (6) an acceptable cylinder coolant heat rejection loss, which is affected by the in-cylinder air turbulence level. The current trend in heavy-duty diesel engine design is to decrease the intake swirl but to increase the fuel injection pressure, number of nozzle holes, and combustion chamber diameter. Diesel engine cylinder head design is reviewed by Gale (1990).

Compared to naturally aspirated engines, turbocharging with intercooling can increase the air–fuel ratio and reduce soot. Soot can also be greatly reduced by increasing fuel injection pressure. NO_x is controlled mainly by using EGR and retarding fuel injection timing. In order to satisfy different air demands within a wide speed range in automotive diesel applications, the fixed-geometry turbine has gradually been replaced by the wastegated turbine or the VGT. Turbocharging also enables EGR to be used because engine delta P can be modulated by turbine area to drive EGR in a high-pressure-loop EGR system. Alternatively, a compressor can be used to pump EGR into the intake manifold in a low-pressure-loop EGR system. Coordinated joint design of turbocharging, intake port and cam profile to match with the EGR system is the key to fulfill the air flow requirements over a wide speed range. More information on the fundamentals of the air system configurations and performance of turbocharged EGR diesel engines can be found in Jacobs *et al.* (2003) and Hochegger *et al.* (2002).

Turbine outlet gas temperature and exhaust manifold gas temperature are two critical parameters in air system design. The former is related to aftertreatment regeneration. The latter affects turbocharger performance and also has been used to gauge the thermal loading on engine components. When the in-cylinder gas temperature increases, the exhaust temperature and the heat flux to the power cylinder components also increase. The following can lead to a decrease in exhaust manifold gas temperature: (1) increase in air–fuel ratio or EGR rate; (2) fuel injection timing advance; and (3) enhanced cylinder cooling. Previously, without an emissions-driven design approach, turbocharging parameters were selected based on a target turbine inlet gas temperature determined by the requirements of acceptable thermal load and engine reliability. The required air–fuel ratio and large valve overlap were then calculated to achieve the target exhaust manifold temperature. Finally, the turbine flow area and pressure ratio were determined based on the required intake boost pressure and turbocharger efficiency. Now the situation is different due to an emissions-driven approach. The turbocharger selection criterion is based on air–fuel ratio and EGR rate to meet emission requirements, with the maximum exhaust manifold gas temperature as a durability design constraint.

13.3 Exhaust gas recirculation (EGR) system configurations

13.3.1 Classification of EGR systems

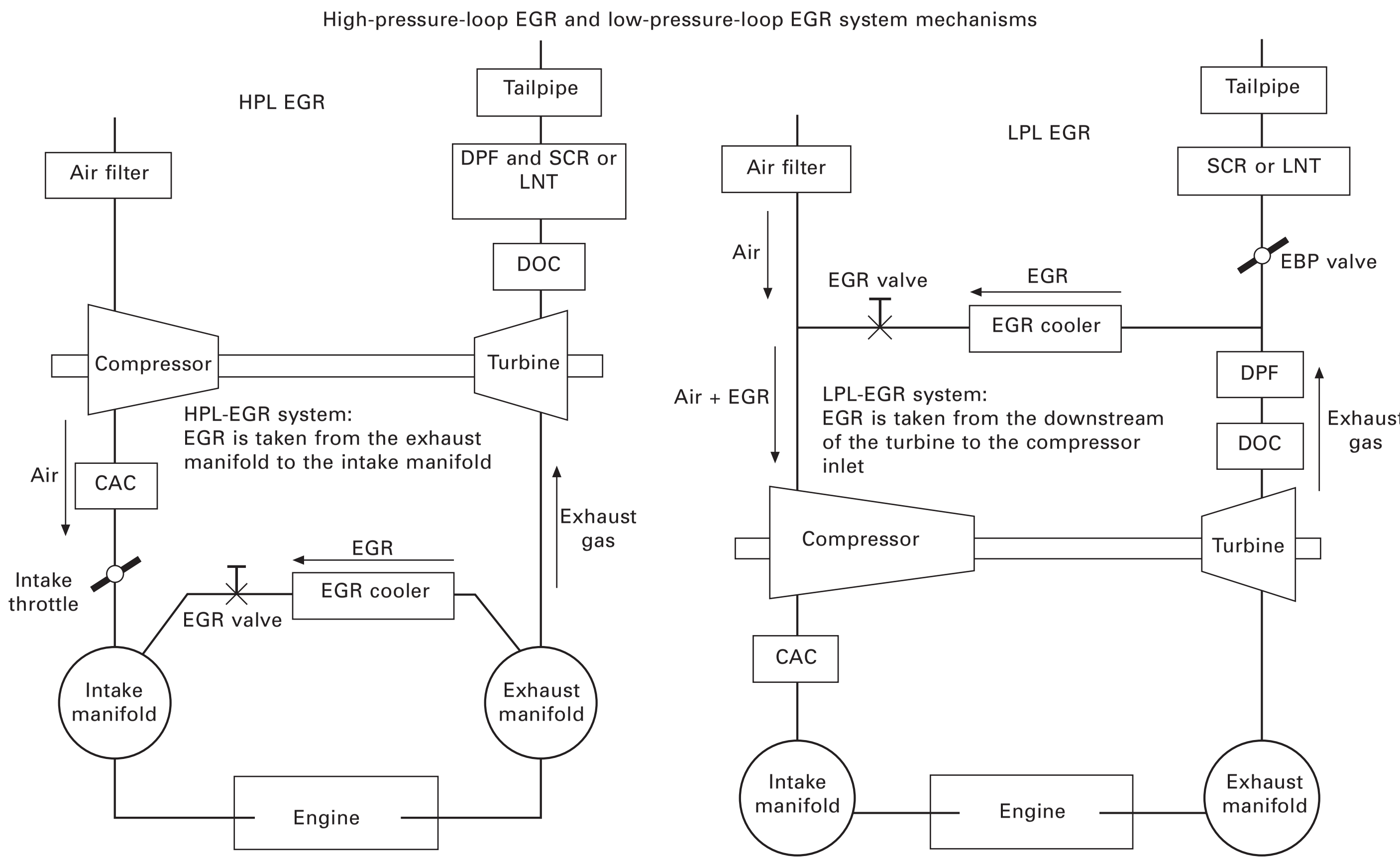
EGR systems can be classified into external and internal EGR systems. The internal EGR, usually uncooled, refers to the trapped in-cylinder residue of combustion product and the reverse gas flow from the exhaust manifold/port to the cylinder. Cooled external EGR is generally more effective than uncooled internal EGR for emissions reduction and fuel economy, although the heat rejection in the external EGR system needs to be handled by the cooling system. The external EGR can be further classified into high-pressure-loop (HPL) EGR, low-pressure-loop (LPL) EGR, and hybrid (or dual-loop, i.e., HPL plus LPL) EGR systems (Fig. 13.4).

Diesel engine EGR mixing has been researched by Siewert *et al*. (2001) and Partridge *et al*. (2002). Diesel engine EGR systems have been extensively investigated by Akiyama *et al*. (1996), Baert *et al*. (1996, 1999), Kohketsu *et al*. (1997), Mattarelli *et al*. (2000), Graf *et al*. (2000), Lundqvist *et al*. (2000), Luján *et al*. (2001), Osborne and Morris (2002), Andersson *et al*. (2002), Chatterjee *et al*. (2003), Maiboom *et al*. (2008), and Shutty (2009).

13.3.2 EGR engine design principles

Engine pumping loss reduction is the most important design objective in EGR system design. The BSFC of modern diesel engine is largely dominated by pumping loss (Fig. 13.2). Pumping loss control is critical for three reasons: (1) to maintain good BSFC in a high-EGR engine within a wide speed range from peak torque to rate power; (2) to increase air–fuel ratio to reduce soot; and (3) to reduce peak cylinder pressure and exhaust manifold pressure to alleviate their design challenges. One example of engine delta P distribution in the engine speed–load domain is shown in Fig. 13.3. The control factors in air system design include mainly turbine area and/or wastegate, EGR valve opening, and intake throttle opening. Wastegating essentially means higher pumping loss. Closing the EGR valve indicates an undesirable increase in EGR circuit flow restriction. Closing the intake throttle results in a large decrease in air–fuel ratio.

The design guidelines for the air system of high-EGR engines to achieve good BSFC include the following:



13.4 Turbocharged EGR engine systems.

- achieve high EGR rate with low flow restriction in the EGR circuit or low engine delta P.
- eliminate local throttle losses (i.e., exergy losses or entropy increases) at the EGR valve, the intake throttle valve, and the turbine wastegate valve.
- use inexpensive ways to achieve high turbocharger efficiency (e.g., shift engine operating points to the high-efficiency region on the compressor map).

The design and calibration principles of EGR engine's air system can be explained by the four core equations introduced in Chapter 4, i.e., equation 4.40 (engine volumetric efficiency), 4.44 (EGR circuit pressure drop), 4.47 (turbocharger power balance), and 4.57 (turbine flow), summarized (combined) as below:

$$\begin{cases} \eta_{vol} = \dfrac{2(\dot{m}_{air} + \dot{m}_{EGR})T_{2a}R_{gas}}{(p_2 - \Delta p_{IntakeThrottle} - \Delta p_{CAC})N_E V_E} \\ p_{EGRin} - p_{EGRout} = f(C_{d,EGR}, \dot{m}_{EGR}, T_{EGRcoolerOut}) \approx C_0 + C_1 \dot{m}_{EGR}^2 \\ 1 - \left(\dfrac{p_2}{p_1}\right)^{\frac{\kappa_c - 1}{\kappa_c}} + \eta_C \eta_T \eta_{TC,mech} \left(\dfrac{\dot{m}_T}{\dot{m}_C}\right)\left(\dfrac{c_{p,T}}{c_{p,C}}\right)\left(\dfrac{T_3}{T_1}\right)\left[1 - \left(\dfrac{p_4}{p_3}\right)^{\frac{\kappa_t - 1}{k_t}}\right] = 0 \\ \dot{m}_T = A_T \cdot \dfrac{p_3}{\sqrt{R_{ex}T_3}} \cdot \sqrt{\dfrac{2\kappa_t}{k_t - 1}} \cdot \sqrt{\left(\dfrac{p_4}{p_3}\right)^{\frac{2}{\kappa_t}} - \left(\dfrac{p_4}{p_3}\right)^{\frac{k_t + 1}{\kappa_t}}} \end{cases} \tag{13.1}$$

These four core equations can be used to solve for any four unknowns (e.g., intake manifold boost pressure p_2, exhaust manifold pressure p_3, turbine area A_T, EGR circuit flow restriction coefficient $C_{d,EGR}$) with a given target of air flow rate and EGR rate. Based on these equations, in any EGR system design, two questions need to be considered:

1. Does the EGR system raise engine delta P unfavorably?
2. Which of the following causes more deterioration in BSFC: EGR pumping work with a pump (if any) or engine pumping loss?

The answer to these questions depends on turbocharger efficiency.

13.3.3 Advantages and challenges of HPL EGR system

The HPL EGR system has traditionally been used in EGR engines. The HPL EGR system uses the engine delta P created by the turbocharger to drive EGR from the exhaust manifold to the intake manifold. A higher EGR rate requires higher engine delta P to drive in HPL EGR. Modern diesel engines need to drive a high EGR rate and sufficient air–fuel ratio from peak torque to rated power in order to meet the stringent emissions requirements relating to NO_x and soot. If a fixed-geometry turbine is used, with the lowest level of EGR system flow restriction (i.e., with the EGR valve fully open), the turbine area is usually sized small enough to drive the required EGR rate at peak torque. The turbine usually needs to be wastegated at rated power in order to avoid over-boosting the engine on peak cylinder pressure and/or exhaust manifold pressure. The smaller the turbine area, the more wastegating is required at high speeds. Compared to a VGT (without wastegating) to deliver the same boost pressure, wastegating has to use a higher turbine pressure ratio which compensates for the loss of the turbine mass flow rate in order to maintain the same turbine power. This is the reason why wastegating causes higher engine delta P and BSFC than VGT. Therefore, by the nature of engine–turbocharger matching with a fixed turbine area, there is a trade-off between low-speed and high-speed performance in pumping loss and system efficiency. VGT may solve the dilemma by using flexible turbine area at different engine speeds, but there are often penalties in turbocharger cost and durability as well as a reduction in turbine efficiency at the very closed or wide open VGT vane opening. The role of intake throttle is to reduce excessively high air–fuel ratio at peak torque or high speed and low load, or slightly increase EGR rate at a penalty of throttling loss. An excessively high air-fuel ratio may cause high NO_x and high engine delta P. This is usually because the turbocharger efficiency is too high. It should be noted that in most applications the air–fuel ratio at peak torque is short compared to the ideal requirement instead of being excessively high.

In summary, when the engine delta P is insufficient to drive EGR at low speed and high load, the following measures can be considered: (1) reducing the flow restriction of the EGR valve and the EGR cooler (e.g., using parallel instead of serial EGR coolers); (2) using a check valve or a reed valve to harvest EGR flow pulsation and prevent EGR backflow loss; (3) using a small turbine area or a VGT to raise the engine delta P; (4) using the intake throttle to lower the intake manifold pressure; (5) using a venturi device to locally reduce the static pressure at the EGR merging location in order to induce the EGR flow into the intake manifold; or (6) using an EGR pump.

The weakness of the HPL EGR system in heavy-duty diesel engines is that a large EGR flow rate often needs to be driven by the engine delta P at peak torque. An improved EGR-driving system needs to provide a better way to pump EGR in order to reach a lower engine delta P and higher turbocharger efficiency. One possible solution is to use the LPL EGR or a dual-loop EGR system to help EGR driving at peak torque or low speeds so that a larger turbine area can be selected in order to avoid high pumping loss at high speeds. However, it should be noted that an excessively large turbine area may not be able to deliver the required air–fuel ratio at high altitude.

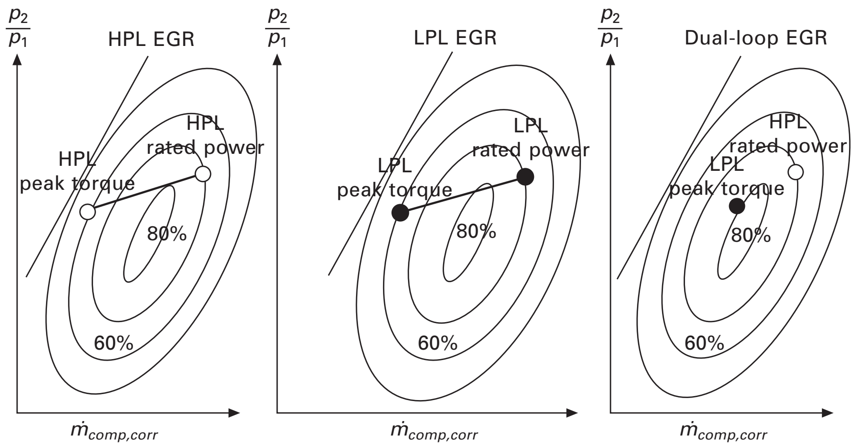
13.3.4 Design principles of LPL EGR system

In a clean LPL EGR system (Fig. 13.4) the EGR flow is driven by the pressure difference between the EGR pickup location (usually downstream of the DPF) and the compressor inlet. EGR cooling is needed in order to prevent the compressor outlet gas temperature from exceeding the allowable limit. The EGR circuit flow restriction needs to be minimized because any artificial increase in the EGR-driving pressure differential beyond the 'naturally available free pressure differential' provided by the system would lead to a large increase in pumping loss. Such an artificial increase in the LPL EGR circuit pressure differential includes closing an EBP valve installed after the DPF or closing an intake throttle valve at the compressor inlet. The latter option usually incurs a larger pumping loss. Equation 13.1 shows that raising the exhaust restriction (p_4) or reducing the intake restriction (p_1) in order to drive EGR in LPL EGR actually cause a large increase in engine delta P.

A LPL EGR system may be promising under the following four conditions:

1. There is a sufficiently large 'naturally available free pressure differential' (i.e., free of extra pumping loss penalty) to drive the EGR flow. Some examples include the high local vacuum created by the suction near the compressor wheel at the compressor inlet, or the high local pressure created by a diffuser at the EGR pickup location in the exhaust pipe.
2. The turbine area can be sized large enough to reduce the engine delta P while still delivering sufficiently high air–fuel ratio at peak torque and high altitude.
3. The engine operating points can be shifted from the low efficiency region to the higher efficiency region on the turbocharger map (e.g., Fig. 13.5).
4. The power consumption of any EGR pump can be more than offset by the BSFC benefit resulting from a lower engine delta P in the LPL EGR system.

It should be noted that the LPL EGR system has a higher turbine flow rate than the HPL EGR system. Therefore, the LPL EGR system demands a larger aftertreatment device in order to maintain an acceptable level of exhaust restriction. The LPL EGR system also requires a larger compressor size (due to increased compressor flow rate), a higher compressor outlet temperature, a higher heat rejection of the charge air cooler, a longer EGR piping route, and worse transient response due to longer EGR purging time and larger turbocharger inertia.

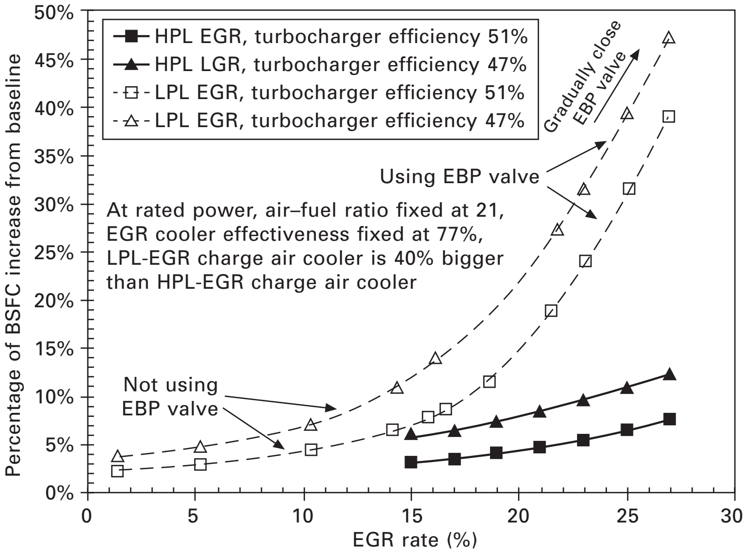


13.5 Regulate engine operating points on compressor map by controlling EGR.

13.3.5 Comparison between HPL and LPL EGR systems

At high speeds and high loads with a high EGR rate, HPL EGR is usually superior to LPL EGR due to its lower turbine power and pumping loss. At low speeds or low loads or a low EGR rate, LPL EGR may exhibit certain advantages if its compressor is better matched at a higher efficiency. Hybrid EGR combines the advantages of both HPL and LPL and provides the lowest pumping loss but at the price of increased design complexity and cost. The need for a hybrid EGR system rises in two scenarios. First, if the turbocharger efficiency is low and a small turbine area has to be used to compensate in order to achieve a required boost pressure, any resulting engine delta P then becomes a free pressure differential which can be utilized to drive EGR. In this case, HPL EGR needs to be used, and the balance of the EGR rate can be supplied by a LPL EGR system. Secondly, flexibly switching between the LPL and HPL EGR operations at different engine speeds/loads may be able to keep the engine operating points in the high efficiency region on the compressor map by regulating the compressor mass flow rate.

Figure 13.6 shows a simulation comparison between HPL and LPL EGR at rated power for a heavy-duty diesel engine. In the LPL EGR system analyzed, a large artificial increase in the LPL EGR circuit pressure differential is created by using an EBP valve to drive EGR. Therefore, there is a large BSFC penalty in the LPL EGR system. This demonstrates that the LPL EGR system, if not designed correctly, may incur BSFC penalty. Also note that turbocharger efficiency has a large impact on the comparison between the EGR systems.



13.6 Performance interaction between EGR system and turbocharging.

13.4 Turbocharger configurations and matching

13.4.1 Turbocharger configurations

Turbocharger fundamentals

In general, exhaust turbocharging (SAE J922, 1995) is superior to mechanical supercharging in terms of thermodynamic efficiency. Turbocharging with intercooling may increase intake air density, engine power, and fuel economy, while reducing combustion noise and emissions. Automotive turbocharger design has evolved from the early fixed geometry turbocharger to today's electronically controlled wastegated turbocharger and VGT with high turbine efficiency of around 80%. The turbine flow passage is designed to cause the exhaust gas to impinge on the turbine blades at an optimum angle and velocity. The compressor diffuser is designed to maximize the total pressure at the compressor outlet. The efficiency of centrifugal compressors is dependent on the design of the blade and the diffuser.

On the compressor map, there is a surge limit curve on the upper left side and a choke limit curve on the right side. The limit curves represent the boundaries between stable and unstable operations. Surge is a self-sustaining but unstable oscillation in flow rate and pressure ratio initiated by a flow reversal within the boundary layers on the compressor blades. Audible coughing and vibration occur when surge happens. Compressor choke happens when the air velocity reaches the speed of sound at the inlet of the compressor wheel or diffuser. The flow rate at which choke happens varies with the tip speed. The precise position of the choke curve on the compressor map is more difficult to determine than the surge line because it depends on the flow structure of the internal boundary layers. Sometimes for the purpose of approximation it is assumed that compressor choke happens when the compressor efficiency contours reduce to around 55%.

Pulse and constant-pressure turbocharging

There are two types of turbocharging: pulse and constant pressure. In pulse turbocharging, in order to minimize the pressure wave interference or gap windage in exhaust scavenging between the cylinders and to raise the average turbine efficiency, the cylinders with a certain firing interval (ideally about 240° crank angle) are grouped together by using a small-volume exhaust manifold to preserve strong exhaust pressure pulses. A pulse converter may be used to eliminate pulse interference. A divided turbine entry may be used to totally separate the pulses of different groups, but may cause a larger pumping loss at high engine speeds than an undivided turbine entry. In VGT, sometimes it is difficult to use a divided turbine entry because of the extra controlling device. When the unsteady energy flow is fed into the turbine, turbine efficiency may decrease.

In constant-pressure turbocharging, the exhaust flows of all the cylinders are collected in one large-volume exhaust manifold to damp out all the pressure fluctuations. A steady energy flow is fed into a single-entry turbine, and the turbine can be matched to the peak efficiency on the steady flow test map. The averaged turbine efficiency over an engine cycle with the unsteady flow of pulse turbocharging is usually lower than the peak efficiency matched in constant-pressure turbocharging. But the disadvantage of efficiency loss is normally more than offset by the advantage of the pulse energy available at the turbine. The overall exhaust energy utilization efficiency is actually characterized by the combination of the pulse energy and the turbine efficiency.

Previous experiments indicated that at a low compressor pressure ratio (e.g., at part load), pulse turbocharging has a higher overall efficiency than constant-pressure turbocharging. But the situation is reversed at very high compressor pressure ratios because the pulsed turbine efficiency is low at very high instantaneous turbine pressure ratios (e.g., above 3:1). During the transient acceleration, pulse turbocharging is superior due to its smaller exhaust manifold volume and faster response of the turbine inlet energy pulse to speed up the turbocharger. In EGR engines, pulse turbocharging used with a reed valve (check valve) can drive more EGR flow than constant-pressure turbocharging with less penalty in engine pumping loss.

Single-stage, two-stage, and twin-parallel turbocharging

Usually single-stage turbocharging can deliver a compressor pressure ratio around 3.5–4.5, depending on its wheel material and design. Compressor speed is limited by the allowable centrifugal force exerted upon the compressor wheel. The maximum compressor speed and the aerodynamic blade shape limit the pressure ratio capability (Arnold, 2004). Moreover, the compressor wheel and casing made of aluminum alloy are limited in material strength by a maximum allowable compressor discharge temperature of 204–210°C (400–410°F). Titanium impeller wheels can resist much higher temperatures and deliver a very high pressure ratio with the simple single-stage turbocharging, but titanium and its manufacturing process are expensive. Another disadvantage of the titanium wheels is their heavier weight and larger inertia.

Despite the disadvantages in cost, weight, cumbersome piping, and packaging space, two-stage turbocharging is superior to single-stage in the following aspects: (1) higher compressor pressure ratio; (2) higher turbo efficiency due to the relatively lower pressure ratio in each stage; (3) feasibility of using inter-stage cooling to reduce the compressor discharge temperature and the compressor power to improve the efficiency; (4) easier matching for the turbocharger for a wide range of engine speed combined with high rated power and high peak torque; (5) improved altitude capability due to a larger margin of choke; (6) faster transient response due to the small high-pressure-stage turbocharger and its low rotor inertia; and (7) improved low-cycle-fatigue life of the compressor wheels and fewer durability problems associated with excessively high turbocharger speed.

In the twin-parallel turbocharger configuration, both turbines can be located very close to the cylinder heads of each bank of the engine in order to minimize the exhaust energy losses caused by manifold volume, pressure drop, and heat transfer.

Wastegate vs. VGT turbocharging

The advantages of VGT over wastegated turbines are summarized as follows: (1) no throttling loss of the wastegate valve; (2) a better ability to achieve a high air–fuel ratio and high peak torque at low engine speeds; (3) a better ability to achieve fast vehicle accelerations without encountering a high pumping loss at high engine speeds; (4) overall lower engine delta P; and (5) a better ability to cover a wider region of low BSFC in the engine speed–load domain. The VGT area change can be in either continuous or discrete settings, depending on the requirements of engine performance and reliability. The maximum efficiency of most variable-geometry turbines occurs at a medium vane opening of around 60–70% open. At the fully open and fully closed vane openings, the efficiency drops rapidly. The challenges of

VGT design include minimizing gas leakage, increasing turbine efficiency, and enhancing reliability (Furukawa *et al*., 1993).

Previous research on turbocharging

Turbomachinery fundamentals are introduced in Japikse and Baines (1997), Wilson and Korakianitis (1998), and Japikse (2009). The detailed theory of turbocharging the non-EGR engines is provided by Watson and Janota (1982) and Watson (1999). A more recent summary on turbocharging the internal combustion engines is provided by Baines (2005b). For the mapping procedures of turbocharger gas stand test data, the reader is referred to SAE J1826 (1995).

Early investigations on turbocharging the non-EGR diesel engines are represented by Freeman and Walsham (1978), Watson *et al*. (1978), Flynn (1979), and Watson (1979). Turbocharging the EGR diesel engine is presented by Bozza *et al*. (1997) and Galindo *et al*. (2007). Advanced turbocharging technologies for EGR diesel engines are discussed by Ludu *et al*. (2001), Arnold *et al*. (2001, 2005), Arnold (2004), Amos (2002), Carter *et al*. (2010), and Tufail (2010).

A turbocharger matching procedure for the wastegated turbochargers is given by Ubanwa and Kowalczyk (1993). VGT performance is investigated by Flaxington and Szczupak (1982), Watson (1982, 1986), Watson and Banisoleiman (1986), Yokota *et al*. (1986), Wallace *et al*. (1986), Hishikawa *et al*. (1988), Kawamoto *et al*. (2001), Arnold *et al*. (2002), Tange *et al*. (2003) and Uchida *et al*. (2006). Two-stage turbocharging on diesel engines has been researched by Ghadiri-Zareh and Wallace (1978), Watson *et al*. (1978), Saulnier and Guilain (2004), Millo *et al*. (2005), Choi *et al*. (2006), and Serrano *et al*. (2008). Twin-parallel turbocharging is presented by Cantemir (2001).

It should be noted that the design and development of non-VNT type of VGT (also called VAT or variable area turbine) has been active since 1970s. The VAT has the advantages of simpler design, better reliability and lower cost, compared with VNT. However, the VAT usually suffers on turbine efficiency. Major VAT designs include Pampreen (1976), Chapple *et al*. (1980), Bhinder (1984), Hirabayashi *et al*. (1986), Okazaki *et al*. (1986), Franklin and Walsham (1986), Franklin (1989), Inoue *et al*. (1989), Ogura *et al*. (1989), Umezaki *et al*. (1989), Ogura and Shao (1995), Shao *et al*. (1996), and Wang (1996). The recent most noticeable and promising VAT design was the variable flow turbine (VFT) developed by Aisin Sekei and used by other companies (Kawaguchi *et al*., 1999; Ishihara *et al*., 2002; Inter-Tech Energy Progress, 2003; Andersen *et al*., 2006; and Ito *et al*., 2007).

Another noticeable and promising turbocharger technology is the swirl jet turbine developed by Anada *et al*. (1997). The swirl jet turbine generates a swirling exhaust gas flow at the turbine outlet by utilizing the wastegated gas flow so that the turbine outlet pressure is effectively reduced locally to enable an increased turbine power and higher turbine speed. At mean time, the turbine inlet pressure and the engine delta P may decrease. By properly controlling the turbine wastegate opening or the swirling jet, a higher air–fuel ratio or lower engine delta P (and lower pumping loss) can be achieved. Similar designs to the swirl jet turbine were reported by Baker *et al.* (2001) and Schmid and Sumser (2007).

Moreover, turbocharging simulation models have been developed by Nasser and Playfoot (1999), Gurney (2001), Galindo *et al.* (2006), Zhuge *et al.* (2009), and Park *et al.* (2010). Supercharging the diesel engine is investigated in Cantore *et al.* (2001). Supercharger testing was introduced by SAE J1723 (1995).

13.4.2 Compressor matching, aerodynamic design, and durability

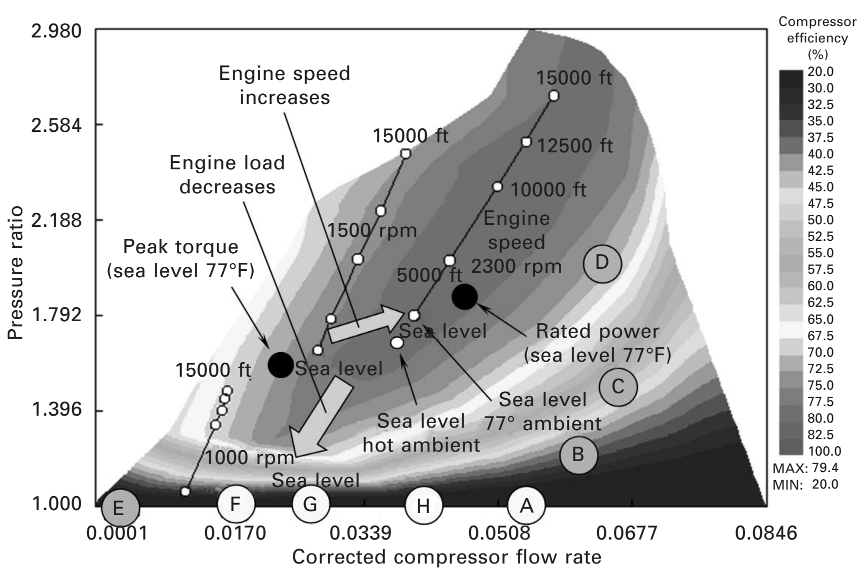
The turbocharger is one of the most expensive components in the engine and therefore deserves close attention from a commercial point of view. The selection of size and efficiency of a turbocharger is based on the requirements of delivering the target air–fuel ratio, EGR rate, and transient acceleration performance. The size of a turbocharger determines the flow range and the pressure ratio. The objective of turbocharger matching is to achieve the best compromise in the engine speed–load range.

The turbocharger test procedure was outlined in SAE J1826 (1995), Turbocharger Gas Stand Test Code. In principle, the compressor map can be generated from a flow bench test by regulating two valves. One is installed at the inlet of a driving turbine, and the other is installed at the compressor outlet. The compressor outlet valve can be gradually closed at each turbine valve opening in order to obtain a series of compressor flow points (from high to low until surge), while the compressor speed remains almost constant. The process can be repeated with different turbine valve openings, and finally a range of compressor flow points at many shaft speeds can be obtained.

The focus of engine–compressor matching is to analyze the engine operating characteristic points or lines at various speed, load, and ambient conditions on the compressor map. Diesel engine characteristic lines include the full-load lug curve, constant-speed curve, and constant-load curve. The slope of the constant-speed curve on the compressor map, which is essentially the ratio between the pressure and the mass flow rate, represents the reciprocal of the engine volumetric efficiency. Therefore, any factors affecting volumetric efficiency may have an impact on the shape of engine speed characteristic lines. These factors include valvetrain design (valve size, timing, and lift profile), valve/port flow discharge coefficient, intake port heating, engine delta P, and intake manifold gas temperature. The characteristics of engine volumetric efficiency are further discussed in Fig. 13.10.

In order to achieve cost-effective designs, a ‘family’ of different blades with various tip widths and eye diameters are produced for one compressor impeller wheel ‘frame size’ (or casing). The compressor frame size is determined by the requirement of engine maximum air flow rate. The design parameters such as wheel inducer diameter, wheel and shroud contours, and diffuser width determine the compressor ‘trim’. These trims offer different flow characteristics of surge and choke as well as variations in efficiency. Compressor efficiency and shaft speed characteristic curves on the map can be further fine-tuned by using different trims with changes in impeller splitter blade outlet angle, diffuser entry angle, and diffuser geometry. For example, changing the diffuser entry angle or blade eye diameter can ‘rotate’ the map curves around the origin of the coordinates so that the compressor surge line moves. It should be noted that changing the aerodynamic design also affects the compressor tip speed and the stress. Turbocharger design details are provided by Rodgers and Rochford (2002), Watson and Janota (1982), and Watson (1999).

A typical compressor map is shown in Fig. 13.7. The engine operating points shown as the smaller white dots at different engine speeds and ambient conditions reflect the engine operation with a single-stage fixed-geometry turbocharger. The compressor pressure ratio of the engine operating points increases as the altitude increases.



13.7 Illustration of compressor map and engine operation.

Not all the locations from A to H shown on the compressor map in Fig. 13.7 can be operated in reality. The engine can run at locations E and D stably. Although operating at locations B and C is possible, they are in choke and the compressor efficiency is very low. Choke is a relative term and is usually defined at a given compressor efficiency (normally around 55%). In the choke region, engine performance becomes very poor. Note that the three main factors affecting the compressor power are air flow rate, compressor efficiency, and pressure ratio. The compressor power becomes very high when the efficiency is low for delivering a required boost pressure. This would require the turbine to deliver a high power with a large expansion ratio, which would increase the exhaust manifold pressure and the pumping loss. The location B has a very poor efficiency (30–35%), but the pressure ratio is low. The location C has a much higher efficiency, but the pressure ratio is higher. Their compressor power may be different. Another consideration when comparing locations B and C is that the turbocharger thrust bearing (axial) loading will be high at location B since there will likely be a high pressure behind the turbine wheel and a much lower pressure behind the compressor wheel. This could potentially cause a thrust bearing failure.

The operating locations F, G, H, and A in Fig. 13.7 cannot be run stably in reality. Without redrawing another compressor map for simplicity, one example is to imagine these four locations are intended for the operation of a small high-pressure-stage (HP) compressor in a two-stage series sequential turbocharger system. The HP compressor can be sized very small to enable good transient performance at part load. At rated power, the HP turbine wastegate needs to be opened in order to avoid over-boosting the engine. In reality, the HP compressor will act like a large flow restriction at higher engine speeds, and all the required engine air flow will not be able to flow through the HP compressor at locations F, G, H, or A. In this case, not only is a HP turbine wastegate required, a HP compressor bypass valve is also needed to bypass the engine air from the HP compressor. If the HP turbine wastegate is open but the HP compressor bypass is closed, the location A cannot be run stably without choke although the HP turbocharger still spins and some small percentage of the flow can still go through the HP compressor. The turbine speed would drop dramatically since there is little or no HP turbine power. The compressor needs to run at the same shaft speed as the turbine (e.g., very low at 10,000 rpm). Such a low speed line on the compressor map would give a very small value of choked flow. The point of compressor pressure ratio equal to one on the 10,000 rpm speed line is the maximum compressor flow that can travel through the HP compressor at that shaft speed. And this value is much less than the flow rates of the higher-shaft-speed points at locations F, G, H, and A. In other words, it is less than the flow rates required to support the engine operation at high speeds/loads. Also note that when the HP compressor bypass is open, the compressor pressure ratio will actually be below 1 since there will be a slight pressure drop across the bypass valve. In this case, the compressor slightly expands the air rather than compressing it.

Centrifugal compressor matching guidelines are summarized as follows:

- The matching should suit the normal operating environment of the engine, and should also be conducted at extreme ambient conditions from sea-level altitude and cold winter to high altitude and hot summer with minimum fueling derating to ensure the matching is within all the design constraints.
- Always use a single-stage turbocharger whenever possible in order to reduce cost, weight, and complexity.
- Provide proper pressure ratio and air flow rate at all engine speed–load modes, including zero-EGR conditions (e.g., fast transient, aftertreatment regeneration, cold climate).
- There should be no compressor surge along the full-load lug curve, especially below the peak torque speed. A proper surge margin (10–15% of the surge flow rate on the compressor map) should be reserved for engine transients, intake flow variations at various ambient temperatures (e.g., cold climate) and high altitude conditions, as well as for the extreme case of high intake restriction caused by air filter blockage.
- The compressor flow range should be wide enough so that there is no compressor choke and over-speeding at all altitude levels and ambient temperatures (i.e., no sonic flow inside the compressor and no operation in the very-low-efficiency region such as less than approximately 55%).
- Provide an acceptable turbine inlet temperature with a sufficiently high air–fuel ratio and EGR rate. High compressor efficiency is desired to reduce the compressor discharge temperature.
- Full-load and typical vehicle driving modes should be matched in the high-efficiency region on the compressor map in order to reduce the pumping loss. Minimum fuel consumption in a given driving cycle can be used as the optimization target for the compressor matching in this case.
- The compressor's moment of inertia should be as low as possible in order to reduce the turbocharger lag to achieve good engine transient acceleration performance.
- Compressor durability life needs to be acceptable. The turbocharger life is determined mainly by the low-cycle-fatigue (LCF) life of the compressor wheel under cyclic loading in relatively slow driving or usage cycles (see Chapter 2 for details). The compressor impeller life is dictated by the following: its maximum tip speed; the difference between the maximum and minimum shaft speed in typical driving cycles or aftertreatment regeneration cycles; and the cycle frequency. Those

factors can be shown in the load cycle diagram of 'turbocharger speed vs. driving time'. In turbocharger matching, the maximum turbocharger speed limit is determined by calculating the compressor fatigue life based on transient vehicle duty cycle, strain, impeller material (aluminum or titanium), and historical warranty data (Ryder *et al*., 2002). Maximum fueling needs to be restricted at high engine speeds in order to limit the maximum turbocharger speed.

- There should be no oil leakage from the bearing area to the compressor section that is caused by an improper pressure differential across the compressor for a given sealing design.

Centrifugal compressor design is discussed in Flaxington and Swain (1999), Japikse (2001), Japikse and Platt (2004); Japikse (1996), Pan *et al*. (1999), Moroz *et al*. (2005); McCutcheon (1978), Came and Bellamy (1982), Japikse (2000), Yamaguchi *et al*. (2002), Griffith (2007), Kuiper (2007), and Jiao *et al*. (2009). Compressor losses (e.g., tip clearance effect) are discussed in Spraker (1991), Sharp *et al*. (1999), Casey (2007), and Nili-Ahmadabadi *et al*. (2008).

13.4.3 Turbine matching, aerodynamic design, and durability

The turbine flow characteristic map shows that the turbine corrected flow rate is a strong function of turbine pressure ratio and a weak function of turbine speed. The turbine map can be generated from a flow bench test by regulating the turbine inlet pressure and the compressor power (load). The process is repeated until each shaft speed line is obtained. Turbine efficiency is a strong function of the blade speed ratio v_T/C_{T0} and a weak function of pressure ratio, usually with high efficiency located at a large pressure ratio and a high flow rate. Denote v_T as the tip speed of a radial turbine rotor,

$$v_T = d_T N_T \tag{13.2}$$

and denote C_{T0} as a theoretical velocity that would be achieved if the gas had expanded isentropically from the turbine inlet condition to the outlet pressure. C_{T0} can be calculated as

$$C_{T0} = \sqrt{2c_p T_{03}\left[1 - \left(\frac{p_4}{p_{03}}\right)^{(\kappa_t - 1)/\kappa_t}\right]} \tag{13.3}$$

Usually, a larger turbocharger has a higher efficiency because of its lower clearance-to-wheel diameter ratio and less associated leakage. It should be noted that the measured turbine map efficiency is the adiabatic efficiency multiplied by the mechanical efficiency. The efficiency map is significantly affected by the heat transfer from the turbine to the compressor via the turbocharger housing at low flow rates. The heat transfer decreases the measured turbine inlet temperature and increases the compressor outlet temperature, which causes the compressor efficiency artificially low and the turbine efficiency artificially high. The turbine efficiencies measured by different turbocharger manufacturers may not be directly comparable because of differences in their measurement methods and gas stand designs. For example, different turbine inlet temperatures can cause different heat transfer effects.

