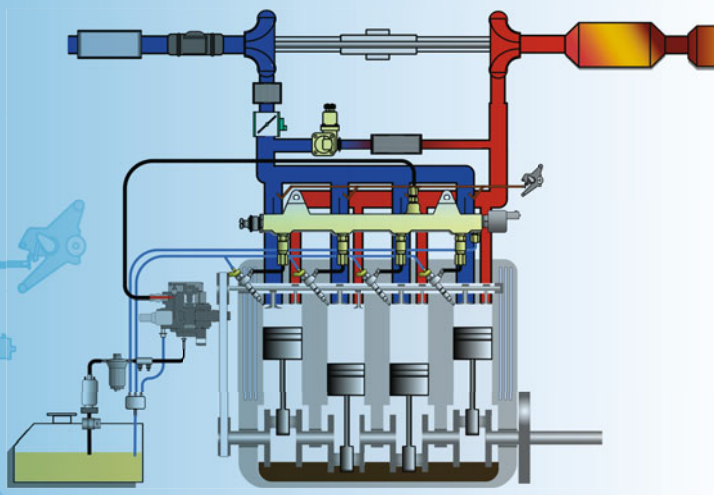
ATZ/MTZ-Fachbuch



Rolf Isermann

Combustion Engine Diagnosis

Model-based Condition Monitoring of Gasoline and Diesel Engines and their Components

Springer Vieweg

ATZ/MTZ-Fachbuch

Die komplexe Technik heutiger Kraftfahrzeuge und Antriebsstränge macht einen immer größer werdenden Fundus an Informationen notwendig, um die Funktion und die Arbeitsweise von Komponenten oder Systemen zu verstehen. Den raschen und sicheren Zugriff auf diese Informationen bietet die Reihe ATZ/MTZ-Fachbuch, welche die zum Verständnis erforderlichen Grundlagen, Daten und Erklärungen anschaulich, systematisch, anwendungsorientiert und aktuell zusammenstellt.
Die Reihe wendet sich an Ingenieure der Kraftfahrzeugentwicklung und Antriebstechnik sowie Studierende, die Nachschlagebedarf haben und im Zusammenhang Fragestellungen ihres Arbeitsfeldes verstehen müssen und an Professoren und Dozenten an Universitäten und Hochschulen mit Schwerpunkt Fahrzeug- und Antriebstechnik. Sie liefert gleichzeitig das theoretische Rüstzeug für das Verständnis wie auch die Anwendungen, wie sie für Gutachter, Forscher und Entwicklungsingenieure in der Automobil- und Zulieferindustrie sowie bei Dienstleistern benötigt werden.

More information about this series at http://www.springer.com/series/12236

Rolf Isermann

Combustion Engine Diagnosis

Model-based Condition Monitoring of Gasoline and Diesel Engines and their Components

Springer Vieweg

Rolf Isermann
Darmstadt, Germany

ATZ/MTZ-Fachbuch
ISBN 978-3-662-49466-0 ISBN 978-3-662-49467-7 (eBook)
DOI 10.1007/978-3-662-49467-7

Library of Congress Control Number: 2017936024

Springer Vieweg

Printed on acid-free paper

This Springer Vieweg imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Germany
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface

During the last decades internal combustion engines have developed to rather complex systems through an increasing number of components and variabilities. Therefore, more actuators, sensors and many more electronic control functions were implemented. This resulted together with constructive and thermodynamic designs in a considerable improvement of performance and reduction of fuel consumption and emissions. However, because of the increased complexity improved monitoring and troubleshooting is a well-known challenge.

The increasing requirements on the coverage and precision of fault detection and diagnosis for internal combustion engines and powertrains therefore need systematic procedures for the development. Compared to the classical on-board diagnosis (OBD) for emission-related faults advanced methods with signal models and process models allow a considerable expansion of the detection and localization performance for engine component faults.

Advanced engine diagnostic methods support also the workshop-based diagnosis (off-board) to find the root causes of faults more quickly and to reduce vehicle downtime. It also opens remote access to be used by service stations and manufacturers. Hence, these methods enable to improve the engines reliability, maintenance and lifetime.

After an introduction to current developments for powertrains, on-board and off-board diagnosis systems and some failure statistics the book gives in Part I a brief survey on advanced supervision, fault detection and diagnosis methods. Part II first describes structures for combustion engine control and diagnosis. Then, model-based diagnosis methods for gasoline and diesel engines are treated for the main components, like the intake system, fuel supply, fuel injection, combustion process, turbocharger, exhaust system and exhaust gas aftertreatment. In general, series production sensors are used.

The fault diagnosis of electrical DC, AC and PMSM motors, electrical, pneumatic and hydraulic actuators is compiled in Part III. Theoretical as well as experimental results from test benches are shown and many experimental studies display the applicability and the diagnosis quality for implemented faults. In the last part IV on fault-tolerant systems, fault-tolerant sensors and actuators are considered.

The book is dedicated to graduate students of mechanical, electrical and mechatronic engineering and computer science and is especially focused for automotive engineers in practice.

The author is grateful to his research associates, who have performed many theoretical and experimental research projects on the subjects of this book, among them S. Leonhardt, C. Ludwig, M. Willimowski, F. Kimmich, A. Schwarte, E. Hartmannshenn, M. Leykauf, S. Clever, C. Eck, A. Sidorow and P. Kessler.

Without their continuous work on new methods and building up and maintaining the combustion-engine test bench, measurement and computer equipment many results of this book would not have been obtained. Great appreciation goes also to our precision mechanics workshop guided by A. Stark.

We also would like to thank the research organization Forschungsgemeinschaft Verbrennungskraftmaschinen (FVV), within the Arbeitsgemeinschaft industrieller Forschungsvereinigungen (AiF) who supported many projects. Several results were obtained in cooperation projects with industrial companies. Among them are Adam Opel AG, GM Europe, and Volkswagen AG. We appreciate these cooperations strongly as they contributed positively to our own research.

Finally, I would like to thank Kerstin Keller, Moritz Neeb, Lisa Hesse and especially Brigitte Hoppe for the laborious and precise text setting, Sandra Schütz for drawing many figures and Springer Verlag for the excellent cooperation.

Darmstadt, February 2017

Rolf Isermann

Contents

List of Symbols

Only frequently used symbols and abbreviations are given:

1. General letter symbols

a	parameters of differential or difference equations
b	parameters of differential or difference equations
c	spring constant, constant, concentration, stiffness
d	damping coefficient
e	equation error, control deviation $e = w - y$, number $e = 2.71828\ldots$
f	fault, frequency ($f = 1/T_\mathrm{p}$, T_p period), function $f(\ldots)$
g	gravitational acceleration, function $g(\ldots)$, impulse response
i	integer, gear ratio, index, $\sqrt{-1}$ (imaginary unit)
j	integer, index
k	discrete number, discrete time $k = t/T_0 = 0, 1, 2, \ldots$ (T_0: sampling time)
l	index, length
m	mass, order number
n	rotational speed, order number, disturbance signal
p	pressure, index, controller parameter, probability density function, process parameter
q	controller parameter
r	index, radius, reference variable, residual
s	Laplace variable $s = \delta + i\omega$, symptom, actuator position
t	continuous time
u	input signal change ΔU
v	specific volume, disturbance signal
w	speed, reference value, setpoint
x	space coordinate, state variable, concentration
y	output signal change ΔY, space coordinate, control variable change ΔY, signal