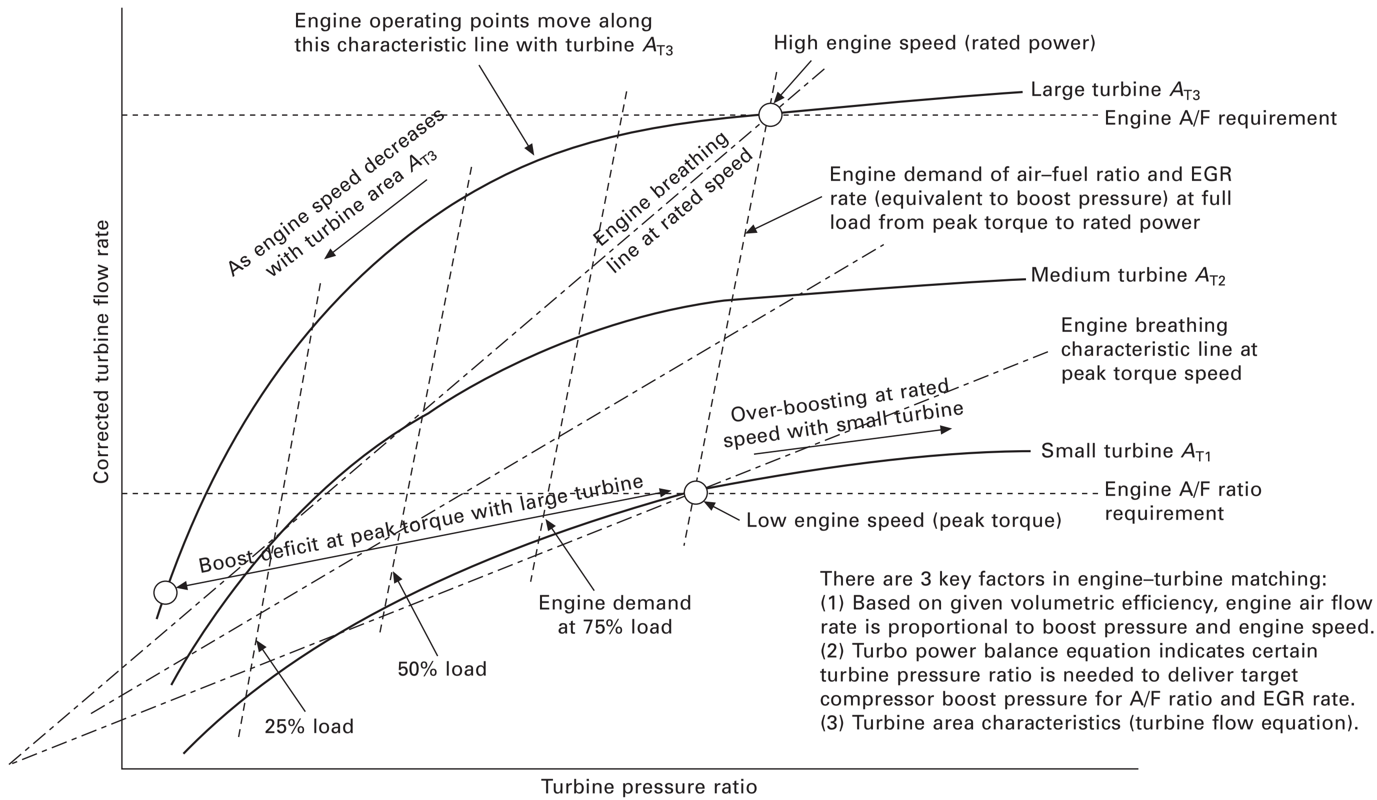
The turbine flow range is determined by an '*A*/*R*' ratio, which is the ratio of the smallest area of turbine housing inlet passage to the radial distance from the shaft centerline to the centroid of the inlet area. A smaller *A*/*R* ratio or turbine cross-sectional area produces a higher tangential gas velocity impinging on the turbine blades. As a result, a higher turbine speed, pressure ratio, and compressor boost pressure can be achieved. However, a small *A*/*R* also increases the exhaust manifold pressure and the pumping loss. Once the compressor frame size and the turbine rotor wheel size are selected, a range of casings with different volute cross-sectional areas or different nozzle stator rings (with various blade angles) can be chosen for the rotor, depending on the engine flow requirement. In VGT, either the *A*/*R* ratio or the nozzle vane is adjustable, discretely or continuously. A VGT may have lower efficiency than a fixed-geometry turbine due to larger clearances and flow disturbances. Figure 13.8 illustrates the principle of matching the turbine with the engine. In the figure, the cross-over point between the line of engine breathing characteristics at a given engine speed and the line of the air–fuel ratio requirement determines the required turbine effective area.

The turbine matching guidelines are summarized as follows:

- The turbine size needs to be small enough to drive EGR flow at low engine speeds (e.g., peak torque), and appropriate to deliver sufficient power to drive the compressor to produce the required air–fuel ratio for peak torque and lower speeds at full load. The turbine effective area, which is controlled by the nozzle ring or volute, needs to be selected to provide the best trade-off between high speeds/loads and low speeds/loads. If a small turbine area is sized for peak torque, wastegating is probably needed at rated power in order to prevent over-boosting where the peak cylinder pressure exceeds the structural limit. Note that the peak compression pressure in the cylinder may be estimated by

  $$p_{compression} \approx p_{boost}\,\Omega^{1.36} \qquad 13.4$$

  where Ω is the engine compression ratio. If the pumping loss associated with a wastegated turbine is not acceptable, the more sophisticated VGT may be considered at a higher cost.
- The necessity of turbine wastegating at various ambient conditions needs to be checked to prevent turbocharger over-speeding.



13.8 Engine–turbine matching theory.

- There should be no turbine choke. In pulse turbocharging, it is necessary to size the turbine frame large enough to account for the highest instantaneous flow rate.
- Appropriate turbine efficiency is desired, being neither too high nor too low, to deliver the turbine power to match the engine needs of air–fuel ratio and EGR rate. Occasionally, the turbine or compressor is deliberately matched in the low-efficiency region for the purpose of holding down the boost pressure and hence the maximum cylinder pressure below the design limit at high engine speeds, or for the purpose of creating a high engine delta P with a small turbine area to drive EGR at low speeds.
- Some special EGR configurations require the turbocharger efficiency to be as high as possible, and in such circumstances the turbocharger may become the limiting factor in engine development.
- There should be no size mismatch or speed mismatch between the compressor and the turbine for good aerodynamic performance. Driving a large amount of EGR at low engine speeds requires a small turbine area to build up a high engine delta P, and the corrected turbine flow rate may be much lower than the corrected compressor flow rate. On the other hand, the requirement of high air flow rate at rated power may demand a large compressor. Matching a relatively large compressor with a very small turbine will give poor turbocharger performance due to their speed mismatch. If their size difference is too large, the turbine efficiency may decrease, and the turbocharger bearings may wear due to excessive axial loading. Possible counteraction includes balancing trim size (turbine exducer or compressor inducer flow area) and impeller size between the turbine and the compressor (Arnold, 2004).
- For fast transient performance, the turbine size (or A/R ratio) or the moment of inertia should be as small as possible in order to reduce turbocharger lag, for example by using light ceramic turbine wheels.
- Pulse turbocharging with proper turbine entry may be considered to better utilize the exhaust energy to increase the air–fuel ratio at part load and improve transient acceleration performance.
- The turbine power split between the stages in two-stage turbocharging needs to be optimized for EGR driving, BSFC, compressor outlet temperature control, and the minimum risk of compressor surge or choke in the two stages. The transition between the two stages at different engine speeds needs to be smooth and should not cause a high engine delta P.
- Turbine durability life needs to be acceptable. The steel turbine wheel and iron housing must withstand the exhaust manifold gas temperature. The temperature limit is governed by the creep and scaling properties of the turbine wheel and housing materials.

Turbine design is introduced in Moustapha *et al.* (2003), Okazaki *et al.* (1982), Hussain and Bhinder (1984), Spraker (1992), Baines (2002, 2005a), Rodgers (2003), and Zhang *et al*. (2007). Turbine losses are discussed in Spraker (1987) and Keshavarz *et al*. (2003). Turbine pulsating flow unsteady performance is addressed in Bhinder and Gulati (1978), Ehrlich *et al*. (1997), Gu *et al*. (2001), Macek *et al*. (2002), Lam *et al*. (2002), Winkler *et al*. (2005), Capobianco and Marelli (2006), and Macek and Vítek (2008). VGT aerodynamic performance modeling is discussed in Kessel *et al*. (1998) and Luján *et al*. (2006).

13.4.4 Turbocharger performance at extreme ambient conditions

The turbocharger performance at high altitude can be explained by the four 'core equations' of the air system, 4.40, 4.44, 4.47, and 4.57 (or combined in equation 13.1). As the air density and air mass flow rate reduce, the turbine inlet temperature increases due to the lower air–fuel ratio. As ambient pressure falls, the turbine pressure ratio increases. Therefore, the turbocharger speed and the compressor pressure ratio increase so that the inlet air density reduction is partially compensated to alleviate fueling/power derating. At hot ambient, the air–fuel ratio decreases, and the operating limits are usually due to smoke and turbine inlet temperature. At cold ambient temperatures, the maximum cylinder pressure or compressor surge may be the limiting factors. In summary, large variations of ambient temperature or altitude level may cause problems in turbocharging such as compressor surge, choke, over-boosting, compressor over-speeding, excessively high exhaust manifold gas temperature, low air–fuel ratio, and high smoke. The situation is further complicated by the EGR strategy used at various ambient conditions. For example, the compressor may choke when the EGR valve is fully closed at high altitude. Moreover, turbocharger operation at high altitude affects the strategy of engine fueling derating.

Turbocharged diesel engine performance at high altitude is analyzed by Dennis (1971), Wu and McAulay (1973), and Bi *et al*. (1996). The aerodynamic performance of centrifugal compressors at high altitude is investigated in Wiesner (1979) and Strub *et al*. (1987).

13.5 Exhaust manifold design for turbocharged engines

Exhaust manifold design is an important part of diesel engine air system design, along with the designs of valvetrain and EGR system as well as turbocharging. The following topics need to be considered in the exhaust manifold design for modern EGR engines: (1) exhaust manifold configurations;

(2) turbine entry (divided or undivided); (3) EGR pickup location and EGR driving potential. The exhaust manifold configurations include divided and undivided manifolds, and the manifolds used for pulse turbocharging and constant-pressure turbocharging. The exhaust manifold design analysis is usually conducted by using the simulation software of one-dimensional gas wave dynamics coupled with engine cycle simulation. High exhaust energy utilization, minimum pumping loss and good part-load and transient performance are key objectives in exhaust manifold design.

The efficiency of the turbocharging system is characterized by

$$\eta_{ts} = \eta_{cyl}\eta_{TC} \tag{13.5}$$

where η_{cyl} is the efficiency of the energy transfer from the cylinder to the turbine inlet, and η_{TC} is the turbocharger efficiency, which is equal to $\eta_T\eta_C\eta_{TC,mech}$. Usually, at the same engine speed, η_{cyl} decreases as the engine load decreases; and at the same engine power, η_{cyl} decreases as the engine speed decreases. Pressure loss is often a more convenient parameter to use than η_{cyl} in engine design. The following measures can increase the efficiency of pulse energy utilization and η_{cyl}:

1. Use fast exhaust valve opening, large exhaust valve diameter, and high valve lift. During the exhaust stroke, it is necessary to reduce the throttling loss of the supersonic flow across the exhaust valve and the flow restriction or friction in the exhaust manifold in order to reduce the pumping loss.
2. Use appropriate exhaust runner grouping and turbine entry design.
3. Optimize the cross-sectional area of the exhaust manifold pipe for the best effect of pressure wave dynamics.
4. Use short exhaust runner and manifold piping without sharp turns or abrupt changes in the cross-sectional area.

Divided exhaust manifold together with pulse turbocharging has been used for some non-EGR engines and EGR engines. Design simplifications to the undivided exhaust manifold systems often require evaluation of their impact on EGR driving and BSFC, for example, comparing a divided exhaust manifold of an inline six-cylinder engine having three larger exhaust pressure pulses in the manifold with an undivided manifold having six smaller pressure pulses. Other examples are to compare the pulse difference caused by divided and undivided turbine entries, or to compare the optimum EGR pickup location for different types of exhaust manifolds and turbine entries. The following aspects need to be considered in the exhaust manifold design.

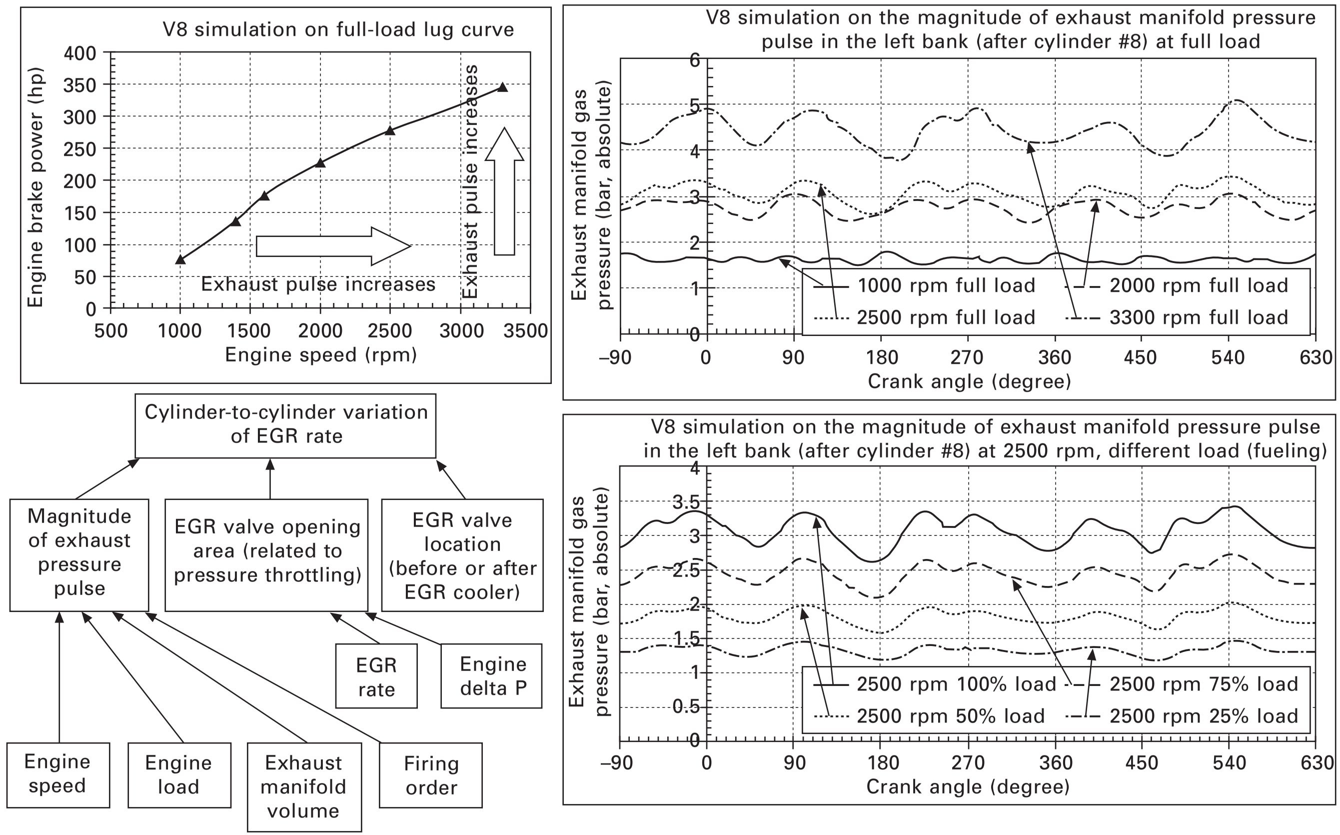
1. The effect of the exhaust pressure pulses on volumetric efficiency and engine delta P.
2. The effect of the exhaust pressure pulses on transient corrected turbine flow rate and turbine efficiency.

3. The relationship between the pressure pulses and the EGR driving capability. This topic needs to be analyzed for two scenarios: with and without a reed valve. For low-EGR engines, at the peak torque condition, the EGR rate can be so low that the gas pressure pulses at the EGR circuit inlet (exhaust manifold back pressure) in the crank angle domain cross over with the intake manifold pressure. In this case, the use of a reed valve may harvest more EGR flow by eliminating the reverse flows at the crank angle locations where the back pressure pulse is lower than the boost pressure. The greater the exhaust pulse, the more effective the reed valve. However, for high-EGR engines, a greater cycle-average pressure differential is required to drive the high EGR flow rate so that the back pressure and boost pressure become far apart from each other and do not cross over in the crank angle domain. In this case, the reed valve cannot help to obtain more EGR and hence does not affect the comparison between the two exhaust manifolds.
4. The margin between the maximum peak of the exhaust pressure pulse and the allowable design limit. The undivided exhaust manifold, having smaller (less violent) pressure pulses, can reach the design limit pressure with a higher cycle-average manifold pressure than the divided manifold, resulting in a higher air–fuel ratio capability at rated power.

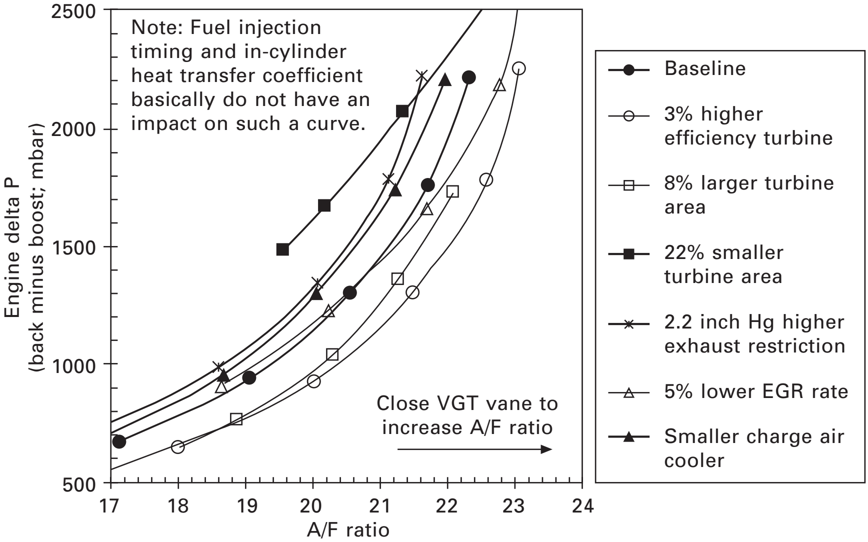
Figure 13.9 illustrates the pulsation characteristics of the exhaust manifold pressure at different engine speeds and loads for a V8 heavy-duty diesel engine. Engine exhaust manifold design and analysis are summarized by Winterbone and Pearson (1999, 2000) and are researched by Tabaczynski (1982), Capobianco *et al*. (1993), Bassett *et al*. (2000), Lakshmikantha and Kec (2002), Fu *et al*. (2005), and He *et al*. (2006). The effect of turbine entry design is studied by Hribernik (1997). Pulse converters were researched by Bassett *et al*. (2000), Yang *et al*. (2006), and Zhang *et al*. (2008b)

13.6 The principle of pumping loss control for turbocharged exhaust gas recirculation (EGR) engines

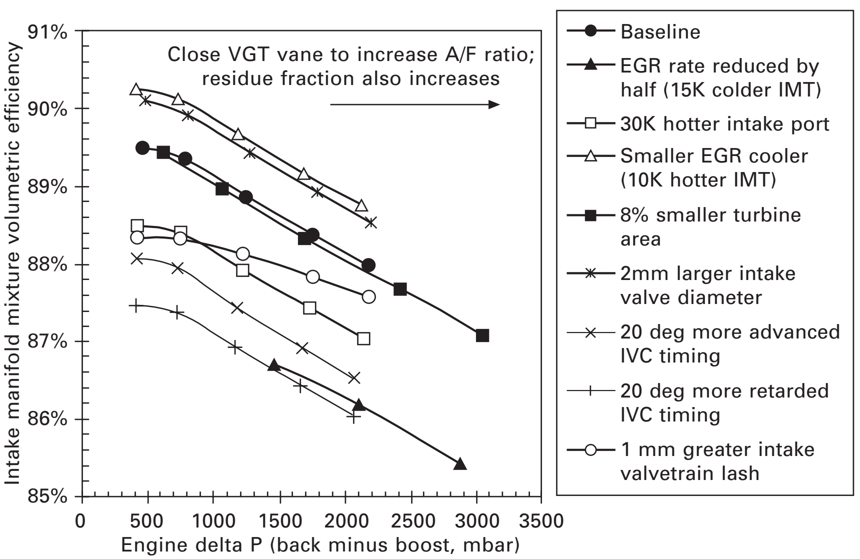
As discussed in Chapter 4, pumping loss consists of the contributions from engine delta P and intake manifold mixture volumetric efficiency. Using simulation data, Figs 13.10 and 13.11 illustrate the design effects on engine delta P and volumetric efficiency, respectively. In the figures, the EGR rate is kept constant (unless especially mentioned) by regulating the EGR valve opening. Figure 13.10 shows that the engine delta P is largely affected by air–fuel ratio when VGT vane opening (turbine area) is regulated. In fact, if the air–fuel ratio is varied by changing the turbocharger efficiency without changing the turbine area, the engine delta P is basically not affected.



13.9 Exhaust manifold pressure pulse and the impact of engine speed and load.



13.10 Design effects on engine delta P at rated power.

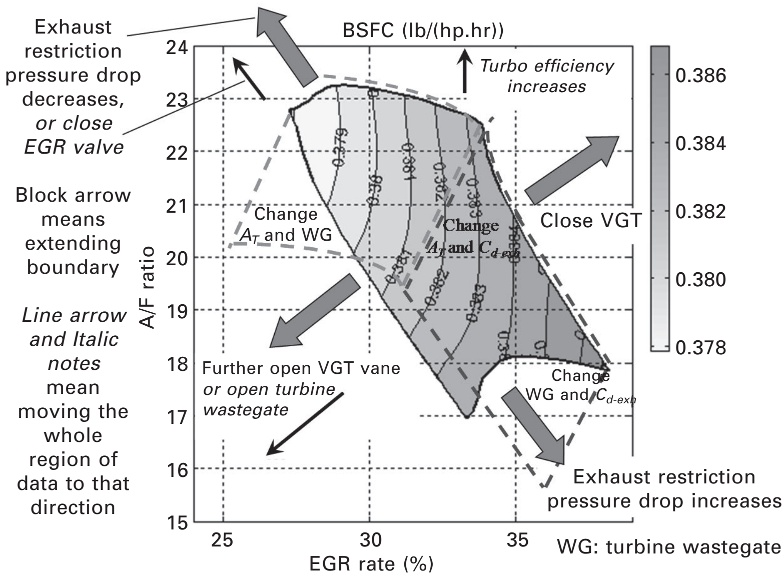


13.11 Design effects on engine volumetric efficiency at rated power.

From Fig. 13.11 it is observed that with a given valvetrain and cylinder head design, the volumetric efficiency is strongly affected by engine delta P because the in-cylinder internal residue fraction can be strongly affected by the pressure differential between the exhaust port and the intake port, depending on the valve overlap size. As the residue fraction increases, the volumetric efficiency decreases. Intake manifold gas temperature also strongly affects the volumetric efficiency. At the same intake manifold mass flow rate, a hotter charge actually results in a higher value of intake manifold volumetric efficiency. The hotter charge can be produced by a smaller EGR cooler or a smaller charge air cooler. Therefore, when comparing the volumetric efficiency values of different engines to try to assess the design effects of the cylinder head, the valvetrain, and the manifolds, it is very important to align the values at similar or comparable intake manifold gas temperatures and engine delta P values in order to avoid misleading conclusions.

In order to meet combustion and emissions requirements, there is a precise design target of air–fuel ratio and EGR rate at each engine speed and load in steady state. Ideally, the EGR valve should be set fully open all the time with minimum flow restriction in order to minimize the pumping loss. But in reality at some speeds and loads the EGR valve has to be set partially closed for the following reasons: (1) insufficient turbocharger efficiency and the consequent use of a small turbine area and the resulting high engine delta P; (2) necessary smooth transition to EGR shut-off conditions; (3) long actuator response time of the EGR valve when the VGT vane is suddenly closed during the fast transient. If the turbocharger efficiency is not high enough, the exhaust manifold pressure must be made high by using a small turbine area to produce a sufficient turbine pressure ratio and turbine power to deliver the target air–fuel ratio. This results in a high engine delta P that forces the EGR valve to be partially closed (throttled) undesirably. On the other hand, if the turbo efficiency is too high, the exhaust manifold pressure must be lowered by using a larger turbine area to reduce the turbine power to prevent over-boosting the air–fuel ratio. In that case, the engine delta P may become too low and cannot drive sufficient EGR. The root cause of the inappropriate engine delta P is the fact that turbocharger efficiency has its inherent characteristics in the turbomachinery design. The efficiency cannot match engine needs flexibly at every speed and load. Generally, closing the EGR valve results in a decrease in the EGR rate and an increase in the air–fuel ratio. Closing the VGT vane or the turbine wastegate may result in an increase in both the EGR rate and the air–fuel ratio. Increasing the turbocharger efficiency results in an increase in the air–fuel ratio basically without changing the EGR rate. Closing the exhaust back-pressure valve or the intake throttle results in a large reduction in the air–fuel ratio and a small increase in the EGR rate. Figures 13.12 and 13.13 use engine cycle simulation data to illustrate the ability of flexibly controlling the air–fuel ratio and the EGR rate with different air system options.

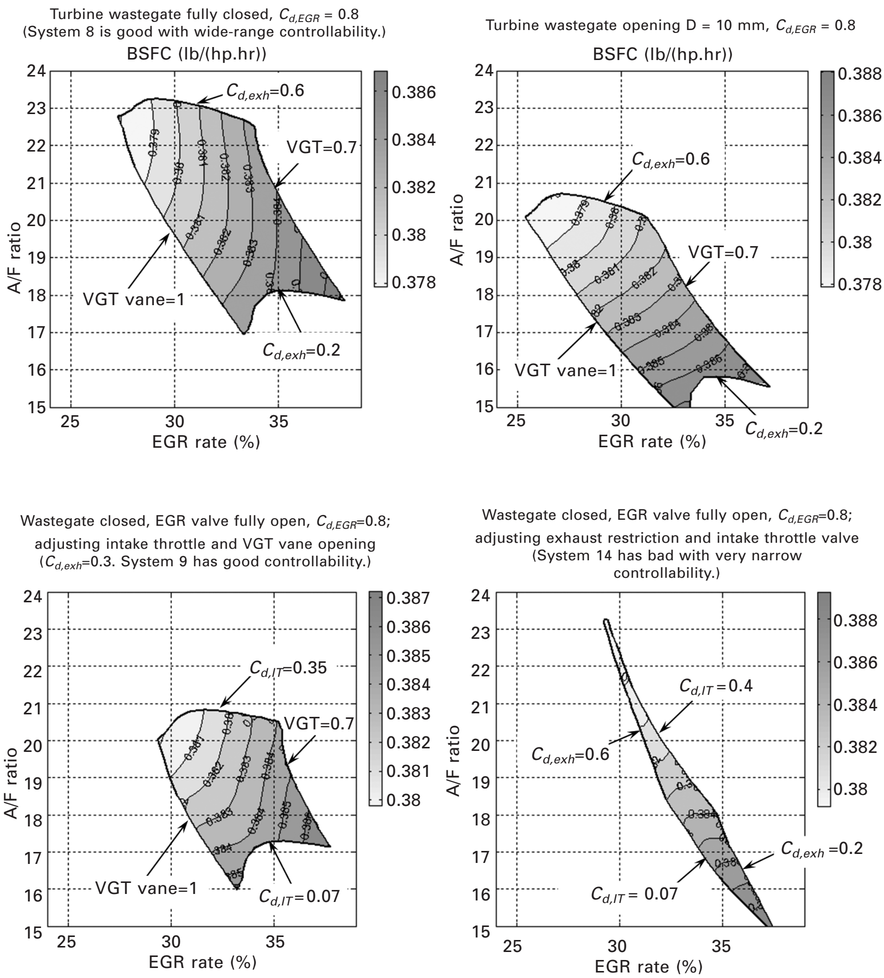
When evaluating how many 'knobs' are needed in air system design or how many controls need to be tuned to achieve air and EGR flows, the mathematical systems formulated in Table 4.1 can be used in any flexible combination to count the number of hardware control parameters to match with the number of unknowns in the system of equations. In this way, the air



13.12 Air system capability to control air–fuel ratio and EGR rate with different hardware.

system control does not become over- or under-constrained. Moreover, the relative effectiveness of each 'knob' (control parameter) can be compared by analyzing the equations in Chapter 4. For example, to reach a given pair of air–fuel ratio and EGR rate values, with low EGR circuit restriction (e.g., EGR valve fully open) and the fully closed turbine wastegate, the required turbine effective area A_T and turbocharger efficiency η_{TC} can be solved as two unknowns by using the equations in system 10. In other words, two design 'knobs' are needed: A_T and $\eta_{T/C}$. However, the turbocharger efficiency normally cannot be adjusted freely in the entire engine speed–load domain due to the turbocharger design constraints. Nor can the exhaust restriction due to the aftertreatment be freely adjusted. In that case, system 2 has to be used to solve for the required turbine area (e.g., VGT vane opening) and EGR valve opening in order to reach a pair of target air–fuel ratio and EGR rate values. That is the situation usually encountered by a calibration engineer during the engine calibration stage once the air system hardware has been finalized. It should be noted that closing the EGR valve raises the engine delta P and results in a high pumping loss.

There are two reasons for using a turbine wastegate: (1) the required turbine effective area computed with system 10 at a low speed is smaller than that at a high speed; (2) the fast transient response demands low turbo inertia, which requires the turbine not to be sized too large so that at high speed a wastegate may be used to bleed off the exhaust. System 3 can be used to solve for $\dot{m}_T$ or essentially the wastegate opening. In system 13, the



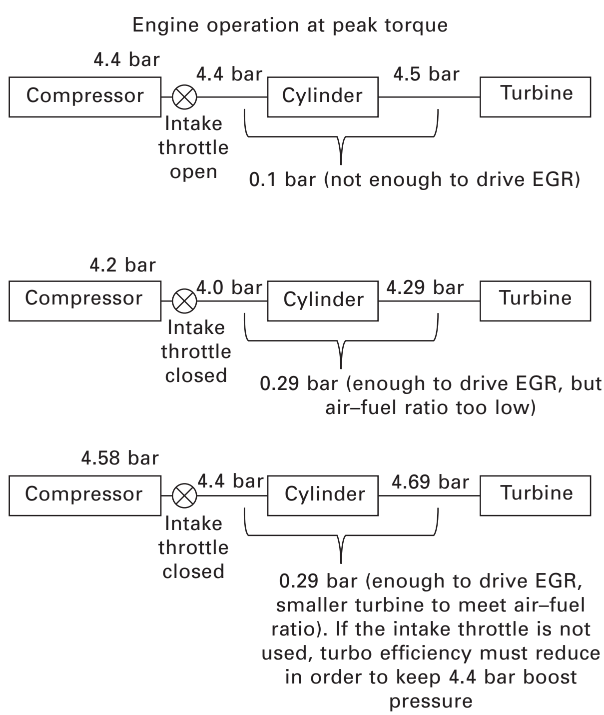
13.13 Air system capability and controllability (BSFC contour maps).

turbine area is known, and the turbine wastegate opening and the turbocharger efficiency are solved as unknowns for the target air–fuel ratio and EGR rate values. If the turbine area, the turbocharger efficiency, and the EGR circuit flow restriction have to be treated as fixed known inputs, turbine wastegate and exhaust back-pressure valve (or intake throttle) need to be used (systems 11 and 12).

Intake throttle is sometimes used to induce more EGR or reduce air–fuel ratio. It should be noted that if the intake throttle is the only 'knob' available to adjust performance, either the EGR rate or the air–fuel ratio can be set as a target, but not both. For example, in system 14, at peak torque, the EGR valve is set fully open, the turbine wastegate is fully closed, and the turbine nozzle area and the turbocharger efficiency are fixed known parameters. If the exhaust restriction $C_{d,exh}$ is also fixed, then the only adjustable 'knob' is the intake throttle. Figure 13.14 provides an illustration to explain the function of intake throttle.

From the above theory, changing any two of the following six parameters can reach a given pair of values of target air–fuel ratio and EGR rate: (1) EGR circuit flow restriction; (2) turbine effective cross-sectional area (or VGT vane opening); (3) turbine wastegate; (4) turbine or compressor efficiency; (5) exhaust restriction or back-pressure valve opening; and (6) intake throttle opening. However, their effects on engine delta P are different. A large turbine area or wastegating is required in order to reduce the engine delta P, meanwhile a less restrictive EGR circuit flow restriction is required in order to maintain the same EGR rate. If the exhaust restriction cannot be reduced, higher turbocharger efficiency is required in order to maintain the same air–fuel ratio. If EGR cooling is enhanced, the same emissions can be achieved with a lower EGR rate, and this may reduce the engine delta P.

The effects of the coordinated air system controls (e.g., VGT and EGR system, intake throttle) for optimizing emissions and performance are explored by Boulouchos and Stebler (1998), Hawley *et al.* (1999), Chi *et al.* (2002), Pfeifer *et al.* (2002), Senoguz *et al.* (2008), and Adachi *et al.* (2009).



13.14 Illustration of the function of intake throttle.

13.7 Turbocompounding

Turbocompounding is an important waste heat recovery technology. It uses a compound turbine at the downstream of the turbocharger turbine to extract mechanical work from the exhaust gas and put this work back to the engine crankshaft in order to increase the thermal efficiency of the engine. It should be noted that the effectiveness of turbocompounding depends on the availability of the exhaust gas, which is low for modern high-EGR engines, and the negative impact of the increased exhaust restriction on the turbocharger turbine performance. Diesel turbocompounding analysis is reported by Assanis and Heywood (1986), Tennant and Walsham (1989), Hopmann and Algrain (2003), Stobart and Weerasinghe (2006), and Hountalas *et al*. (2007).

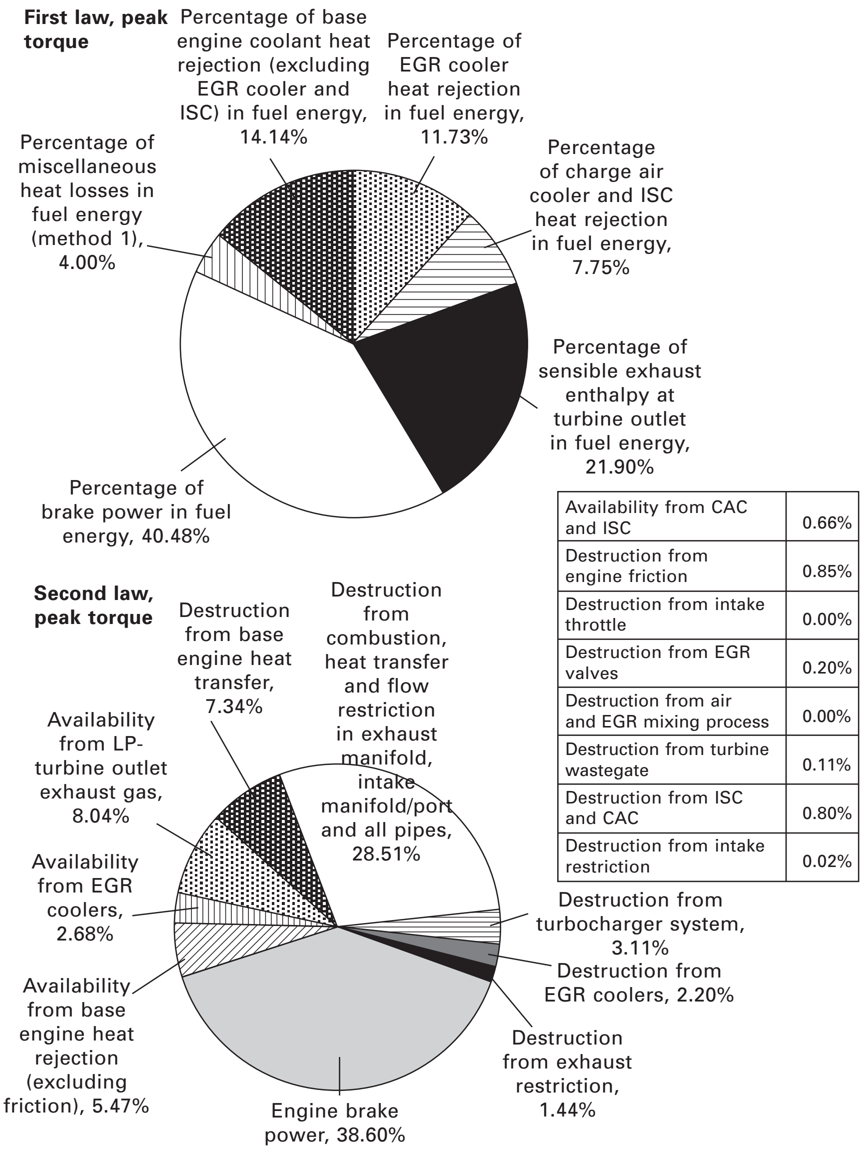
13.8 Thermodynamic second law analysis of engine system

While the first law of thermodynamics (i.e., energy balance in 'quantity') has been widely used in diesel engine system design and analysis, the second law of thermodynamics (i.e., the 'quality' of the usable energy) is in fact equally or even more important in system design. During the course of engine system design, it is important to understand the ultimate potential of design improvement in system efficiency and to compare the quality of system design from an energy quality point of view. Modern diesel engines have the following areas that need guidance from the second law analysis:

1. The potential of waste heat recovery from the exhaust and cooling systems (e.g., turbine outlet exhaust gas, EGR heat, charge air cooler heat).
2. The ultimate potential of design improvement in engine air system.
3. The root causes of high losses in the engine system.
4. The comparison of system/component losses between different hardware configurations (e.g., different turbocharger systems, EGR systems, air system control valves, aftertreatment systems that have different exhaust restriction levels).

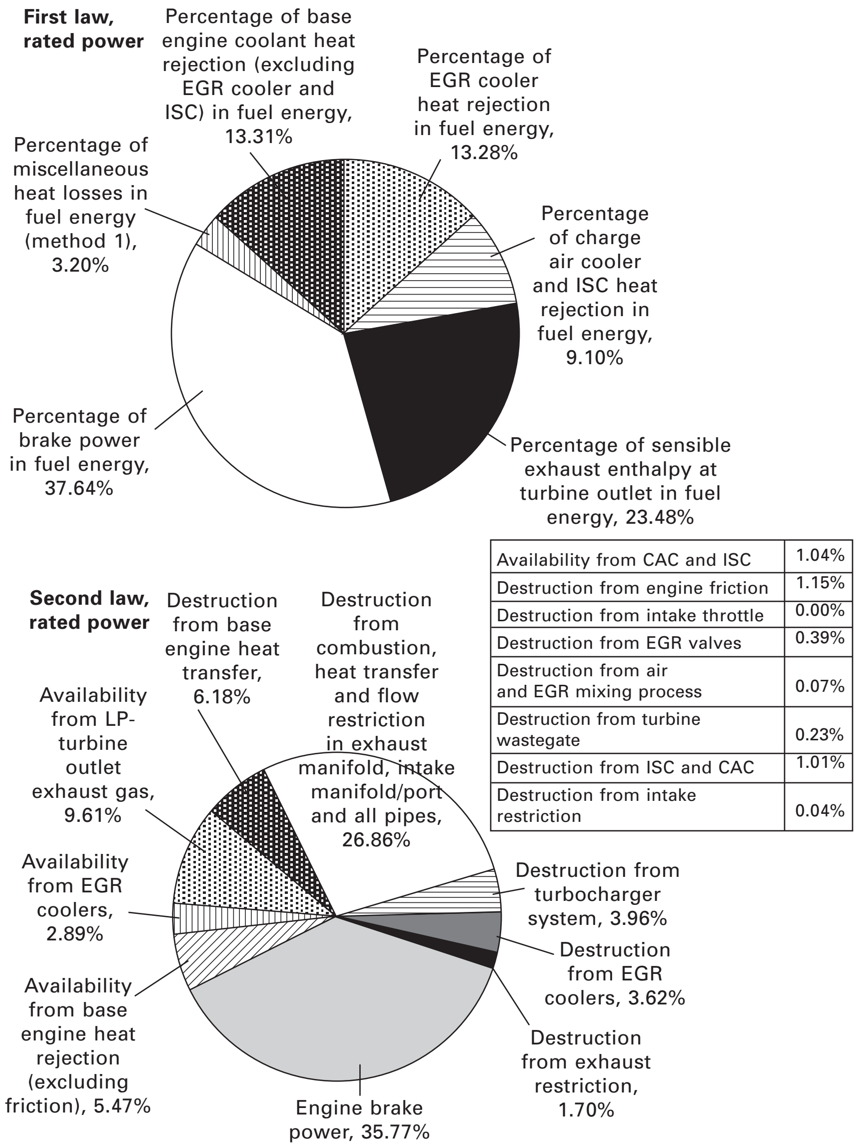
Advanced engine system analysis should be conducted using the second law of thermodynamics to address these issues. The second law analysis is able to quantify various losses occurring in different subsystems (e.g., combustion, air, cooling, friction) and point out the root causes of the system deficiency. For cooling system components, the losses include the destructions from heat transfer, flow restriction, etc. For air system components, the losses include the destructions from component efficiency losses (e.g., low turbocharger efficiency), flow losses (e.g., exhaust and intake restrictions), and throttle losses at all the air system control valves (e.g., EGR valve, turbine wastegate valve, EBP valve, intake throttle valve, etc.).