z	space coordinate, disturbance variable change ΔZ, z-transform variable $z = \exp T_0 s$
$\hat{x}$	estimated or observed variable
$\tilde{x}$	estimation error
$\bar{x}$	average, steady-state value
x_0	amplitude
x_{00}	value in steady state (identification methods)
x_d	desired value
A	area
B	magnetic flux density
C	capacitance
D	damping ratio, diameter
E	module of elasticity, energy, potential, bulk modulus
F	filter transfer function, force
G	weight, transfer function
H	magnetic field strength, height
I	electrical current, mechanical momentum, torsion, second moment of area
J	moment of inertia, loss function
K	constant, gain
L	inductance
M	torque
N	discrete number, windings number
P	power, probability
Q	generalized force, heat
R	electrical resistance, covariance or correlation function
S	spectral density, sum, performance criterion
T	absolute temperature, time constant
T_0	sampling time
U	input variable, manipulated variable (control input), voltage
V	volume
X	space coordinate
Y	output variable, space coordinate, control variable
Z	space coordinate, disturbance variable
$\mathbf{a}$	vector
$\mathbf{A}$	matrix
$\mathbf{A}^\mathrm{T}$	transposed matrix
$\mathbf{I}$	identity matrix
$\boldsymbol{\theta}$	parameter vector
$\mathbf{P}$	covariance matrix
$\boldsymbol{\psi}$	data vector
α	coefficient, angle

β	coefficient, angle
γ	specific weight, correcting factor
δ	decay factor, impulse function
ϕ	correlation function, validity function
η	efficiency
ϑ	temperature
λ	thermal conductivity, forgetting factor, failure rate
μ	friction coefficient, permeability, membership function
ν	kinematic viscosity, index
π	number $\pi = 3.14159\ldots$
ρ	density
σ	standard deviation, σ^2 variance
τ	time
φ	angle
ω	angular frequency, $\omega = 2\pi/T_\mathrm{p}$: T_p period
Δ	change, deviation
θ	parameter
Π	product
Σ	sum

2. General mathematical abbreviations

$\exp(x)$	$= e^x$
$E\{\ldots\}$	expectation of a statistical variable
dim	dimension
adj	adjoint
det	determinant
Re	real part
Im	imaginary part
$\dot{Y}$	dY/dt (first derivative)
var[]	variance
cov[]	covariance
$\mathcal{F}$	Fourier transform
$\mathcal{L}$	Laplace transform
rms (...)	root of mean squared of...

3. Letter symbols for internal combustion engines

3.1 Geometry and time (DIN 1304, ISO 31)

A	area	m^2
a	acceleration	$\mathrm{m/s}^2$
b	breadth, width	m
c	absolute velocity	m/s
d	diameter	m

D	characteristic diameter	m
f	frequency	Hz
g	acceleration of free fall, gravitational acceleration	m/s^2
h	height	m
l	length	m
n	rotational speed	1/s, rpm
r	radius	m
s	actuator position	m
u	peripheral velocity	m/s
V	volume	m^3
v	specific volume	m^3/kg
w	relative velocity	m/s
ω	angular velocity	rad/s

3.2 Mechanics (DIN 1304, ISO 31)

a	specific work	J/kg
m	mass	kg
$\dot{m}$	mass flow rate	kg/s
p	pressure	Pa
C_d	orifice discharge coefficient	1
E	energy	J
F	force	N
J	moment of inertia	kg m^2
L	angular momentum	kg m^2/s
M	torque	Nm
P	power	W
W	work	J
$\nu, \nu = \eta/\rho$	kinematic viscosity	m^2/s
η	dynamic viscosity	Pa s
η	efficiency	1
ρ	mass density	kg/m^3
Π	pressure ratio	1

3.3 Thermodynamics and heat transfer (DIN 1304, ISO 31)

c_p	specific heat capacity at constant pressure	J/(kg K)
c_v	specific heat capacity at constant volume	J/(kg K)
Gr	Grashof number	1
H	enthalpy	J/kg
h	specific enthalpy	J/kg
L	characteristic length	m
n	polytropic exponent	1
Nu	Nusselt number	1
Pr	Prandtl number	1

Q	heat	J
$\dot{Q}$	heat flow rate	W
q	specific heat	J/kg
R	specific gas constant	J/(kg K)
Re	Reynolds number	1
S	entropy	J/K
s	specific entropy	J/(kg K)
T	thermodynamic temperature	K
U	internal energy	J
u	specific internal energy	J/kg
x	mass fraction	1
α	coefficient of heat transfer	W/(m^2 K)
κ	isentropic exponent	1
λ_{th}	thermal conductivity	W/(m K)
ϑ	temperature	K

3.4 Engine specific symbols

α_{ped}	accelerator pedal position	%
b_{f}	fuel consumption	kg/h
b_{sfc}	effective specific fuel consumption	g/kWh
c_{NOx}	nitrogen oxide concentration	g/m^3
c_{op}	opacity	%
c_{pa}	soot concentration	g/m^3
Δt_{pi}	timing of pilot injection	ms
$\Delta\varphi_{\mathrm{pi}}$	crank angle of pilot injection (difference angle to main injection)	°CS
ϵ	compression ratio	1
h	valve lift	m
H_{l}	lower fuel heating value	J/kg
φ	crank angle	°CS
φ_{mi}	crank angle of main injection	°CS
φ_{ign}	ignition angle	°CS
λ	air-fuel ratio (excess air factor)	1
λ_{a}	air expenditure	1
λ_{P}	connecting rod ratio	1
l_{r}	relative filling	1
L_{st}	stoichiometric air requirement	kg/kg
m_{f}	injection mass	mg/cyc
$\dot{m}_{\mathrm{f}}$	injection mass flow	kg/h
m_{mi}	main injection mass	mg/cyc
m_{pi}	pilot injection mass	mg/cyc
$\dot{m}_{\mathrm{eng,in}}$	gas flow into the engine	kg/h
$\dot{m}_{\mathrm{eng,out}}$	gas flow out of the engine	kg/h
m_{air}	air mass per cycle	mg/cyc

M_{aux}	auxiliaries torque	Nm
M_{cyl}	torque of one cylinder	Nm
M_{eng}	crankshaft mean torque at flywheel	Nm
M_{f}	friction torque	Nm
M_{g}	gas force torque	Nm
M_{i}	indicated torque	Nm
M_{l}	load torque	Nm
M_{m}	dynamic masses torque	Nm
M_{drg}	drag torque, motoring torque	Nm
M_{v}	valve train torque	Nm
$\dot{m}_{\mathrm{air}}$	air mass flow	kg/h
q_{f}	injection quantity	mm^3/cyc
q_{mi}	main injection quantity	mm^3/cyc
q_{pi}	pilot injection quantity	mm^3/cyc
r_{egr}	exhaust gas recirculation ratio	1
V_{c}	clearance volume per cylinder	m^3
V_{d}	displaced volume per cylinder	m^3
V_{D}	total displacement (all cylinders)	m^3
z	number of cylinders	1

3.5 Combustion pressure analysis

$dp_{\mathrm{cyl}}/d\varphi$	pressure gradient	Pa/°CS
$dQ_{\mathrm{f}}/d\varphi$	heat release rate	J/°CS
φ_{Q5}	location of mass fraction burned 5%	°CS
φ_{Q50}	location of mass fraction burned 50%	°CS
φ_{Q95}	location of mass fraction burned 95%	°CS
p_{mi}	mean indicated pressure	Pa
p_{m}	motored cylinder pressure (no injection)	Pa
p_{mep}	brake mean effective pressure	Pa
Q_{f}	heat release	J