Figures 13.15–13.17 present a comprehensive analysis of the comparison between the thermodynamic first law and second law analyses for a modern heavy-duty diesel engine based on engine cycle simulation data. The second



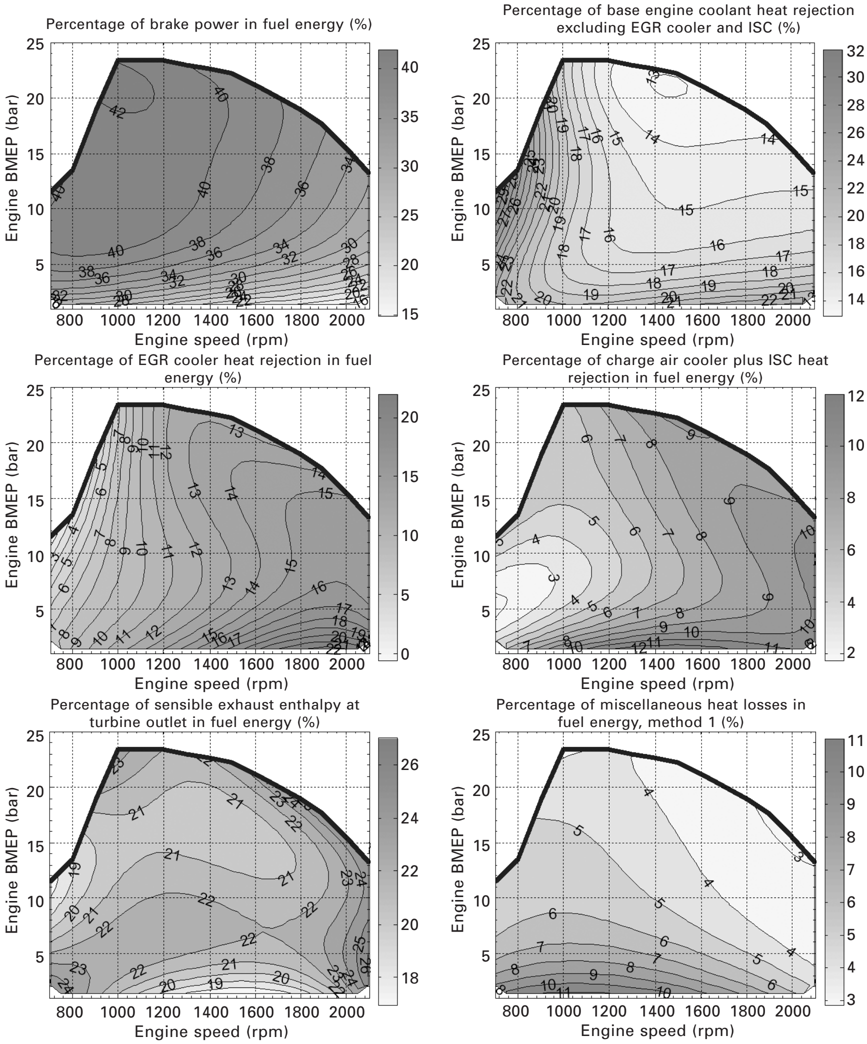
Availability from CAC and ISC	0.66%
Destruction from engine friction	0.85%
Destruction from intake throttle	0.00%
Destruction from EGR valves	0.20%
Destruction from air and EGR mixing process	0.00%
Destruction from turbine wastegate	0.11%
Destruction from ISC and CAC	0.80%
Destruction from intake restriction	0.02%

13.15 Thermodynamic first law and second law analyses at peak torque and rated power.



13.15 Continued

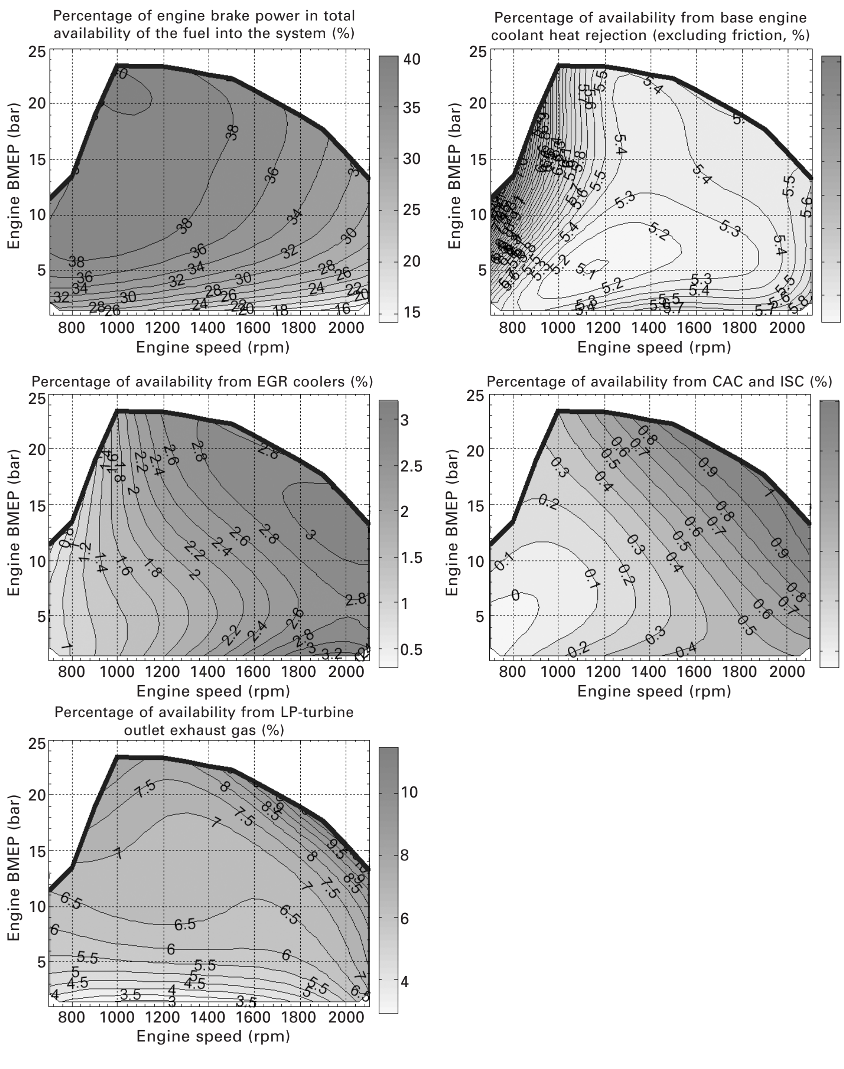
law analysis is conducted with the 'dead state' datum defined at 1 atm (1.013 bar) and 298.15 K (77°F). Figure 13.15 illustrates the breakdown percentages of the energy and availability/destruction distributions at the peak torque and rated power conditions. It is observed that the second law analysis is able to quantify more realistically the potential of each waste heat source than the first law analysis. For example, the first law analysis shows that at



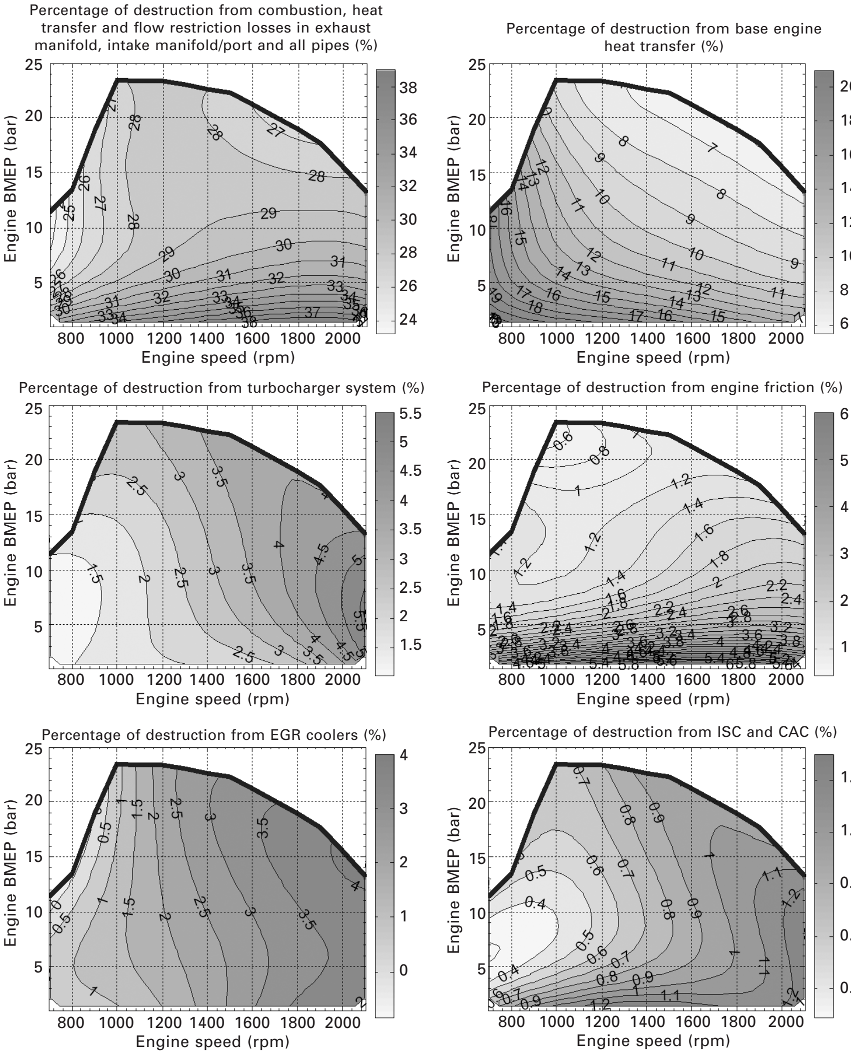
13.16 Thermodynamic first law analysis in engine speed–load domain.

rated power the EGR cooler heat rejection, the charge air cooler plus inter-stage cooler heat rejection, and the turbine-outlet exhaust gas energy occupy 13.28%, 9.10%, and 23.48% of the total fuel energy, respectively. The second law analysis points out that the availabilities of these three terms are 2.89%, 1.04%, and 9.61% in the total availability of the fuel, respectively.

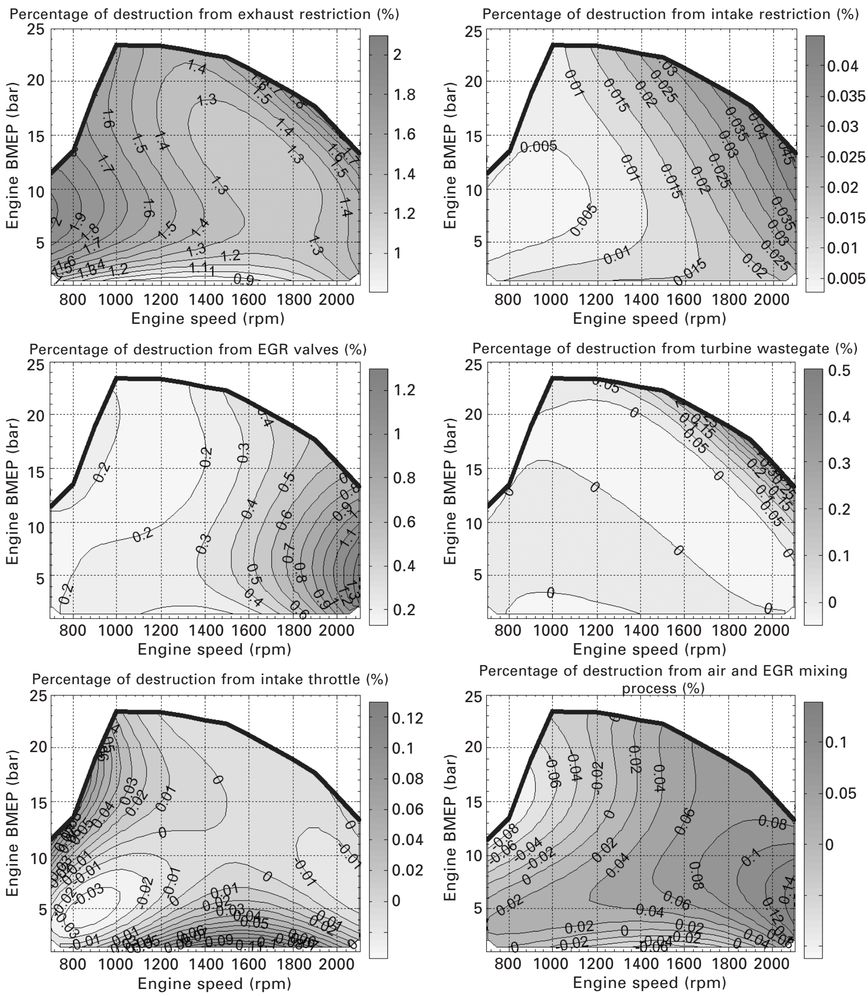
Moreover, it is observed that the second law analysis is able to reveal the source of system losses, compare the destructions of various subsystems/components and processes, and indicate the potential of design improvement for those losses. In contrast, the first law analysis is not able to achieve these. For example, the second law analysis indicates that at rated power the destruction in the turbocharger system, exhaust restriction, engine mechanical friction, EGR valve throttling, and turbine wastegate throttling are 3.96%, 1.70%, 1.15%, 0.39%, and 0.23% of the total availability of the fuel, respectively. This shows that the damage to the system caused by the high exhaust restriction in modern diesel engines' aftertreament devices is about the same as the total losses of engine friction, EGR throttling loss, and



13.17 Thermodynamic second law analysis in engine speed–load domain.



13.17 Continued



13.17 Continued

turbine wastegate throttling loss. And this destruction caused by the exhaust restriction is as high as 43% of the destruction of the turbocharger system.

This second law analysis indicates that large potential to improve diesel engine efficiency resides in the waste heat recovery from the turbine outlet exhaust gas and EGR gas, followed by improvement in turbocharger efficiency and reduction in exhaust restriction, cooler restriction, mechanical friction, and the throttle losses at the EGR valve and the turbine wastegate. If an intake throttle is used, it may create large or non-negligible destruction in the system. It should be noted that for a poorly designed air system the flow losses and throttle losses in the system can be much higher than the data presented in this case. Figures 13.16 and 13.17 present the complete analysis of the first law and second law breakdown maps in the engine speed–load domain. It illustrates that the second law analysis provides an in-depth understanding of the engine system behavior of the energy conversion processes, and is a powerful approach to guide the system and component design in the right direction to maximize the engine system efficiency.

Good reviews of the fundamentals of the second law of thermodynamics and engine analysis are provided by Flynn (2001) and Rakopoulos and Giakoumis (2006, 2009). A complete list of good references on the engine applications of the second law of thermodynamics are provided in Section 13.9.1.

13.9 References and bibliography

Adachi T, Aoyagi Y, Kobayashi M, Murayama T, Goto Y and Suzuki H (2009), 'Effective NO_x reduction in high boost, wide range and high EGR rate in a heavy duty diesel engine', SAE paper 2009-01-1438.

Akiyama M, Sakai H, Yamada T and Kanesaka H (1996), 'An elegant solution for vehicular diesel's emission and economy – hybrid EGR system', SAE paper 960842.

Amos R S (2002), 'Trends in turbocharger application', IMechE paper C602/038/2002.

Anada S, Kawakami T and Shibata N (1997), 'Development of SJ (swirl jet) turbocharger for diesel engine vehicles', SAE 970341.

Andersson S, Akerlund C and Blomquist M (2002), 'Low pressure EGR calibration strategies for reliable diesel particulate filter regeneration on HDD engines', SAE paper 2002-01-2780.

Andersen J, Karlsson E and Gawell A (2006), 'Variable turbine geometry on SI engine', SAE paper 2006-01-0020.

Arnold S D (2004), 'Turbocharging technologies to meet critical performance demands of ultra-low emissions diesel engines', SAE paper 2004-01-1359.

Arnold S, Slupski K, Groskreutz M, Vrbas G, Cadle R and Shahed S M (2001), 'Advanced turbocharging technologies for heavy-duty diesel engines', SAE paper 2001-01-3260.

Arnold S, Groskreutz M, Shahed S M and Slupski K (2002), 'Advanced variable geometry turbocharger for diesel engine applications', SAE paper 2002-01-0161.

Arnold S, Balis C, Jeckel D, Larcher S, Uhl P and Shahed S M (2005), 'Advances in turbocharging technology and its impact on meeting proposed California GHG emission regulations', SAE paper 2005-01-1852.

Assanis D N and Heywood J B (1986), 'Development and use of a computer simulation of the turbocompounded diesel engine system for engine performance and component heat transfer studies', SAE paper 860329.

Baert R S G, Beckman D E and Verbeek R P (1996), 'New EGR technology retains HD diesel economy with 21st century emissions', SAE paper 960848.

Baert R S G, Beckman D E and Veen A (1999), 'Efficient EGR technology for future HD diesel engine emission targets', SAE paper 1999-01-0837.

Baines N (2002), 'Radial and mixed flow turbine options for high boost turbochargers', *Seventh International Conference on Turbochargers and Turbocharging*, May 14–15, 2002. Available at: http://www.conceptsnrec.com/tech_papers.html

Baines N (2005a), 'Radial turbines: an integrated design approach', Concepts NREC, available at: http://www.conceptsnrec.com/tech_papers.html

Baines N C (2005b), *Fundamentals of Turbocharging*, Concepts NREC, White River Junction, VT.

Baker G L, Henderson G H and Free P D (2001), 'Multiple nozzle ejector for wastegated turbomachinery', US patent 6,202,413.

Banisoleiman K, Smith L A and French B A (1991), 'The interaction of diesel engine turbocharging and tuned inlet manifold systems under steady state and transient operation', *Proc. IMechE, Part A: Journal of Power and Energy*, 205(4), 269–281.

Bassett M D, Winterbone D E and Pearson R J (2000), 'Modelling engines with pulse converted exhaust manifolds using one-dimensional techniques', SAE paper 2000-01-0290.

Bhinder F S (1984), 'Some unresolved problems in the design of turbochargers', SAE paper 840018.

Bhinder F S and Gulati P S (1978), 'A method for predicting the performance of centripetal turbines in non-steady flow', IMechE paper C77/78.

Bi X, Zhang G and Zheng X (1996), 'Predicting vehicle turbocharged diesel engine performance at altitude', SAE paper 961826.

Boulouchos K and Stebler H (1998), 'Combustion features and emissions of a DI-diesel engine with air path optimization and common rail fuel injection', SAE paper 981931.

Bozza F, Cameretti M C, Senatore A and Tuccillo R (1997), 'Experimental investigation and numerical modelling of an advanced turbocharged DI diesel engine', SAE paper 970057.

Came P M and Bellamy A G (1982), 'Design and performance of advanced large turbochargers', IMechE paper C37/82.

Cantemir C-G (2001), 'Twin turbo strategy operation', SAE paper 2001-01-0666.

Cantore G, Mattarelli E and Fontanesi S (2001), 'A new concept of supercharging applied to high speed DI diesel engines', SAE paper 2001-01-2485.

Capobianco M and Marelli S (2006), 'Turbocharger turbine performance under steady and unsteady flow: test bed analysis and correlation criteria', *8th International Conference on Turbochargers and Turbocharging,* Woodhead Publishing Limited, Cambridge, pp. 193–206.

Capobianco M, Gambarotta A and Nocchi M (1993), 'Unsteady flow phenomena and volume effects in automotive engines manifolds', SAE paper 931897.

Carter J, Sharp N K and Tennant H (2010), 'Turbocharging technologies for heavy-duty diesel engines', in Zhao H (ed.), *Advanced Direct Injection Combustion Engine Technologies and Development, Volume 2: Diesel Engines*, Woodhead Publishing Limited, Cambridge.

Casey M V (2007), 'Accounting for losses and definitions of efficiency in turbomachinery stages', *Proc. IMechE, Part A: Journal of Power and Energy*, 221, 735–743.

Chapple P M, Flynn P F and Mulloy J M (1980), 'Aerodynamic design of fixed and variable geometry nozzleless turbine casings', *Transactions of the ASME*, *Journal of Engineering for Power*, 102, 141–147.

Chatterjee S, Conway R, Viswanathan S, Blomquist M, Klüsener B and Andersson S (2003), 'NO_x and PM control from heavy duty diesel engines using a combination of low pressure EGR and continuously regenerating diesel particulate filter', SAE paper 2003-01-0048.

Chi Y, Cheong J, Kim C and Choi K (2002), 'Effects of VGT and injection parameters on performance of HSDI diesel engine with common rail FIE system', SAE paper 2002-01-0504.

Choi C, Kwon S and Cho S (2006), 'Development of fuel consumption of passenger diesel engine with 2-stage turbocharger', SAE paper 2006-01-0021.

Dennis J W (1971), 'Turbocharged diesel engine performance at altitude', SAE paper 710822.

Ehrlich D A, Lawless P B and Fleeter S (1997), 'On-engine turbocharger turbine inlet flow characterization', SAE paper 971565.

Flaxington D and Szczupak D T (1982), 'Variable area radial-inflow turbine', IMechE paper C36/82.

Flaxington D and Swain E (1999), 'Turbocharger aerodynamic design', *Proc. IMechE, Part C: Journal of Mechanical Engineering Science*, 213, 43–57.

Flynn P F (1979), 'Turbocharging four-cycle diesel engines', SAE paper 790314.

Flynn P F (2001), 'How chemistry controls engine design', *Proceedings of the 2001 Fall Technical Conference of the ASME Internal Combustion Engine Division, Diesel Combustion and Emissions, Fuel injection and Sprays*, Volume 1, pp. 1–9.

Franklin P C and Walsham B E (1986), 'Variable geometry turbochargers in the field', IMechE paper C121/86.

Franklin P C (1989), 'Performance development of the Holset variable geometry turbocharger', SAE paper 890646.

Freeman P F and Walsham B E (1978), 'A guide to some analytical turbocharger matching techniques', IMechE paper C59/78.

Fu H, Chen X, Shilling I and Richardson S (2005), 'A one-dimensional model for heat transfer in engine exhaust systems', SAE paper 2005-01-0696.

Furukawa H, Yamaguchi H, Takagi K and Okita A (1993), 'Reliability on variable geometry turbine turbocharger', SAE paper 930194.

Gale N F (1990), 'Diesel engine cylinder head design: the compromises and techniques', SAE paper 900133.

Galindo J, Luján J M, Guardiola C and Lapuente G S (2006), 'A method for data consistency checking in compressor and variable-geometry turbine maps', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 220, 1465–1473.

Galindo J, Luján J M, Climent H and Guardiola C (2007), 'Turbocharging system design of a sequentially turbocharged diesel engine by means of a wave action model', SAE paper 2007-01-1564.

Ghadiri-Zareh M S and Wallace F J (1978), 'Variable geometry vs two stage turbocharging of high output diesel engines', IMechE paper C63/78.

Graf G, Hrauda G and Bartsch P (2000), 'Layout of a high load EGR system for LD, MD and HD truck engines by means of simulation', SAE paper 2000-01-0225.

Griffith R C (2007), 'Series turbocharging for the Caterpillar heavy-duty on-highway truck engines with ACERT™ technology', SAE paper 2007-01-1561.

Gu F, Engeda A and Benisek E (2001), 'A comparative study of incompressible and compressible design approaches of radial inflow turbine volutes', *Proc. IMechE, Part A: Journal of Power and Energy*, 215, 475–486.

Gurney D (2001), 'The design of turbocharged engines using 1D simulation', SAE paper 2001-01-0576.

Hawley J G, Wallace F J, Cox A, Horrocks R W and Bird G L (1999), 'Reduction of steady state NO_x levels from an automotive diesel engine using optimised VGT/EGR schedules', SAE paper 1999-01-0835.

He Y, Battiston P A and Alkidas A C (2006), 'Thermal studies in the exhaust manifold of a turbocharged V6 diesel engine operating under steady-state conditions', SAE paper 2006-01-0688.

Hirabayashi Y, Sumi Y and Nishiguchi F (1986), 'Development of Nissan variable geometry JET turbocharger', SAE paper 860105.

Hishikawa A, Okazaki Y and Busch P (1988), 'Development of variable area radial turbines for small turbochargers', SAE paper 880120.

Hochegger W, Hrauda G, Prenninger P and Seitz H F (2002), 'Effect of various EGR systems on HDT-turbocharging', IMechE paper C602/040/2002.

Hopmann U and Algrain M C (2003), 'Diesel engine electric turbo compound technology', SAE paper 2003-01-2294.

Hountalas D T, Katsanos C O and Lamaris V T (2007), 'Recovering energy from the diesel engine exhaust using mechanical and electrical turbocompounding', SAE paper 2007-01-1563.

Hribernik A (1997), 'Advantages and disadvantages of the application of a single entry turbine for pulse system turbocharging of 4-cylinder engine', SAE paper 972775.

Hussain M and Bhinder F S (1984), 'Experimental study of the performance of a nozzle-less volute casing for turbocharged turbines', SAE paper 840571.

Inoue K, Kubota O, Kishi N and Yano S (1989), 'A high power, wide torque range, efficient engine with a newly developed variable geometry turbocharger', SAE paper 890457.

Inter-Tech Energy Progress, Inc. (2003), 'Variable flow turbocharger developed by Aisin', *Engine Technology Progress in Japan*, *Compression-Ignition Engine Technology*, No. 22003015.

Ishihara H, Adachi K and Kono S (2002), 'Development of VFT part 2', SAE paper 2002-01-2165.

Ito N, Ohta T, Kono R, Arikawa S and Matsumoto T (2007), 'Development of a 4-cylinder gasoline engine with a variable flow turbo-charger', SAE paper 2007-01-0263.

Jacobs T, Assanis D and Filipi Z (2003), 'The impact of exhaust gas recirculation on performance and emissions of a heavy-duty diesel engine', SAE paper 2003-01-1068.

Japikse D (1996), *Centrifugal Compressor Design and Performance*, Concepts ETI, Inc., Wilder, VT.

Japikse D (2000), 'Decisive factors in advanced centrifugal compressor design and development', presented at the *International Mechanical Engineering Congress & Exposition* (*IMechE*), November available at: http://www.conceptsnrec.com/tech_papers.html

Japikse D (2001), 'Progress and potential in agile engineering for turbomachinery', *ASME 2001 Fluids Engineering Division Summer Meeting*, New Orleans, LA, June 1.

Japikse D (2009), 'Turbomachinery performance modeling', SAE paper 2009-01-0307.

Japikse D and Baines N C (1997), *Introduction to Turbomachinery*, Concepts ETI, Inc. and Oxford University Press, White River Junction, VT.

Japikse D and Platt M J (2004), 'Optimization in component design and redesign', *10th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery*, Honolulu, HI, March 7–11.

Jiao K, Li X, Sun H, Schram T, Krivitzky E and Larosiliere L M (2009), 'Numerical investigation of advanced compressor technologies to meet future diesel emission regulations', SAE paper 2009-01-1469.

Kawaguchi J, Adachi K, Kono S and Kawakami T (1999), 'Development of VFT (variable flow turbocharger)', SAE paper 1999-01-1242.

Kawamoto A, Takahashi Y, Koike T and Nakamura F (2001), 'Variable geometry system turbocharger for passenger car diesel engine', SAE paper 2001-01-0273.

Keshavarz A, Chapman K S, Shultz J and Kuiper D G (2003), 'Effect of turbine blade tip clearance on turbocharger performance: an experimental investigation', *Proceedings of 2003 Fall Technical Conference of the ASME Internal Combustion Engine Division*, ICEF2003-703, Erie, PA, September 7–10.

Kessel J-A, Schaffnit J and Schmidt M (1998), 'Modelling and real-time simulation of a turbocharger with variable turbine geometry (VTG)', SAE paper 980770.

Kohketsu S, Mori K, Sakai K and Hakozaki T (1997), 'EGR technologies for a turbocharged and intercooled heavy-duty diesel engine', SAE paper 970340.

Kuiper D (2007), 'Turbocharger design and performance analysis, part 1 of 2 (compressor analysis)', *Proceedings of the Gas Machinery Research Council's 2007 Gas Machinery Conference*, Dallas, TX, October 1–3.

Lakshmikantha M and Kec M (2002), 'Optimization of exhaust systems', SAE paper 2002-01-0059.

Lam J K-W, Roberts Q D H and McDonnell G T (2002), 'Flow modelling of a turbocharger turbine under pulsating flow', IMechE paper C602/025/2002.

Ludu A, Athenstaedt G and Dexter S G (2001), 'Turbocharging a medium speed diesel engine to suit different applications', *ASME Internal Combustion Engine Division (publication) ICE*, 37 (2), 1–7.

Luján J M, Payri F, Vera F and Guardiola C (2001), 'Modelling, effect and behaviour of the EGR venturi in a heavy-duty diesel engine', SAE paper 2001-01-3227.

Luján J M, Serrano J R, Cervello C, Arnau F J and Soltani S (2006), 'A one-dimensional model for variable and fixed geometry radial turbines for turbochargers', *8th International Conference on Turbochargers and Turbocharging,* Woodhead Publishing Limited and CRC Press LLC, pp. 97–117.

Lundqvist U, Smedler G and Stalhammar P (2000), 'A comparison between different EGR systems for HD diesel engines and their effect on performance, fuel consumption and emissions', SAE 2000-01-0226.

Macek J and Vítek O (2008), 'Simulation of pulsating flow unsteady operation of a turbocharger radial turbine', SAE paper 2008-01-0295.

Macek J, Vávra J and Vítek O (2002), '1-D model of radial turbocharger turbine calibrated by experiments', SAE paper 2002-01-0377.

Maiboom A, Tauzia X and Hetet J-F (2008), 'Influence of high rates of supplemental cooled EGR on NO_x and PM emissions of an automotive HSDI diesel engine using an LP EGR loop', *International Journal of Energy Research*, 32, 1383–1398.

Mattarelli E, Bianchi G M and Ivaldi D (2000), 'Experimental and numerical investigation on the EGR system of a new automotive diesel engine', SAE 2000-01-0224.

McCutcheon A R S (1978), 'Aerodynamic design and development of a high pressure ratio turbocharger compressor', IMechE paper C73/78.

Millo F, Mallamo F and Mego G G (2005), 'The potential of dual stage turbocharging and Miller cycle for HD diesel engines', SAE paper 2005-01-0221.

Millo F, Mallamo F, Pautasso E and Mego G G (2006), 'The potential of electric exhaust gas turbocharging for HD diesel engines', SAE paper 2006-01-0437.

Moroz L, Govoruschenko Y, Pagur P and Romanenko L(2005), 'Integrated conceptual design environment for centrifugal compressors flow path design', *Proceedings of 2008 ASME International Mechanical Engineering Congress and Exposition*, IMECE2008-69122, Boston, MA, October 31–November 6.

Moustapha H, Zelesky M F, Baines N C and Japikse D (2003), *Axial and Radial Turbines*, Concepts NREC, White River Junction, VT.

Nasser S H and Playfoot B B (1999), 'A turbocharger selection computer model', SAE paper 1999-01-0559.

Nili-Ahmadabadi M, Hajilouy-Benisi A, Durali M and Ghadak F (2008), 'Investigation of a centrifugal compressor and study of the area ratio and tip clearance effects on performance', *Journal of Thermal Science*, 17(4), 314–323.

Ogura M, Umezaki E and Tomita T (1989), 'Study of variable scroll type turbocharger (performance test in diesel engine)', SAE paper 891873.

Ogura M and Shao J (1995), 'Research development of variable nozzle turbocharger, part 1: a new axially movable turbocharger', SAE paper 952102.

Okazaki T, Uchiyama K, Yamada K and Tsubouchi H (1982), 'Development of high speed small turbochargers for passenger cars', IMechE paper C33/82.

Okazaki Y, Matsudaira N and Hishikawa A (1986), 'A case of variable geometry turbocharger development', IMechE paper C111/86.

Osborne R P and Morris T (2002), 'Aspects of EGR control on turbocharged light-duty diesel engines', IMechE paper C602/013/2002.

Pampreen R C (1976), 'The use of variable inlet guide vanes for automotive gas turbine engine augmentation and load control', SAE paper 760285.

Pan D, Whitfield A and Wilson M (1999), 'Design considerations for the volutes of centrifugal fans and compressors', *Proc. IMechE, Part C: Journal of Mechanical Engineering Science*, 213, 401–410.

Park S, Matsumoto T and Oda N (2010), 'Numerical analysis of turbocharger response delay mechanism', SAE paper 2010-01-1226.

Partridge W P, Lewis S A, Ruth M J, Muntean G G, Smith R C and Stang J H (2002), 'Resolving EGR distribution and mixing', SAE paper 2002-01-2882.

Pfeifer A, Smeets M, Herrmann H-O, Tomazic D, Richert F and Schloßer A (2002), 'A new approach to boost pressure and EGR rate control development for HD truck engines with VGT', SAE paper 2002-01-0964.

Rodgers C (2003), 'The characteristics of radial turbines for small gas turbines', GT2003-38026, *Proceedings of the ASME Turbo Expo 2003: Power for Land, Sea, and Air*, June 16–19, Atlanta, Georgia, USA.

Rodgers C and Rochford K (2002), 'Small turbocharger turbomachinery', IMechE paper C602/003/2002.

Ryder O, McKenzie D J A and Johnson S (2002), 'Turbo matching techniques for highly cyclic bus applications', IMechE paper C602/006/2002.

SAE Surface Vehicle Recommended Practice J922 (1995), 'Turbocharger nomenclature and terminology'.

SAE Surface Vehicle Standard J1723 (1995), 'Supercharger testing standard'.

SAE Surface Vehicle Recommended Practice J1826 (1995), 'Turbocharger gas stand test code'.

Saulnier S and Guilain S (2004), 'Computational study of diesel engine downsizing using two-stage turbocharging', SAE paper 2004-01-0929.

Schmid W and Sumser S (2007), 'Exhaust gas turbocharger for an internal combustion engine', US patent 7,162,872.

Senoguz M T, Bozkurt H A, Karadeniz C and Adil K N (2008), 'Modeling and testing of the throttle body usage in a diesel engine to improve NO_x emissions', SAE paper 2008-01-2646.

Serrano J R, Arnau F J, Dolz V, Tiseira A, Lejeune M and Auffret N (2008), 'Analysis of the capabilities of a two-stage turbocharging system to fulfill the US2007 anti-pollution directive for heavy duty diesel engines', *International Journal of Automotive Technology*, 9(3), 277–288.

Shao J, Ogura M, Liu Y and Yoshino M (1996), 'Performance development of a new axially moveable vane turbocharger', SAE paper 961747.

Sharp C A, Khair M K and Gorel A (1999), 'The effect of a turbocharger clearance control coating on the performance and emissions of a 2-stroke diesel engine', SAE paper 1999-01-3665.

Shutty J (2009), 'Control strategy optimization for hybrid EGR engines', SAE paper 2009-01-1451.

Siewert R M, Krieger R B, Huebler M S, Baruah P C, Khalighi B and Wesslau M (2001), 'Modifying an intake manifold to improve cylinder-to-cylinder EGR distribution in a DI diesel engine using combined CFD and engine experiments', SAE paper 2001-01-3685.

Spraker W A (1987), 'Contour clearance losses in radial in-flow turbines for turbochargers', ASME paper 87-ICE-53.

Spraker W A (1991), 'Clearance and Reynolds number effects on the efficiency of radial flow compressors', SAE paper 910417.

Spraker W A (1992), 'The aerodynamic design of variable power turbines for turbochargers', SAE paper 920041.

Stobart R and Weerasinghe R (2006), 'Heat recovery and bottoming cycles for SI and CI engines – a perspective', SAE paper 2006-01-0662.

Strub R A, Bonciani L, Borer C J, Casey M V, Cole S L, Cook B B, Kotzur J, Simon H and Strite M A (1987), 'Influence of the Reynolds number on the performance of centrifugal compressors', *Transactions of the ASME*, *Journal of Turbomachinery*, 109, 541–544.

Sumi Y and Yamane K (1986), 'Recent development of high-response turbochargers for passenger cars', IMechE paper C107/86.

Tabaczynski R J (1982), 'Effects of inlet and exhaust system design on engine performance', SAE paper 821577.

Tange H, Ikeya N, Takanashi M and Hokari T (2003), 'Variable geometry diffuser of turbocharger compressor for passenger vehicles', SAE paper 2003-01-0051.

Tennant D W H and Walsham B E (1989), 'The turbocompound diesel engine', SAE paper 890647.

Tufail K (2010), 'Turbocharging and air-path management for light-duty diesel engines', in Zhao H (ed.), *Advanced Direct Injection Combustion Engine Technologies and Development, Volume 2: Diesel Engines*, Woodhead Publishing Limited, Cambridge, UK.

Ubanwa B and Kowalczyk M (1993), 'Optimization procedure for matching exhaust waste-gate turbocharger to automotive diesel engine', SAE paper 931102.

Uchida H, Kashimoto A and Iwakiri Y (2006), 'Transient performance prediction of the turbocharging system with the variable geometry turbochargers', *8th International Conference on Turbochargers and Turbocharging,* Woodhead Publishing Limited and CRC Press LLC, pp. 341–350.

Umezaki E, Ogura M and Tomita T (1989), 'Study of variable scroll type turbocharger (determination of shape of scroll)', SAE paper 891874.

Wallace F J, Howard D, Roberts E W and Anderson U (1986), 'Variable geometry turbocharging of a large truck diesel engine', SAE paper 860452.

Wang R (1996), *Study on the Flow in VGT Scroll and the Performance of Turbocharger*, doctoral dissertation (in Chinese), Shanghai Jiao-Tong University, Shanghai, China.

Watson N (1978), 'Turbocharger system options for vehicle engines', IMechE paper C61/78.

Watson N (1979), 'Turbochargers for the 1980s – current trends and future prospects', SAE paper 790063.

Watson N (1982), 'Resonant intake and variable geometry turbocharging systems for a V8 diesel engine', IMechE paper C40/82.

Watson N and Banisoleiman L (1986), 'Performance of a highly rated vehicle diesel engine with a variable geometry turbocharger', IMechE paper C103/86.

Watson N (1999), 'The theory of turbocharging', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 27–72.

Watson N and Janota M S (1982), *Turbocharging the Internal Combustion Engine*, Macmillan Publishers Limited, London.

Watson N, Marzouk M and Baazaari Z (1978), 'An evaluation of two stage turbocharging for efficient high-output diesel engines', *ASME Energy Technology Conference & Exhibition*, Houston, TX, ASME paper 78-DGP-2.

Westin F and Angstrom H E (2002), 'A method of investigating the on-engine turbine efficiency combining experiments and modeling', IMechE paper C602/029/2002.

Wiesner F J (1979), 'A new appraisal of Reynolds number effects on centrifugal compressor performance', *Transactions of the ASME, Journal of Engineering for Power*, 101, 384–396.

Wilson D and Korakianitis T (1998), *The Design of High-Efficiency Turbomachinery and Gas Turbines*, Prentice Hall, New Jersey.

Winkler N, Ångström H-E and Olofsson U (2005), 'Instantaneous on-engine twin-entry turbine efficiency calculations on a diesel engine', SAE paper 2005-01-3887.

Winterbone D E and Alexander G I (1985), 'Efficiency of the manifolds of turbocharged engines', *IMechE Proc Instn Mech Engrs*, 199(D2), 137–149.

Winterbone D E and Pearson R J (1999), *Design Techniques for Engine Manifolds*, SAE International, Warrendale, PA.

Winterbone D E and Pearson R J (2000), *Theory of Engine Manifold Design*, Professional Engineering Publishing, London.

Wu T and McAulay K J (1973), 'Predicting diesel engine performance at various ambient conditions', SAE paper 730148.

Yamaguchi S, Yamaguchi H, Goto S, Nakao H and Nakamura F (2002), 'The development of effective casing treatment for turbocharger compressors', IMechE paper C602/008/2002.

Yang S, Wang L-S, Gu H, Deng K and Cui Y (2006), 'MIXPC turbocharging system for diesel engines', SAE paper 2006-01-3390.

Yokota K, Hattori H, Shimizu M and Furukawa H (1986), 'A highly BMEP diesel engine with variable geometry turbocharger', IMechE paper C119/86.

Zhang J, Zhuge W, Hu L and Li S (2007), 'Design of turbocharger variable nozzle', GT2007-27562, *Proceedings of the ASME Turbo Expo 2007: Power for Land, Sea, and Air*, May 14–17, Montreal, Canada, pp. 1313–1319.

Zhang Z, Deng K, Wang Z and Zhu X (2008a), 'Experimental study on the three-phase sequential turbocharging system with two unequal size turbochargers', SAE paper 2008-01-1698.

Zhang Z, Gu H, Guo Z and Deng K (2008b), 'Modular multi-purpose pulse converter

turbocharging system for four-stroke diesel engines', *International Journal of Energy Research*, 32(6), 569–580.
Zhuge W, Zhang Y, Zheng X, Yang M and He Y (2009), 'Development of an advanced turbocharger simulation method for cycle simulation of turbocharged internal combustion engines', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 223, 661–672.

13.9.1 References and bibliography on second law of thermodynamics

Alkidas A C (1988), 'The application of availability and energy balances to a diesel engine', *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power*, 110, 462–469.
Alkidas A C (1989), 'The use of availability and energy balances in diesel engines', SAE paper 890822.
Bozza F, Nocera R, Senatore A and Tuccillo R (1991), 'Second law analysis of turbocharged engine operation', SAE paper 910418.
Bueno A V, Velásquez J A and Milanez L F (2004), 'Exergy based diagnosis of in-cylinder diesel engine', SAE paper 2004-01-3258.
Caton J A (2000a), 'A review of investigations using the second law of thermodynamics to study internal-combustion engines', SAE paper 2000-01-1081.
Caton J A (2000b), 'Operating characteristics of a spark-ignition engine using the second law of thermodynamics: effects of speed and load', SAE paper 2000-01-0952.
Caton J A (2002), 'A cycle simulation including the second law of thermodynamics for a spark-ignition engine: implications of the use of multiple-zones for combustion', SAE paper 2002-01-0007.
Caton J A (2005), 'Use of a cycle simulation incorporating the second law of thermodynamics: results for spark-ignition engines using oxygen enriched combustion air', SAE paper 2005-01-1130.
Edwards K D, Wagner R M and Graves R L (2008), 'Identification of potential efficiency opportunities in internal combustion engines using a detailed thermodynamic analysis of engine simulation results', SAE paper 2008-01-0293.
Endo T, Kawajiri S, Kojima Y, Takahashi K, Baba T, Ibaraki S, Takahashi T and Shinohara M (2007), 'Study on maximizing exergy in automotive engines', SAE paper 2007-01-0257.
Fijalkowski S and Nakonieczny K (1997), 'Operation of exhaust systems of turbocharged diesel engines identified by means of exergy analysis', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 211, 391–406.
Flynn P F, Hoag K L, Kamel M M and Primus R J (1984), 'A new perspective on diesel engine evaluation based on second law analysis', SAE paper 840032.
Gallo W L R (1992), 'Thermodynamic evaluation of variable valve timing', SAE paper 921479.
Gerpen J H V and Shapiro H N (1990), 'Second-law analysis of diesel engine combustion', *Transactions of the ASME, Journal of Engineering for Gas Turbines and Power*, 112, 129–137.
Giakoumis E G (2007), 'Cylinder wall insulation effects on the first- and second-law balances of a turbocharged diesel engine operating under transient load conditions', *Energy Conversion and Management*, 48(11), 2925–2933.

Giakoumis E G and Andritsakis E C (2007), 'Irreversibility production during transient operation of a turbocharged diesel engine', *International Journal of Vehicle Design*, 45(1/2), 128–149.

Kyritsis D C and Rakopoulos C D (2001), 'Parametric study of the availability balance in an internal combustion engine cylinder', SAE paper 2001-01-1263.

Li J, Zhou L, Pan K, Jiang D and Chae J (1995), 'Evaluation of the thermodynamic process of indirect injection diesel engines by the first and second law', SAE paper 952055.

Lipkea W H and DeJoode A D (1989), 'A comparison of the performance of two direct injection diesel engines from a seond law perspective', SAE 890824.

McKinley T L and Primus R J (1988), 'An assessment of turbocharging systems for diesel engines from first and second law perspectives', SAE paper 880598.

Nakonieczny K (2002), 'Entropy generation in a diesel engine turbocharging system', *Energy*, 27(11), 1027–1056.

Naterer G F (2008), *Entropy Based Design and Analysis of Fluids Engineering Systems*, CRC Press, Boca Raton, FL.

Primus R J (1984), 'A second law approach to exhaust system optimization', SAE paper 840033.

Primus R J and Flynn P F (1985), 'Diagnosing the real performance impact of diesel engine design parameter variation (a primer in the use of second law analysis)', *International Symposium on Diagnostics and Modeling of Combustion in Reciprocating Engines*, Tokyo, Japan, pp. 529–538.

Primus R J, Hoag K L, Flynn P F and Brands M C (1984), 'An appraisal of advanced engine concepts using second law analysis techniques', SAE paper 841287.

Rakopoulos C D and Giakoumis E G (2004a), 'Availability analysis of a turbocharged diesel engine operating under transient load conditions', *Energy*, 29, 1085–1104.

Rakopoulos C D and Giakoumis E G (2004b), 'Parametric study of transient turbocharged diesel engine operation from the second-law perspective', SAE paper 2004-01-1679.

Rakopoulos C D and Giakoumis E G (2005), 'Second-law analysis of indirect injection turbocharged diesel engine operation under steady-state and transient conditions', SAE paper 2005-01-1131.

Rakopoulos C D and Giakoumis E G (2006), 'Second-law analyses applied to internal combustion engines operation', *Progress in Energy and Combustion Science*, 32, 2–47.

Rakopoulos C D and Giakoumis E G (1997), 'Speed and load effects on the availability balances and irreversibilities production in a multi-cylinder turbocharged diesel engine', *Applied Thermal Engineering*, 17(3), 299–313.

Rakopoulos C D and Giakoumis E G (2009), *Diesel Engine Transient Operation – Principles of Operation and Simulation Analysis*, Springer Verlag London Limited, London.

Shapiro H N and Gerpen J H V (1989), 'Two zone combustion models for second law analysis of internal combustion engines', SAE paper 890823.

Teh K-Y, Miller S L and Edwards C F (2008), 'Thermodynamic requirements for maximum internal combustion engine cycle efficiency. Part 1: optimal combustion strategy', *International Journal of Engine Research*, 9, 449–465.

Teh K-Y, Miller S L and Edwards C F (2008), 'Thermodynamic requirements for maximum internal combustion engine cycle efficiency. Part 2: work extraction and reactant preparation strategies', *International Journal of Engine Research*, 9, 467–481.

Velásquez J A and Milanez L F (1994), 'Analysis of the irreversibilities in diesel engines', SAE paper 940673.

14
Diesel engine system dynamics, transient performance, and electronic controls

Abstract: This chapter forms the foundation of dynamic engine system design based on transient performance and engine controls. The system dynamics approach adopted shares a common ground with the engine controls community. The chapter begins with discussions on the differences between steady-state and transient performance and the effects of hardware design and software control strategy on transient turbocharged engine performance. It further covers mean-value models, crank-angle-resolution high-fidelity real-time models, air path model-based controls, and fuel path controls. It also illustrates virtual sensor modeling with an example of the exhaust manifold gas temperature sensor. Furthermore, it is pointed out that analytical controller design is an important area where engine system design can make significant contributions.

Key words: system dynamics, transient performance, engine controls, mean-value model, real-time model, model-based controls, sensor, controller design.

14.1 Overview of diesel engine transient performance and controls

14.1.1 The roles of engine hardware design and software controls

Engine controls fulfill the requirements derived from performance, emissions, and durability. Moreover, electronic controls enable flexible hardware designs in modern engines, such as variable valve actuation (VVA) and high-bandwidth-controlled variable combustion systems. Over the years, electronic controls, consisting of software and calibration, have become a core part of engine fundamental architecture. When engine hardware design and electronic controls are integrated, two challenges are faced: (1) What hardware should be controlled (i.e., which sensors and controllers should be able to achieve the settings quickly and reliably)? (2) How should the hardware be controlled (i.e., how to create the optimum algorithm and seamlessly integrate it in the hierarchical structure of the system design with the best trade-offs to minimize transient emission spikes and pumping loss)?

Engine air system hardware needs to be designed to have acceptable transient capability to minimize turbocharger lag, transient pumping loss, fuel consumption and emissions, and to ensure fast warm-up. Transient simulation plays a critical role in the development process, from predicting vehicle transients to evaluating electronic control strategies. The commonly encountered vehicle transients include load response, vehicle launch, acceleration, and driving cycles. Moreover, predicting the transients can help to analyze both performance and durability issues, and hence reveal their parametric dependency. For example, compressor wheel fatigue life can be evaluated from the transient load calculations for a turbocharger, and the impact of gear shifting during driving cycles on component durability life can be accessed as well.

14.1.2 The difference between steady-state and transient performance

The transient acceleration or deceleration process of naturally aspirated diesel engines can be approximated by a continuous series of steady-state operating conditions. However, in turbocharged diesel engines, during transients the turbine power is not equal to the compressor power, and the turbocharger speed is affected by turbocharger inertia and turbocharger power imbalance. There is a turbocharger lag during which the compressor boost pressure gradually changes to reach a new steady state. The EGR circuit has a certain volume and there is a transient dynamic response of an EGR purging and filling process in the intake manifold. The transient response is usually in the order of several engine cycles. There is also an air–fuel ratio smoke limit for the maximum fueling, depending on the available air flow. During transients, air–fuel ratio, EGR rate and in-cylinder metal wall temperature (due to thermal inertia) are all different from those in the steady state. The resulting deteriorated combustion efficiency and pumping loss cause differences in emissions and fuel economy between steady-state and transient. Valve overlap also has a large impact on transient acceleration performance. With a large overlap, at the beginning of fast acceleration, the exhaust manifold pressure can be much higher than the intake manifold pressure for various reasons (e.g., EGR valve closed). The high engine delta P results in a large reverse flow of residue gases from the exhaust manifold into the cylinder and the intake manifold. The increased residue gas fraction reduces the air–fuel ratio and retards the vehicle acceleration. In contrast, a small valve overlap helps transient acceleration for the same reason.

14.1.3 Controlling engine transient performance

Steady-state emissions testing establishes appropriate NO_x, PM, HC, and CO throughout the engine speed–load domain by setting the engine calibration parameters including fuel injection timing and pressure, VGT vane opening, EGR valve opening, and intake throttle valve opening. Coordinated controls of the EGR valve and the VGT and coordinated controls of the EGR valve and the intake throttle valve have been studied extensively (Nieuwstadt, 2003). Boost pressure (or air–fuel ratio) and EGR rate are mapped in the speed–load domain during engine steady-state calibration. But the combined control of the EGR valve and the VGT for transient EGR rate and air–fuel ratio is challenging. Steady-state engine calibration may set the VGT vane and the EGR valve opening positions to establish an air–fuel ratio and EGR rate, but those steady-state air–fuel ratios and EGR rates are not achievable during fast transients due to turbocharger lag, even if the transient gains are applied to the steady-state set points of the positions in the lookup table of the calibration maps.

Compared to steady-state emissions levels, there are three strategies for designing the transient emission profiles in engine controls: (1) overshoot spike vs. time; (2) undershoot or slow approach vs. time; and (3) a compromise to closely match a predetermined target without excessive overshoot or undershoot. These strategies depend on the transient fueling rate changes. Each strategy has different transient trade-offs between NO_x and soot and different characteristics of transient engine delta P and pumping loss. For example, a transient NO_x spike is normally caused by shutting off EGR during fast acceleration. Design of each transient emissions profile with proper combination of hardware and electronic controls is important. The difference on emissions margin between steady-state and transient can be predicted at the early stage of engine development. In addition, the requirements of transient emissions and driving cycle fuel economy can be met through proper designs.

Meeting a composite emissions target in the US supplemental emissions test (SET) 13-mode steady-state does not mean the engine can meet the same emissions target during an FTP transient cycle. In fact, as the NO_x emission regulation becomes more stringent, the relative difference between the FTP transient emissions and the steady-state emissions becomes larger. It would translate to an unreasonably low steady-state emissions development target in order to meet the FTP transient emissions target. Therefore, a dynamic high EGR rate and adjustments of fuel injection timing and injection pressure during transients must be used to reduce the transient emissions. During transient operation, the purpose of engine air system controls is to calculate and regulate the EGR valve and the turbocharger actuator settings to achieve minimum emissions and maintain drivability. The settings may be obtained from the position lookup tables or may be calculated by model-based control algorithms. Various engine control approaches can be used with different effects on engine transients. The conventional approach uses various gain-based transient controllers to regulate the EGR valve and the VGT vane opening. But even with the most sophisticated map-based gain settings, it is still difficult to achieve a prescribed emissions trade-off during speed and load transients. Model-based control is promising in shaping the transient emissions profiles as a function of time.

14.1.4 Analysis of hardware design for engine transients

Turbocharged diesel engines cannot respond to a sudden change in speed or load as fast as naturally aspirated diesel engines because the change in compressor air flow lags behind the change in the fueling rate. The reasons for turbocharger lag include: (1) because the manifold has certain volumes, it takes time (usually several engine cycles) to gradually build up the gas pressure in the exhaust manifold and the intake manifold; (2) during fast acceleration, after the EGR valve is closed it takes time to purge EGR out of the intake manifold; and (3) because the turbocharger rotor assembly has a certain moment of inertia, it takes time for the turbine to gradually accelerate the compressor to a higher speed under the difference between the turbine power and the compressor power. For smoke control, fueling rate and engine power are restricted during the fast acceleration transients according to the available intake boost pressure. Reducing the volume of the intake manifold and the exhaust manifold can reduce turbocharger lag. For example, pulse turbocharging with a smaller manifold volume has a better transient response than constant-pressure turbocharging. Matching the turbine size with a small flow area for low speed–load conditions can help build up the turbine power faster during transients and reduce turbocharger lag. Note that at high speeds and high loads, a wastegated turbine or VGT needs to be used to prevent over-boosting. Another way to reduce turbocharger lag is to use low turbo inertia, such as: (1) reducing the turbine size; (2) using two smaller turbochargers to replace one big unit; (3) using a low-inertia small high-pressure-stage turbocharger in a two-stage turbocharger system; and (4) using a ceramic turbine rotor. Other methods that may reduce turbocharger lag and improve the smoke limit and transient response include: (1) placing the EGR valve close to the intake manifold to minimize the EGR purging time; (2) reducing the heat losses in the cylinder and the exhaust manifold; (3) using a small valve overlap; (4) improving the transient combustion efficiency; (5) retarding fuel injection timing to increase the turbine inlet exhaust temperature; and (6) using boost-assisting devices such as injecting extra air or mechanical supercharging during fast acceleration.

One important consideration for turbocharger selection is controlling compressor surge during transients. On the compressor map, the operating trace of a fast speed-increase transient is located at the right side of the steady-state point. The operating trace of the transient of a fast load-increase or fast speed-decrease is located on the left side of the steady-state point, possibly resulting in compressor surge. The compressor surge during fast transients happens when the compressor flow rate, which responds quickly to engine speed change, decreases much faster than the lagged compressor boost pressure. The boost pressure is affected by turbocharger shaft speed and turbo inertia. There is a need to simulate the fast deceleration to check compressor surge in turbocharger matching.

In addition to the turbocharger, other engine hardware design evaluations for transient acceleration performance usually include the effects of engine inertia, charge air cooler or inter-stage cooler volume and cooling medium temperature, manifold volume and piping, EGR circuit volume and EGR purging time, and the location of the aftertreatment components.

Engine electronic controls are reviewed by Stobart *et al.* (2001). Reviews on diesel engine controls are provided by Shigemori (1988), Gant and Alves (1990), Anderson (1991), Winterbone and Jai-In (1991), Hirschlieb *et al.* (1995), Guzzella and Amstutz (1998), Guzzella and Onder (2004), and Grondin *et al.* (2004). Overviews on engine model-based controls are provided by Hafner (2001), Lehner *et al.* (2001), Smith *et al.* (2007), Turin *et al.* (2007, 2008), Šika *et al.* (2008), Stobart *et al.* (1998), Atkinson *et al.* (2009), and Guzzella (2010). Engine control process development is decribed by Kämmer *et al.* (2003), Baumann *et al.* (2004), and Erkkinen and Breiner (2007).

14.2 Turbocharged diesel engine transient performance

Turbocharged diesel engine transient performance has been extensively researched experimentally and analytically since the 1970s. The transient performance of variable-geometry turbochargers was investigated by Lundstrom and Gall (1986), Pilley *et al.* (1989), Brace *et al.* (1999), and Filipi *et al.* (2001). Dynamic optimization and the differences between the steady-state and transient conditions were researched by Wijetunge *et al.* (1999). They addressed the subsystem coordination and interaction during the dynamic transients. The transient performance of the EGR circuit in a diesel engine was evaluated by Serrano *et al.* (2005). Diesel engine transient operation was systematically summarized in a book authored by Rakopoulos and Giakoumis (2009).

Computer simulations on turbocharged diesel engine transient performance were conducted by Ledger and Walmsley (1971), Watson and Marzouk (1977), Winterbone *et al.* (1977), Marzouk and Watson (1978), Watson (1981), Ma and Gu (1990), Qiao *et al.* (1992), Ma and Agnew (1994) and a group of researchers at the Universidad Politecnica de Valencia in Spain (Payri *et al.*, 1999, 2002; Benajes *et al.*, 2000, 2002), as well as a group of researcher at National Technical University of Athens in Greece (Rakopoulos *et al.*, 1997b, 2004, 2005, 2007; Rakopoulos and Giakoumis, 2006, 2007, 2009; Theotokatos and Kyrtatos, 2001).

Combustion heat release rate models were explored by Watson *et al.* (1980), Felsch *et al.* (2009), and Serrano *et al.* (2009a, 2009b). Unlike previous research that focused on the impact of engine hardware on transient performance, Watson (1984) proposed a modeling methodology that linked the relationship between transient performance and engine controls. The needs of engine performance simulation with electronic controls have started to emerge ever since.

14.3 Mean-value models in model-based controls

Mean-value real-time transient models are the primary simulation approach used in gasoline and diesel engine controls. In order to reduce computing time, the mean-value model uses engine maps such as the volumetric efficiency map and the exhaust manifold temperature map as a function of other dependent parameters such as engine speed, load, and boost pressure. The model does not have the resolution at a crank-angle level. The engine thermodynamic and flow states are represented by their respective single mean values over the engine cycle. Important progress in this area are detailed as follows. Real-time modeling of diesel engines was started in the 1980s by Shamsi (1980), Hendricks (1989), and Jensen *et al.* (1991). The real-time mean-value models on gasoline engines are summarized by Hendricks *et al.* (1996) on intake manifold dynamics, Chevalier *et al.* (2000) on the validity of the mean-value models, and Buckland *et al.* (2000) on the application for direct injection gasoline engines. Important research on the turbocharged diesel engine modeling for nonlinear engine controls was performed by Kao and Moskwa (1995). More advanced modeling of mean-value models after the 1990s was investigated by Moraal and Kolmanovsky (1999), Allmendinger *et al.* (2001), Eriksson (2002), Jung *et al.* (2002), Chung *et al.* (2005), Fiorani *et al.* (2006), Eriksson (2007), Pettiti *et al.* (2007), Chen (2008), and Olin (2008). A lot of the above-mentioned modeling work was conducted using MATLAB/Simulink or other programming languages. GT-POWER is a leading software tool used for engine system simulation, commercially available from Gamma Technologies. Mean-value modeling in GT-POWER including controller simulations has been presented by Silvestri *et al.* (2000), Papadimitriou *et al.* (2005), He (2005), He *et al.* (2006), and He and Lin (2007).

The above discussion is for the engine air system. On the fuel system side, a real-time model of fuel injection dynamics used for hardware-in-the-loop (HIL) test is presented in Woermann *et al.* (1999).

14.4 Crank-angle-resolution real-time models in model-based controls

The high-fidelity crank-angle-resolution real-time model is more advanced than the mean-value model. Instead of using the prescribed cycle-averaged 'mean value' maps, it can predict the in-cylinder cycle process details at a crank-angle level without losing the real-time capability. This type of model is the direction for future development. The real-time crank-angle-resolution model without manifold gas wave dynamics is presented in Schulze *et al.* (2007). The real-time crank-angle-resolution model including the manifold gas wave dynamics features is developed by Pacitti *et al.* (2008). Other related work has been reported by Wurzenberger *et al.* (2009).

14.5 Air path model-based controls

14.5.1 Lookup table approach in air system controls for engine transients

In the traditional lookup table controls, the set point is usually either an actuator position such as the EGR valve duty cycle or a performance parameter such as MAF (mass air flow), MAP (manifold air pressure), lambda (equivalent air–fuel ratio) or air–fuel ratio, and intake manifold oxygen concentration. During transients, the air system component actuator such as the EGR valve or VGT vane can be driven to a pre-determined position, which is obtained from the steady-state calibration at a given engine speed and load and then superpositioned with the transient PID gains and the transient algorithms with a feedback PID control. The moving part within the actuator has transient response characteristics against time. The transient delay produces a non-instant response to the demand from the engine control unit, and the delay is due to a characteristic time constant. In the MAF control, the actuator such as the EGR valve is modulated to achieve a preset air–fuel ratio through a PI or PID controller fed by the feedback difference between the required MAF and the sensor signal. The sensor signal can be either the actual measured signal or a calculated signal from a virtual sensor.

The lookup table approach is not flexible enough to handle real-world engine variations. One example of the limitation of the lookup table is the aftertreatment dynamic dosage control, e.g., hydrocarbon dosing in active deNO_x catalyst or urea dosing in SCR. The lookup table approach cannot compensate for the interferences from factors such as production variation, component aging, and engine acceleration or deceleration transients. During transient turbocharger lag or valve lag, the dosage rate determined from the steady-state calibration lookup tables may not be suitable and could result in an unacceptable slip amount.

14.5.2 Model-based controls for engine transients

Traditional engine controls used lookup tables. The calibration complexity increases exponentially as new functions and the number of associated control tables or maps increase in the development of modern turbocharged EGR engines. Electronic controls have been evolving toward the mathematical model-based controls, either open loop or closed loop, which have become an important part of system design and diagnosis. The success of online model-based controls relies largely on the accuracy of the thermodynamic cycle performance models. The models can be built for various operating conditions such as new and aged engines, normal and extreme climates. Model-based air system controls use the sensors and the actuators to sense and control the gas flows inside and outside the cylinder, such as using engine valve actuation, EGR valve, intake throttle valve, exhaust back-pressure valve, and turbine vane actuator. Model-based cooling system controls use the sensors and actuators for coolant flows. In order to provide flexible cooling as needed to minimize driving power consumption and enhance engine performance, the desirable cooling system in the future is to use more electronic controllers in addition to the current pump, fan, and thermostat. The gas-side data and coolant-side data may be linked by a model of heat rejection and other engine performance parameters. In the coordinated model-based control of EGR valve and turbine actuator, according to the theory in Chapter 4, the valve opening and the actuator setting may be calculated by specifying two performance targets among the following three parameters: fresh air flow rate, EGR rate, and intake manifold boost pressure. The engine performance parameters used in model-based controls can be either the measured signals from real sensors, or the simulated data from virtual sensors.

The opening position of valves (e.g., EGR valve, intake throttle valve, exhaust back-pressure valve) can be calculated based on target performance parameters. For example, the gas mass flow rate through an EGR valve orifice can be modeled by using the equation of isentropic compressible flow of the ideal gas:

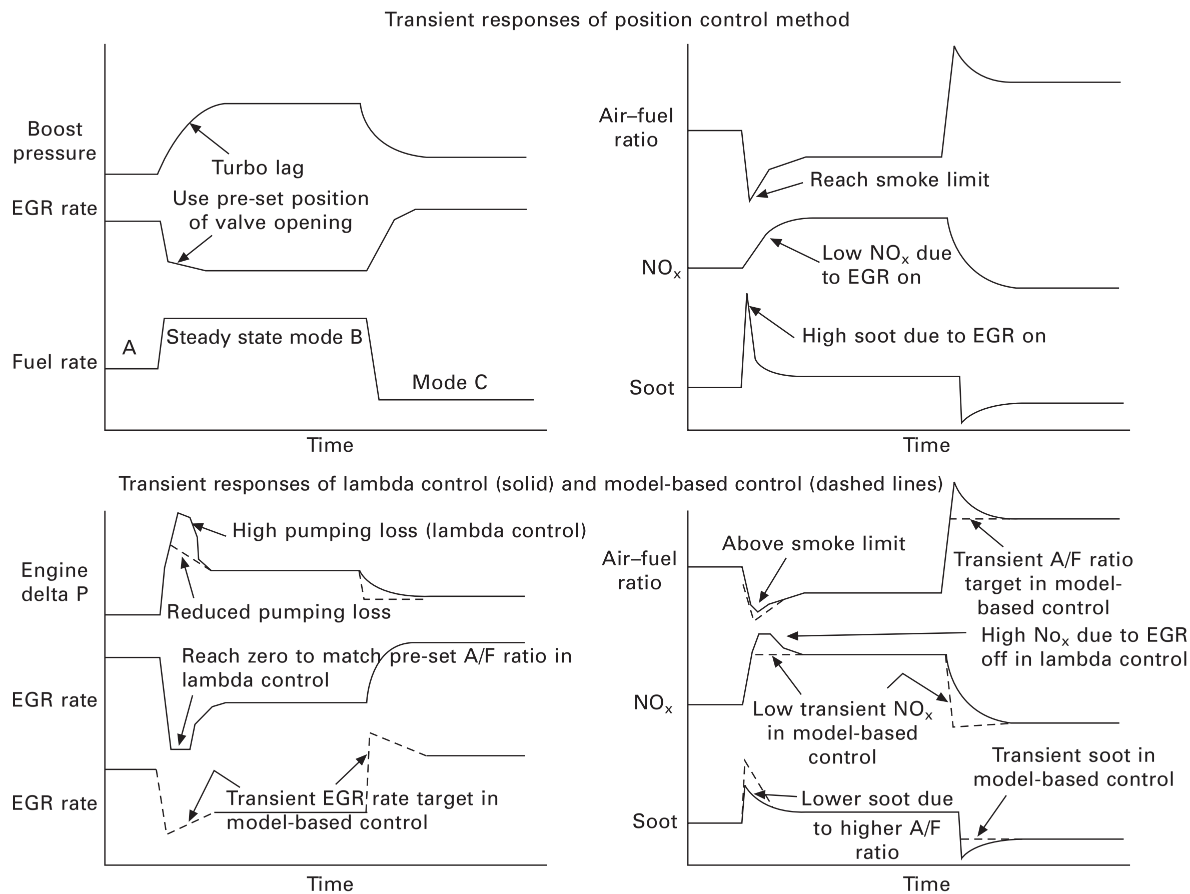
$$\dot{m}_{EGR} = C_{EGR} \frac{A_{EGR}\, p_{upstream}}{\sqrt{R_{gas} T_{upstream}}} \times \sqrt{\frac{2\kappa_t}{\kappa_t - 1}} \cdot \sqrt{\left(\frac{p_{downstream}}{p_{upstream}}\right)^{2/\kappa_t} - \left(\frac{p_{downstream}}{p_{upstream}}\right)^{(\kappa_t+1)/\kappa_t}} \qquad 14.1$$

where A_{EGR} is the theoretical effective flow area of the EGR valve orifice at a given opening and C_{EGR} is a variable correction coefficient. A_{EGR} is obtained from the correlation between valve lift and valve flow area. C_{EGR}

is calibrated by measurement data at various engine flow conditions and is used to correct any inaccuracy in the theoretical valve flow area. To meet the valve actuator control demand, the required valve flow area can be calculated first from the desired EGR mass flow rate at any given engine speed and load by rearranging equation 14.1. Then, the valve area can be converted to the valve opening position or lift. In model-based controls, in order to achieve a desirable transient EGR rate (not necessarily the minimum deviation from the steady-state EGR rate), a target transient rate profile may be flexibly defined to a certain shape as a function of the rate of transient speed–load change. For example, the target transient EGR rate can be imposed by a fixed EGR percentage multiplied by the measured total engine gas mass flow rate. It is worth noting that in model-based control the dynamic behavior of the component can also be modeled.

To illustrate the effect of different engine control methods on transient performance, a step increase in the fueling rate followed by a step decrease is shown in Fig. 14.1. As a result, engine speed and brake torque change through three steady-state modes, named A, B, and C. Such an event is a typical representation of real-world driving cycles. Due to turbocharger lag, there are inevitable delays in the exhaust manifold pressure and the intake manifold boost pressure during the fast transient of fueling rate change. The instant fueling increase results in a fast decrease in the air–fuel ratio at the beginning of the transient event, possibly reaching the smoke limit. The level of EGR rate commanded by the engine controls has a direct impact on the air–fuel ratio, NO_x and soot emissions during the transient. For example, in the method of EGR valve position control, the EGR valve opening undergoes sharp changes from the steady-state calibration opening of mode A to those of modes B and C. The resulting EGR flow reduces the air–fuel ratio and increases the transient soot while keeping the transient NO_x low. On the other hand, in the method of lambda (i.e., equivalent air–fuel ratio) control or MAF control, in order to maintain the steady-state setting of air–fuel ratio during transients, the EGR valve is commanded to close when the fueling rate suddenly increases, and to open more when the fueling rate suddenly decreases. As a result, during a step load increase, less or even zero EGR flow is obtained. Consequently, a higher air–fuel ratio and lower transient soot are achieved, but the transient NO_x and pumping loss (due to a higher engine delta P) may be higher compared with the position control.

The transient soot spike is an inevitable phenomenon during fast accelerations due to turbocharger lag. The task of the EGR control is to regulate the EGR flow to minimize the transient emission spikes or to achieve the best NO_x–soot trade-off with the best compromise on drivability. Meanwhile, EGR control needs to be coordinated with turbocharger control. The intake manifold boost pressure can then be properly controlled. Changes in intake and exhaust restrictions, production variations, and changes in ambient conditions can



14.1 Illustration of transient engine controls and air system performance.

be promptly responded to. It is difficult for the position-based control to accomplish these tasks. The MAF or MAP control can only partially fulfill the requirements. A promising model-based control may finally compute the actuator position such as the EGR valve opening or the VGT vane position accurately with a calibrated model based on the preset desirable engine performance parameters. The preset parameters may include the desirable transient EGR flow rate profile or the dynamic transient fuel injection control parameters, or the desirable NO_x or soot limit if the real-time NO_x or soot model is available. Another advantage of model-based control is that its steady-state engine calibration set points and transient calibration gains are much less hardware-dependent than in the position-based controls, because the set points are more fundamental engine performance parameters, such as air–fuel ratio and EGR rate, rather than the valve opening or the vane opening of a particular EGR valve or VGT. Figure 14.1 illustrates the concept of model-based EGR rate control or air–fuel ratio control in order to reduce transient NO_x and pumping loss. It also demonstrates how to optimize the trade-off between NO_x and soot during load increase and decrease.

The VGT effective area opening or turbine wastegate opening can be modeled similarly with an orifice flow equation (4.57) and by using the turbine power equation. Model-based turbocharger control may reduce turbocharger lag and prevent transient compressor surge. A detailed theory of model-based VGT and EGR nonlinear controls was outlined by Ammann *et al*. (2003). In summary, hardware design needs to match electronic control strategies so that the inherent transient difficulties (e.g., turbocharger lag or high transient pumping loss) can be best alleviated.

Air path control strategies for turbocharged diesel engines usually include the research topics of coordinated EGR–VGT control algorithms, controls by MAP, MAF or exhaust manifold pressure, and controller designs. These topics were extensively researched by Watson and Banisoleiman (1988), Winterbone and Jai-In (1988), Gissinger *et al*. (1990), Duffy *et al*. (1999), Shirawaka *et al*. (2001), Wijetunge *et al*. (2004), Nieuwstadt *et al*. (2000), Osborne and Morris (2002), Nieuwstadt (2003), Ammann *et al*. (2003), Kolmanovsky and Stefanopoulou (2000, 2001), Kolmanovsky *et al*. (1999), Yokomura *et al*. (2004), Mueller *et al*. (2005), Kobayashi *et al*. (2005), Darlington *et al*. (2006), Black *et al*. (2007), Luján *et al*. (2007), Das and Dhinagar (2008), Plianos and Stobart (2008), and Moulin *et al*. (2009).

14.6 Fuel path control and diesel engine governors

Fuel path control is another important area in diesel engine controls. The modeling work includes three main areas: (1) engine speed and governor controls; (2) real-time modeling of fuel system hydraulic dynamics for HIL; and (3) fuel delivery unevenness detection and misfire detection, and their model-based controls. The fuel path dynamics of a diesel engine is very different from that in a port-injection gasoline engine because the diesel engine does not have such issues as wall-wetting and mixing/evaporation during the fuel transporting process.

Mechanical governors and modern electric governors have always been used in diesel engines to control engine speed (SAE J830, 1999). Diesel fuel control and governor designs are part of the fuel system development. Diesel engine governors and fuel path control were investigated by Gant (1984), Okazaki *et al*. (1990), Bazari (1990), Rakopoulos *et al*. (1997a), Mruthunjaya and Dhariwal (2000), Stefanopoulou and Smith (2000), Larisch and Sobieszczanski (2001), Makartchouk (2002), Chatlatanagulchai *et al*. (2009), and Deng *et al*. (2010).

Fuel delivery unevenness detection and misfire detection are important for fuel injection quantity control, especially real-time correction of fuel injection failures. The unevenness refers to the difference in fuel quantity between the cylinders or cycles. Macián *et al*. (2005, 2006a, and 2006b) explored this area of fault diagnostics and proposed improved control algorithms and controller designs. The conventional diagnosis techniques use the instantaneous crankshaft speed during an engine cycle as input information. Although these techniques are able to detect misfire at low engine speeds, they are not very effective at high speeds, especially at low loads (Macián *et al*., 2006a). Other engine system performance parameters were explored as better alternatives to replace or supplement the crankshaft speed signal for the purpose of fault diagnosis and control of the injection and combustion processes in order to better detect fuel delivery unevenness and misfire at high engine speeds. These parameters explored included exhaust manifold pressure, instantaneous turbocharger speed, and the mean temperature in the exhaust port. Macián *et al*. (2006b) reported that using instantaneous turbocharger speed as input signal was effective for fuel quantity control in order to achieve satisfactory correction on fuel injection unevenness. A system-level fault diagnosis model can be developed in engine system design by combining the thermodynamic performance and engine dynamics submodels to facilitate fuel system control development.

14.7 Torque-based controls

Torque-based controls are widely used in both gasoline and diesel engines to facilitate the overall powertrain control demands. Engine system design models can play a key role in developing the more accurate torque-based control models because the diesel engine system design engineers analyze engine torques in their daily work and have a thorough understanding of the engine torque behavior during steady-state and transient conditions. System engineers can complement control engineers very well in this area.

For example, the friction torque and pumping loss torque calculations/ estimates used in the torque-based controls can be refined and enhanced by the system design. Moreover, more advanced parametric dependency of the engine indicated torque and brake torque can be developed by the engine system design engineers for the torque-based control models in order to accurately calculate the engine torques under various operating or aging conditions. Torque-based controls are discussed by Ginoux and Champoussin (1997), Müller and Schneider (2000), Greff and Günther (2001), Heintz *et al*. (2001), Lee *et al*. (2001), Park and Sunwoo (2003), Wang and Chu (2005), Katsumata, *et al*. (2007), and Livshiz *et al*. (2004, 2008) for gasoline engines; and Maloney (2004), Li *et al*. (2002), Grünbacher *et al*. (2003), Chauvin *et al*. (2004), Brahma *et al*. (2008), Tian *et al*. (2008), and Oh *et al*. (2009) for diesel engines.

14.8 Powertrain dynamics and transient controls

14.8.1 Transient performance simulation

Transient engine testing is much more difficult and expensive than steady-state testing. Using engine cycle performance simulation to study transients is a very effective approach, especially for real-time transient modeling. The engine crankshaft lumped model has the governing equation 5.20. In transient engine torque simulation, the 'filling-and-emptying' method is usually used to model the zero-dimensional manifold gas dynamics in order to reduce the computing time. There are several key issues related to model accuracy in transient cycle simulation. They are listed below.

1. During the fast acceleration transient, a sudden increase in fueling causes a rapid decrease in the air–fuel ratio that may result in incomplete combustion. Combustion efficiency as a function of air–fuel ratio was usually assumed as input in the thermodynamic cycle transient simulation. Engine testing or sophisticated combustion simulation that tries to quantify such efficiency change remains challenging.
2. Another difficulty is related to the turbocharger. The instantaneous turbocharger speed N_{TC} (revolution per second) is given by

$$4\pi^2 I_{TC}\frac{\mathrm{d}N_{TC}}{\mathrm{d}t} = \frac{\dot{W}_T - \dot{W}_C - \dot{W}_{f,TC}}{N_{TC}} \qquad 14.2$$

where I_{TC} is the moment of inertia of the turbocharger, $\dot{W}_T$ is the turbine power, $\dot{W}_C$ is the compressor power, and $\dot{W}_{f,TC}$ is the turbocharger bearing friction power. The turbine instantaneous temperature, pressure and efficiency affect N_{TC} and turbocharger lag. Transient turbocharging performance modeling is usually approximated by a quasi-steady-state approach, which computes the instantaneous varying turbocharger parameters within an engine cycle by looking up the steady-state turbocharger maps. Understanding the difference between the steady-state turbine efficiency and the transient efficiency remains challenging. There are large discrepancies between the conclusions from different authors (Capobiano *et al*., 1989; Westin and Angstrom, 2002) who obtained their data using different measurement methods (e.g., pulsating gas stands or on engine *in-situ*) or CFD simulations. As to the turbine flow rate, it was reported that the turbine under unsteady conditions has a 3–6% higher swallowing flow capacity (i.e., the corrected mass flow rate) than that measured at steady state (Capobiano *et al*., 1989). The heat losses, volume, and thermal inertia of the exhaust manifold also have significant effects on the accuracy of turbocharger lag and engine transient performance simulation.

3. Transient engine simulation and accurate prediction of driving cycle fuel economy demand accurate engine control models that reflect the engine calibration set points in the speed–load domain, actual transient control strategies, and the dynamic response of the sensors and the actuators.
4. Accurate simulations of cold start, warm-up and hot start transients are important for analyzing transient emission cycles.
5. Transient simulation also requires accurate models of mechanical losses from main friction components (i.e., piston–ring–liner assembly, bearings and valvetrain) and parasitic losses from the accessories (Taraza *et al*., 2007).

14.8.2 Real-time high-fidelity versus mean-value transient simulations

Real-time high-fidelity simulation with a detailed instantaneous in-cylinder process and manifold gas wave dynamics is currently available. Mean-value models have been used to conduct real-time simulations for engine control design and vehicle driving cycle analysis prior to the real-time high-fidelity modeling. The mean-value models use a simplified map-based approach without calculating the in-cylinder cycle process at a crank-angle-level resolution. Typically, volumetric efficiency, indicated efficiency, and exhaust energy fraction are built in maps as functions of certain parameters such as engine speed, fueling rate, intake manifold pressure, and air–fuel ratio. The mean-value model has some disadvantages: (1) tremendous efforts need to be spent to run the DoE data upfront to build the maps that are actually hardware-specific and primitive; and (2) the model cannot predict the turbocharger lag transient conveniently and accurately. Mean-value models may use a large time step in the order of one engine cycle, while a detailed in-cylinder process model requires 1–5° crank angle resolution. The theory on the mean-value model is provided in Schulten and Stapersma (2003). There is a trend to develop and use the real-time high-fidelity simulation to replace the mean-value models for future transient analysis in order to increase the predictability of the model. Moreover, in engine control simulations, software-in-the-loop (SIL) and hardware-in-the-loop (HIL) are used to validate the algorithms for both steady state and transient operations.

14.9 Sensor dynamics and model-based virtual sensors

14.9.1 Classification of engine sensors

The model-based controls of valves and turbochargers mentioned above are essentially virtual actuators. In their equations, it is noted that the engine temperature, pressure, and flow parameters can be from either actual measurement or virtual sensor modeling. Model-based virtual sensors may replace some actual physical sensors in the engine and aftertreatment systems in order to reduce cost or to enable flexible controls. For instance, switching combustion mechanisms between different speed–load regions can be achieved by predicting the in-cylinder parameters based on the combustion process. The development of virtual sensors relies largely on thermodynamic performance modeling.

Engine sensors can be classified into two categories, the physical sensors (or real sensors) and the virtual sensors. The physical engine sensors usually include: the crank position sensor and the cam position sensor for sensing engine speed, TDC position, fuel injection timing and duration; the accelerator pedal position sensor (SAE J1843, 2009); the pressure sensors for manifold absolute air pressure (MAP), barometric ambient, DPF inlet pressure, oil, and fuel; the temperature sensors for intake manifold air, exhaust manifold gas, DPF inlet and outlet gases, oil, fuel, and coolant; the mass air flow (MAF) sensor; the exhaust oxygen or lambda sensor; the NO_x and NH_3 sensors used in aftertreatment; and the oil and coolant level sensors.

The virtual sensors can be classified by algorithm into two types: physics model based, and empirical data based (e.g., response surface models or neural network models). Moreover, the virtual sensors can be divided into steady-state virtual sensors and dynamic transient virtual sensors. They require different computing algorithms. An example of the virtual sensors is provided in Section 14.9.3.

14.9.2 Development needs for engine sensors

There are abundant needs for future virtual sensors in both steady-state and transient for the next generation of 'intelligent' diesel engines. The needs in future development include the following:

- steady-state virtual MAF and MAP sensors
- other steady-state virtual sensors of gas flow, pressure, and temperature which can be calculated using the 19 equations (4.39–4.57) governing the engine air system described in Chapter 4
- a virtual engine torque sensor in torque-based controls of powertrain central torque-demand coordination
- transient virtual sensors for in-cylinder real-time quantities, for example air temperature, pressure and charge equivalence ratio during the compression stroke in HCCI burning control
- engine-out NO_x, PM/soot, and HC virtual sensors based on the in-cylinder cycle process calculation with heuristic macro-parameter-dependent emissions models
- virtual sensors for aftertreatment flow parameters and outlet emissions built based on heuristic models
- combining all of the above to enable a whole engine model with affordable computing time for integrated virtual sensing. A fully validated physical model can then be used in real-time predictive algorithms of engine controls.

Engine performance and emissions modeling is the foundation for building these advanced virtual sensors. This area opens many challenges and opportunities for system integration analysis and design.

Automotive sensors are introduced in Westbrook and Turner (1994). An overview of diesel engine sensors (physical sensors) is provided by Challen and Stobart (1998). Physics-based virtual sensors are discussed by Grimes *et al*. (2005). Overviews on neural network based virtual sensing and sensors are presented by Atkinson *et al*. (1998) and Nareid *et al*. (2005). EGR and residue gas sensing has been researched by Müller *et al*. (2001) and Leroy *et al*. (2009). Diesel engine air mass flow sensing has been investigated by Höckerdal *et al*. (2008). Diesel exhaust temperature sensing is evaluated by Culbertson *et al*. (2008) and Hori and Todo (2009). A virtual sensor of barometric pressure for altitude detection is presented in Olin and Maloney (1999). The diesel engine exhaust pressure sensor for DPF operation is addressed by Ueno *et al*. (2008).

NO_x sensors have been investigated by Kato *et al*. (1999) and Orban *et al*. (2005). NO_x virtual sensors are presented by Re *et al*. (2005) and Subramaniam *et al*. (2008). Particulate matter sensors are discussed in Allan *et al*. (2003), Warey and Hall (2005), Hauser (2006), Diller *et al*. (2008, 2009), and Cai and Ma (2009).

Cylinder pressure sensors are discussed in Ulrich *et al*. (2001) and Hasegawa *et al*. (2006). Virtual cylinder pressure sensors are explored in Palma *et al*. (2004) and Wang *et al*. (2005).

14.9.3 Virtual sensor modeling of exhaust manifold gas temperature

The virtual sensor for exhaust manifold gas temperature is important for engine durability and EGR rate control. Its modeling is based on the thermodynamic first law energy balance equation 12.2. Assuming

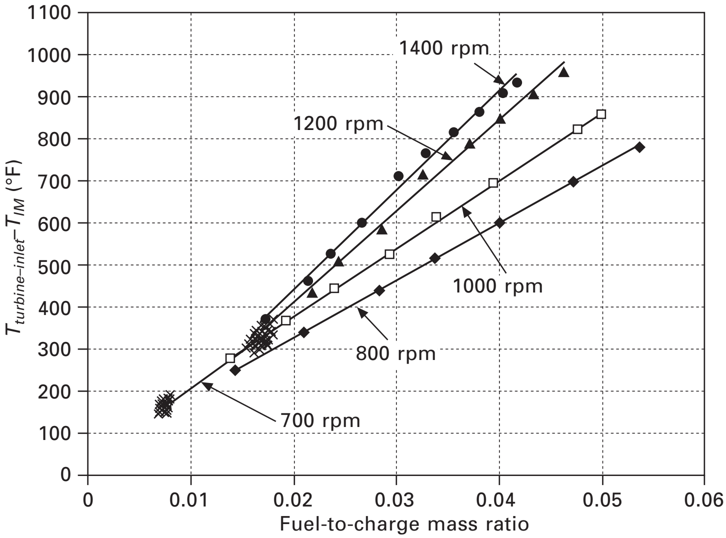
$$\begin{cases} \dot{W}_E = \eta_{th} \cdot (q_{LHV} \dot{m}_{fuel}) \\ \dot{Q}_{base-coolant} = B \cdot (q_{LHV} \dot{m}_{fuel}) \\ \dot{Q}_{miscellaneous,2} = G_2 \cdot (q_{LHV} \dot{m}_{fuel}) \end{cases} \tag{14.3}$$

where η_{th}, B and G_2 are constants or known functions of engine speed, load, air–fuel ratio, and fuel injection timing, the steady-state exhaust manifold gas temperature can be calculated using equation 12.2 as:

$$T_{turbine-inlet} = \frac{\left(\frac{\dot{m}_{IM}}{\dot{m}_{fuel}}\right) c_{p,in} T_{IM} + q_{LHV}(1 - \eta_{th} - B - G_2)}{\left(\frac{\dot{m}_{IM}}{\dot{m}_{fuel}} + 1\right) c_{p,ex}} \tag{14.4}$$

where $\dot{m}_{IM}$ is the intake manifold mixture flow rate of both fresh air and EGR, and $c_{p,ex}$ and $c_{p,in}$ are the equivalent average specific heats of the exhaust flow and the intake manifold flow, respectively. It is observed that the exhaust manifold gas temperature is a function of the 'mass-to-fuel' ratio $(\dot{m}_{IM}/\dot{m}_{fuel})$. In equation 14.4, one method to estimate the engine brake thermal efficiency η_{th} is to calculate the engine brake power using the indicated torque and the engine friction obtained from the torque-based controls. The prediction of the engine indicated torque and the brake torque has been used increasingly as a part of the coordinated shift control in automotive powertrains. In equation 14.4, an accurate estimation of B is a challenge. The value of B depends on engine speed, load, charge mass-to-fuel ratio, fuel injection timing, intake manifold gas temperature, coolant temperature, etc. One approach is to use the equations of the in-cylinder cycle process along with the Woschni heat transfer coefficient α_g outlined in Chapter 4 to first calculate the transient instantaneous exhaust manifold gas temperature, and then average it over an engine cycle to obtain the steady-state exhaust manifold gas temperature. The other approach is to develop heuristic models for the steady-state values of B in order to build its sensitivity to other engine performance and operating parameters.