3.6 Subscripts for internal combustion engines

1	state variables in front of the compressor
3	state variables in front of the turbine (in the exhaust manifold)
4	state variables after the turbine
5	state variables after the DPF
2c	state variables after the compressor
2i	state variables in the intake manifold
2ic	state variables after the intercooler
a	ambient
afi	air filter
afl	air flaps
air	air

b	burned
c	compressor
cam	camshaft
cas	crankshaft
cd	combustion duration
cg	combustion gas
cl	cooling medium, coolant
cool	cooling
cr	common rail
cyl	cylinder
dpf	diesel particulate filter
ds	delivery start
ec, EC	exhaust closes
eff	effective
eg	exhaust gas
egr	exhaust gas recirculation
egrc	EGR cooler
egrv	EGR valve
eng	engine
eo, EO	exhaust opens
es	exhaust system
eth	exhaust throttle valve
ev	exhaust valve
f	fuel, friction
geo	geometrical
H_2O	coolant water
hpegr	high pressure exhaust gas recirculation
hpp	high pressure pump
ic	intake closes
id	injection duration
igd	ignition delay
in	streaming in
int	intake
inj	injectors
io	intake opens
iv	intake valve
lpegr	low pressure exhaust gas recirculation
mc	main combustion
meas	measured
mi	main injection
mv	metering valve
oil	oil
osc	oscillating
out	streaming out
p	piston

pc	pilot combustion
pcv	pressure control valve
pi	pilot injection
r	rod
rail	rail system
red	reduced
rot	rotating
sim	simulated
soc	start of combustion
soi	start of injection
sw	swirl flap
t	turbine
tc	turbocharger
th	throttle
tum	tumble
u	unburned
vac	vacuum system
vgt	variable geometry turbocharger
w	wall
wg	waste gate turbocharger

3.7 Abbreviations for internal combustion engines

AFR	air-to-fuel ratio
ASAM	Association for Standardization of Automation and Measuring Systems
BDC	bottom dead center
CI	compression ignition engine
CR	common rail
DOC	diesel oxidation catalyst
DPF	diesel particulate filter
ECU	electronic control unit
EGR	exhaust gas recirculation
HFM	hot film measurement
ICE	internal combustion engine
MIL	malfunction indicator lamp
NSC	NO_x storage catalyst
OSEK	Offene Systeme und deren Schnittstellen für die Elektronik im Kraftfahrzeug (open systems and their interfaces for the electronics in vehicles)
PF	particulate filter
PM	particulate matter
SCR	selective catalytic reduction
SI	spark ignition engine
TDC	top dead center
VGT	variable geometry turbocharger

VVT	variable valve train

4. Abbreviations for identification and signal-analysis and fault-diagnosis methods

ACF	Auto Correlation Function
APRBS	Amplitude-modulated PRBS
ARMA	Auto Regressive Moving Average process
CCF	Cross Correlation Function
DFT	Discrete Fourier Transform
DSFC	Square root filtering in covariance form
DSFI	Square root filtering in information form
ELS	Extended least squares
ETA	Event Tree Analysis
FDD	Fault Detection and Diagnosis
FDI	Fault Detection and Isolation
FMEA	Failure Mode and Effects Analysis
FFT	Fast Fourier Transform
FTA	Fault Tree Analysis
HA	Hazard Analysis
HCCI	Homogeneous Charge Compression Ignition
LS	Least Squares
MIMO	Multiple Input Single Output
MISO	Multiple Input Multiple Output
MLP	Multilayer Perceptron
MTBF	Mean Time Between Failures
MTTF	Mean Time To Failure = $1/\lambda$
MTTR	Mean Time To Repair
NN	Neural Net
PCA	Principal Component Analysis
PRBS	Pseudo Random Binary Signal
RBF	Radial Basis Function
RLS	Recursive Least Squares
SISO	Single Input Single Output

1

Introduction

The increasing *electrification* and *electronification* is a dominant feature of modern automotive developments. This is demonstrated by an increasing part of electrics/electronics (E/E) of the manufacturing costs from about 20 % in 1995 to 35 % in 2014. The electrics comprise primarily the electrical energy flows to the consumers through the energy board net. Frequently, former mechanical, pneumatic or hydraulic actuated components of chassis and powertrain are replaced by electrical ones. The electronics are primarily used for many control functions. However, they enable also to implement advanced diagnostic functions.

1.1 Combustion engine control and diagnosis developments

Many automotive developments in the last three decades have been possible through an increasing number of *mechatronic components* in the powertrain and the chassis. Figure 1.1.1 gives some examples for engines, drive trains, suspensions, brakes and steering systems. *Mechatronic systems* are characterized by an integration of mechanics and electronics, where the integration is between the components (hardware) and the information-driven functions (software). This development has a considerable influence on the design and operation of the *powertrain* consisting of the combustion engine and the drive train and the *chassis* with suspension, steering and braking systems. In the case of hybrid drives this includes also the electrical motor and the battery.

The mechatronic components replace formerly pure mechanical, hydraulic or pneumatic parts and use sensors with electrical outputs, actuators with electrical inputs and digital electronics for control. The available electrical sensor measurements open the access to internal functions and thus enable new possibilities not only for control but also for fault detection and diagnosis.

In former time the *supervision of internal combustion* engines consisted of limit checking of few variables such as the oil pressure, coolant temperature and board-net voltage. The introduction of analog control in 1967 and digital control in 1979 then opened new ways for on-board checks and electrical workshop testers. In 1988 an

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7_1

on-board diagnosis (OBD) was required by state regulations and emission laws in USA and 2000 in Europe to monitor the emission related engine components. Since then the allowable emission limit values have been reduced by law in periods of 4 to 5 years. This had a great influence on the design of the combustion engines and also on the performance and extent of the on-board diagnostic functions. The combustion engines obtained several *variabilities* such as variable valve trains, variable intake systems, variable exhaust gas recirculation, variable turbo chargers and controlled exhaust gas aftertreatment. Because of the increasing complexity and electronic control functions trouble shooting became a difficult task. *Advanced engine diagnostics* therefore support to find the root causes of failures more quickly and reduce vehicle down-time.

A further development of an advanced fault diagnosis is a *remote access* to detect faults early, guide the customer and prepare maintenance and repair. In addition a computer supported diagnosis system allows the manufacturers to obtain data from the field and to improve the products continuously, see e.g. Alfes et al (2014).

The next sections looks at the *development for gasoline and diesel engines* in more detail together with the legislative emission requirements, see also Isermann (2014).

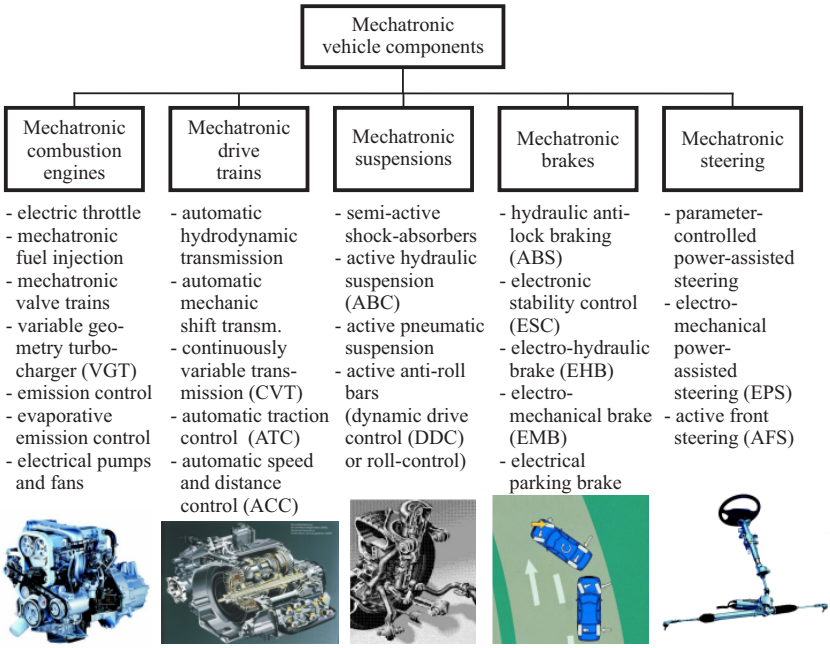


Fig. 1.1.1. Mechatronic components and systems for automobiles and engines.