Steady-state engine performance test data reveal that there is a correlation between the temperature difference across the cylinder and the fuel-to-charge mass ratio (Fig. 14.2), and the slope of the correlation line changes



14.2 Illustration of virtual sensor – correlation between exhaust manifold gas temperature and fuel-to-charge ratio.

at different engine speeds. Such a correlation can be explained as follows. If it is assumed that $\dot{m}_{IM}$ is much greater than $\dot{m}_{fuel}$, and $c_{p,in} \approx c_{p,ex}$, equation 14.4 is simplified to:

$$\begin{aligned} T_{turbine\text{–}inlet} - T_{IM} &\approx \frac{q_{LHV}(1 - \eta_{th} - B - G_2)}{c_{p,ex}}\left(\frac{\dot{m}_{fuel}}{\dot{m}_{IM}}\right) \\ &= C\left(\frac{\dot{m}_{fuel}}{\dot{m}_{IM}}\right) \end{aligned} \qquad 14.5$$

where the coefficient C is a function of engine speed. A simplified empirical model of the virtual sensor of exhaust manifold gas temperature can be developed as follows:

$$T_{turbine\text{–}inlet} - T_{IM} = (C_1 + C_2 N_E) + (C_3 + C_4 N_E)\left(\frac{\dot{m}_{fuel}}{\dot{m}_{IM}}\right) \qquad 14.6$$

where C_1, C_2, C_3 and C_4 are the model tuning constants, and N_E is the engine speed. It should be noted that although the model equation 14.6 can predict the steady-state exhaust temperatures reasonably well, it may not be able to accurately predict the exhaust temperatures during fast transients due to the lack of thermal-lag related factors in the model.

14.10 On-board diagnostics (OBD) and fault diagnostics

On-board diagnostics (SAE J1699-2, 1998; J1699-3, 2009; J1930, 2008) is an engine control topic, and has been researched by Grimaldi and Mariani (1999), Geraldo (2006), Vitale *et al*. (2007), Millo *et al*. (2009), and Fischer *et al*. (2009).

Engine fault diagnostics is closely related to engine system design and electronic controls because many parameters used in the diagnosis are system-level performance parameters such as the instantaneous gas pressure, temperature, flow, and the component rotational speeds. A system-level dynamics model is usually required for model-based diagnostics. The area of engine fault diagnostics is reviewed by Haddad (1984) and further investigated by Schwarte and Isermann (2002), Macián *et al*. (2004, 2005, 2006a, and 2006b), and Yan *et al*. (2007). Diesel engine protection is reviewed by Fouch and Gross (1991).

14.11 Engine controller design

The theoretical foundation of modern control engineering is summarized in the classical textbooks authored by Ogata (2002, 2004), Palm III (2005),Dorf and Bishop (2007), and DiStefano *et al*. (1990). Internal combustion engine control and modeling is introduced in Guzzella and Onder (2004) and Guzzella and Sciarretta (2005).

Engine controller design itself generally does not belong to the scope of diesel engine system design. It falls within the job function of a control engineer. However, dynamic engine system design cannot be implemented without considering controller design. Diesel engine controls (e.g., air path control, fuel system control, speed control, etc.) are nonlinear controls. Linear design and reduced-order design are the common techniques used in controller designs. PI, PD and PID controllers are still the most widely used controllers in engines. Nonlinear numerical simulation of engine system dynamics may provide a virtual confirmation of the controller behavior and the control strategy, but the simulation itself is not a controller design tool. There is a need to bridge the gap between engine system simulation and controller design in order to promote the simulation-based controller design for nonlinear controllers, especially for the air path and powertrain controls.

Engine controls and controller designs are presented in Chin and Coats (1986), Weisman (1987), Tsai and Goyal (1986), Tuken *et al*. (1990), Scotson and Heath (1996), Balfour *et al*. (2000), Christen *et al*. (2001), and Malkhede *et al*. (2005). An adaptive torque controller design is reported by Fullmer *et al*. (1992). PID controller design theories are presented by Strom and Hagglund (1995, 2005). A PID controller for diesel engine speed control has been investigated by Mruthunjaya and Dhariwal (2000). Idle speed control was studied by Memering and Meckl (1994, 2002). Diesel engine air system (EGR–VGT) controller designs have been researched by Stefanopoulou *et al*. (2000) and Utkin *et al*. (2000). A PID controller tuning for diesel engine EGR–VGT control is investigated in Wahlström *et al*. (2008). A PID controller design for a gasoline engine is introduced by Lauber *et al*. (2002). The multi-input multi-output (MIMO) controller has been researched by Stefanopoulou and Smith (2000) to reduce transient smoke for a marine diesel engine. The comparison between a PID controller and a neural network controller is provided by Tsuchiya *et al*. (2003). A fuzzy logic controller for diesel engines is reported by Plianos *et al*. (2007). Robust control for marine diesel engines is summarized by Xiros (2002).

14.12 Software-in-the-loop (SIL) and hardware-in-the-loop (HIL)

Software-in-the-loop is introduced by Philipp *et al*. (2005) and Mitts *et al*. (2009). Hardware-in-the-loop has been actively applied to engine/powertrain control development and investigated in research (Isermann *et al*., 1998; Rolfsmeier *et al*., 2003; Nabi *et al*., 2004; Shayler *et al*., 2005; Köhl and Jegminat, 2005; Steiber *et al*., 2005; Wanpal *et al*., 2006; Schuette and Ploeger, 2007; Wang *et al*. 2009; and Dhaliwal *et al*., 2009).

14.13 Cylinder-pressure-based controls

Good references on cylinder-pressure-based controls for diesel engines are provided by Johnson *et al*. (1999), Nakayama *et al*. (2003, 2008), Klein *et al*. (2007), and Husted *et al*. (2007). Due to limited space, the details are not discussed here.

14.14 Homogeneous charge compression ignition (HCCI) controls

HCCI control is an important area of combustion control. Detailed discussions on HCCI are given in Section 9.8 in Chapter 9. For completeness, selected references are provided here (Olsson *et al*., 2001; Zhao and Asmus, 2003; Strandh *et al*., 2005; Bengtsson *et al*., 2006; Narayanaswamy and Rutland, 2006; Chauvin *et al*., 2006; Wang, 2007; Tuneståll and Johansson, 2007; Kumar *et al*., 2007).

14.15 References and bibliography

Allan W D E, Freeman R D, Pucher G R, Faux D, Bardon M F and Gardiner D P (2003), 'Development of a smoke sensor for diesel engines', SAE paper 2003-01-3084.

Allmendinger K, Guzzella L, Seiler A and Loffeld O (2001), 'A method to reduce the calculation time for an internal combustion engine model', SAE paper 2001-01-0574.

Ammann M, Fekete N P, Guzzella L and Glattfelder A H (2003), 'Model-based control of the VGT and EGR in a turbocharged common-rail diesel engine: theory and passenger car implementation', SAE paper 2003-01-0357.

Anderson R C (1991), 'Electronic diesel engine controls for industrial applications. a systems perspective', SAE paper 912684.

Atkinson C M, Long T W and Hanzevack E L (1998), 'Virtual sensing: a neural network-based intelligent performance and emissions prediction system for on-board diagnostics and engine control', SAE paper 980516.

Atkinson C M, Allain M, Kalish Y and Zhang H (2009), 'Model-based control of diesel engines for fuel efficiency optimization', SAE paper 2009-01-0727.

Balfour G, Dupraz P, Ramsbottom M and Scotson P (2000), 'Diesel fuel injection control for optimum driveability', SAE paper 2000-01-0265.

Baumann J, Torkzadeh D D, Kiencke U, Schlegl T and Oestreicher W (2004), 'Integration of a common rail diesel engine model into an industrial engine software development process', SAE paper 2004-01-0900.

Bazari Z (1990), 'The transient performance analysis of a turbocharged vehicle diesel engine with electronic fuelling control', SAE paper 900236.

Benajes J, Luján J M and Serrano J R (2000), 'Predictive modelling study of the transient load response in a heavy-duty turbocharged diesel engine', SAE paper 2000-01-0583.

Benajes J, Luján J M, Bermudez V and Serrano J R (2002), 'Modelling of turbocharged diesel engines in transient operation, part 1: insight into the relevant physical phenomena', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 216, 431–441.

Bengtsson J, Strandh P, Johansson R, Tunestål P and Johansson B (2006), 'Multi-output control of a heavy duty HCCI engine using variable valve actuation and model predictive control', SAE paper 2006-01-0873.

Black J, Eastwood P G, Tufail K, Winstanley T, Hardalupas Y and Taylor A M K P (2007), 'Diesel engine transient control and emissions response during a European Extra-Urban Drive Cycle (EUDC)', SAE paper 2007-01-1938.

Brace C J, Cox A, Hawley J G, Vaughan N D, Wallace F W, Horrocks R W and Bird G L (1999), 'Transient investigation of two variable geometry turbochargers for passenger vehicle diesel engines', SAE paper 1999-01-1241.

Brahma I, Sharp M C and Frazier T R (2008), 'Estimation of engine torque from a first law based regression model', SAE paper 2008-01-1014.

Buckland J, Cook J, Kolmanovsky I and Sun J (2000), 'Technology assessment of boosted direct injection stratified charge gasoline engines', SAE paper 2000-01-0249.

Cai W and Ma L (2009), 'Numerical investigation of an optical sensor for modern diesel engines', SAE paper 2009-01-1514.

Capobiano M, Gambarotta A and Cipolla G (1989), 'Influence of the pulsating flow operation on the turbine characteristics of a small internal combustion engine turbocharger', *IMechE Conference Transactions of the Seventh International Conference on Turbochargers and Turbocharging*, IMechE paper C372/019/1989.

Caraceni A, Cristofaro F D, Ferrara F, Scala S and Philipp O (2003), 'Benefits of using a real-time engine model during engine ECU development', SAE paper 2003-01-1049.

Challen B J and Stobart R K (1998), 'Some more diesel engine sensors', SAE paper 980167.

Chatlatanagulchai W, Aroonsrisopon T and Wannatong K (2009), 'Robust common-rail pressure control for a diesel-dual-fuel engine using QFT-based controller', SAE paper 2009-01-1799.

Chauvin J, Petit N, Rouchon P, Corde G, Moulin P and Castagné M (2004), 'Observer design for torque balancing on a DI engine', SAE paper 2004-01-1370.

Chauvin J, Petit N, Rouchon P, Paris E M, Corde G and Vigild C (2006), 'Air path estimation on diesel HCCI engine', SAE paper 2006-01-1085.

Chen J-S (2008), 'Mean value engine model using object oriented modeling', SAE paper 2008-01-0982.

Chevalier A, Müller M and Hendricks E (2000), 'On the validity of mean value engine models during transient operation', SAE paper 2000-01-1261.

Chin Y-K and Coats F E (1986), 'Engine dynamics: time-based versus crank-angle based', SAE paper 860412.

Christen U, Vantine K J and Collings N (2001), 'Event-based mean-value modeling of DI diesel engines for controller design', SAE paper 2001-01-1242.

Chung N, Kim S and Sunwoo M (2005), 'Nonlinear dynamic model of a turbocharged diesel engine', SAE paper 2005-01-0017.

Corti E and Solieri L (2005), 'Rapid control prototyping system for combustion control', SAE paper 2005-01-3754.

Culbertson D P, Harvey D D and Kovacevich S A (2008), 'The development of active thermocouples for diesel exhaust temperature measurement', SAE paper 2008-01-2492.

Darlington A, Glover K and Collings N (2006), 'A simple diesel engine air-path model to predict the cylinder charge during transients: strategies for reducing transient emissions spikes', SAE paper 2006-01-3373.

Das H B and Dhinagar S J (2008), 'Airpath modelling and control for a turbocharged diesel engine', SAE paper 2008-01-0999.

Deng J, Winward E, Stobart R and Desai P R (2010), 'Modeling techniques to support fuel path control in medium duty diesel engines', SAE paper 2010-01-0332.

Dhaliwal A, Nagaraj S C and Ali S (2009), 'Hardware-in-the-loop simulation for hybrid electric vehicles – an overview, lessons learned and solutions implemented', SAE paper 2009-01-0735.

Diller T T, Hall M J and Matthews R D (2008), 'Further development of an electronic particulate matter sensor and its application to diesel engine transients', SAE paper 2008-01-1065.

Diller T T, Osara J, Hall M J and Matthews R D (2009), 'Electronic particulate matter sensor – mechanisms and application in a modern light-duty diesel vehicle', SAE paper 2009-01-0647.

DiStefano J J, Stubberud A R and Williams I J (1990), *Schaum's Outline of Theory and Problems of Feedback and Control Systems*, 2nd edn, McGraw-Hill, New York.

Dorf R C and Bishop R H (2007), *Modern Control Systems*, 11th edn, Pearson Education, Upper Saddle River, NJ.

Duffy K P, Miller R L, Bowyer R, Bromnick P A, Mason J, Stobart R, Wilson R P, Middlemiss I D and Parsons M A (1999), 'Model-based EGR control development for an HSDI engine', *ASME 1999 Fall Technical Conference*, ASME paper No. 99-ICE-204, ICE-Vol. 33(1), 21–32.

Eriksson L (2002), 'Mean value models for exhaust system temperatures', SAE paper 2002-01-0374.

Eriksson L (2007), 'Modeling and control of turbocharged SI and DI engines', *Oil & Gas Science and Technology – Rev. IFP*, 62(4), 523–538.

Erkkinen T and Breiner S (2007), 'Automatic code generation – technology adoption lessons learned from commercial vehicle case studies', SAE paper 2007-01-4249.

Felsch C, Hoffmann K, Vanegas A, Drews P, Barths H, Abel D and Peters N (2009),'Combustion model reduction for diesel engine control design', *International Journal of Engine Research*, 10(6), 359–387.

Filipi Z, Wang Y and Assanis D (2001), 'Effect of variable geometry turbine (VGT) on diesel engine and vehicle system transient response', SAE paper 2001-01-1247.

Fiorani P, Gambarotta A, Tonetti M and Lanfranco E (2006), 'A real-time model for the simulation of transient behaviour of automotive diesel engines', SAE paper 2006-01-3007.

Fischer M, Boettcher J, Kirkham C and Georgi R (2009), 'OBD of diesel EGR using artificial neural networks', SAE paper 2009-01-1427.

Fouch W H and Gross D M (1991), 'Advances in diesel protection systems', SAE paper 911877.

Fullmer R R, Tuken T and Gerpen J H V (1992), 'Adaptive torque control of a diesel engine for transient test cycles', SAE paper 920238.

Gant G C (1984), 'The governing of diesel engines', in Haddad S D and Watson N (eds), *Principles and Performance in Diesel Engineering*, Ellis Horwood Limited, Chichester.

Gant G C and Alves G J (1990), 'Progress in electronic control of large diesel engines', *Transactions of the ASME*, *Journal of Engineering for Gas Turbines and Power*, 112(3), 280–286.

Geraldo G (2006), 'Differences between on board diagnostic systems (EOBD, OBD-II, OBD-BR1 and OBD-BR2)', SAE paper 2006-01-2671.

Ginoux S and Champoussin J-C (1997), 'Engine torque determination by crankangle measurements: state of the art, future prospects', SAE paper 970532.

Gissinger G L, Frank P M and Wunnennberg J (1990), 'Model-based electronic diesel engine and turbocharger control', SAE paper 900595.

Greff A and Günther T (2001), 'A new approach for a multi-fuel, torque based ECU concept using automatic code generation', SAE paper 2001-01-0267.

Grimaldi C N and Mariani F (1999), 'Prediction of engine operational parameters for on board diagnostics using a free model technology', SAE paper 1999-01-1224.

Grimes M R, Verdejo J R and Bogden D M (2005), 'Development and usage of a virtual mass air flow sensor', SAE paper 2005-01-0074.

Grondin O, Stobart R, Chafouk H, Maquet J (2004), 'Modelling the compression ignition engine for control: review and future trends', SAE paper 2004-01-0423.

Grünbacher E, Langthaler P, Steinmaurer G, del Re L and Kokal H (2003), 'Adaptive inverse torque control of a diesel engine using adaptive mapping update', SAE paper 2003-01-0397.

Guzzella L (2010), 'Advanced control and engine management for future light-duty diesel engines', in Zhao H (ed.), *Advanced Direct Injection Combustion Engine Technologies and Development, Volume 2: Diesel Engines*, Woodhead Publishing Limited, Cambridge, UK.

Guzzella L and Amstutz A (1998), 'Control of diesel engines', *IEEE Control Systems*, October, 53–71.

Guzzella L and Onder C H (2004), *Introduction to Modeling and Control of Internal Combustion Engine Systems*, Springer-Verlag, Berlin.

Guzzella L and Sciarretta A (2005), *Vehicle Propulsion Systems – Introduction to Modeling and Optimization*, Springer-Verlag, Berlin.

Haddad S D (1984), 'Condition monitoring and fault diagnosis in diesel engines', in Haddad S D and Watson N (eds), *Principles and Performance in Diesel Engineering*, Ellis Horwood Limited, Chichester.

Hafner M (2001), 'Model based determination of dynamic engine control function parameters', SAE paper 2001-01-1981.

Hasegawa M, Shimasaki Y, Yamaguchi S, Kobayashi M, Sakamoto H, Kitayama N and Kanda T (2006), 'Study on ignition timing control for diesel engines using in-cylinder pressure sensor', SAE paper 2006-01-0180.

Hauser G (2006), 'Smoke particulate sensors for OBD and high precision measuring', SAE paper 2006-01-3549.

He Y (2005), 'Development and validation of a 1D model of a turbocharged V6 diesel engine operating under steady-state and transient conditions', SAE paper 2005-01-3857.

He Y and Lin C-C (2007), 'Development and validation of a mean value engine model for integrated engine and control system simulation', SAE paper 2007-01-1304.

He Y, Lin C-C and Gangopadhyay A (2006), 'Integrated simulation of the engine and control system of a turbocharged diesel engine', SAE paper 2006-01-0439.

Heintz N, Mews M, Stier G, Beaumont A J and Noble A D (2001), 'An approach to torque-based engine management systems', SAE paper 2001-01-0269.

Hendricks E (1989), 'Mean value modelling of large turbocharged two-stroke diesel engines', SAE paper 890564.

Hendricks E, Chevalier A, Jensen M, Sorenson S C, Trumpy D and Asik J (1996), 'Modelling of the intake manifold filling dynamics', SAE paper 960037.

Hirschlieb G, Schiller G and Stottler S (1995), 'Engine control', in Jurgen R (ed.), *Automotive Electronics Handbook*, McGraw-Hill, New York.

Höckerdal E, Eriksson L and Frisk E (2008), 'Air mass-flow measurement and estimation in diesel engines equipped with EGR and VGT', SAE paper 2008-01-0992.

Hori T and Todo Y (2009), 'High accuracy exhaust gas temperature sensor with anti-resonance structure', SAE paper 2009-01-0641.

Husted H, Kruger D, Fattic G, Ripley G and Kelly E (2007), 'Cylinder pressure-based control of pre-mixed diesel combustion', SAE paper 2007-01-0773.

Isermann R, Sinsel S and Schaffnit J (1998), 'Modeling and real-time simulation of diesel engines for control design', SAE paper 980796.

Jensen J-P, Kristensen A F, Sorenson S C, Houbak N and Hendricks E (1991), 'Mean value modeling of a small turbocharged diesel engine', SAE paper 910070.

Johnson W P, Li S C and Oppenheim A K (1999), 'Potential for closed loop air–fuel ratio management of a diesel engine', SAE paper 1999-01-0517.

Jung M, Ford R G, Glover K, Collings N, Christen U and Watts M (2002), 'Parameterization and transient validation of a variable geometry turbocharger for mean-value modeling at low and medium speed-load points', SAE paper 2002-01-2729.

Kämmer A, Liebl J, Krug C, Munk F and Reuss H-C (2003), 'Real-time engine models', SAE paper 2003-01-1050.

Kao M and Moskwa J J (1995), 'Turbocharged diesel engine modeling for nonlinear engine control and state estimation', *Transactions of the ASME*, *Journal of Dynamic Systems, Measurement, and Control*, 117, 20–30.

Kato N, Kokune N, Lemire B and Walde T (1999), 'Long term stable NO_x sensor with integrated in-connector control electronics', SAE paper 1999-01-0202.

Katsumata M, Kuroda Y and Ohata A (2007), 'Development of an engine torque estimation model: integration of physical and statistical combustion model', SAE paper 2007-01-1302.

Klein P, Grüter R and Loffeld O (2007), 'Real-time estimation of the exhaust gas recirculation ratio based on cylinder pressure signals', SAE paper 2007-01-0493.

Kobayashi M, Aoyagi Y, Misawa M, Odaka M and Goto Y (2005), 'Proposal for electronic control system using intake air oxygen concentration in high boost and high EGR system in multi-cylinder diesel engine', *Review of Automotive Engineering*, 26(4), 399–403.

Köhl S and Jegminat D (2005), 'How to do hardware-in-the-loop simulation right', SAE paper 2005-01-1657.

Kolmanovsky I V and Stefanopoulou A G (2000), 'Evaluation of turbocharger power assist system using optimal control techniques', SAE paper 2000-01-0519.

Kolmanovsky I V and Stefanopoulou A G (2001), 'Optimal control techniques for assessing feasibility and defining subsystem level requirements: an automotive case study', *IEEE Transactions on Control Systems Technology*, 9(3), 524–534.

Kolmanovsky I, Nieuwstadt M and Moraal P (1999), 'Optimal control of variable geometry turbocharged diesel engines with exhaust gas recirculation', *Proceedings of the ASME Dyanmic Systems and Control Division*, 67, 265–273.

Kumar R, Zheng M, Asad U and Reader G T (2007), 'Heat release based adaptive control to improve low temperature diesel engine combustion', SAE paper 2007-01-0771.

Larisch J and Sobieszczanski M (2001), 'Impact of the proportional factors of electronic speed governor on diesel engine performance', SAE paper 2001-01-0265.

Lauber J, Guerra T-M and Perruquetti W (2002), 'IC engine: tracking control for an inlet manifold with EGR', SAE paper 2002-01-2156.

Ledger J D and Walmsley S (1971), 'Computer simulation of a turbocharged diesel engine operating under transient load conditions', SAE 710177.

Lee B, Rizzoni G, Guezennec Y, Soliman A, Cavalletti M and Waters J (2001), 'Engine control using torque estimation', SAE paper 2001-01-0995.

Lehner C, Parker G, Arici O and Johnson J (2001), 'Design and development of a model based feedback controlled cooling system for heavy duty diesel truck applications using a vehicle engine cooling system simulation', SAE paper 2001-01-0336.

Leroy T, Bitauld M, Chauvin J and Petit N (2009), 'In-cylinder burned gas rate estimation and control on VVA diesel engines', SAE paper 2009-01-0366.

Li J, Yang M, Zhou M and Li X (2002), 'Advanced torque estimation and control algorithm of diesel engines', SAE paper 2002-01-0198.

Livshiz M, Kao M and Will A (2004), 'Validation and calibration process of powertrain model for engine torque control development', SAE paper 2004-01-0902.

Livshiz M, Kao M and Will A (2008), 'Engine torque control variation analysis', SAE paper 2008-01-1016.

Luján J M, Climent H, Guardiola C and Garcia-Ortiz J V (2007), 'A comparison of different algorithms for boost pressure control in a heavy-duty tubocharged diesel engine', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 221, 629–640.

Lundstrom R R and Gall J M (1986), 'A comparison of transient vehicle performance using a fixed geometry, wastegated turbocharger and a variable geometry turbocharger', SAE paper 860104.

Ma J and Agnew B (1994), 'Numerical solution and study of thermal effect of transient processes of a turbocharged diesel engine', SAE paper 941731.

Ma J and Gu H (1990), 'A mathematical simulation and test to transient performance of turbo-charged diesel engine', SAE paper 901671.

Macián V, Luján J M, Bermudez V and Guardiola C (2004), 'Exhaust pressure pulsation observation from turbocharger instantaneous speed measurement', *Measurement Science and Technology*, 15(6), 1185–1194.

Macián V, Galindo J, Luján J M and Guardiola C (2005), 'Detection and correction of injection failures in diesel engines on the basis of turbocharger instantaneous speed frequency analysis', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 219, 691–701.

Macián V, Luján J M, Guardiola C and Perles A (2006a), 'A comparison of different methods for fuel delivery unevenness detection in diesel engines', *Mechanical Systems and Signal Processing*, 20(8), 2219–2231.

Macián V, Luján J M, Guardiola C and Yuste P (2006b), 'DFT-based controller for fuel injection unevenness correction in turbocharged diesel engines', *IEEE Transactions on Control Systems Technology*, 14(5), 819–827.

Makartchouk A (2002), *Diesel Engine Engineering: Thermodynamics, Dynamics, Design, and Control*, Marcel Dekker, New York.

Malkhede D N, Seth B and Dhariwal H C (2005), 'Mean value model and control of a marine turbocharged diesel engine', SAE paper 2005-01-3889.

Maloney P J (2004), 'Embedded torque estimator for diesel engine control application', SAE paper 2004-01-1371.

Marzouk M and Watson N (1978), 'Load acceptance of turbocharged diesel engines', IMechE paper C54/78.

Memering D W and Meckl P H (1994), 'A comparison of control techniques applied to diesel engine idle speed regulation', *ASME 16th Annual Fall Technical Conference, October, Lafayette, Indiana,* ICE-Vol. 22, *Heavy Duty Engines: A Look at the Future*, pp. 57–66.

Memering D W and Meckl P H (2002), 'Comparison of adaptive control techniques applied to diesel engine idle speed regulation', *Transactions of the ASME*, *Journal of Dynamic Systems*, *Measurement*, *and Control*, 124, 682–688.

Millo F, Ferraro C V, Giacominetto P F and Cianflone F (2009), 'An experimental investigation on OBD II techniques for fuel injection system monitoring in a common rail passenger car diesel engine', SAE paper 2009-01-0240.

Mitts K J, Lang K, Roudier T and Kiskis D L (2009), 'Using a co-simulation framework to enable software-in-the-loop powertrain system development', SAE paper 2009-01-0520.

Moraal P and Kolmanovsky I (1999), 'Turbocharger modeling for automotive control applications', SAE paper 1999-01-0908.

Moulin P, Grondin O and Fontvieille L (2009), 'Control of a two stage turbocharger on a diesel engine', *Proceedings of Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, Shanghai, P. R. China, December 16–18, pp. 5200–5206.

Mruthunjaya U and Dhariwal H C (2000), 'Investigation of control of speed of diesel engines using PID governor', SAE paper 2000-01-0264.

Mueller V, Christmann R, Muenz S and Gheorghiu V (2005), 'System structure and controller concept for an advanced turbocharger/EGR system for a turbocharged passenger car diesel engine', SAE paper 2005-01-3888.

Müller R and Schneider B (2000), 'Approximation and control of the engine torque using neural networks', SAE paper 2000-01-0929.

Muller M, Olin P M and Schreurs B (2001), 'Dynamic EGR estimation for production engine control', SAE paper 2001-01-0553.

Nabi S, Balike M, Allen J and Rzemien K (2004), 'An overview of hardware-in-the-loop testing systems at Visteon', SAE paper 2004-01-1240.

Nakayama S, Fukuma T, Matsunaga A, Miyake T and Wakimoto T (2003), 'A new

dynamic combustion control method based on charge oxygen concentration for diesel engines', SAE paper 2003-01-3181.
Nakayama S, Ibuki T, Hosaki H and Tominaga H (2008), 'An application of model based combustion control to transient cycle-by-cycle diesel combustion', SAE paper 2008-01-1311.
Narayanaswamy K and Rutland C J (2006), 'A modeling investigation of combustion control variables during DI-diesel HCCI engine transients', SAE paper 2006-01-1084.
Nareid H, Grimes M and Verdejo J (2005), 'A neural network based methodology for virtual sensor development', SAE paper 2005-01-0045.
Nieuwstadt M (2003), 'Coordinated control of EGR valve and intake throttle for better fuel economy in diesel engines', SAE paper 2003-01-0362.
Nieuwstadt M J, Kolmanovsky I V and Moraal P E (2000), 'Coordinated EGT-VGT control for diesel engines: an experimental comparison', SAE paper 2000-01-0266.
Ogata K (2002), *Modern Control Engineering*, 4th edn, Prentice Hall, Upper Saddle River, NJ.
Ogata K (2004), *System Dynamics*, 4th edn, Pearson Prentice Hall, Upper Saddle River, NJ.
Oh S, Kim D, Kim J, Oh B, Lee K and Sunwoo M (2009), 'Real-time IMEP estimation for torque-based engine control using an in-cylinder pressure sensor', SAE paper 2009-01-0244.
Okazaki T, Akagi J and Yawata Y (1990), 'Komatsu high pressure fuel injection pump "KP21"', SAE paper 901669.
Olin P M (2008), 'A mean-value model for estimating exhaust manifold pressure in production engine applications', SAE paper 2008-01-1004.
Olin P M and Maloney P J (1999), 'Barometric pressure estimator for production engine control and diagnostics', SAE paper 1999-01-0206.
Olsson J-O, Tunestal P and Johansson B (2001), 'Closed-loop control of an HCCI engine', SAE paper 2001-01-1031.
Orban J E, Naber S J, Sharp C A, Khair M K and McGill R N (2005), 'Long-term aging of NO_x sensors in heavy-duty engine exhaust', SAE paper 2005-01-3793.
Osborne R P and Morris T (2002), 'Aspects of EGR control on turbocharged light-duty diesel engines', IMechE paper C602/013/2002.
Pacitti G C, Amphlett S, Miller P, Norris R and Truscott A (2008), 'Real-time crank-resolved engine simulation for testing new engine management systems', SAE paper 2008-01-1006.
Palm III W J (2005), *System Dynamics*, McGraw-Hill, New York.
Palma G, Scognamiglio O and Lavorgna M (2004), 'Low cost virtual pressure sensor', SAE paper 2004-01-1367.
Papadimitriou I, Warner M, Silvestri J, Lennblad J and Tabar S (2005), 'Neural network based fast-running engine models for control-oriented applications', SAE paper 2005-01-0072.
Park S and Sunwoo M (2003), 'Torque estimation of spark ignition engines via cylinder pressure measurement', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 217, 809–817.
Payri F, Reyes E and Serrano J R (1999), 'A model for load transients of turbocharged diesel engines', SAE paper 1999-01-0225.
Payri F, Benajes J, Galindo J and Serrano J R (2002), 'Modelling of turbocharged diesel engines in transient operation. Part 2: wave action models for calculating the transient operation in a high speed direct injection engine', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 216, 479–493.

Pettiti M, Pilo L and Millo F (2007), 'Development of a new mean value model for the analysis of turbolag phenomena in automotive diesel engines', SAE paper 2007-01-1301.

Philipp O, Buhl M, Diehl S, Huber M, Roehlich S and Thalhauser J (2005), 'Engine ECU function development using software-in-the-loop methodology', SAE paper 2005-01-0049.

Pilley A D, Noble A D, Beaumont A J, Needham J R and Porter B C (1989), 'Optimization of heavy-duty diesel engine transient emissions by advanced control of a variable geometry turbocharger', SAE paper 890395.

Plianos A and Stobart R (2008), 'Modeling and control of diesel engines equipped with a two-stage turbo-system', SAE paper 2008-01-1018.

Plianos A, Stobart R and Achir A (2007), 'Real-time adaptive predictive control of the diesel engine air-path based on fuzzy parameters estimation', SAE paper 2007-01-0971.

Qiao J, Dent J C and Garner C P (1992), 'Diesel engine modelling under steady and transient conditions using a transputer based concurrent computer', SAE paper 922226.

Rakopoulos C D and Giakoumis E G (2006), 'Review of thermodynamic diesel engine simulations under transient operating conditions', SAE paper 2006-01-0884.

Rakopoulos C D and Giakoumis E G (2007), 'Study of the transient operation of low heat rejection turbocharged diesel engine including wall temperature oscillations', SAE paper 2007-01-1091.

Rakopoulos C D and Giakoumis E G (2009), *Diesel Engine Transient Operation – Principles of Operation and Simulation Analysis*, Springer-Verlag London Limited, London.

Rakopoulos C D, Giakoumis E G and Hountalas D T (1997a), 'A simulation analysis of the effect of governor technical characteristics and type on the transient performance of a naturally aspirated IDI diesel engine', SAE paper 970633.

Rakopoulos C D, Hountalas D T, Mavropoulos G C and Giakoumis E G (1997b), 'An integrated transient analysis simulation model applied in thermal loading calculations of an air-cooled diesel engine under variable speed and load conditions', SAE paper 970634.

Rakopoulos C D, Giakoumis E G, Hountalas D T and Rakopoulos D C (2004), 'The effect of various dynamic, thermodynamic and design parameters on the performance of a turbocharged diesel engine operating under transient load conditions', SAE paper 2004-01-0926.

Rakopoulos C D, Michos C N and Giakoumis E G (2005), 'Study of the transient behavior of turbocharged diesel engines including compressor surging using a linearized quasi-steady analysis', SAE paper 2005-01-0225.

Rakopoulos C D, Giakoumis E G and Dimaratos A M (2007), 'Evaluation of various dynamic issues during transient operation of turbocharged diesel engine with special reference to friction development', SAE paper 2007-01-0136.

Re L, Langthaler P, Furtmueller C, Winkler S and Affenzeller M (2005), 'NO_x virtual sensor based on structure identification and global optimization', SAE paper 2005-01-0050.

Rolfsmeier A, Richert J and Leinfellner R (2003), 'A new calibration system for ECU development', SAE paper 2003-01-0131.

SAE Surface Vehicle Standard J830 (1999), 'Fuel injection equipment nomenclature'.

SAE Surface Vehicle Recommended Practice J1699-2 (1998), 'OBD-II related SAE specification verification test procedures'.

SAE Surface Vehicle Recommended Practice J1699-3 (2009), '(R) Vehicle OBD II compliance test cases'.

SAE Surface Vehicle Recommended Practice J1843 (2009), 'Accelerator pedal position sensor for use with electronic controls in medium- and heavy-duty vehicle applications – truck and bus'.

SAE Surface Vehicle Recommended Practice J1922 (2008), 'Powertrain control interface for electronic controls used in medium and heavy-duty diesel on-highway vehicle applications'.

SAE Surface Vehicle Recommended Practice J1930 (2008), '(R) Electrical/electronic systems diagnostic terms, definitions, abbreviations, and acronyms – equivalent to ISO/TR 15031-2'.

Schuette H and Ploeger M (2007), 'Hardware-in-the-loop testing of engine control units – a technical survey', SAE paper 2007-01-0500.

Schulten P J M and Stapersma D (2003), 'Mean value modeling of the gas exchange of a 4-stroke diesel engine for use in powertrain applications', SAE paper 2003-01-0219.

Schulze T, Wiedemeier M and Schuette H (2007), 'Crank angle-based diesel engine modeling for hardware-in-the-loop applications with in-cylinder pressure sensors', SAE paper 2007-01-1303.

Schwarte A and Isermann R (2002), 'Model-based fault detection of diesel intake with common production sensors', SAE paper 2002-01-1146.

Scotson P G and Heath W P (1996), 'Crank-angle based control of high speed diesel engines', SAE paper 960619.

Serrano J R, Climent H, Arnau F J and Traumat G (2005), 'Global analysis of the EGR circuit in a HSDI diesel engine in transient operation', SAE paper 2005-01-0699.

Serrano J R, Arnau F J, Dolz V and Piqueras P (2009a), 'Methodology for characterisation and simulation of turbocharged diesel engines combustion during transient operation. Part 1: Data acquisition and post-processing', *Applied Thermal Engineering*, 29, 142–149.

Serrano J R, Climent H, Guardiola C and Piqueras P (2009b), 'Methodology for characterisation and simulation of turbocharged diesel engines combustion during transient operation. Part 2: Phenomenological combustion simulation', *Applied Thermal Engineering*, 29, 150–158.

Shamsi S S (1980), 'Development of a real-time digital computer simulation of a turbocharged diesel engine', SAE paper 800521.

Shayler P J, Allen A J and Roberts A L (2005), 'Running real-time engine model simulation with hardware-in-the-loop for diesel engine development', SAE paper 2005-01-0056.

Shigemori M (1988), 'Overview on electronic controls for vehicular diesel engines', *Proceedings of International Congress on Transportation Electronics – Convergence '88*, Dearborn, Michigan, pp. 295–303.

Shirawaka T, Miura M, Itoyama H, Aiyoshizawa E and Kimura S (2001), 'Study of model-based cooperative control of EGR and VGT for a low-temperature, premixed combustion diesel engine', SAE paper 2001-01-2006.

Šika Z, Valášek M, Florián M, Macek J and Polášek M (2008), 'Multilevel predictive models of IC engine for model predictive control implementation', SAE paper 2008-01-0209.

Silvestri J, Morel T, Lang O, Schernus C and Rauscher M (2000), 'Advanced engine/drivetrain/vehicle modeling techniques', *The 9th Aachener Kolloquium Fahrzeug- und Motorentechnik*.

Smith P F, Prabhu S M and Friedman J (2007), 'Best practices for establishing a model-based design culture', SAE paper 2007-01-0777.

Sokolov A and Perkovic A (2000), 'Entropy generation in flow restrictions: modeling and identification for engine control', SAE paper 2000-01-0937.

Stefanopoulou A G and Smith R (2000), 'Maneuverability and smoke emission constraints in marine diesel propulsion', *Control Engineering Practice*, 8, 1023–1031.

Stefanopoulou A G, Kolmanovsky I and Freudenberg J S (2000), 'Control of variable geometry turbocharged diesel engines for reduced emissions', *IEEE Transactions on Control Systems Technology*, 8(4), 733–745.

Steiber J, Surampudi B, Treichel B and Kluger M (2005), 'Vehicle HIL, the near term solution for optimizing engine and transmission development', SAE paper 2005-01-1050.

Stobart R K, May A, Challen B J and Morel T (1998), 'Modeling for diesel engine control: the CPower environment', SAE paper 980794.

Stobart R K, Challen B J and Bowyer R (2001), 'Electronic controls – breeding new engines', SAE paper 2001-01-0255.

Strandh P, Bengtsson J, Johansson R, Tunestål P and Johansson B (2005), 'Variable valve actuation for timing control of a homogeneous charge compression ignition engine', SAE paper 2005-01-0147.

Strom K J and Hagglund T (1995), *PID Controllers: Theory, Design and Tuning*, 2nd edn, ISA – The Instrumentation, Systems, and Automation Society, Research Triangle Park, NC.

Strom K J and Hagglund T (2005), *Advanced PID Control*, ISA – The Instrumentation, Systems, and Automation Society, Research Triangle Park, NC.

Subramaniam M N, Tomazic D, Tatur M and Laermann M (2008), 'An artificial neural network-based approach for virtual NO_x sensing', SAE paper 2008-01-0753.

Taraza D, Henein N A, Ceausu R and Bryzik W (2007), 'Engine friction model for transient operation of turbocharged, common rail diesel engines', SAE paper 2007-01-1460.

Theotokatos G and Kyrtatos N P (2001), 'Diesel engine transient operation with turbocharger compressor surging', SAE paper 2001-01-1241.

Tian S, Li J, Yang F and Ouyang M (2008), 'A control oriented simplified transient torque model of turbocharged diesel engines', SAE paper 2008-01-1708.

Tsai S-C and Goyal M R (1986), 'Dynamic turbocharged diesel engine model for control analysis and design', SAE paper 860455.

Tsuchiya T, Morishita S, Enomoto T, Sasaki H and Yoshimura M (2003), 'Revolution control of generator diesel engine by neural network controller', SAE paper 2003-01-0365.

Tuken T, Fullmer R R and Gerpen J V (1990), 'Modeling, identification, and torque control of a diesel engine for transient test cycles', SAE paper 900235.

Tunestal P and Johansson B (2007), 'HCCI control', in Zhao H (ed.), *HCCI and CAI Engines for the Automotive Industry*, Woodhead Publishing Limited, Cambridge.

Turin R, Chang M-F, Zhang R and Dagci O H (2007), 'A three-pillar framework for model-based engine control system development', SAE paper 2007-01-1624.

Turin R, Zhang R and Chang M-F (2008), 'Systematic model-based engine control design', SAE paper 2008-01-0994.

Ueno M, Izumi T, Watanabe Y and Baba H (2008), 'Exhaust gas pressure sensor', SAE paper 2008-01-0907.

Ulrich O, Wlodarczyk R and Wlodarczyk M T (2001), 'High-accuracy low-cost cylinder pressure sensor for advanced engine controls', SAE paper 2001-01-0991.

Utkin V I, Chang H-C, Kolmanovsky I and Cook J A (2000), 'Sliding model control for

variable geometry turbocharged diesel engines', *Proceedings of the American Control Conference*, Chicago, IL, June, pp. 584–588.
Vitale G, Siebenbrunner P, Hülser H, Bachler J and Pfahl U (2007), 'OBD algorithms: model-based development and calibration', SAE paper 2007-01-4222.
Wahlström J, Eriksson L and Nielsen L (2008), 'Controller tuning based on transient selection and optimization for a diesel engine with EGR and VGT', SAE paper 2008-01-0985.
Wang J (2007), 'Hybrid robust control for engines running low temperature combustion and conventional diesel combustion modes', SAE paper 2007-01-0770.
Wang Y and Chu F (2005), 'Application of non-linear observers to on-line estimation of indicated torque in automotive engines', *Proc. IMechE, Part D: Journal of Automobile Engineering*, 219, 65–75.
Wang J, Roecker R C and Roberts C E (2005), 'Virtual cylinder pressure sensor (VCPS) with individual variable-oriented independent estimators', SAE paper 2005-01-0059.
Wang Y M, Zhang F J, Wang Y T and Liu B L (2009), 'Hardware-in-the-loop research on the electronic control unit for diesel engines', *Journal of Beijing Institute of Technology* (English edition), 18(4), 416–421.
Wanpal A, Babu M G, Kankariya N, Mundhra K, Sundaresan S A and Deshpande A S (2006), 'ECU testing and verification using hardware-in the-loop', SAE paper 2006-01-1444.
Warey A and Hall M J (2005), 'Performance characteristics of a new on-board engine exhaust particulate matter sensor', SAE paper 2005-01-3792.
Watson N (1981), 'Transient performance simulation and analysis of turbocharged diesel engines', SAE paper 810338.
Watson N (1984), 'Dynamic turbocharged diesel engine simulator for electronic control system development', *Transactions of the ASME*, *Journal of Dynamic Systems, Measurement, and Control*, 106, 27–45.
Watson N and Banisoleiman K (1988), 'A variable-geometry turbocharger control system for high output diesel engines', SAE paper 880118.
Watson N and Marzouk (1977), 'A non-linear digital simulation of turbocharged diesel engines under transient conditions', SAE paper 770123.
Watson N, Pilley A D and Marzouk M (1980), 'A combustion correlation for diesel engine simulation', SAE paper 800029.
Weber F and Guzzella L (2000), 'Control oriented modeling of a pressure wave supercharger', SAE paper 2000-01-0567.
Weisman S M (1987), 'A control system design for heavy duty diesel engines', ASME paper 87-ICE-33.
Westbrook M H and Turner J D (1994), *Automotive Sensors*, Institute of Physics Publishing, Bristol and Philadelphia, PA.
Westin F and Angstrom H E (2002), 'A method of investigating the on-engine turbine efficiency combining experiments and modelling', IMechE paper C602/029/2002.
Wijetunge R S, Brace C J, Hawley J G, Vaughan N D, Horrocks R W and Bird G L (1999), 'Dynamic behaviour of a high speed direct injection diesel engine', SAE paper 1999-01-0829.
Wijetunge R S, Hawley J G and Vaughan N D (2004), 'Application of alternative EGR and VGT strategies to a diesel engine', SAE paper 2004-01-0899.
Winterbone D E and Jai-In S (1988), 'Control studies of an automotive turbocharged diesel engine with variable geometry turbine', SAE paper 880485.
Winterbone D E and Jai-In S (1991), 'The application of modern control theory to a

turbocharged diesel engine powerplant', *Proc. IMechE, Part I: Journal of Systems and Control Engineering*, 205, 69–83.

Winterbone D E, Thiruarooran C and Wellstead P E (1977), 'A wholly dynamic model of a turbocharged diesel engine for transient function evaluation', SAE paper 770124.

Woermann R J, Theuerkauf H J and Heinrich A (1999), 'A real-time model of a common rail diesel engine', SAE paper 1999-01-0862.

Wurzenberger J C, Heinzle R, Schuemie A and Katrasnik T (2009), 'Crank-angle resolved real-time engine simulation – integrated simulation tool chain from office to testbed', SAE paper 2009-01-0589.

Xiros N (2002), *Robust Control of Diesel Ship Propulsion*, Springer-Verlag London Limited, London.

Yan W, Scacchioli A and Rizzoni G (2007), 'Model based fault diagnosis for engine under speed control', SAE paper 2007-01-0775.

Yokomura H, Kouketsu S, Kotooka S and Akao Y (2004), 'Transient EGR control for a turbocharged heavy duty diesel engine', SAE paper 2004-01-0120.

Zhao F and Asmus T (2003), 'HCCI control and operating range extension', in Zhao F, Asmus T W, Assanis D N, Dec J E, Eng J A and Najt P M (eds), *Homogeneous Charge Compression Ignition (HCCI) Engines – Key Research and Development Issues*, SAE International, Warrendale, PA.

15
Diesel engine system specification design and subsystem interaction

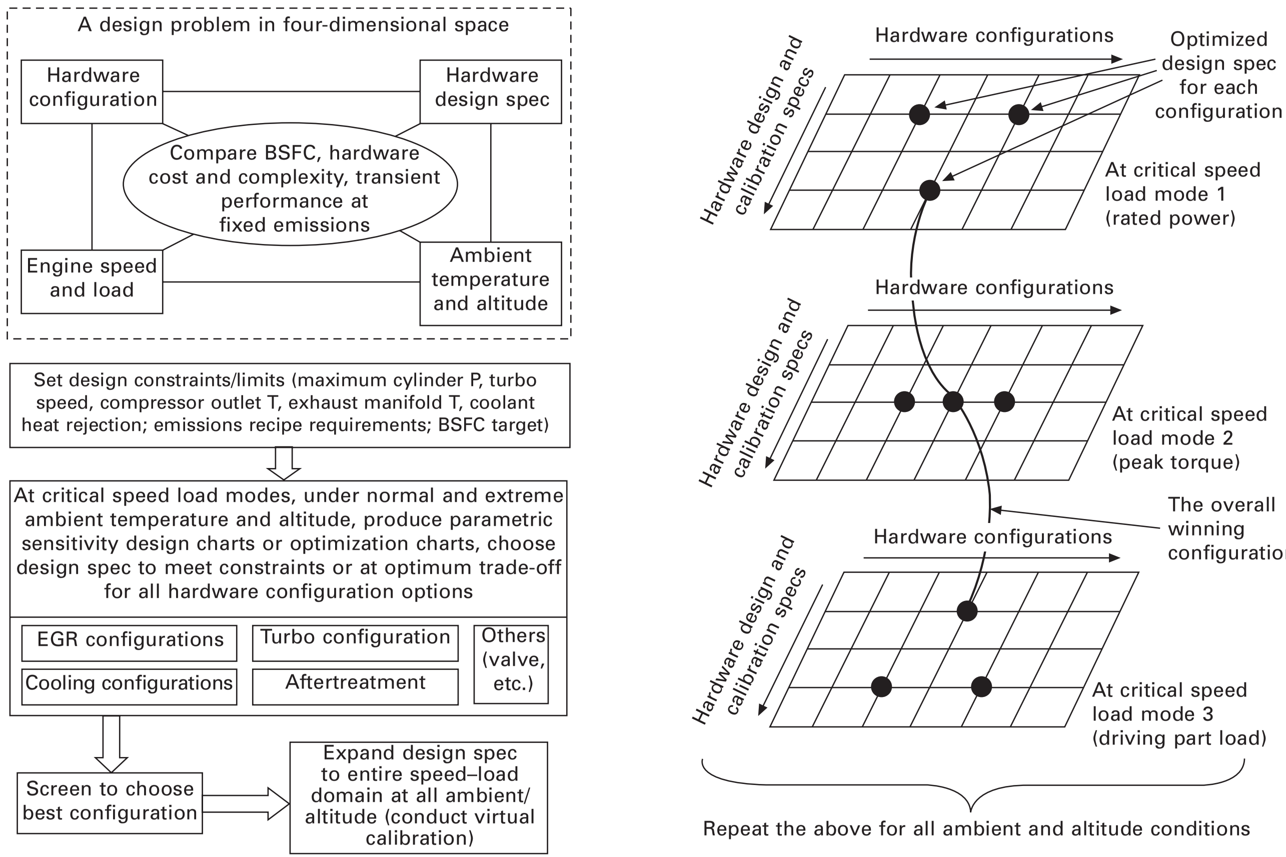
Abstract: This chapter illustrates the process of engine system design analysis in a four-dimensional design space that consists of hardware configuration, hardware design specification, engine speed and load, and ambient temperature and altitude. It goes on by summarizing the design guidelines for critical modes (e.g., peak torque, rated power) including considerations relating to system design constraints. Many simulation examples of subsystem interaction and optimization are presented.

Key words: system design specification, four-dimensional design space, subsystem interaction, optimization, sensitivity analysis.

15.1 The process of system design analysis

As mentioned in Chapter 1, the main objective of diesel engine system design is to produce system performance design specifications and the justification of the optimum design point by using advanced simulation tools to construct parametric sensitivity or optimization charts. The need for using advanced software is obvious for three reasons: (1) primitive hand calculations of engine air system performance are not accurate enough for precise system design; (2) the required nonlinear constrained optimization of DoE emulators for generating parametric design charts can be computed only with specialized software; and (3) the high demand of complex transient analysis in hardware selection cannot be conducted without using simulation software.

Figure 15.1 illustrates the complex process of engine system analysis in a four-dimensional design space as follows: (1) hardware configuration (e.g., single-stage or two-stage turbocharger); (2) hardware design specification (both steady-state and transient; both normal operation and aftertreatment regeneration; e.g., turbine area and EGR valve opening); (3) engine speed and load; and (4) ambient temperature and altitude. The process starts with setting the design constraints and identifying all possible configuration options to reach the target emissions recipe or its equivalent air system requirements (air–fuel ratio, EGR rate and intake manifold gas temperature, detailed in Chapter 7). The DoE method is normally used to quickly rule out some options and then to optimize the design specification of each subsystem for minimum BSFC subject to the design constraints. The subsystems include turbocharger, valvetrain, ports, manifolds, EGR circuit, exhaust restriction,



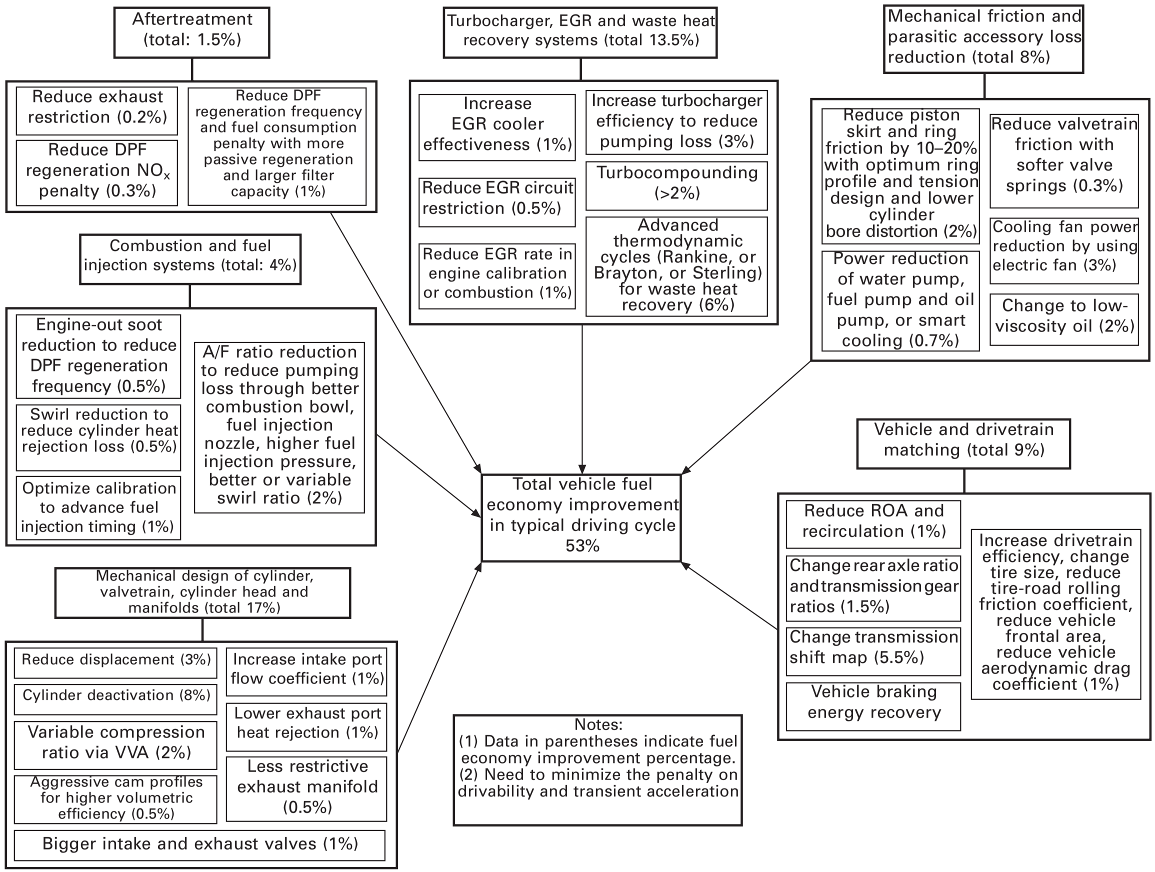
15.1 Design space and design process of engine system specification.

and cooling. Sensitivity design charts are produced at this stage. At each speed and load mode, the DoE factors may include some or all of the following: exhaust restriction, charge air cooler size, EGR cooler size, compressor or turbine efficiency (if turbocharger maps are not used) or efficiency multiplier (if turbocharger maps are used), turbine effective area, turbine wastegate opening, EGR circuit flow restriction coefficient, certain design factors affecting engine volumetric efficiency (e.g., valve size, valve timing, port flow discharge coefficient), and start-of-combustion or start-of-injection timing. The analysis output includes critical instantaneous parameters such as the in-cylinder details and the gas flow pulsation in the pipes. The output also includes all the cycle-average parameters describing the entire engine system performance such as air–fuel ratio, EGR rate, oxygen mass fraction, BSFC, peak cylinder pressure, peak cylinder bulk gas temperature, pumping loss, engine delta P, volumetric efficiency, manifold pressures and gas temperatures, heat rejections, cooler effectiveness and flow restriction, turbocharger flow rate, pressure ratio and temperatures, and turbocharger efficiency (if turbocharger maps are used).

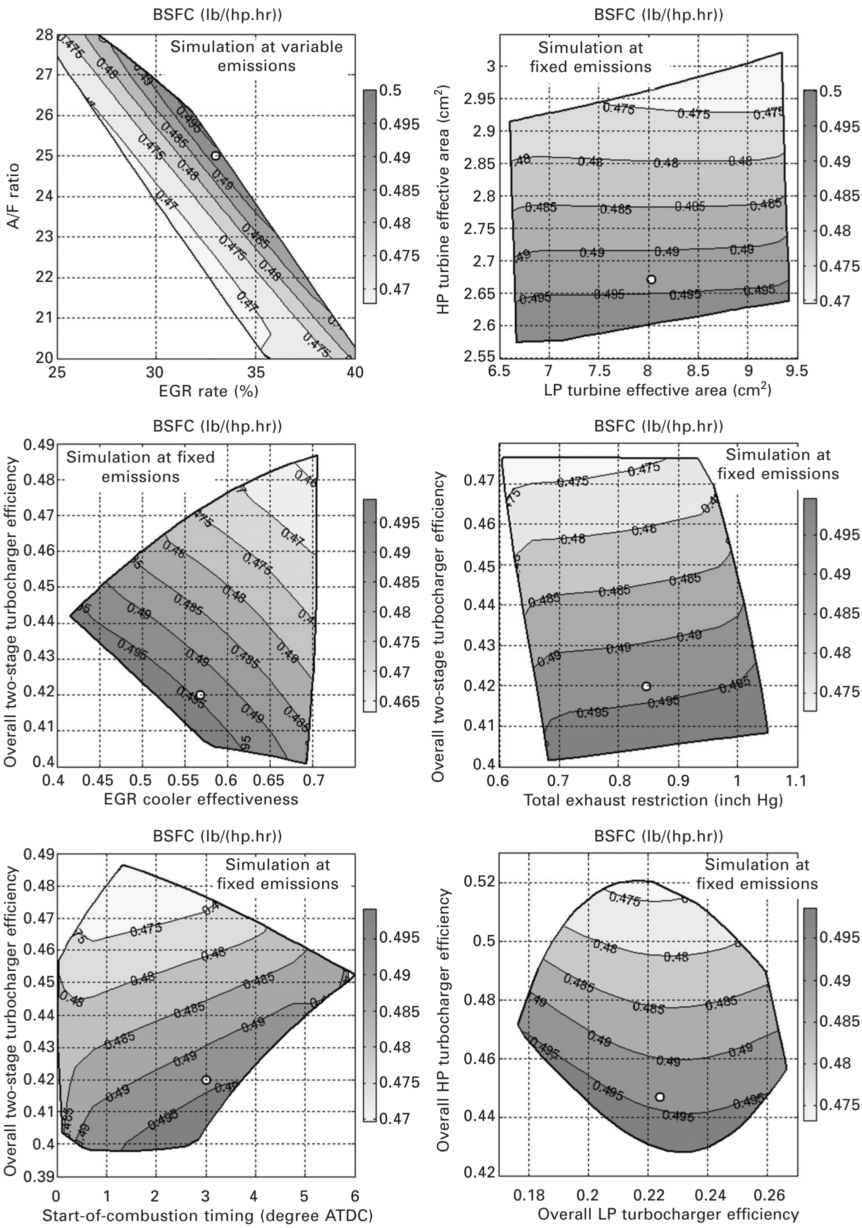
In the concept design phase, technology evaluation is important in order to choose the right path for the engine product development. Diesel engine technology reviews and evaluations are provided by Merrion (1994), Regueiro and Chen (1997), Chen *et al.* (1997), Khair (1997), Hikosaka (1997), Schindler (1997), Hountalas (2000), and Conley and Taylor (2002). Diesel engine research program reviews are summarized by Rickeard *et al.* (1996), Eberhardt (1999), Singh *et al.* (2000), and Thompson *et al.* (2004). System-level design and optimization techniques are described by Page and Edgar (1998), Burtt and James (2004), Berard *et al.* (2000), and Fussey *et al.* (2001).

15.2 Roadmap of fuel economy improvement

In the efforts to improve fuel economy, the role of engine system design is to lead the way in defining a reasonable roadmap and cascade it to each subsystem to implement design changes. The roadmap, as shown in Fig. 15.2 (for improving a non-optimized high pumping loss engine), usually consists of six areas: (1) aftertreatment; (2) combustion and fuel system; (3) mechanical design of cylinder, valvetrain, cylinder head, and manifolds; (4) turbocharger, EGR, and waste heat recovery systems; (5) mechanical friction and parasitic losses of accessories; and (6) vehicle drivetrain matching. Improvements can be made either through incremental change or by adopting new technologies. Figure 15.3 shows a simulation analysis on the methods to improve BSFC by improving the design or calibration parameters in the air system, cooling system, and the combustion system.



15.2 Example of fuel economy improvement roadmap.



15.3 Summary on the impact on BSFC at typical part-load operation (2000 rpm 30% load).

15.3 Critical mode design at various ambient conditions

15.3.1 System design constraints

In engine system specification design there are two categories of limits, usually occurring at full load: the hardware design limit and a lower limit used for the calibration test setting. The necessity for the two limits is based on the statistical distribution of hardware variations. The 'nominal' engine design or calibration set point (e.g., intake manifold pressure or exhaust manifold gas temperature) under the standard normal ambient condition should be lower than the 'calibration limit' by a certain margin. The full-load conditions of the hardware in the system specification need to be designed below the 'design limit' in order to cover the worst ambient conditions in emissions certification and real-world driving. Different design limits may occur at different ambient conditions or engine speeds, for example (also illustrated in Section 3.4.5):

- In order to be compliant with the US EPA NTE emission requirements, the EGR cooler and the charge air cooler need to be sized for sea-level altitude and hot ambient (e.g., 38°C or 100°F) to reach a target intake manifold gas temperature.
- The compressor inter-stage cooler in a two-stage turbocharger needs to be sized with a sufficient cooling capacity and an appropriate cooler sink temperature (i.e., the cooling medium temperature) to ensure proper operation of the high-pressure-stage compressor under high ambient temperatures at sea level or high altitude conditions. Specifically, the outlet air temperature of the compressor needs to be below the material limits of the compressor wheel, the housing, and the charge air cooler inlet tube.
- The engine coolant heat rejection and the associated EGR strategy need to be designed to ensure that at high altitude and hot ambient (e.g., 38°C) the engine outlet or radiator inlet coolant temperature does not exceed the durability limit. High altitude refers to, for example, 1676 meters or 5500 feet – the US EPA NTE limit, or 10,000 feet in practical real-world applications.
- The turbocharger control and the EGR strategy need to be tailored for high altitude and hot ambient conditions to ensure that soot emission is still below the smoke limit with a lower air–fuel ratio. Additionally, the exhaust manifold gas temperature should be within the limit imposed by the durability of the cylinder head and the turbine.
- The compressor needs to be sized large enough to ensure the turbocharger speed is still below the maximum limit at high altitude without fueling derating.

- The maximum cylinder gas pressure and temperature as well as the peak heat flux need to be designed below the combined limits of mechanical and thermal stresses for the power cylinder components.
- The turbocharger and EGR controls need to be designed to ensure that the peak cylinder pressure along the full-load lug curve does not exceed the structural limit in a cold climate. Moreover, any compressor surge at cold ambient should be avoided. The full-load lug curve refers to US SET modes A100, B100, C100, peak torque and rated power.
- During the firing and engine braking operations in both steady state and transient, the exhaust manifold pressure should be below the structural limit. The engine delta P should be kept below the point where the exhaust valve floats off the valve seat.
- For high EGR engines in extremely humid climates, the intake manifold gas temperature needs to be designed not to exceed the durability limits of condensation control and corrosion resistance.
- All the hardware design constraints need to be satisfied during the maximum engine braking operation.

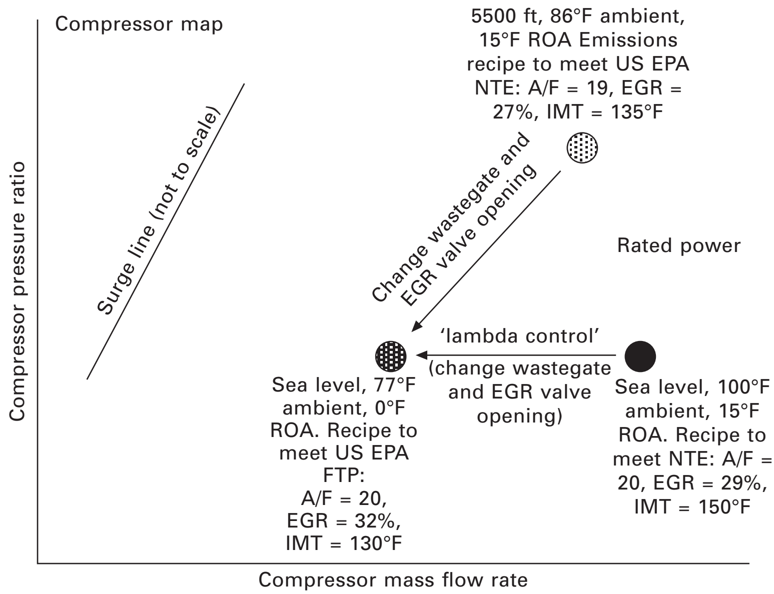
15.3.2 Design from extreme ambient to standard lab condition

The ambient conditions used in hardware sizing usually include the following:

1. Sea-level altitude and normal ambient temperature (e.g., 25°C or 77°F).
2. The US EPA's heavy-duty emissions NTE limits (sea-level altitude, from cold ambient to 38°C or 100°F hot ambient).
3. The 1676 meters or 5500 feet altitude, from cold ambient to 30°C or 86°F, with a humidity range.
4. The real-world driving environments (e.g., 5500 feet on a hot day, a very high altitude level such as 8000–10,000 feet on a hot day, and a cold ambient condition at sea level, for example –18°C or 0°F).

It is also necessary to consider the following factors: hot air recirculation, turning on air-conditioning, and the rise-over-ambient elevated compressor inlet air temperature due to in-vehicle underhood heating.

Figure 15.4 shows an example of the system design sequence from the extreme ambient and in-vehicle conditions to the standard lab engine condition. The EGR rate required at the hot ambient or high altitude is affected by the emissions requirement, intake air temperature, air–fuel ratio, and oxygen concentration. The difference in the required EGR rate between the sea level and the extreme ambient condition depends on emissions regulation level, particular engine design, and the fuel system capability.



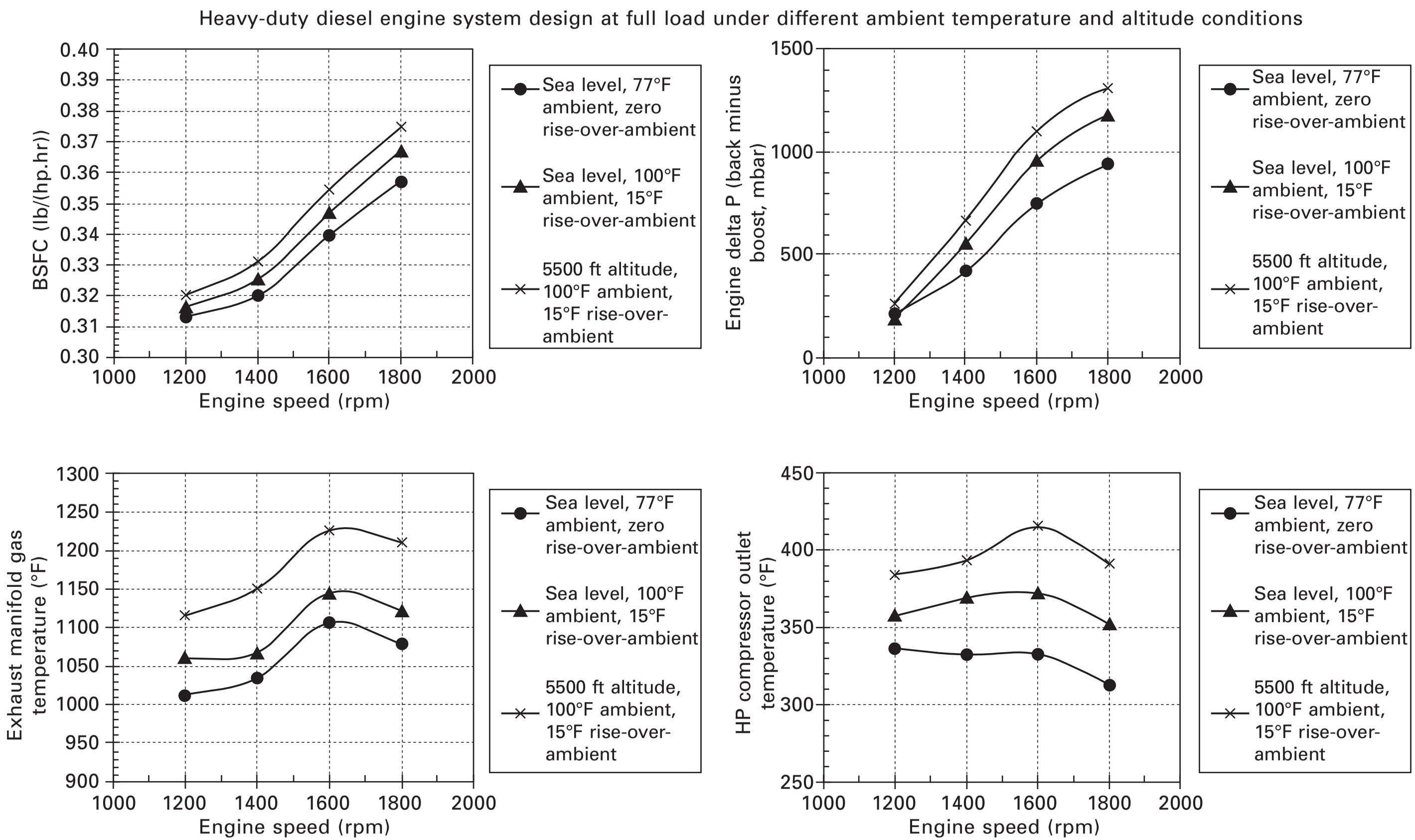
15.4 Illustration of engine system design from extreme ambient to standard lab condition.

The achievable EGR rate and air–fuel ratio at the same altitude but different ambient temperatures are determined by the engine control strategy of the air system. For example, the VGT vane opening at sea level hot ambient can be controlled by exhaust manifold pressure or intake manifold pressure. In the 'lambda control', air–fuel ratio is maintained at a constant level at a given altitude and is controlled by regulating the EGR valve opening.

Figure 15.5 shows typical behavior of engine performance at different ambient conditions at the full-load lug curve. Note that the exhaust manifold gas temperature and compressor outlet air temperature are higher at the high altitude and hot ambient condition than the sea level normal ambient condition. The engine delta P can also be higher if the EGR is reduced and the turbine wastegate is closed in order to increase the air–fuel ratio at high altitude. The higher BSFC at the extreme ambient conditions is caused mainly by a reduction in the indicated power in the compression and expansion strokes due to the decrease in air–fuel ratio and secondarily affected by the increase in engine delta P or pumping loss.

15.3.3 Design for varying exhaust flow rate

In addition to the variations in ambient pressure and temperature conditions, hardware matching is also complicated by the large variations in the EGR flow and the exhaust restriction in the aftertreatment system. For instance,



15.5 Engine system design at standard lab condition and extreme ambient conditions.

EGR rate control may vary from zero in very cold conditions or during fast transients to very high at the normal ambient. Moreover, large fluctuations in the exhaust restriction may occur due to soot loading changes in the DPF. The compressor size needs to be large enough to cover all these variations.

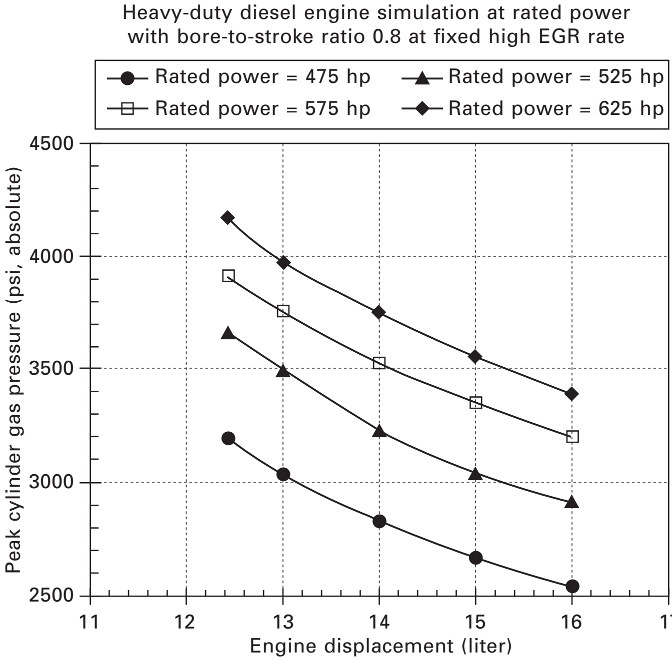
15.3.4 Critical mode design – rated power

When designing the maximum capacity of the hardware and comparing different configuration options in heavy-duty applications, critical speed and load modes are selected such as rated power, peak torque, and typical part-load driving conditions identified by the vehicle matching analysis. In light-duty applications, the rated power and peak torque conditions are rarely encountered in real-world driving, and usually very little EGR is applied under the full-load conditions for light-duty emissions certification. In light- or medium-duty applications, two practical and frequently used operating modes are: high speed and high load; and low speed and high load. They both have high EGR demand and need to be added to the critical mode analysis to size the hardware.

It should be emphasized that an important design target is to minimize the engine pumping loss and fuel consumption for both steady-state and transient operations in the proper speed–load region. The minimum pumping loss is realized by a low engine delta P design in the EGR and turbocharger systems, and an equally important high volumetric efficiency design in the valvetrain, ports, and manifolds. For the durability design targets, over-speed and over-fueling may be simulated to determine the engine's maximum potential capability and determine the safety margins.

Successful design of the rated power condition is important for heavy-duty engines. All the design constraints mentioned earlier apply to rated power. Moreover, at high speed, the reciprocating inertia force is another constraint for structure strength. The commonly used methods of designing a series of engines with different power ratings or applications include the following: (1) raising the rated speed; (2) using turbocharging to achieve different levels of air density and the corresponding power ratings without changing the engine cylinder bore and stroke; (3) increasing the cylinder bore diameter or stroke to increase the engine displacement and the power rating; and (4) increasing or decreasing the number of cylinders to achieve different power ratings. Other methods of increasing the power rating include enhancing volumetric efficiency by valvetrain or port design, reducing mechanical friction, and raising the peak cylinder pressure limit. Figure 15.6 shows the impact of engine displacement and power rating on peak cylinder pressure.

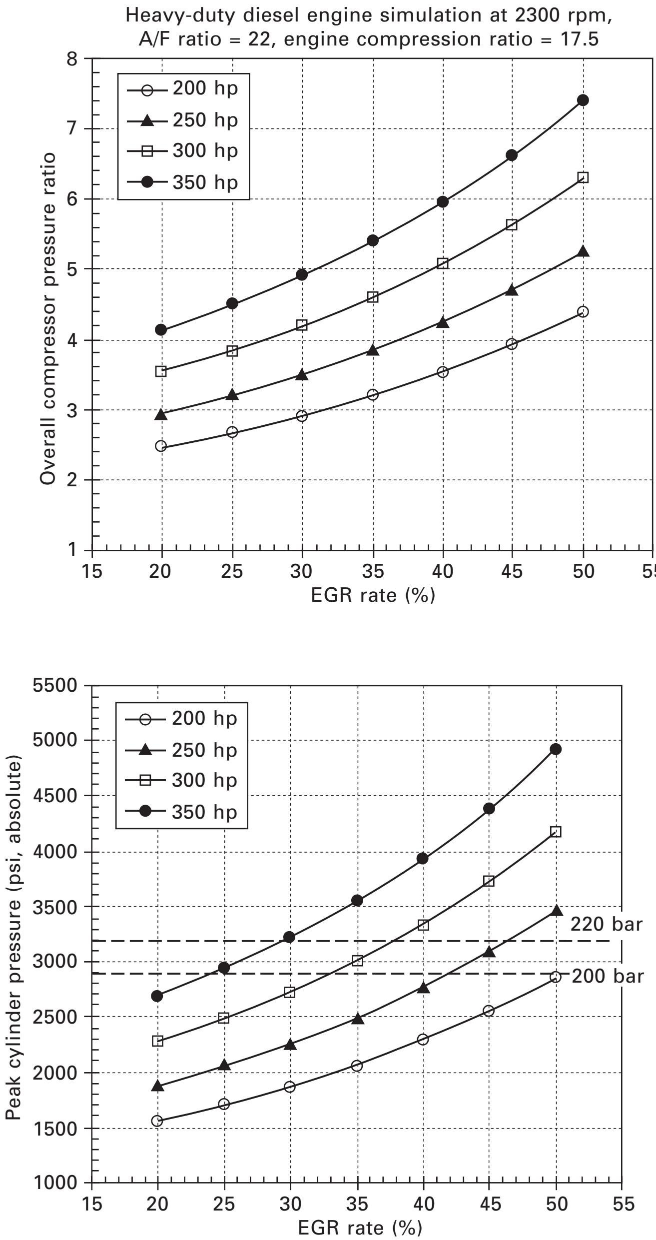
At a given engine speed, there are trade-offs between the air–fuel ratio, EGR rate, and power rating under a fixed constraint of peak cylinder pressure. Figure 15.7 shows these trade-offs.



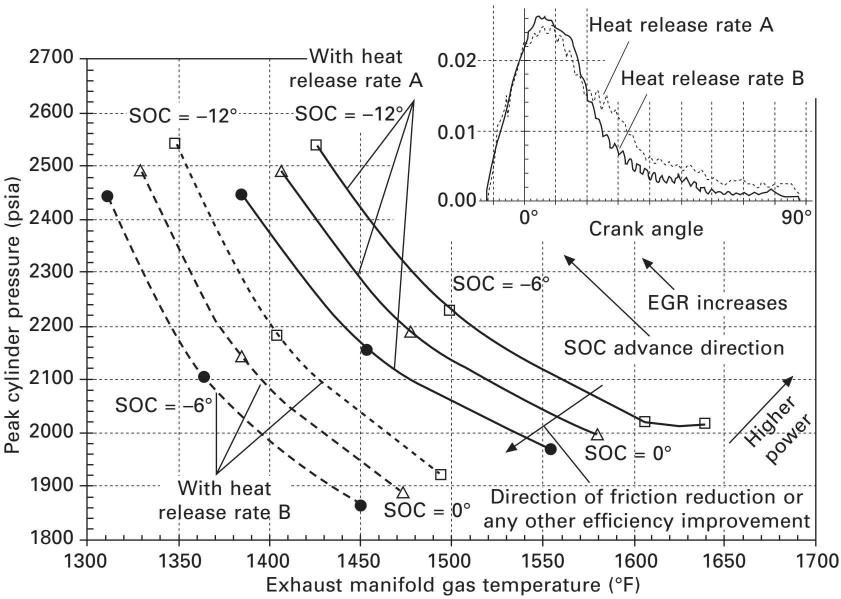
15.6 Effect of engine displacement and power rating on peak cylinder pressure.

There is also another fundamental trade-off at a given speed/load between the peak cylinder pressure and the exhaust manifold gas temperature, both being design constraints. The trade-off is affected by the shape of the heat release rate, fuel injection timing, and engine friction. The simulation data in Fig. 15.8 show the measures needed to bring the cylinder pressure and exhaust temperature under control at rated power (i.e., moving the curves to the lower left direction). Reducing the engine compression ratio may reduce the peak cylinder pressure at the expense of lower indicated thermodynamic efficiency and unaided cold start capability.

The rated power condition is used to define the compressor and turbine maximum flow ranges, the cooler size, the maximum heat rejection, the intake restriction, and the exhaust restriction. In high-EGR engines, usually the engine delta P at rated power has to be very high due to the small turbine area sized for the required EGR rate and air–fuel ratio at peak torque. As a consequence, the EGR valve has to be partially closed in order to prevent excessive EGR. Therefore, the rated power condition should not be used to define the minimum flow restriction of the EGR circuit.



15.7 Estimate on compressor pressure ratio and peak cylinder pressure at different EGR rate and power rating.



15.8 Engine fundamental trade-offs at a fixed speed and power.

15.3.5 Critical mode design – peak torque

The peak torque condition has the maximum mean effective pressure, and is prone to compressor surge (especially at high altitude) due to the low engine air flow rate and the relatively high boost pressure. It is also usually the most difficult mode to drive EGR due to insufficient engine delta P. The minimum flow restriction of the EGR circuit (with EGR valve fully open, with or without a check valve) and the minimum required turbine area are often determined based on the EGR-driving requirement at peak torque, although the lower speeds at full load also need to be checked to ensure the minimum turbine area is acceptable. In fact, strictly speaking, the minimum required turbine area should be determined at the most demanding operating mode in terms of a combination of the requirements of EGR driving, air–fuel ratio and engine torque, rather than simply 'peak torque'. In some vehicle applications, the peak torque or a medium speed between the peak torque and rated power may produce the maximum engine outlet coolant temperature. In that speed–load region, although the coolant heat rejection is lower than that at the rated power, the coolant flow rate is relatively low so that the engine coolant temperature may reach the highest value.

15.3.6 Critical mode design – part load and other modes

At part load, it is important to check pumping loss, fuel economy, the EGR driving capability with the selected turbocharger, and the capability to control the air or cooling system in order to raise the turbine outlet exhaust temperature for aftertreatment thermal management such as DPF regeneration.

15.4 Subsystem interaction and optimization

15.4.1 Types of sensitivity analysis

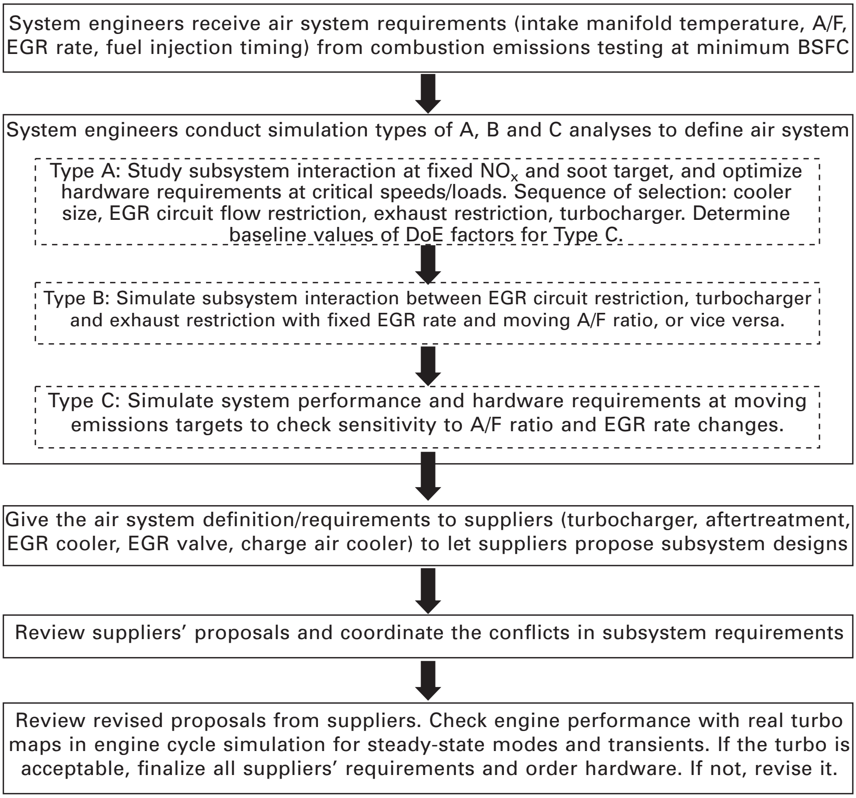
In the sensitivity analysis by parametric sweeping or optimization, there are generally three types of simulation as follows:

1. Type A: Simulation at fully fixed emissions constraints (i.e., NO_x and soot) to compare BSFC and hardware cost. The simulation results are presented in the domain of the characteristic parameter of one subsystem vs. the characteristic parameter of another subsystem. The values of factors and responses are presented as contour maps in the domain.
2. Type B: Simulation at partially fixed emissions constraints (e.g., fixed NO_x or soot; or fixed EGR rate or air–fuel ratio). The simulation results are presented in the subsystem parameter domain similar to type A.
3. Type C: Simulation at variable emissions or at variable air–fuel ratio and EGR rate. The simulated results are presented in the domain of 'air–fuel ratio vs. EGR rate' that represents the capability of the air system.

Types A and B can be used to conveniently study the interaction between the two subsystems plotted on the two axes of a map to facilitate the selection of a system design point from the map. Type C is particularly useful when the emissions recipe or design constraints are uncertain, and the required air or EGR flows are moving targets. In type A analysis, the emissions constraints can usually be approximated by one of the following methods: (1) a pair of fixed air–fuel ratio and EGR rate; (2) a pair of fixed peak cylinder gas temperature and air–fuel ratio; (3) a pair of fixed peak cylinder gas temperature and intake manifold oxygen mass fraction; or (4) a pair of fixed NO_x and soot predicted by emissions models. Figure 15.9 illustrates an analysis process to analyze subsystem interactions and optimize the system.