1.1.1 On the historical development of gasoline engines control

The historical development of gasoline (or spark ignition, SI) engines during the last 50 years with view on their control and diagnosis is depicted in Fig. 1.1.2. Until about 1965 the engines were mechanically controlled with transistor-triggered electromechanical coil ignition. Fuel injection systems for the intake manifold with electronic analog control began to replace the carburetors in 1967. Since about 1970 increasingly more functions are controlled electronically, first with transistor technology. This development required more sensors (knock sensors, air flow and air pressure sensors) with electrical outputs and actuators with electrical inputs (fuel injectors). A large influence on the developments had the state regulations and emission laws, for the United States the Clean Air Act (CARB) in California (1983) and since 1993 for US states in different tiers. This resulted in laws for low emission vehicles (LEV), ultra low emission vehicles (ULEV) and super ultra low emission vehicles (SULEV). The corresponding European regulations are EURO 1 (1992), EURO 2 (1996), EURO 3 (2000), EURO 4 (2005), EURO 5 (2009), and EURO 6 (2014). These regulations were supplemented by the requirements for an on-board diagnosis in the United States OBD I (1988), OBD II (1994) etc. and EOBD (2000) for Europe.

Gasoline engines received catalytic converters with λ-control (1976) and microprocessor control in 1979. The electrical throttle was introduced in 1986, the direct injection about 1999 and since 2000 gasoline engines are supplied with variable valve trains for valve timing and lift control. Present gasoline engines are characterized by electromagnetic or piezoelectric injectors, high-pressure injection pumps (120 bar), homogeneous and stratified combustion, mechanical or turbo charging and increased specific power (downsizing). Figure 1.1.3 depicts the development of sensors and actuators for gasoline engines. Today's SI engines have about 15–25 sensors and 6–8 main manipulated variables and are controlled with a powerful microcomputer control unit (ECU) with 80–120 look-up tables and many control algorithms.

1.1.2 On the historical development of diesel engines control

The historical development of diesel (or compression ignition, CI) engines with regard to their control and diagnosis is shown in Fig. 1.1.4. Around 1960 diesel engines had a swirl chamber, mechanically controlled piston injection pumps and fly-weight overspeed control. Microprocessor control with direct injection and distributor pump (900 bar) and wastegate turbo chargers appeared about 1989. Further steps were exhaust gas recirculation (EGR), oxidation catalyst and turbochargers with variable geometry (1992). First common-rail injection systems with direct injection (1500 bar) with VGT turbochargers reduced further fuel consumption and emissions and resulted in good dynamic torque generation. Today's diesel engines are characterized by high pressure (2000 bar), common-rail multiple injection, piezo-injectors, twin turbochargers or VGT chargers, high EGR rates, DeNO_x-catalyst, particulate filters with regeneration, and selective catalytic reduction (SCR).

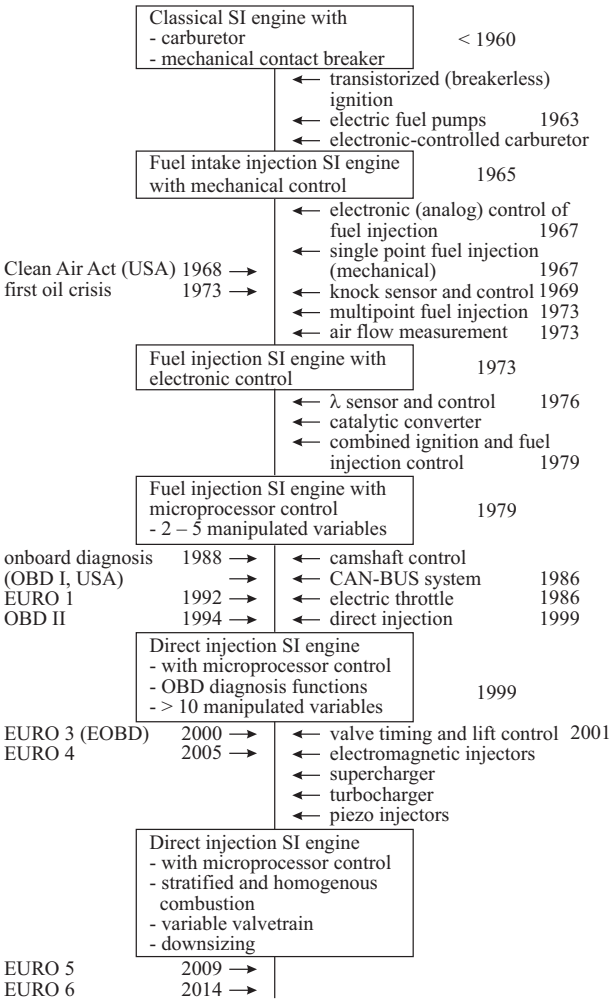


Fig. 1.1.2. Historical development of gasoline engine control and diagnosis.

The development of sensors and actuators for diesel engines is summarized in Fig. 1.1.5. Present diesel engines need about 15–20 sensors, 5–9 main manipulated variables and an ECU with more than 100 look-up tables and many control algorithms.

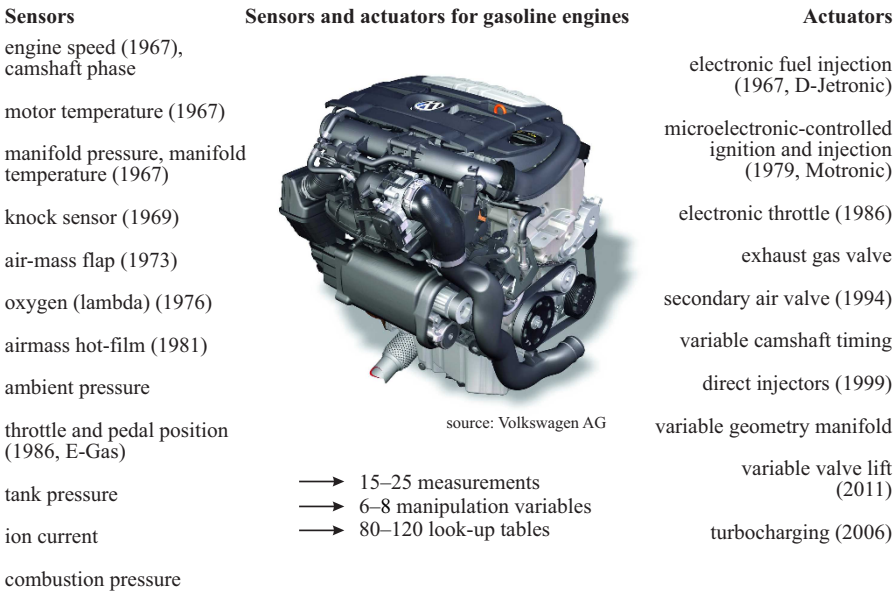


Fig. 1.1.3. Sensors and actuators for gasoline engines (SI).

1.2 Current engine developments

1.2.1 Gasoline engines

Current developments for the further improvement of gasoline engines are, for example, variable valve trains, downsizing and modified combustion processes.

Variable valve trains (VVT: variable valve timing) permit the improvement of the gas exchange. The conventional phase shifting of the inlet valves primarily increases the torque through early or late opening in dependence on torque and speed. In order to reduce the gas flow losses through the throttle the valves require variable timing as well as variable lift. Then the fresh air mass can be controlled by the inlet valves. In addition the residual gases can be influenced by changing the overlapping of inlet and outlet valves to improve the emissions through internal exhaust gas recirculation. Manipulation of the valve lift in two steps or continuously gives more freedom for controlling the load without throttling, see, e.g. van Basshuysen and Schäfer (2004), Braess and Seiffert (2005), Köhler and Flierl (2012). A comparison of different designs of VVT, Schulz and Kulzer (2006), shows that the fuel consumption can be improved with phase actuation by 3–4 %, lift switching by 8–10 %, continuously variable lift by 8–10 %, and with full variable hydraulic or electrical VVT by 14–16 %. However, the complexity is relatively high for the full variable VVT. Therefore, the first mentioned three mechanical VVT‘s are a good compromise.

The reduction of the displacement, i.e. *downsizing* for a given well powered vehicle leads to a smaller specific fuel consumption (less throttling) in part load, as the consumption in the torque-speed diagram shows. However, in order to increase

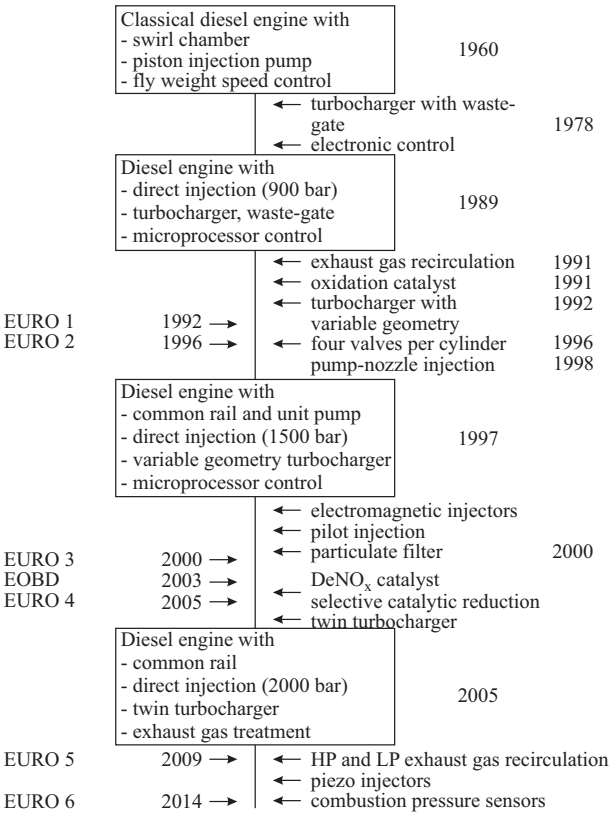


Fig. 1.1.4. Historical development of diesel engine control and diagnosis.

the torque for small speeds and to reach a certain power at higher loads and speeds exhaust turbocharging or supercharging with a mechanical compressor is required. This means, for example, to reduce the displacement from 2 l to 1.3 l and an increase of the mean pressure from 6 to 9 bar. A comparison of different gasoline engines shows that the downsizing factor should be at least 1.3 and should be combined with a change of the operation point to higher torques by increasing the transmission ratio in the drive train (*downspeeding*) to result in a fuel reduction of about 11 %, Königstein et al (2008).

The optimization of the *combustion process* has of course a large influence. Compared to the conventional intake manifold injection and stoichiometric combustion with $\lambda = 1$ and three-way catalyst the direct injection into the cylinders allows considerable saving of fuel consumption for SI engines. Together with a VVT a reduction of 10 % is possible. High pressure injectors (120 bar) with piezoelectric actuation gives a better spraying and makes a stratified, lean combustion with $\lambda > 1$ in part load possible, resulting in about 15 % fuel saving, Weingärtner et al (2007), Berns (2007). A homogeneous charge compression ignition (HCCI) with an increase of the

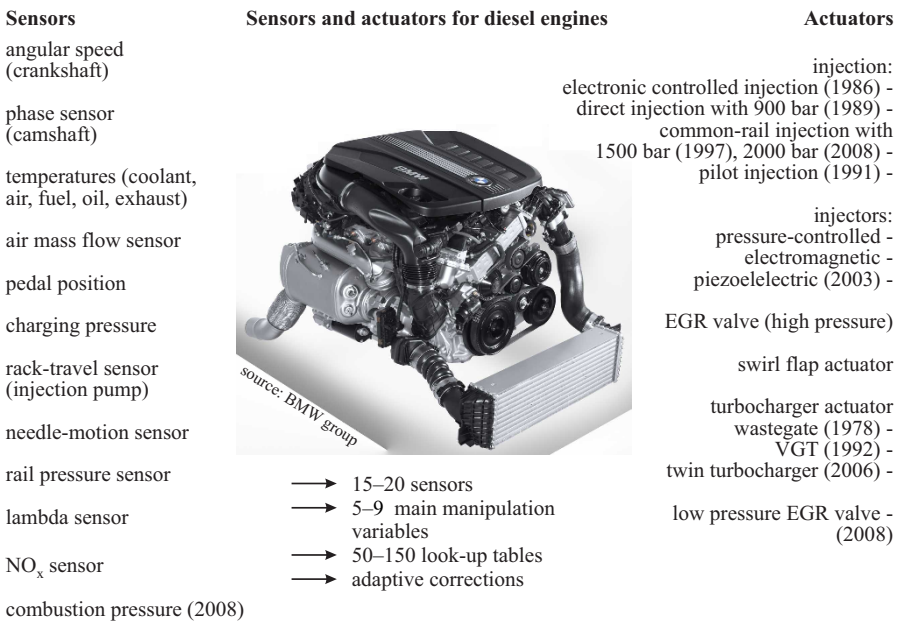


Fig. 1.1.5. Sensors and actuators for diesel engines (CI).

gas temperature by increased residual gases can be obtained, for example, through early closing of the outlet valve and early injection. The combination of early closing the outlet valve and late opening of the inlet valve enables a recompression and a first injection, which can be applied for part load up to 40 %, Alt et al (2008), Backhaus (2008). However, this requires a combustion pressure measurement and control and full variable VVT. A reduction of fuel consumption of about 13–19 % is expected and a NO_x-catalyst becomes unnecessary.

1.2.2 Diesel engines

Of current interest for the further development of diesel engines are a reduction of fuel consumption, NO_x and particulates. This can be reached by further improvements of the common-rail direct injection, combustion processes, charging and exhaust-gas treatment.

Some steps for the *common-rail direct injection* are higher pressures (2200 bar) and multiple injections in order to improve the combustion, emissions and noise. Solenoid and fast piezoelectric injectors allow different combinations of pre-, main- and post-injection pulses. An increase of the exhaust-gas recirculation rate with strong cooling results in low NO_x emissions. However, a too strong increase of the EGR reduces the turbocharger power. Therefore a low pressure EGR after the particulate filter through a cooler to the compressor inlet can be added. Then high EGR rates with a good mixture of fresh air and exhaust-gas and low temperature through an intercooler may lead to a good cylinder filling, Berns et al (2004), Hadler et al

(2008), Körfer et al (2014). This requires several catalysts and regeneration phases and more sensors in the exhaust path, Bauer et al (2007).