Figures 15.10–15.16 present simulation examples of type A analysis with different subsystems plotted on the horizontal and vertical axes in each figure to illustrate the subsystem interactions at approximately fixed emissions. These examples are produced by optimization with DoE emulators. The BSFC of each point in the data domain has been minimized within the factor range of the DoE subject to the optimization constraints. A hardware design point (e.g., the white dot on the maps) can be selected based on the best trade-offs between different subsystems supported by these sensitivity contours maps.

Figures 15.17 and 15.18 illustrate type B analysis. Figures 15.19–15.21 illustrate type C analysis showing the sensitivity of engine performance to

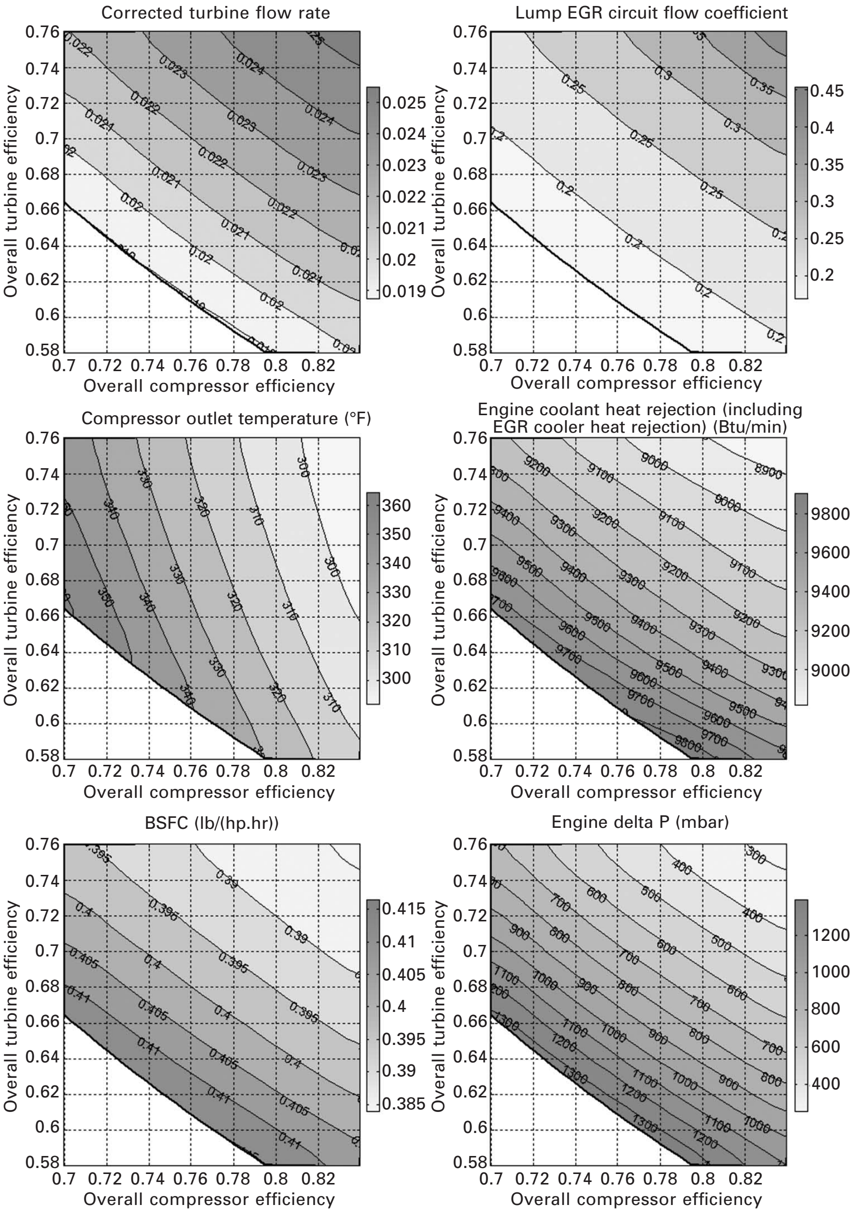


15.9 Process of engine system design to optimize subsystem interaction.

the change in emissions target. A system design point can be selected based on the best trade-offs between the tentative emissions target and the design constraints from these analysis maps.

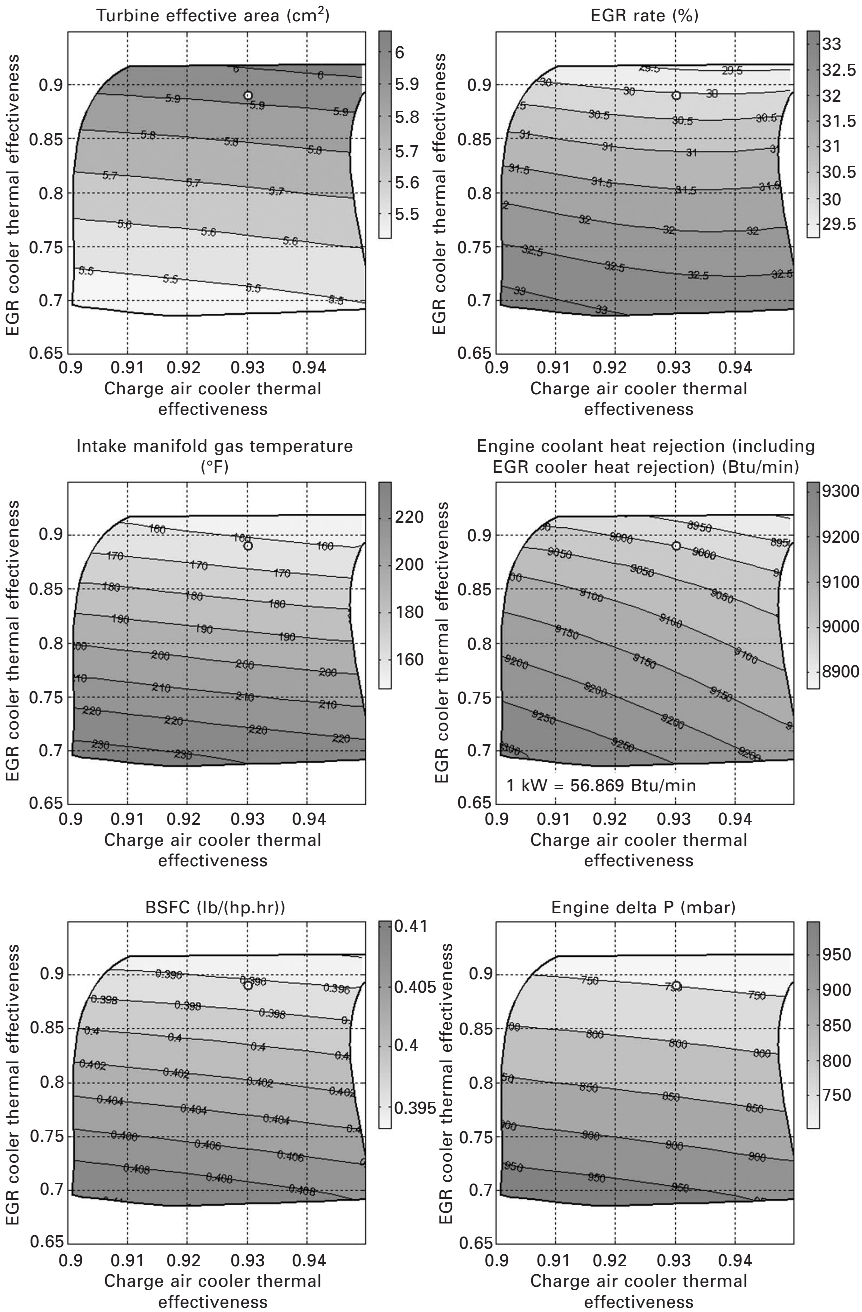
15.4.2 Subsystem interaction and optimization

In engine air system design, there are strong dependencies among the following four parameters: exhaust restriction, EGR circuit flow restriction, turbine effective area, and turbocharger efficiency. In hardware selection at each speed and load mode for a pair of target air–fuel ratio and EGR rate, when the EGR valve is set fully open to minimize the engine delta P, the required turbine area and turbocharger efficiency are unambiguously determined (Table 4.1). But the reality is that the turbocharger efficiency cannot reach all the desirable values computed as such at all speeds and loads. At some modes, if the actual turbocharger efficiency is too low, the air–fuel ratio will become

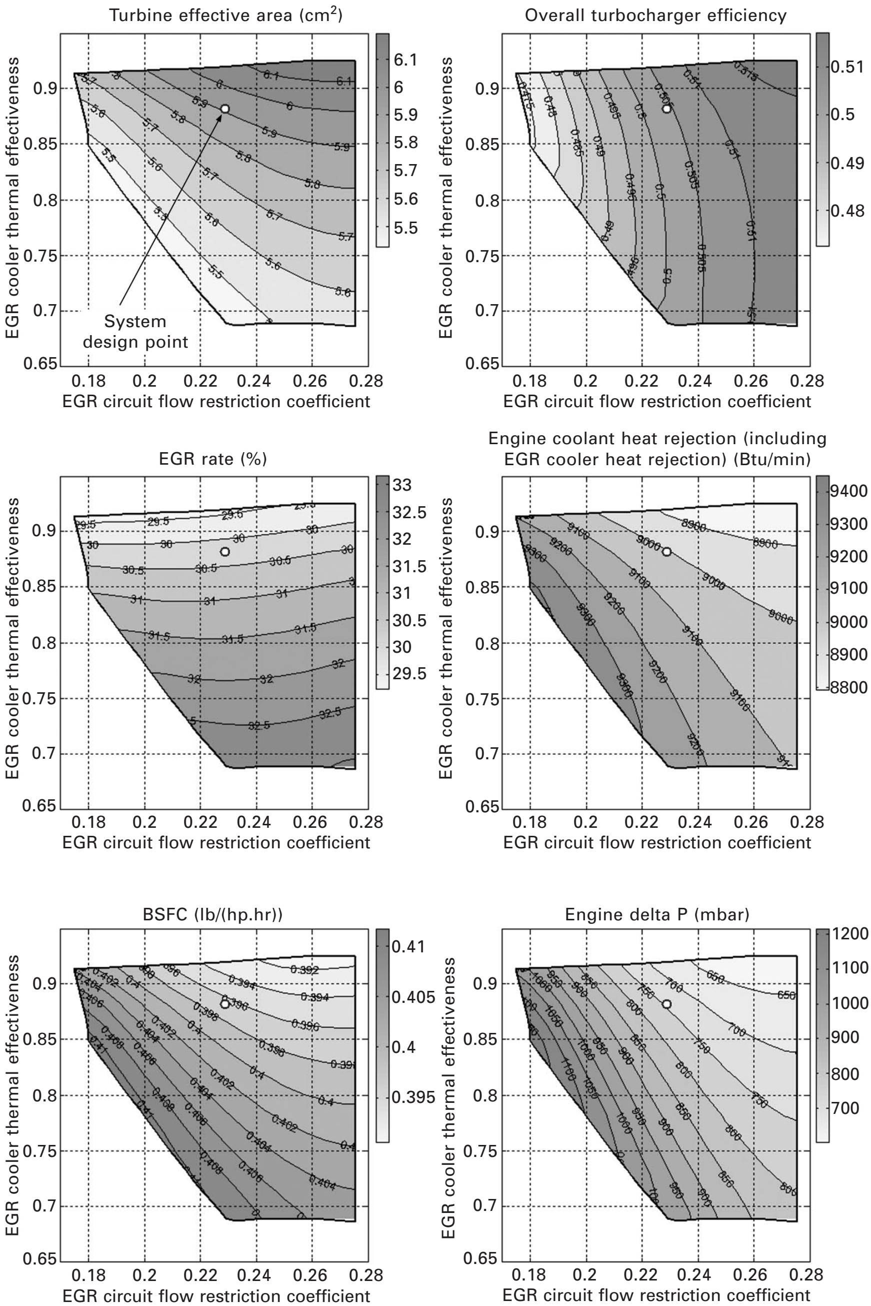


15.10 HD diesel subsystem interaction at rated power, type A1 analysis with fixed A/F ratio = 22 and EGR rate = 31%.

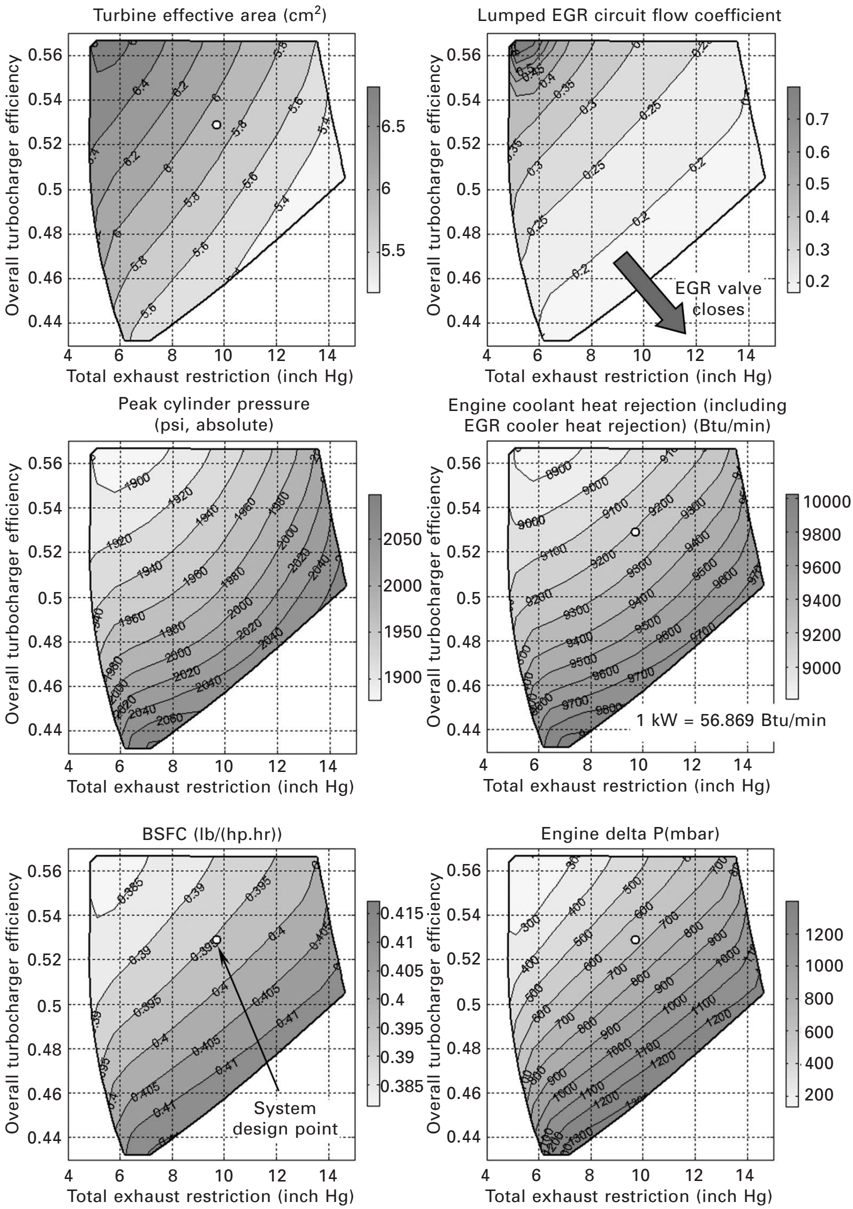
too low. In order to compensate for that, a smaller turbine area must be used and the EGR valve must be correspondingly partially closed to accommodate the increased engine delta P. On the other hand, if the actual turbocharger



15.11 HD diesel subsystem interaction at rated power, type A2 analysis with fixed A/F ratio = 22 and peak cylinder temperature = 2429°F.

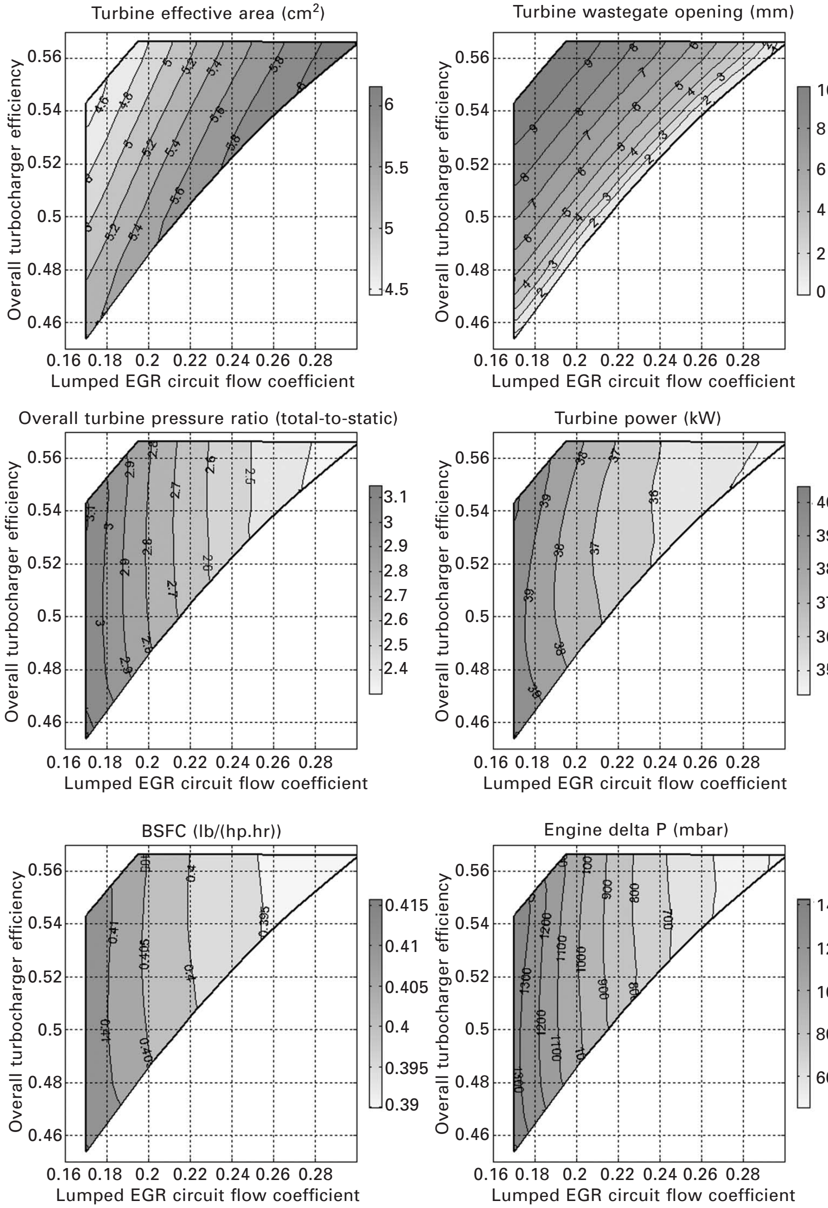


15.12 HD diesel subsystem interaction at rated power, type A3 analysis with fixed A/F ratio = 22 and peak cylinder temperature = 2429°F.



15.13 HD diesel subsystem interaction at rated power, type A4 analysis with fixed A/F ratio = 22 and peak cylinder temperature = 2439°F.

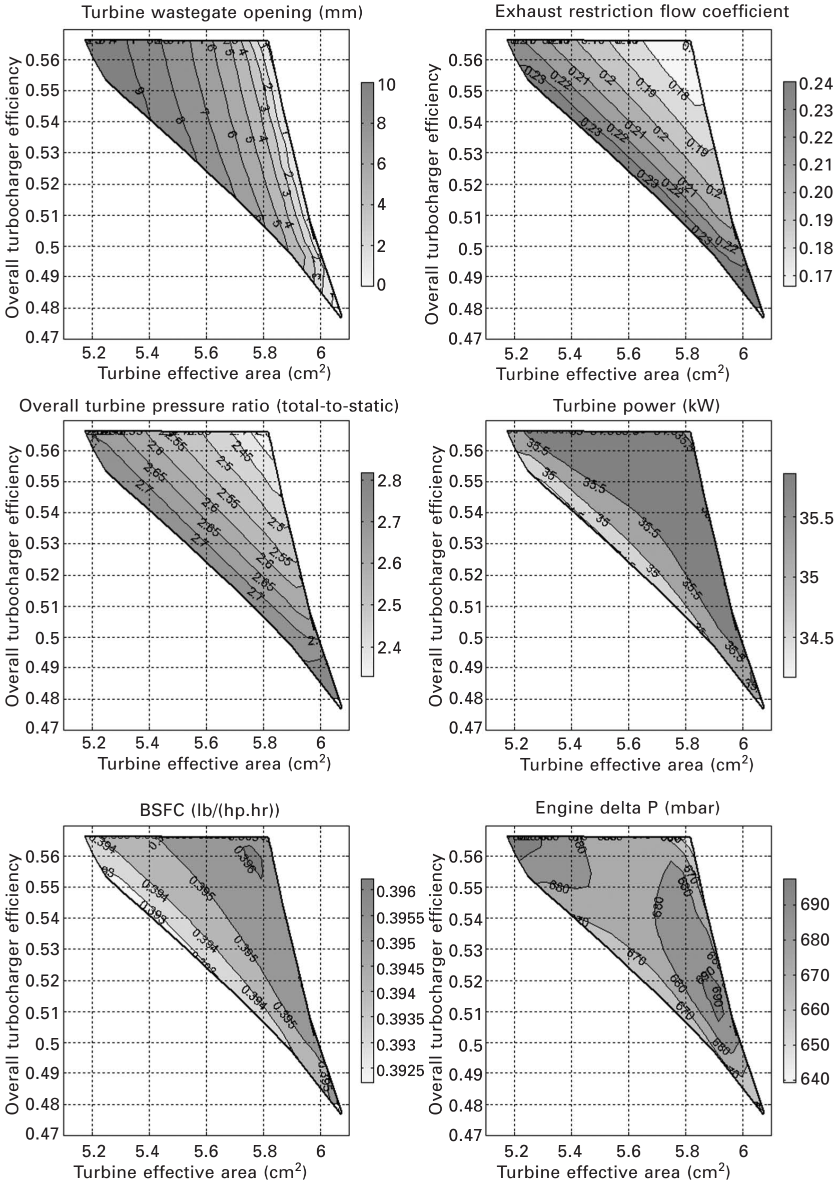
efficiency is too high, the air–fuel ratio will become too high. In order to reduce the air–fuel ratio to prevent over-boosting, a larger turbine area or wastegating must be used to slow down the compressor, and the engine delta



15.14 HD diesel subsystem interaction at rated power, type A5 analysis with fixed A/F ratio = 22 and peak cylinder temperature = 2439°F (exhaust restriction = 9.7 inch Hg, IMT = 150°F).

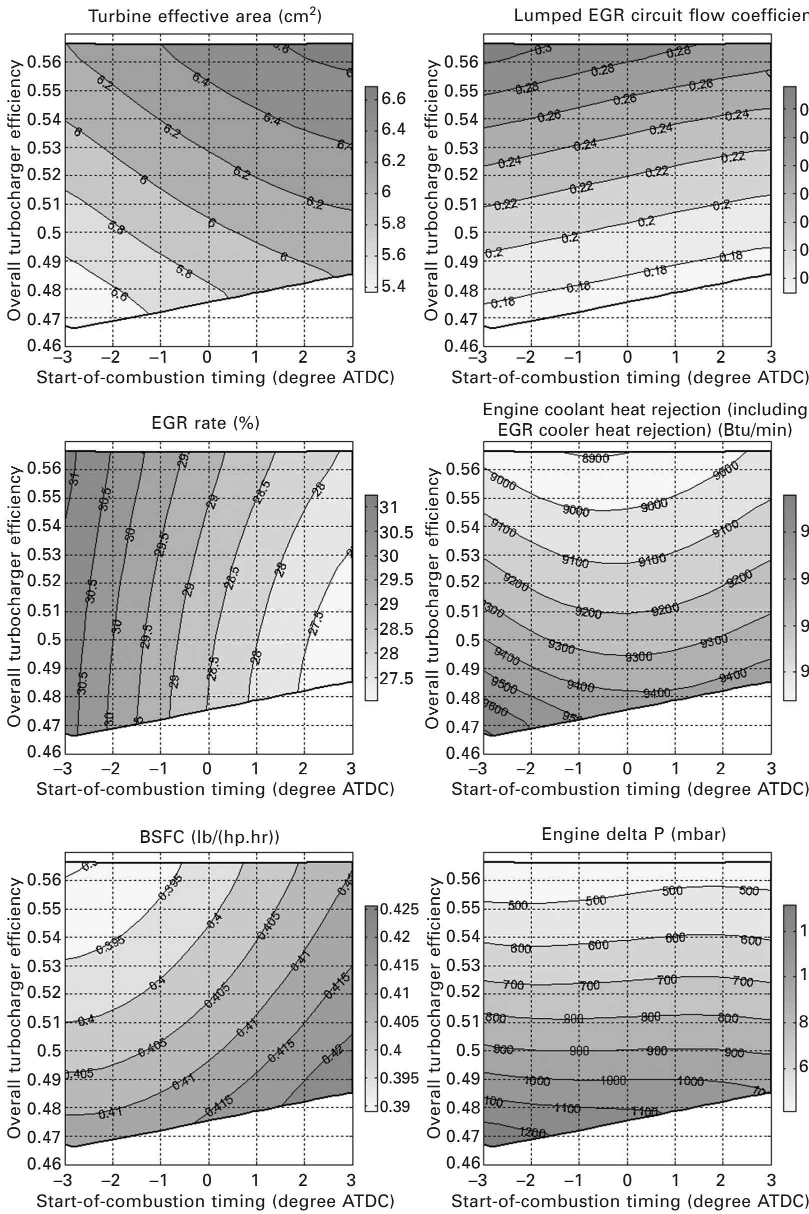
P may become insufficient to drive enough EGR even if the EGR valve is set fully open. These effects are shown in Figs 15.19–15.21.

As for the interaction with aftertreatment, the exhaust restriction flow

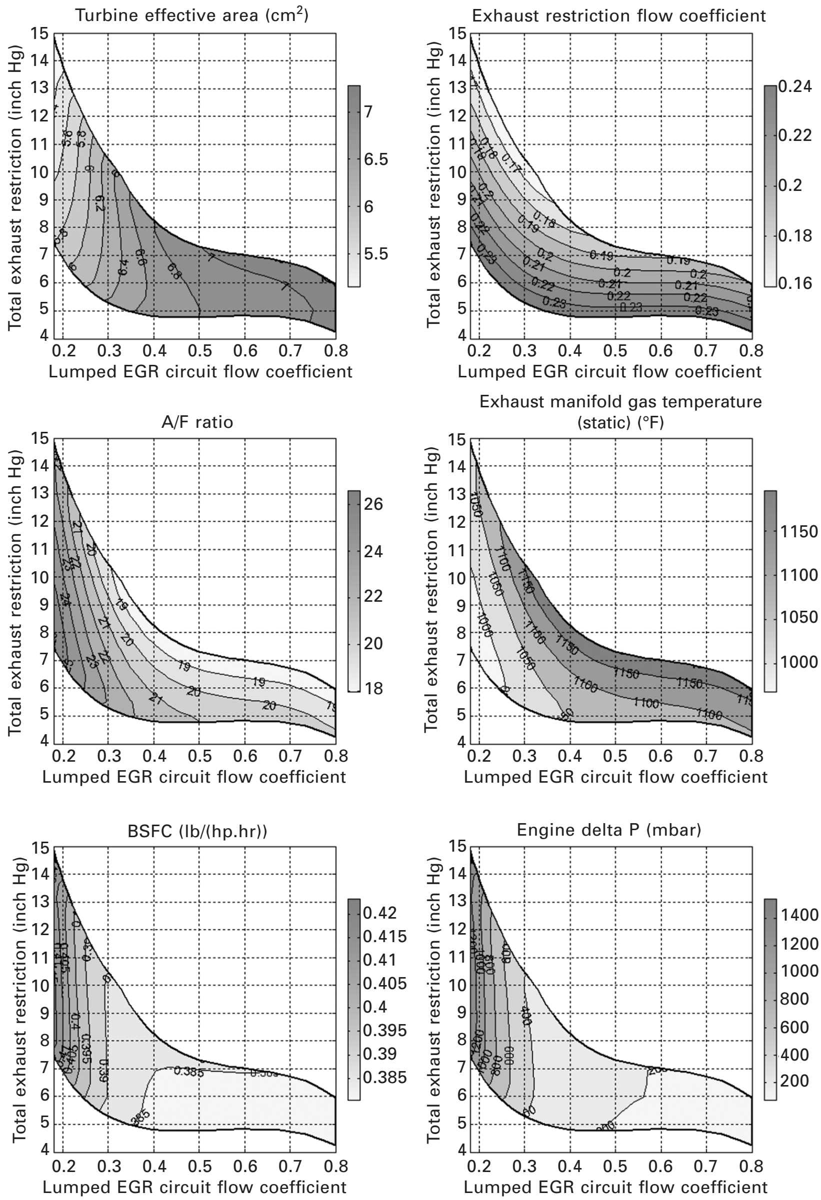


15.15 HD diesel subsystem interaction at rated power, type A6 analysis with fixed A/F ratio = 22 and peak cylinder temperature = 2439°F (EGR valve fully open).

coefficient for given aftertreatment hardware changes only when the DPF soot load changes. When the exhaust restriction or the turbine outlet pressure increases, the turbine pressure ratio and the air–fuel ratio decrease. To

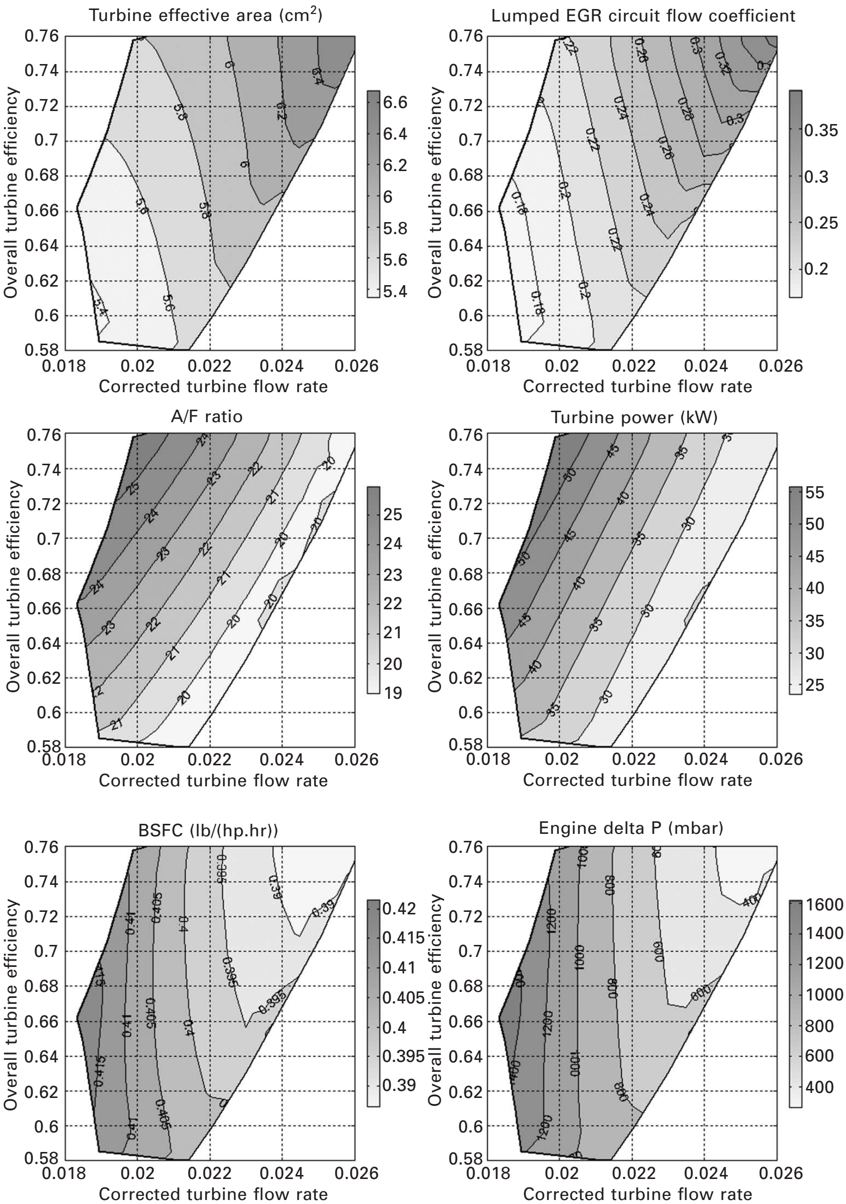


15.16 HD diesel subsystem interaction at rated power, type A9 analysis with fixed A/F ratio = 22 and peak cylinder temperature = 2439°F (fixed exhaust restriction flow coefficient, fixed EGR cooler and CAC size).



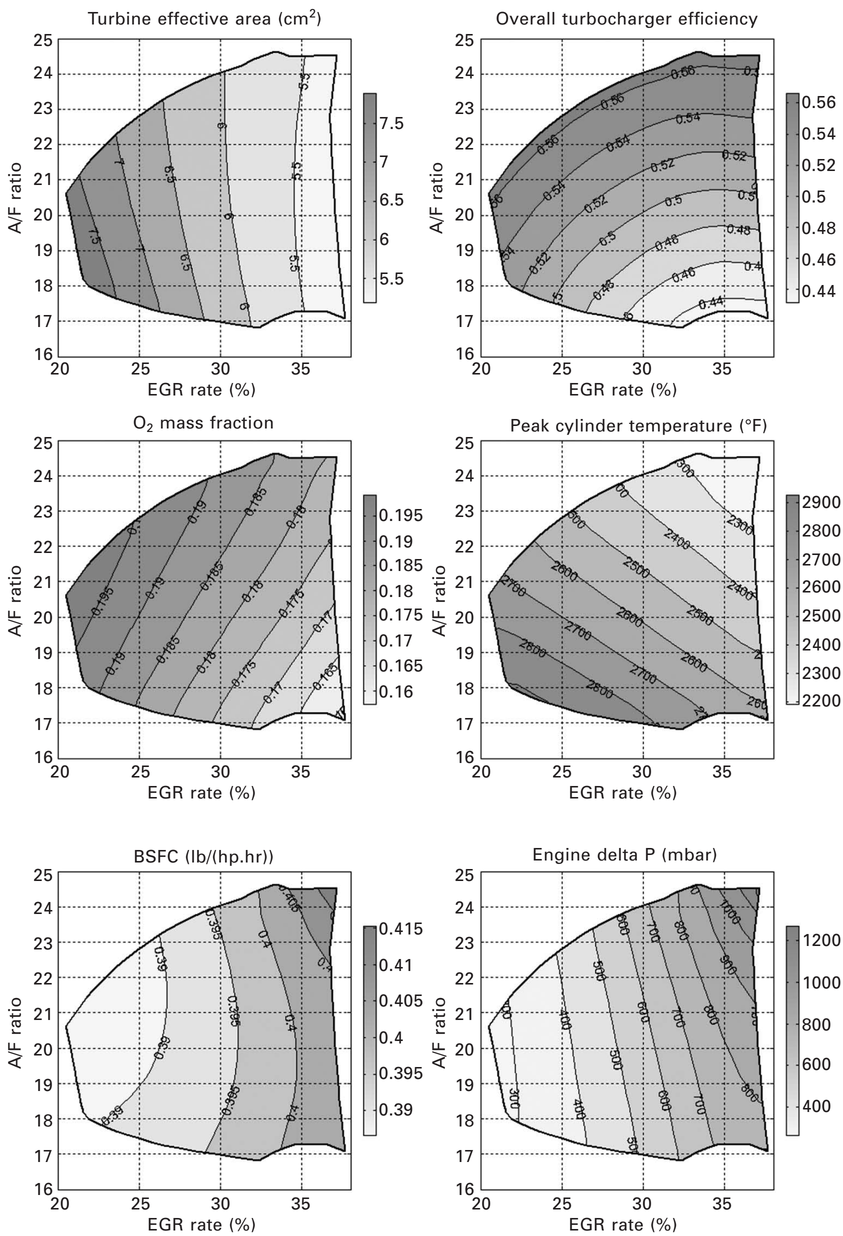
15.17 HD diesel A/F ratio effect at rated power, type B1 analysis with EGR rate = 31%.

compensate for that, a smaller turbine area needs to be used to restore the turbocharger speed and the boost pressure. As a result, the EGR valve has to be closed more, and as a result the engine delta P and the pumping loss increase. Figure 15.13 illustrates such an effect.

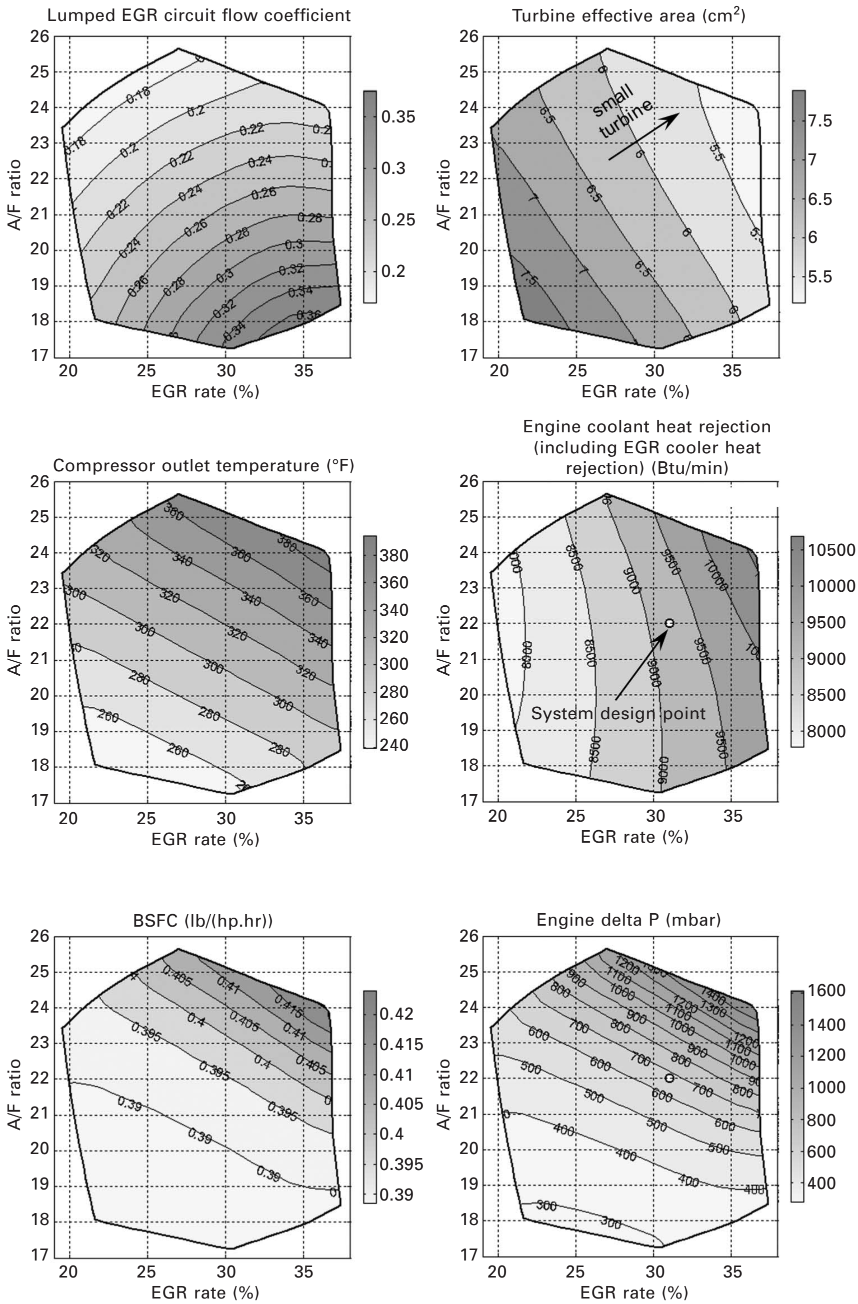


15.18 HD diesel A/F ratio effect at rated power, type B2 analysis with EGR rate = 31%.

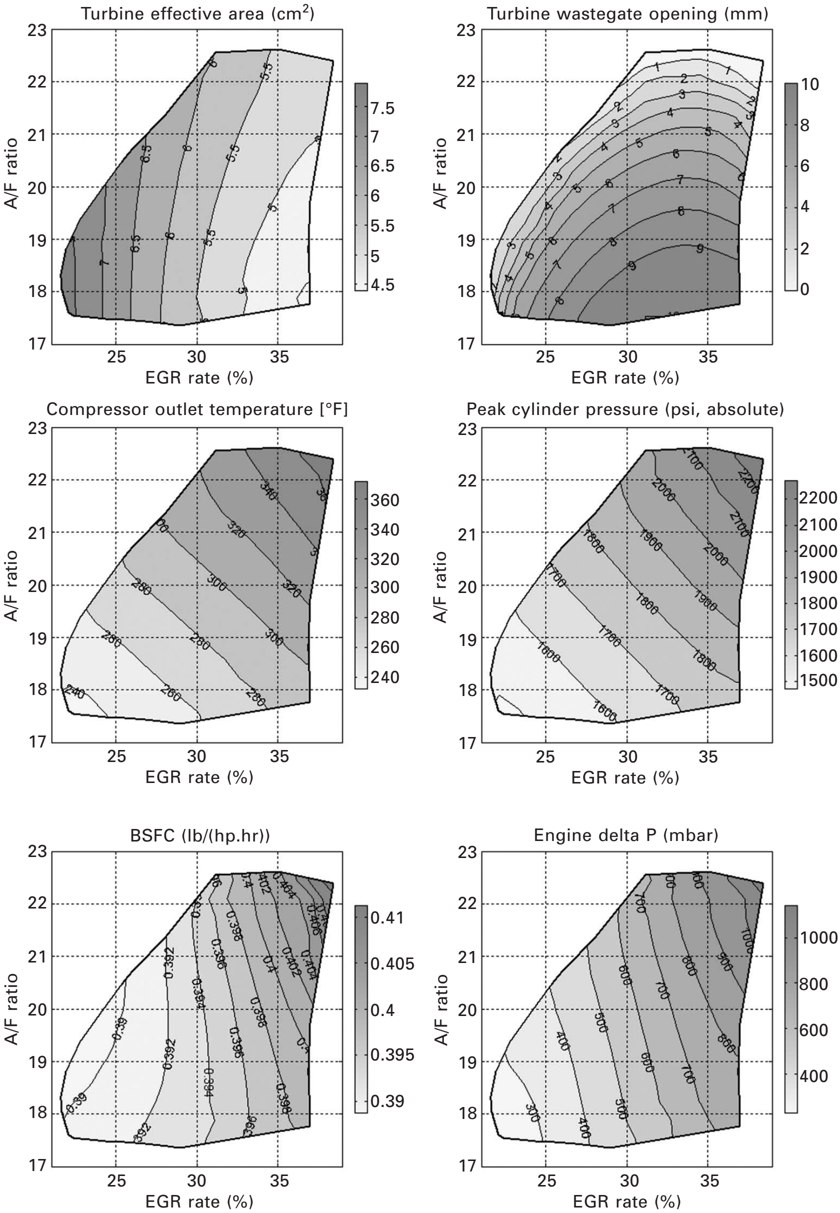
Cooling system design has an impact on pumping loss through intake manifold gas temperature and the flow restriction of EGR cooler or charge air cooler. Higher cooler effectiveness and lower restriction result in less required



15.19 HD diesel air system capability at rated power, type C1 analysis with IMT = 150°F, exhaust restriction = 9.7 inch Hg, EGR valve fully open, and turbine wastegate fully closed.