A *modification of the combustion process* is the homogeneous compression ignition (HCCI), which can, e.g. be realized by an early injection with a high EGR rate in the part-load area. This leads to a strong reduction of NO_x and particulates. However, it needs a combustion feedback control with combustion pressure measurement because of the narrow possible operation limits and concentration differences in the individual cylinders, see e.g. Alt et al (2008), Backhaus (2008).

The use of *two turbochargers* with a small and large diameter enables an operation with better efficiencies, a high medium charging pressure over a larger speed range and results in improved acceleration at low speeds. The turbochargers are switched with pneumatic flaps, Steinparzer et al (2007). Also diesel engines allow a certain downsizing by increasing the specific power.

Especially large efforts go into the *exhaust aftertreatment*, for example, through oxidation catalyst converters and particulate filters for minimization of CO, HC, NO_x and particulates. An alternative is the selective catalytic reduction (SCR) with the injection of dissolved urea, especially for heavy duty vehicles. The combination of oxidation catalyst, particulate filter, NO_x-storage catalyst and H_2S-catalyst results in a reduction of NO_x by 90 % without additives, however requires model based control and several additional sensors, and three different regeneration cycles, Hadler et al (2008). See also a comparison of different exhaust aftertreatment systems in Samuelsen et al (2014).

Summarizing, *gasoline and diesel engines* show several development lines, to improve the torque generation and to decrease fuel consumption, emissions and noise. Their present development can be characterized by:

- reduction of fuel consumption and CO_2 emissions
- reduction of specific emissions (HC, CO, NO_x, particulates, dust)
- powerful exhaust gas aftertreatment systems
- good driving behavior
- increased specific power (downsizing, charging)
- reduction of friction
- auxiliaries: minimization of energy consumption
- reduction of oscillations and noise.

With regard to the increasing variabilities and control functions the engines are supplied with *mechatronic components*. Figure 1.2.1 depicts some of these components. Examples are the electrical throttle in the intake system, the high pressure common-rail injection system with solenoid or piezoelectric injectors, variable cam-shaft, variable valve trains and variable geometry turbochargers. The mechatronic components can be subdivided in *actuators* with electrical, pneumatic or hydraulic auxiliary energy, switching injection valves and solenoid valves and *electrical drives, pumps and fans*. Some of them are decentralized mechatronic components with local integration of sensors and electronics and some are centrally controlled by the engine control unit (ECU), see e.g. Robert Bosch GmbH (2011), Isermann (2014).

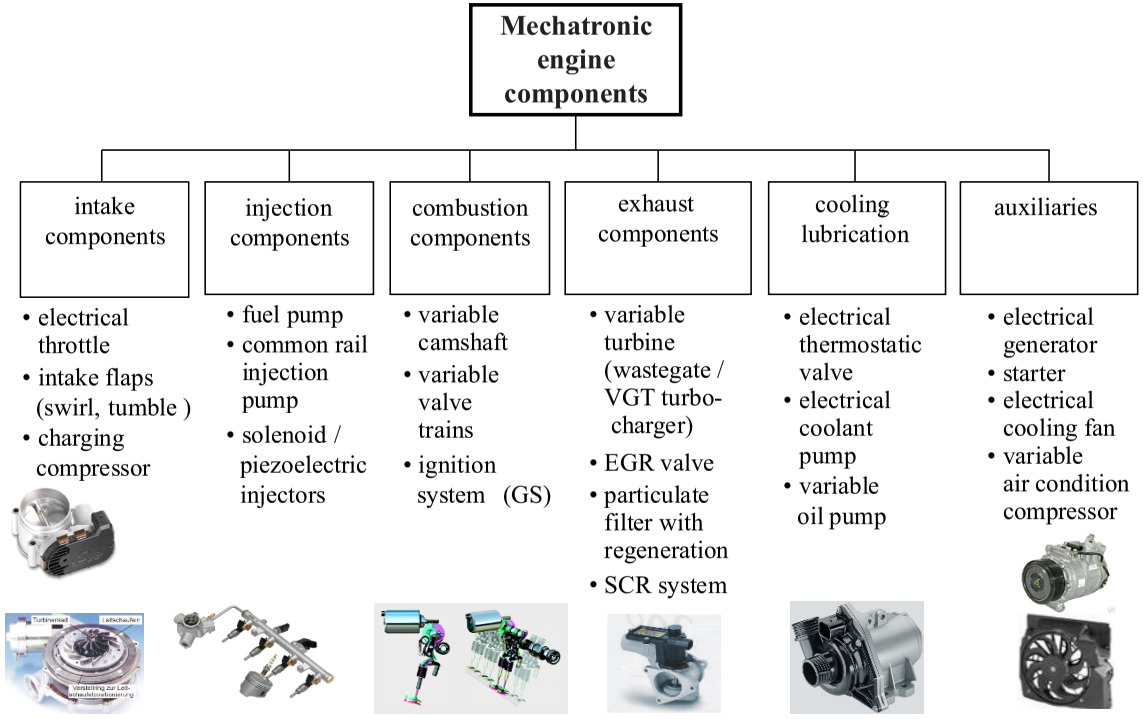


Fig. 1.2.1. Mechatronic components of internal combustion engines.

1.2.3 Alternative drives

A next step of the electrification of vehicles is the development of *hybrid drives*. One reason for this combination of combustion engines and electrical drives is a further saving of fuel consumption and emissions through the operation of the combustion engines in the ranges of better specific fuel consumption, regenerative braking and electrical driving. Micro hybrids have typically a starter/generator with one clutch and start-stop control. Mild and full hybrids in *parallel configuration* are characterized through a stronger starter/generator or a generator/motor with two clutches. *Serial configurations* operate with a combustion engine driving a generator and a separated electric drive. Full hybrids with *power split systems* have a separated generator and electrical motors coupled via a planetary gear. Depending on the hybridization degree the hybrid drives require a high-voltage traction board net, a DC/DC converter for the supply of the chassis board net, power electronics and a very capable battery as energy storage.

All these developments mean a strong increase of electronics and electrics and many control functions and imply a mechatronic overall design. A major influence on the efficiency has the further improvement of specific capacity, lifetime, reliability, safety and cost of the batteries, for example, in the form as NiMH or lithium-ion types. The increase of components also underlines the significance of fault diagnosis functions.

1.3 On-board and off-board diagnosis

The strong increase of sensors, actuators, mechatronic components and electronic control units in modern vehicles have led to a considerable degree of complexity, of both, the power train and the chassis. Therefore these developments were paralleled by implementing supervision and diagnosis functions, both on-board during driving and off-board for service.

The supervision and fault diagnosis performed in the electronic control units can be divided in the diagnosis of electric and electronic (E/E) components and the diagnosis of engine subsystems. The E/E components are attached to the ECUs and can therefore be monitored and diagnosed directly. The fault diagnosis of engine subsystems frequently needs special signal analysis methods or the evaluation of several available signals.

An overview of *supervision* and *fault diagnosis* for automobiles is given in Fig. 1.3.1. The ECUs of the various subsystems have programmed routines for fault detection and, if possible, for fault diagnosis. *Fault detection* means that somewhere in the considered component a fault is detected by measurements. If the fault location and fault size is determined, this is called *fault diagnosis*, see the terminology in the Appendix. Detected or diagnosed faults are continuously stored as a fault code in the non-volatile part of a data memory, together with additional information on operating and environment conditions and on the fault status, e.g. permanent or sporadic. Important faults are displayed to the driver.

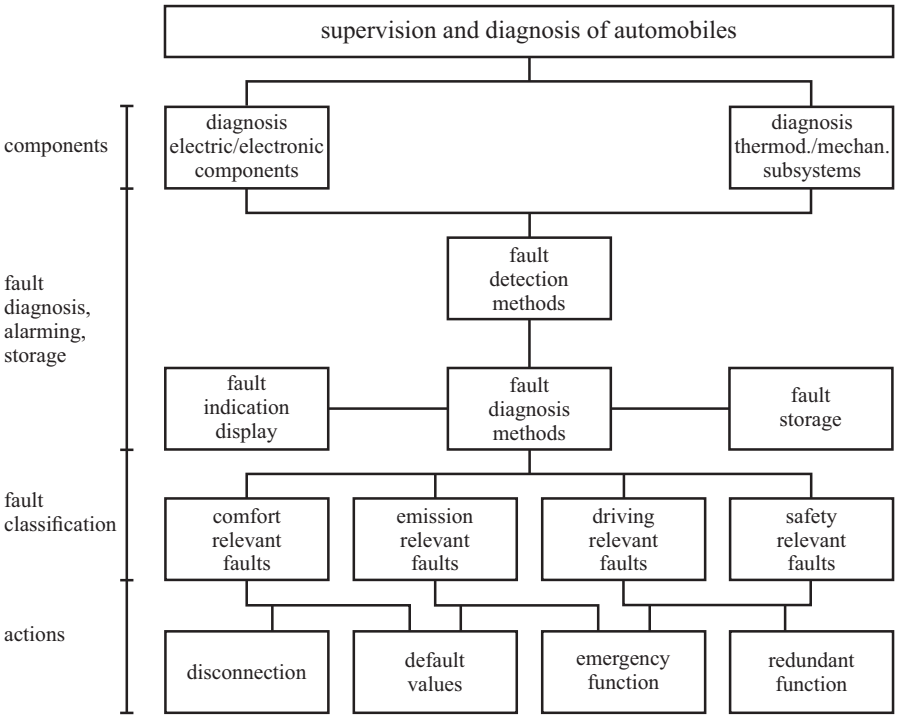


Fig. 1.3.1. Overview of automotive supervision and fault diagnosis, fault classification and relevant actions.

Further actions then depend on the *severity of the faults*. If only the comfort is influenced, the function may be disconnected or default values may be provided (e.g. mirror control). Emission related faults must, according to the OBD legislation, be indicated to the driver after the appearance in a third driving cycle (e.g. misfires or catalyst aging) and default values (e.g. closing of the EGR valve) or emergency functions (e.g. limitation of fuel injection) are provided. A consequence for driving with relevant faults is an *emergency function* (e.g. limp home function with limited fuel injection in the case of a non active electrical throttle or too low turbocharger pressure). Safety-relevant faults are usually covered by a fault-tolerant design in using a redundant component (e.g. double position sensors for the electrical throttle, the steering angle, the acceleration pedal or the two circuit principle of hydraulic brakes).

A survey of conventionally applied methods for fault detection and diagnosis of *E/E components* and vehicle subsystems is depicted in Fig. 1.3.2. The status of sensors, connectors and connecting cables to the ECU are observed by an evaluation of the input signals. Thus short-circuits to the battery voltage or ground or cable interruptions can be detected immediately. The measured value is compared to a

permissible range and the plausibility is checked by means of a comparison with other related measurements. This procedure allows to check the sensor outputs.

Many diagnosis functions are implemented in the ECUs for so called *self-diagnosis*. All components like the microprocessor, the storage and bus systems are continuously checked. Some test procedures are already started when switching to "ignition on" and also after "engine off". Safety relevant microprocessors have a parallel second processor to check the output of the first processor.

The actuators are checked with regard to short-circuits or interruptions, just as sensors. Parameter offsets of the control error in the case of position control, or monitoring of settling times give hints for actuator faults. Also plausibility checks with other measurements and application of test signals and observation of the sign and size of influenced outputs are applied.

The diagnosis of the communication between ECUs and some sensors and actuators is part of the bus protocol. Thus a CAN bus protocol already detects transmission errors. Additionally the time intervals between transmission can be monitored.

The board net including the generator, alternator and the battery is usually monitored by a measurement of the voltage in the engine ECU. Fuses, which protect against short-circuits and generators are usually provided with an overvoltage protection. A practice-oriented description of these applied OBD functions can be found in Rokosch (2006) and Schäffer (2010).

The diagnosis functions of the *vehicle subsystems* can be divided into those for the combustion engine, the transmission and the various chassis systems, like steering, braking, suspensions and tires. The most comprehensive diagnosis system is implemented for the engine and takes up about 50 % of the ECU's software. This is especially due to the always increasing legislative requirements to limit the emissions. The on-board diagnosis OBD I by the Californian Air Resource Board (CARB) in 1988 required that all emission relevant electrical components of an ICE are monitored and its faults are stored in the ECU and displayed by a malfunction indicator lamp (MIL). In 1994 followed the second stage of diagnosis legislation, known as OBD II. Since then all emission related components have to be monitored and the diagnosis system has to be certified for a driving cycle. Other US states followed. A European version is EOBD. It is based on OBD II and is required for passenger cars and light commercial vehicles ($<$3.5 t) with gasoline engines since 2000, with diesel engines since 2003 (EU4) and for heavy commercial vehicles since 2005. The used methods are testing of electrical circuits, limit checks of measured values and plausibility checks.

The fault detection in *automatic hydraulic transmissions* has to be based on measurements of the engine speed, the output shaft speed, selector-lever position and positions of the program selector. Except testing of electrical circuits, limit value checks and plausibility checks are applied. In the case of severe faults a hydraulic emergency operation is provided.

The subsystems of the *chassis* frequently have their own ECUs and individual diagnosis functions. They include the diagnosis routines from the E/E components.

Off-board diagnosis is carried out in service stations with *special workshop testers*. These testers are connected to a diagnosis plug on the driver side and operate

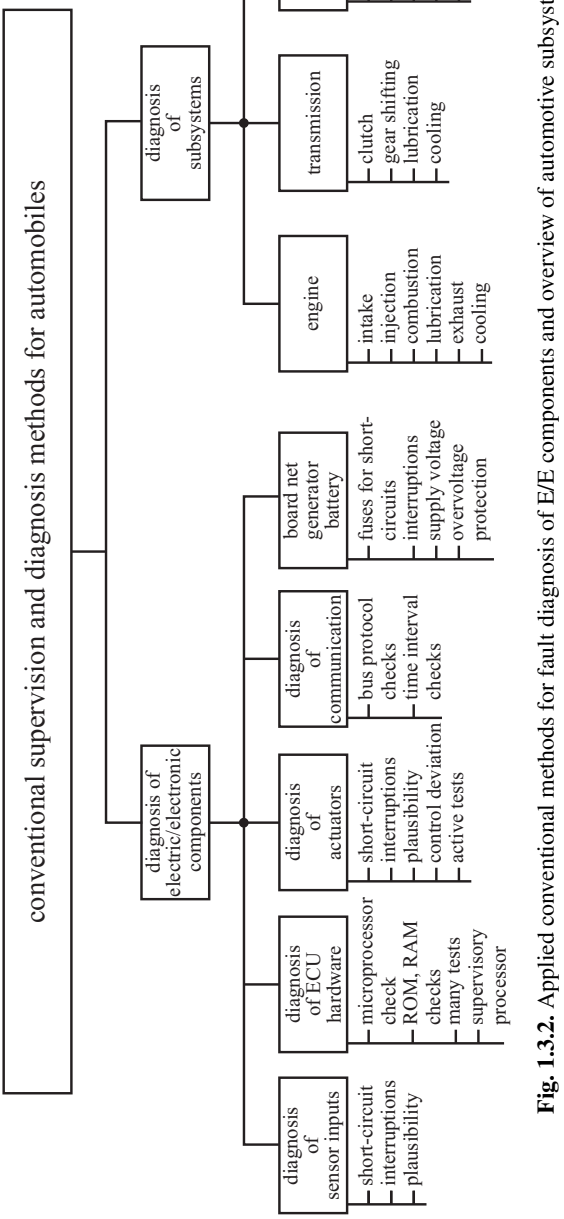


Fig. 1.3.2. Applied conventional methods for fault diagnosis of E/E components and overview of automotive subsystems.

over a serial interface. The communication protocols are standardized in ISO 9141-2 for European passenger cars and SAE J 1850 for American passenger cars. After initiating the ECU, detection of the baud rate and reading of the key bytes of the transmission protocol, the fault memories of different ECU's are read out. The fault storage can be deleted and actuators can be triggered.