15.20 HD diesel air system capability at rated power, type C2 analysis with IMT = 150°F, exhaust restriction = 9.7 inch Hg, overall turbocharger efficiency = 53%, and turbine wastegate fully closed.



15.21 HD diesel air system capability at rated power, type C3 analysis with IMT = 150°F, exhaust restriction = 9.7 inch Hg, overall turbocharger efficiency = 53%, and EGR valve fully open.

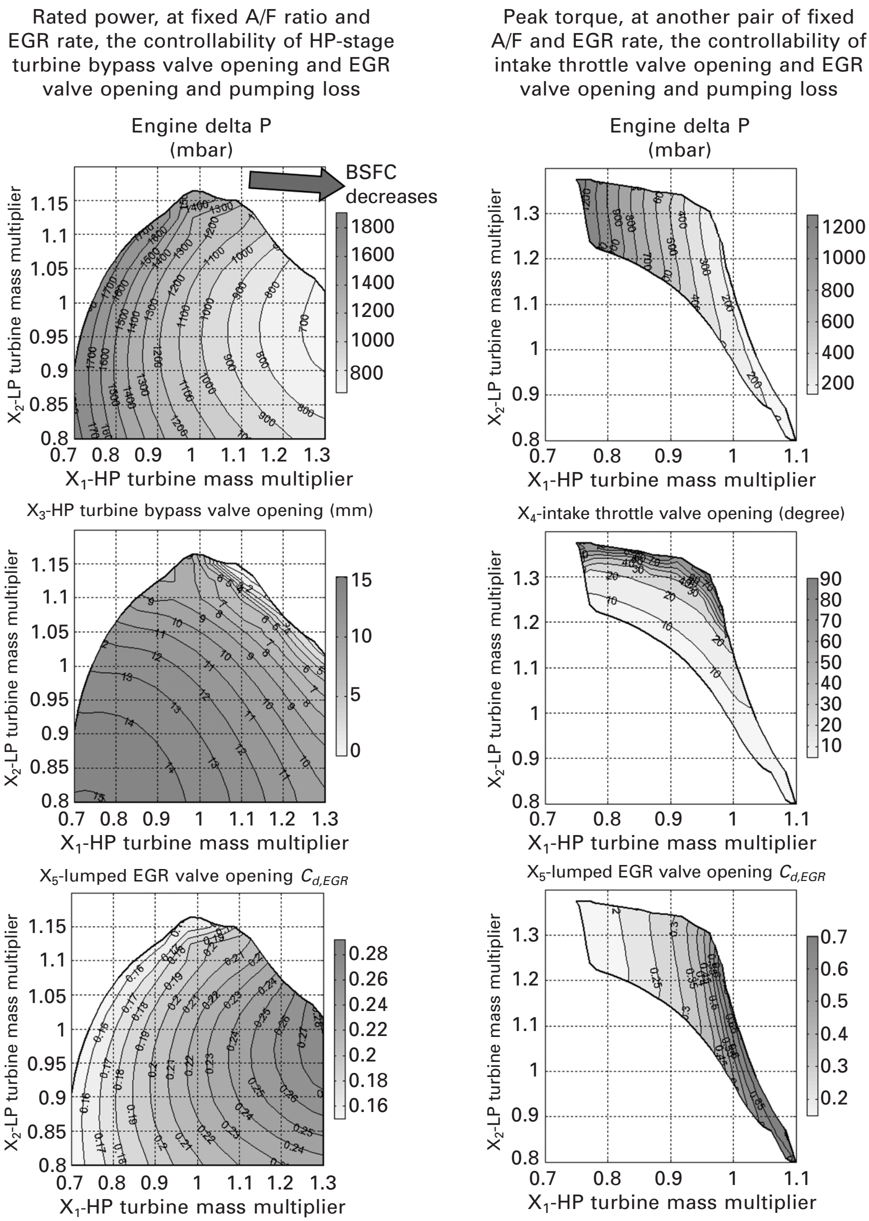
engine delta P to drive a target air–fuel ratio and EGR rate, especially at peak torque where the EGR valve is usually fully open. Then, a larger turbine area can be selected to reduce the pumping loss. There are trade-offs between cooler effectiveness, cooler flow restriction, and packaging size. Figures 15.11–15.12 illustrate the effect of cooler size in engine system design.

Two-stage turbocharging has been adopted for high power-rating engines with high EGR rate due to the high boost pressure required. A lower pressure at the low-pressure-stage turbine inlet generally results in an increase in the high-pressure-stage turbine power and a decrease in the low-pressure-stage turbine power. An advanced approach to analyzing the turbine area sizing of the two stages and the interaction between the two stages is to conduct minimum-BSFC optimization under fixed emissions (or their approximation as a fixed air–fuel ratio and EGR rate) at rated power and peak torque. Such an optimization is illustrated in Fig. 15.22. In the figure, the horizontal axis represents the high-pressure-stage turbine area, and the vertical axis represents the low-pressure-stage turbine area. It is observed that the EGR–turbocharger hardware capability and the range of controllability are very different at different speed–load modes. A compromise usually needs to be made to select the best turbocharger sizes for both stages. A similar dilemma exists regarding using a small turbine or using an intake throttle. Using a small turbine may help drive EGR at peak torque but causes higher pumping loss at rated power when the turbine is wastegated or bypassed. In contrast, using an intake throttle at peak torque may increase EGR rate slightly (with a penalty of a large reduction in the air–fuel ratio) so that a larger turbine may be selected for peak torque. Although the BSFC at rated power becomes lower due to the larger turbine, using intake throttle results in a BSFC penalty at peak torque. The wider the engine speed range, the more difficult the trade-off. The best solution is to use a VGT. If a VGT is not used, a wise compromise needs to be made on the turbine size based on the vehicle's frequent operating modes (low engine speed vs. high speed) and transient acceleration requirements.

To summarize, although the engine hardware selection process is interactive between subsystems, the following can be used as a general guideline for the sequence: (1) exhaust restriction; (2) charge air cooler and EGR cooler capacity; (3) EGR circuit flow restriction; and (4) turbocharger. The above discussions are for engine systems with a conventional valvetrain (cam). For subsystem interaction involving VVA, the engine performance can be best analyzed by using the air system capability chart shown in Fig. 9.12(c) in Chapter 9.

After the air system hardware has been specified for critical modes in the normal and extreme ambient conditions, a simulation of virtual performance and emissions calibration in the entire engine speed–load domain needs to be conducted with proper parameter smoothing over the entire domain (an example is shown in Fig. 7.8). Transient powertrain simulation also needs to be carried out to evaluate the hardware's transient capability and electronic control algorithms.

A bibliography of other applications of diesel engines is listed for completeness.



15.22 Simulation optimization of two-stage turbocharging.

15.5 References and bibliography

Asad U and Shafique U A (2001), 'Dual-fuel system for heavy duty truck engines', ASME paper 2001-ICE-423.

Berard F, Cotta A, Stokes J, Thring R and Wheals J (2000), 'An integrated powertrain (IPT) model – stage one', SAE paper 2000-01-0864.

Boggs D L, Belaire R C, Bartunek B, Durnholz M and Ecker H-J (1997), 'A small displacement DI diesel engine concept for high fuel economy vehicles', SAE paper 972680.

Burtt D J and James I B (2004), 'Use of system design techniques in the optimization of integrated powertrains incorporating an IVT', SAE paper 2004-01-0352.

Chen S K, Mather D and Reitz R (1997), 'Current and advanced design concepts for high power-density mid-range truck diesel engines, part II: the development of advanced design process for combustion system optimization', SAE paper 972689.

Conley J and Taylor S (2002), 'Technological evaluation of fuel efficiency improvement concepts to meet future regulatory requirements in the North American market', SAE paper 2002-01-2809.

Demirbas A (2008), *Biodiesel: A Realistic Fuel Alternative for Diesel Engines*, Springer-Verlag London Limited, London.

Eberhardt J J (1999), 'Overview of the DOE heavy vehicle technologies R&D program', SAE paper 1999-01-2235.

Fussey P M, Goodfellow C L, Oversby K K, Porter B C and Wheals J C (2001), 'An integrated powertrain (IPT) model – stage 2: systems integration, supervisory control and simulation of emissions control technology', SAE paper 2001-01-0938.

Georgi B, Hunkert S, Liang J and Willmann M (1997), 'Realizing future trends in diesel engine development', SAE paper 972686.

Hikosaka N (1997), 'A view of the future of automotive diesel engines', SAE paper 972682.

Hountalas D T (2000), 'Available strategies for improving the efficiency of DI diesel engines – a theoretical investigation', SAE paper 2000-01-1176.

Khair M K (1997), 'Technical and synergistic approaches towards the twenty-first century diesel engine', SAE paper 972687.

May M P, Vertin K, Ren S, Gui X, Myburgh I and Schaberg P (2001), 'Development of truck engine technologies for use with Fischer–Tropsch fuels', SAE paper 2001-01-3520.

Merrion D F (1994), *Diesel Engine Design for the 1990s*, the 40th L. Ray Buckendale Lecture, SAE SP-1011, SAE paper 940130, SAE International, Warrendale, PA.

Page D L and Edgar B L (1998), 'A systems-level optimization strategy for diesel engines', SAE paper 981914.

Pahl G (2008), *Biodiesel: Growing a New Economy*, 2nd edn, Chelsea Green Publishing, White River Junction, VT.

Rakopoulos C D, Hountalas D T, Zannis T C and Levendis Y A (2004), 'Operational and environmental evaluation of diesel engines burning oxygen-enriched intake air or oxygen-enriched fuels: a review', SAE paper 2004-01-2924.

Regueiro J F and Chen S K (1997), 'Current and advanced design concepts for high power-density mid-range truck diesel engines, part I: technology review', SAE paper 972688.

Rickeard D J, Bonetto R and Signer M (1996), 'European programme on emissions, fuels and engine technologies (EPEFE) – comparison of light and heavy duty diesel studies', SAE paper 961075.

Schindler K-P (1997), 'Why do we need the diesel?', SAE paper 972684.

Scott W M (1984), 'The passenger car diesel engine – present and future', in Haddad S D and Watson N (eds), *Design and Applications in Diesel Engineering*, Ellis Horwood Limited, Chichester.

Silva D F C (2006), 'DME as alternative diesel fuel: overview', SAE paper 2006-01-2916.
Singh G, Graves R L, Storey J M, Partridge W P, Thomas J F, Penetrante B M, Brusasco R M, Merritt B T, Vogtlin G E, Aardahl C L, Habeger C F and Balmer M L (2000), 'Emission control research to enable fuel efficiency: Department of Energy heavy vehicle technologies', SAE paper 2000-01-2198.
Taylor R I (2000), 'Heavy duty diesel engine fuel economy: lubricant sensitivities', SAE paper 2000-01-2056.
Thompson N, Ntziachristos L, Samaras Z, Aakko P, Wass U, Hausberger S and Sams T (2004), 'Overview of the European "particulates" project on the characterization of exhaust particulate emissions from road vehicles: results for heavy duty engines', SAE paper 2004-01-1986.
Yang Z (2002), 'A study of two-stroke engine braking for HD diesel engines', SAE paper 2002-01-0487.

15.5.1 References and bibliography on other applications of diesel engines

Light-duty diesel engines

Andersson J D, Wedekind B G A, Hall D, Stradling R and Wilson G (2001), 'DETR/SMMT/CONCAWE particulate research programme: light duty results', SAE paper 2001-01-3577.
Cooper B, Penny I, Beasley M, Greaney A and Crump J (2006), 'Advanced diesel technology to achieve Tier 2 Bin 5 emissions compliance in US light-duty diesel applications', SAE paper 2006-01-1145.
Cowland C, Gutmann P and Herzog P L (2004), 'Passenger vehicle diesel engines for the US', SAE paper 2004-01-1452.
Hublin M, Gadd P G, Hall D E and Schindler K P (1996), 'European programmes on emissions, fuels and engine technologies (EPEFE) – light duty diesel study', SAE paper 961073.
Krieger R B, Siewert R M, Pinson J A, Gallopoulos N E, Hilden D L, Monroe D R, Rask R B and Solomon A S P (1997), 'Diesel engines: one option to power future personal transportation vehicles', SAE paper 972683.
Majewski W A and Khair M K (2006), *Diesel Emissions and Their Control*, SAE International, Warrendale, PA.
McDonald J (2004), 'Progress in the development of Tier 2 light-duty diesel vehicles', SAE paper 2004-01-1791.
Mital R, Li J, Huang S C, Stroia B J, Yu R C, Anderson J A and Howden K (2003), 'Diesel exhaust emissions control for light duty vehicles', SAE paper 2003-01-0041.
Morey C J and Mark J (2000), 'Diesel passenger vehicles – can they meet air quality needs and climate change goals?', SAE paper 2000-01-1599.
Ntziachristos L, Mamakos A, Samaras Z, Mathis U, Mohr M, Thompson N, Stradling R, Forti L and de Serves C (2004), 'Overview of the European "particulates" project on the characterization of exhaust particulate emissions from road vehicles: results for light-duty vehicles', SAE paper 2004-01-1985.
Ntziachristos L, Mamakos A, Samaras Z, Rexeis M and Hausberger S (2006), 'Diesel particle exhaust emissions from light duty vehicles and heavy duty engines', SAE paper 2006-01-0866.

Rickeard D J, Bonetto R and Signer M (1996), 'European programme on emissions, fuels and engine technologies (EPEFE) – comparison of light and heavy duty diesel studies', SAE paper 961075.

Schulte H S-H (1999), 'Passenger car engines', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 524–551.

Zelenka P, Egert M and Cartellieri W (2000), 'Ways to meet future emissions standards with diesel engine powered sport utility vehicles (SUV)', SAE paper 2000-01-0181.

Off-road diesel engines

Eberhardt J J (1999), 'Overview of the DOE heavy vehicle technologies R&D program', SAE paper 1999-01-2235.

Gui X, Dou D and Winsor R (2010), 'Non-road diesel engine emissions and technology options for meeting them', *ASABE Distinguished Lecture Series*, *Tractor Design*, No. 34, ASABE Publication Number 913C0110.

Jackson D R and Blevins W G (1983), 'Possibilities for locomotive diesel fuels – a railway overview', ASME conference paper.

Jankovic S, Balasubramanian S and Roy C (2001), 'Correlation between agricultural tractor's diesel engine on road service load and regulations for exhaust emission testing', SAE paper 2001-01-3345.

Killmann I G and Tholen P (1999), 'Aircooled engines', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 424–448.

Majewski W A and Khair M K (2006), *Diesel Emissions and Their Control*, SAE International, Warrendale, PA.

Niemi S A, Laurén M J and Murtonen T T (2002), 'Effect of waste-gate turbocharging on the exhaust particulate matter of an off-road diesel engine', SAE paper 2002-01-2159.

Niven H (1999), 'Locomotives', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 588–612.

Norén O and Pettersson O (2001), 'Development of relevant work-cycles and emission factors for off-road machines', SAE paper 2001-01-3637.

Pixley M (2000), 'Off-highway noise – diverging demands of the future', SAE paper 2000-01-2548.

Puranen A and Tanska T (1994), 'A study on off-road diesel engines emission reduction', SAE paper 941755.

Sattler M L (2002), 'Technologies for reducing NO_x emission from off-road diesel vehicles: an overview', *EM*: *Air and Waste Management Association's Magazine for Environmental Managers*, February, 20–29.

Shafi A A (2003), 'An overview of modern diesel multiple units today', *Rail International*, 34(9), 32–43.

Treiber P J H and Sauerteig J E (1991), 'Present and future European exhaust emission regulations for off-road diesel engines', SAE paper 911808.

Ullman T L, Webb C C, Jackson C C and Doorlag M H (1999), 'Nonroad engine activity analysis and transient cycle generation', SAE paper 1999-01-2800.

Weaver C S and McGregor D B (1994), 'Emissions control technology for locomotive engines', SAE paper 940453.

Marine diesel engines

Er I D (2002), 'Overview of NO_x emission controls in marine diesel engines', *Energy Sources*, *Part A*: *Recovery, Utilization, and Environmental Effects*, 24(4), 319–327.
Geist M A, Holtbecker R and Chung S Y (1997), 'Marine diesel NO_x reduction technique – a new Sulzer Diesel Ltd approach', SAE paper 970321.
Markle S P and Brown A J (1995), 'Naval diesel engine duty cycle development', SAE paper 950733.
Mueller E, Kollmann T and Reid T (2006), 'Performance development of the supercharged Mercury marine outboard engines', SAE paper 2006-01-0015.
Rakopoulos C D, Michos C N and Giakoumis E G (2005), 'Study of the transient behavior of turbocharged diesel engines including compressor surging using a linearized quasi-steady analysis', SAE paper 2005-01-0225.
Woodyard D (1999), 'Marine engine applications', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 632–661.

Alternative fuels

Alleman T L and McCormick R L (2003), 'Fischer–Tropsch diesel fuels – properties and exhaust emissions: a literature review', SAE paper 2003-01-0763.
Hansen A C, Zhang Q and Lyne P W L (2005), 'Ethanol-diesel fuel blends – a review', *Bioresource Technology*, 96(3), 277–285.
May M P, Vertin K, Ren S, Gui X, Myburgh I and Schaberg P (2001), 'Development of truck engine technologies for use with Fischer–Tropsch fuels', SAE paper 2001-01-3520.
Rakopoulos C D, Hountalas D T, Zannis T C and Levendis Y A (2004), 'Operational and environmental evaluation of diesel engines burning oxygen-enriched intake air or oxygen-enriched fuels: a review', SAE paper 2004-01-2924.
Ribeiro N, Pinto A C, Quintella C M, da Rocha G O, Teixeira L S G, Guarieiro L L N, Rangel M do C, Veloso M C C, Rezende M J C, da Cruz R S, de Oliveira A M, Torres E A and de Andrade J B (2007), 'The role of additives for diesel and diesel blended (ethanol or biodiesel) fuels: a review', *Energy and Fuels*, 21(4) 2433–2445.
Richard R (1999), 'Dual fuel engines', in Challen B and Baranescu R (eds), *Diesel Engine Reference Book*, 2nd edn, SAE International, Warrendale, PA, pp. 614–630.
SAE Surface Vehicle Information Report J1297 (2007), 'Alternative automotive fuels'.
Silva D F C (2006), 'DME as alternative diesel fuel: overview', SAE paper 2006-01-2916.
Tipler W (1984), 'Fuels for diesel engines', in Haddad S D and Watson N (eds), *Principles and Performance in Diesel Engineering*, Ellis Horwood Limited, Chichester.
Weiskirch C, Kaack M, Blei I and Eilts P (2008), 'Alternative fuels for alternative and conventional diesel combustion systems', SAE paper 2008-01-2507.

Biodiesel

Babu A K and Devaradjane G (2002), 'Vegetable oils as fuel for diesel engines: an overview', *ASME Heat Transfer Division Publication*, *HTD*, 372(3), 313–318.

Babu A K and Devaradjane G (2003), 'Vegetable oils and their derivatives as fuels for CI engines: an overview', SAE paper 2003-01-0767.
Balat M and Balat H (2008), 'A critical review of bio-diesel as a vehicular fuel', *Energy Conversion and Management*, 49(10), 2727–2741.
Demirbas A (2008), *Biodiesel: A Realistic Fuel Alternative for Diesel Engines*, Springer Verlag, London.
Devlin C C, Passut C A, Campbell R L and Jao T-C (2008), 'Biodiesel fuel effect on diesel engine lubrication', SAE paper 2008-01-2375.
Jokai R and Duo W L (2008), 'The impact biodiesel blend levels have on engine performance', SAE paper 2008-01-2610.
Kawano D, Ishii H, Goto Y, Noda A and Aoyagi Y (2006), 'Application of biodiesel fuel to modern diesel engine', SAE paper 2006-01-0233.
Knothe G (2010), 'Biodiesel and renewable diesel: a comparison', *Progress in Energy and Combustion Science*, 36, 364–373.
Kumar M S, Bellettre J and Tazerout M (2009), 'The use of biofuel emulsions as fuel for diesel engines: a review', *Proc. IMechE*, *Part A: Journal of Power and Energy*, 223(7), 729–742.
McCrady J, Hansen A and Lee C-F (2007), 'Modeling biodiesel combustion using GT-Power', *Technical Papers of 2007 ASABE Annual International Meeting*, June 17–20, Minneapolis, Minnesota.
McCrady J, Hansen A and Lee C-F (2008), 'Combustion and emissions modeling of biodiesel using GT-Power', *Technical Papers of 2008 ASABE Annual International Meeting*, June 29 – July 2, Providence, Rhode Island.
Murugesan A, Umarani C, Subramanian R and Nedunchezhian N (2009), 'Bio-diesel as an alternative fuel for diesel engines – a review', *Renewable and Sustainable Energy Reviews*, 13(3), 653–662.
Pahl G (2008), *Biodiesel: Growing a New Energy Economy*, 2nd edn, Chelsea Green Publishing, White River Junction, VT.
Pryde E H (1983), 'Vegetable oils as diesel fuels: overview', *Journal of the American Oil Chemists' Society*, 60(8), 1557–1558.
Szybist J P and Boehman A L (2003), 'Behavior of a diesel injection system with biodiesel fuel', SAE paper 2003-01-1039.
Tatur M, Nanjundaswamy H, Tomazic D, Thornton M and McCormick R L (2009), 'Biodiesel effects on US light-duty Tier 2 engine and emission control systems – part 2', SAE paper 2009-01-0281.
Tompkins B T, Esquivel J and Jacobs T J (2009), 'Performance parameter analysis of a biodiesel-fuelled medium duty diesel engine', SAE paper 2009-01-0481.
Trapel E, Mayer C, Schulz C and Roth P (2005), 'Effects of bio diesel injection in a DI diesel engine on gaseous and particulate emission', SAE paper 2005-01-2204.
Williams A, McCormick R L, Hayes R R, Ireland J and Fang H L (2006), 'Effect of biodiesel blends on diesel particulate filter performance', SAE paper 2006-01-3280.

Two-stroke diesel engines

Achates Power: www.achatespower.com
Arias J R, Varela E, Pérez R, Navarro E and Moreno E (2001), 'Numerical simulation of the scavenging process in a two stroke turbocharged diesel engine', SAE paper 2001-01-1094.

Cantore G and Mattarelli E (2007), 'A new concept for ultra-compact automotive HSDI diesel engines', SAE paper 2007-01-1253.

De Marco C A, Mattarelli E, Paltrinieri F and Rinaldini C A (2007), 'A new combustion system for 2-stroke HSDI diesel engines', SAE paper 2007-01-1255.

Duret P (ed.) (1993), *A New Generation of Two-Stroke Engines for the Future?* Editions Technip, Paris.

EcoMotors International: www.ecomotors.com

Franke M, Huang H, Liu J P, Geistert A and Adomeit P (2006), 'Opposed piston opposed cylinder (opoc™) 450 hp engine: performance development by CAE simulations and testing', SAE paper 2006-01-0277.

Franco A and Martorano L (1995), 'Principles for optimization of air cooling system applied to the two-stroke engine', SAE paper 951798.

Heywood J B and Sher E (1999), *The Two-Stroke Cycle Engine: Its Development, Operation, and Design*, Taylor & Francis, Philadelphia, PA.

Hirsch N R, Schwarz E E and McGough M G (2006), 'Advanced opposed-piston two-stroke diesel demonstrator', SAE paper 2006-01-0926.

Hofbauer P (2005), 'Opposed piston opposed cylinder (opoc™) engine for military ground vehicles', SAE paper 2005-01-1548.

Kalkstein J, Röver W, Campbell B, Zhong L, Huang H, Liu J P, Tatur M, Geistert A and Tusinean A (2006), 'Opposed piston opposed cylinder (opoc™) 5/10 kW heavy fuel engine for UAVs and APUs', SAE paper 2006-01-0278.

Kang K Y, Lee J W, Jeong Y I, Huh K Y and Jeong H T (1996), 'The characteristics of scavenging flow in a poppet-valve type 2-stroke diesel engine by using RSSV system', SAE paper 960368.

Kubota Y, Hayashi S, Kajitani S and Sawa N (1991), 'Step response of engine torque and speed in small two-stroke diesel engines', SAE paper 911851.

Kubota Y, Hayashi S, Kajitani S and Sawa N (1992), 'Step response of engine torque and speed to load change in small two-stroke cycle diesel engines', SAE paper 920428.

McGough M G and Fanick E R (2004), 'Experimental investigation of the scavenging performance of a two-stroke opposed-piston diesel tank engine', SAE paper 2004-01-1591.

Melchior J F (1993), 'Presentation of the MT 135 marine diesel two-stroke engine', in Duret P (ed.), *A New Generation of Two-Stroke Engines for the Future?* Editions Technip, Paris, France.

Nomura K and Nakamura N (1993), 'Development of a new two-stroke engine with poppet-valves: Toyota S-2 engine', in Duret P (ed.), *A New Generation of Two-Stroke Engines for the Future?* Editions Technip, Paris, France.

Nuti M (1998), *Emissions from Two-Stroke Engines*, SAE International, Warrendale, PA.

Peng L, Tusinean A, Hofbauer P and Deylami K (2005), 'Development of a compact and efficient truck APU', SAE paper 2005-01-0653.

Pirault J-P and Flint M (2010), *Opposed Piston Engines – Evolution, Use, and Future Applications*, SAE International, Warrendale, PA.

Sattler M L (2002), 'Technologies for reducing NO_x emission from off-road diesel vehicles: an overview', *EM*: *Air and Waste Management Association's Magazine for Environmental Managers*, February , 20–29.

Wyczalek F A (1991), 'Two-stroke engine technology in the 1990s', SAE paper 910663.

16
Concluding remarks and outlook for diesel engine system design

Abstract: This chapter summarizes the key points of each earlier chapter and provides an outlook on the critical research topics in diesel engine system design. It emphasizes the importance of durability prediction, system reliability analysis, emissions prediction, engine transient modeling, air system integration, and virtual engine calibration. It is concluded by a vision of system integration for future diesel powertrain design.

Key words: diesel engine system design (DESD), engine performance and system integration (EPSI), durability prediction, system reliability, emissions prediction, transient modeling, air system integration, virtual engine calibration.

The stringent emissions regulations and ever-increasing customer demands on power, fuel economy, NVH, and reliability will impose greater challenges on future diesel technologies. Different aspects of engine design including direct fuel injection, advanced combustion and air systems, aftertreatment, electronic controls, and powertrain need to be seamlessly integrated. Fast-paced engine product development requires a precise emissions-driven system design approach to integrate different technical areas in order to reduce engine development time and cost. Being an emerging technical field, diesel engine system design (DESD), particularly in the area of engine performance and system integration (EPSI), will become increasingly important.

This book summarizes the theory and analytical approaches of diesel engine system design, and elaborates its relationship with other related technical fields. It also provides a new perspective regarding the traditional mechanical design areas from the viewpoint of engine system design.

Diesel engine system design requires a systems engineering approach to coordinate different design attributes and subsystems. Because the primary application of diesel engine system design is industrial product design, reliability is the ultimate goal and needs to be accounted for at the system design stage. This requires the system designer to consider a systems engineering approach together with a robust engineering and reliability engineering philosophy to properly handle the variability and reliability optimization problems. Design for just a nominal target is not sufficient as a system design solution. Design for variability and design for reliability must be considered. The four attributes (performance, durability, packaging, and cost) in system design need to be balanced. These concepts set the logic and development path for diesel engine system design from a systems engineering perspective, and they are the topics addressed in Chapter 1.

Engine system design is a performance-driven activity subject to durability constraints. Properly understanding how the durability constraints are derived in order to maximize their potential is critical for a system engineer. The system design solutions are often produced right on the edge of the durability constraints in order to maximize the engine performance capability. The research on analytical and experimental analysis for durability and reliability issues should not be overlooked. This is the topic of Chapter 2.

Advanced optimization techniques are used by system engineers in their daily work to address the complex interactions between engine subsystems. Single- and multi-objective optimizations, deterministic and nondeterministic (probabilistic) optimizations (i.e., DoE RSM and Monte Carlo simulation), and variability/reliability-based optimization are all required. These topics are discussed in Chapter 3.

The four cornerstones in diesel engine system design are static design, dynamic design, and the first and second laws of thermodynamics. Chapter 4 lays out a foundation to explain the parametric relationship between the system design parameters in the engine in-cylinder cycle process. It also elaborates the engine gas flow network with the focus on pumping loss, which is the key for modern high-EGR turbocharged diesel engines. Chapter 4 provides the system-level mathematical formulae (e.g., the four core equations, 4.40, 4.44, 4.47, and 4.50) that are helpful for understanding the interactions between subsystems and other topics addressed in the following chapters. The fundamental theory presented in Chapter 4 can be used to explore the options of low-pumping-loss engine design and model-based engine controls.

The three central tasks in engine system design are engine–vehicle matching, engine–aftertreatment matching and engine–turbocharger matching. The matching is surrounded by many boundary conditions such as the requirements from combustion, emissions, cooling, structure, etc. Engine–vehicle is the first design interface a system engineer needs to consider in a 'top-down' approach. Chapter 5 emphasizes this concept by illustrating the parametric relationship between the vehicle/powertrain level and the engine level. The vehicle integration theory in Chapter 5 can be used to build a truly 'top-down' optimization design approach targeting real-world driving profiles.

As a part of the considerations for vehicle and engine system performance, engine brake needs to be considered at the engine system design stage. This topic should not be ignored in engine air system hardware sizing. Chapter 6 summarizes the performance design topics for all the retarders, including vehicle braking performance, engine brake design and performance optimization, braking thermodynamic cycles, and the interactions with valvetrain, VVA, and turbocharger. The technical foundation of engine brake performance is based on thermodynamic cycle processes and engine valve flow characteristics. Thoroughly understanding engine brake performance and its cycle simulation will provide a valuable learning experience for a system engineer in order to comprehend the related simulation and design techniques, at least before the more sophisticated combustion/emissions topics come into the play.

Chapter 7 establishes the system design logic from emissions to the overall engine system especially the air system. It links the three sequential functions in an engine development cycle: combustion/emissions development, engine system design, and engine calibration. Emission strategies, emissions modeling, combustion system design, combustion modeling, advanced combustion mode (e.g., HCCI), combustion control, fuel system optimization, and calibration/control strategies are among the general topics to consider. The calibration function is the one usually ignored in a poor design practice. In fact, the unique requirements of engine calibration must be considered in the early stage of system design by using 'virtual calibration' to ensure the engine system specifications generated by the system design team will correctly reflect the engine reality down the road. Chapter 8 addresses engine–aftertreatment integration and aftertreatment calibration to handle the engine-out emissions produced by in-cylinder combustion.

The valvetrain is a unique subsystem having the dual characteristics of air system and mechanical system. It is a part of the air system because it controls engine gas flows, volumetric efficiency and engine delta P. It also possesses many design issues related to kinematics and dynamics. Chapter 9 starts to bridge the thermodynamic/thermo-fluid topics addressed in the earlier chapters with the mechanical design considerations such as dynamics and vibration in order to provide system design engineers with a balanced view on all the attributes. Valvetrain integration generally includes valvetrain system configurations and performance/design optimization, VVA, cylinder deactivation, activation systems and parasitic losses, cam design, and valvetrain dynamics. VVA is very important for diesel engines. The Miller cycle, the WR (wastegating reduction) or WE (wastegating elimination) intake VVA is one of the very few measures that can reduce both NO_x emissions and fuel consumption. VVA performance is discussed in great detail in Chapter 9.

Engine friction is a typical performance-related topic based on mechanical design considerations. Friction modeling is not only important for the accuracy of an engine system design model, but also critical for fuel economy improvement. Chapter 10 thoroughly elaborates the engine friction theory.

Noise, vibration, and hardness (NVH) (mainly noise), addressed in Chapter 11, is a very important performance attribute in diesel engine system design and it simply cannot be ignored. Engine noise is very critical for the competitiveness of diesel engine products. There is a big potential for the system design function to help optimize the NVH development process in order to save cost and improve the NVH of the engine product. Although very challenging from a modeling perspective, system NVH has the potential to become one of brightest points in diesel engine system design. NVH and friction represent the two most important mechanical topics a system engineer needs to understand. Sometimes they are related, for example, piston-assembly dynamics covers both piston slap noise and piston friction.

Heat rejection is probably the last but one of the most important topics a system design engineer needs to comprehend before he/she starts to conduct steady-state system design by using engine cycle simulation tools. Accurate heat rejection prediction is important for cooling system design. In-cylinder heat rejection is one of the most challenging research topics in the history of internal combustion engines. Chapter 12 presents a new view on the way to handle heat rejection calculations from the system design perspective. The theory is built based on the engine energy balance by using the first law of thermodynamics and a theoretical analysis of engine miscellaneous heat losses.

Chapter 13 addresses a core area in diesel engine system design: the air system, including the turbocharger, the manifolds, and the EGR system. It outlines the design guidelines and illustrates the application of the pumping loss theory. It points out the causes of engine subsystem interactions by applying the mathematical theory developed in Chapter 4 to engine air system design. Comprehensive analysis of the second law of thermodynamics for modern turbocharged EGR diesel engines is also conducted in Chapter 13 to illustrate this powerful approach in engine system design.

Chapter 14 forms the foundation of dynamic system design which is based on transient engine performance and electronic controls. The modeling methodology used by the engine controls community is from a system perspective due to the obvious needs to control the entire engine system. It often requires fast computing speeds in the models. The system dynamics approach is the methodology used by diesel engine system design for dynamic system design, and it shares a common ground with the engine controls community. In addition to the evaluation of engine hardware and control strategies for transient performance, it is foreseeable that analytical controller design may become an area where engine system design can make significant contributions. Chapter 14 also addresses virtual sensors and illustrates that engine system design may greatly contribute to their model development.

Chapter 15 presents a large number of simulation examples of subsystem interaction and optimization.

As outlined in each earlier chapter, in order to make simulations more accurate and the system integration process more effective, key research topics in diesel engine system design need to be identified not only within each subsystem itself but also at all the interfaces between vehicle, engine, aftertreatment, combustion, turbocharger, and EGR system. The key challenges are summarized below.

1. Analytical durability prediction to refine the system design constraints (addressed in Chapter 2) needs to be greatly promoted by developing inter-disciplinary analysis tools.
2. The design for variability and the design for reliability addressed in

Chapters 1 and 3 with the Monte Carlo simulation need to be widely applied in engine system design to supplement the design specification data of using the design-for-target approach.

3. The engine system reliability theory (Chapter 2) and reliability-based design optimization (Chapter 3) need to be fully explored and applied to diesel engine system design areas.
4. Design of Experiments (DoE), neural network and the Monte Carlo simulation are the standard data processing techniques used in system design (Chapter 3). More advanced multi-objective optimization methods need to be researched to better account for the trade-offs between different design attributes for system design decision making.
5. Advanced and fast combustion models of heat release rate need to be developed for better prediction of emissions, BSFC, and exhaust temperature in engine cycle simulation models (Chapter 4). The combustion model can also be used for effective prediction of combustion noise (Chapter 11).
6. High-fidelity vehicle powertrain dynamics simulation for the transient details will be helpful in the conventional and hybrid powertrain performance modeling (Chapter 5). Transient corrections on the steady-state engine maps (especially the emissions maps) used in vehicle driving cycle simulations need to be accurately accounted for because they largely influence the prediction accuracy compared to real-world transient driving events.
7. Innovative engine braking mechanisms (Chapter 6) need to be continuously explored to (1) fully utilize the potential of compression-release and braking gas recirculation (BGR) processes subject to the design constraints of peak cylinder pressure and exhaust manifold gas temperature; (2) reduce the braking component loading under high cylinder pressure; and (3) minimize the compression brake noise from the source of excitation.
8. Effective prediction of the difference between the steady-state emissions and the transient emissions remains a great challenge. This particular challenge has become very important for today's low-emission engines (Chapter 7).
9. Challenging research also remains in combustion, emissions (Chapter 7), and aftertreatment (Chapter 8) areas to develop the heuristic system emissions models that are suitable for the needs of engine system design.
10. Complete system optimization of the entire valvetrain design needs to be promoted to effectively handle the large amount of design parameters involved in the conventional cam-driven valvetrain (Chapter 9).
11. Innovative system integration solutions via VVA and other air system control valves need to be further explored, considering the interaction with turbocharging, in order to reduce pumping loss and achieve superior fuel economy with minimum system cost (Chapter 9). The importance

of volumetric efficiency and pumping loss should be emphasized. Their relationship is not only affected by valve timing and engine delta P but also complicated by manifold gas wave dynamics (e.g., Fig. 16.1). Different engine configurations (e.g., I6 vs. V8) may have different effectiveness of using a particular air system technology (e.g., wastegate elimination intake-VVA). For instance, it is observed from Fig. 16.1 that the I6 engine can achieve 26.4% pumping loss reduction due to its more uniform pumping loops across all cylinders in the *p-V* diagram, while the V8 engine can only achieve 13.9% reduction due to its drastically different, large cylinder-to-cylinder variations in the pumping loops, which are caused by manifold gas wave dynamics. Diesel engine system design largely relies on such a type of engine cycle simulations.

12. Real-time capable and accurate engine system friction models with crank-angle resolution are highly desirable for engine fault diagnostics and transient performance prediction (Chapter 10).
13. Advanced high-fidelity (crank-angle-resolution) models need to be further developed in engine system NVH research for component and overall engine noise predictions (Chapter 11).
14. In the heat transfer area (Chapter 12), further research is required to model EGR cooler soot fouling, transient, heat transfer and the exhaust manifold gas temperature.
15. Innovative air system design solutions will continue to be one of the hottest topics in diesel engine system design. It includes innovative air/gas control valves in the engine gas flow network and their interaction and integration with the turbocharger, the EGR system, and advanced valvetrains. In order to achieve reliable and cost-effective engine system designs, it is also important to evaluate the performance and durability of different turbocharging systems, for example, single-stage vs two-stage, wastegated vs VGT, and different types of VGT (i.e., VNT and vaneless VAT) (Chapter 13).
16. A wide application of the second law of thermodynamics in diesel engine system design will greatly enhance the quality of the work and provide more profound insight into the problems in the energy systems in the powertrain. It includes the system availability modeling for hybrid powertrains (Chapter 5), waste heat recovery (Chapter 12), air/EGR pumping and throttle losses in the air system (Chapter 13).
17. As to electronic controls (Chapter 14), owing to the need to accurately simulate the engine load response, vehicle driving cycles and the aftertreatment regeneration transients, engine system design will require tremendous efforts in transient powertrain performance modeling combined with developing model-based controls, despite the great challenges in transient simulations. Dynamic system design theory will result in a more advanced approach in analytical nonlinear controller design for modern diesel engine transient performance (Chapter 14).

18. Virtual sensors are developed based on either physics models (largely based on thermal/fluid models of pressure, temperature, flow, emissions, and engine torque) or empirical models (based on surface fit or neural network). It is expected that diesel engine system design can contribute greatly in this area by collaborating with engine controls closely.
19. Changing the engine operating cycle from four-stroke to two-stroke provides promising but challenging opportunities to reduce engine-out emissions and improve fuel economy for modern light-duty and heavy-duty diesel engines. The inherent characteristics of lower in-cylinder gas temperature and higher internal EGR rate of the two-stroke operation make strict in-cylinder NO_x control an appealing opportunity. However, the difficulties of external EGR pumping under a negative engine delta P and managing the scavenging gas flow present challenges for engine system design engineers to tackle.
20. All the above-mentioned advances ultimately should lead to a successful engine system specification design and the optimization of hardware subsystem interaction as well as 'virtual engine calibration' in a four-dimensional design space (Chapter 15). Properly integrating those techniques and presenting the system design solutions concisely for a highly complex multi-dimensional design problem will remain a challenge for system engineers to overcome.

Diesel engine system design is a multi-disciplinary field that integrates the knowledge of thermal fluids, dynamics, and controls to optimize various design attributes. The concept of advanced analytical design and system integration should not be limited to the emerging advanced technologies but should rather penetrate all traditional mechanical design areas to enhance the quality of every subsystem design. Effective competitive benchmarking comparison supported by advanced modeling will need to be widely conducted in each area of the system design.

Diesel engine system design not only provides a production design solution, but also participates and integrates modern technologies emerging from each subsystem of engine design. For example, diesel HCCI, alternative fuels, advanced concepts for energy savings, improved thermodynamic cycles, advanced turbocharging, cylinder downsizing or deactivation, hybrid EGR system, variable compression ratio, variable swirl, variable valve actuation, flexible cooling, hybrid diesel powertrain and energy recovery, closed-loop combustion control, and mapless engine controls are the important areas related to engine performance enhancement. Evaluating the impact of new engine technologies on each other and optimally integrating them is the role of engine system design. The question will be 'In order to design a good cost-effective system, do we really need to adopt many of these technologies in one engine? If not, how should we simplify the package?' It is very important to use a system integration approach to study engine performance for future powertrains.

Index