A software standardization for the information in vehicle diagnosis systems was introduced by ASAM in 2004 as Open Diagnostic Data Exchange (ODX). It describes the data communication for diagnosis and software programs as a common basis for manufacturers, suppliers, development and after-sales service. ODX is specified with UML (Unified Modeling Language) and uses XML (Extensible Markup Language) for the diagnosis data. Because ASCII-code is used, the data is easily portable. The readable information is also important, see e.g. Marscholik and Subke (2011). According to the ISO 22901-1 the data protocols like UDS (ISO 14229) are standardized. A standardization of the diagnosis functions is in development, and called OTX (Open Test sequence eXchange), see Chap. 3.

1.4 Failure statistics

Failure statistics of automobiles are published from automobile clubs and similar organizations. As an example Table 1.4.1 shows causes of breakdowns of passenger cars. They are based on about 2.5 millions breakdowns and service helps per year for 2005 and stem from the service personnel. About 45.5 % belong to the power train (engine, tank system, exhaust system, clutch and gear), 35 % to general electrics, 12.7 % to the vehicle (chassis 4.2 %, brake system 1.4 %, suspension 0.6 %, steering system 0.3 %) and 6.8 % to others. Most frequent are (and increasing from 2000 to 2005) failures of the electrics, like battery, generator, V-belt, starter blockage, loosened cables or burned fuses. The ignition system showed defects frequently because of the immobilizer (theft protection), ECU's, spark plugs or marten bites. Detected failures of the engine were broken belts or chains of the camshaft, oil pump defects, too less oil and overheating.

Table 1.4.1. Breakdown of passenger cars, ADAC Pannenstatistik 2005, ADAC (2005).

components	general electrics	ignition	motor	cooling system	fuel system	fuel injection	wheels, tires
percentage of failures [%]	35.0	14.6	7.7	5.6	5.5	6.0	6.2

components	clutch gear	chassis	exhaust system	brake system	suspension	steering system	others
percentage of failures [%]	4.1	4.2	2.0	1.4	0.6	0.3	6.8

Table 1.4.2. Failures of the cooling system of passenger cars, ADAC Pannenstatistik 1999, ADAC (2002).

components	V-belt	flexible pipes (heating, cooling)	cylinder head seals	water pump	coolant liquid	thermo-stat	cooler expansion vessel
percentage of failures [%]	29	20	16	10	6	6	6

The cooling system shares about 4–6 % of all vehicle breakdowns. Table 1.4.2 indicates that most of the failures are caused by the V-belt, the flexible pipes for heating and cooling, cylinder head (seals) and water pumps. Through coolant leak detection almost half of all failures can be detected, e.g. by measuring the coolant liquid level.

The statistics for causes of breakdowns and service helps in the year 2011 based on 2.3 million events are summarized in Table 1.4.3 for the power train and in Table 1.4.4 for the vehicle chassis with some more details. The documented percentage of failures is 48.7 % for the power train and 47.4 % for the vehicle chassis. The numbers are similar than that of Table 1.4.1 for 2005. Most failures of the *power train* have their origin in the fuel system (8.8 %), the ignition system (7.9 %) and auxiliary drives (8.6 %). The electronic control system yields 5.5 % and the cooling system (without V-belt) 4.1 %. The exhaust pipe with EGR and turbocharger, the engine hardware and the drive train contribute with about 4 % each.

The largest groups of failures of the *vehicle chassis* stems with 36 % from electrics and electronics with a major part 28.7 % from the battery, see Table 1.4.4. Dominant failures in the suspension system are the tires with 7.5 %, the brake system with 2.3 %, including the brake pads and the parking brake, and in the steering system with 0.3 % the auxiliary power support.

Some of the failures are safety relevant and therefore very important with regard to accidents. For example, hydraulic brakes of passenger cars in Germany (2000) are responsible for about 900 accidents with injuries from a total of about 36000. This means that 2.5 % of all accidents are due to technical brake failures. Accidents during rain, snow and ice make about 48.5 %, Storey (1996), DEKRA Automobil GmbH (2001). The reasons for technical hydraulic brake failures are leakages or gas enclosures. 60 % are due to lacking maintenance (porous flexible brake tubes, corroded brake lines, cut seals at cylinder pistons) and 10 % due to wrong assembling and repair.

These numbers show that the supervision and fault diagnosis of engines and vehicles should be improved. The increased numbers of sensors with electrical outputs and actuators with electrical inputs give a much better access to important variables and therefore chances to apply modern fault detection and diagnosis methods.

The lifetime requirements for the individual components of vehicles are rather different. Table 1.4.5 shows some basic data usually used for reliability considerations and Table 1.4.6 depicts the resulting events per lifetime of some components.

Table 1.4.3. Break down of passenger cars because of power train failures, ADAC (2011) (Basis: 2.239.635 break down events).

	POWER TRAIN : 47.4 %													
component groups	engine hardware			exhaust path, turbocharger			fuel system			ignition system	lubri-cation	drive train		
component	engine mechanics	valve train, belt	cylind. head	EGR	turbo charger	exhaust pipe catalyst	fuel supply	gasoline: throttle carburetor	diesel: inject-ion	ignition coil, plug	oil supply	clutch	trans-mission	drive shaft
percentage of failures [%]	2	1.4	0.8	0.8	1.1	1.7	5.3	1.2	2.3	7.9	0.8	2.2	1.4	0.3
	4.2			3.6			8.8			7.9	0.8	3.9		

component groups	cooling system	auxiliary drives			control system	
component		auxil-iaries	starter	generator	ECU	sensors
percentage of failures [%]	4.1	1.1	3.7	3.8	2.0	3.5
	4.1	8.6			5.5	

Table 1.4.4. Break down of passenger cars because of vehicle failures, ADAC (2011) (Basis: 2.239.635 break down events).

VEHICLE CHASSIS: 48.7 %																			others
component groups	steering system		brake system					suspension system		electrics & electronics					heating & air condi-tioning		chassis general	tank	-
compon-ent	steering mechan-ics	power support	cylind. brake-booster, linkages	brake fluid	mechan-ics brake pad	park-ing brake	ABS ESC	sus-pens-ion	tires	bat-tery	general electrics	ignition lock	immo-bi-lizer	air bag	heat-ing	air condi-tion	body exter-ior	empty tank	-
percentage of failures [%]	0.05	0.28	0.08	0.04	0.65	0.6	0.2	0.5	7.0	28.7	3.5	1.9	1.8	0.1	0.2	0.2	1.1	1.1	-
	0.3		2.3					7.5		36					0.4		1.1	1.1	3.9

Hence, the number of reliable functions per lifetime of a vehicle is about 50 for airbags, 10000 for air conditioning, and 500000 for ABS.

Table 1.4.5. Basic data for vehicle life.

characteristic variables	requirements
time in service	10 years
operation time	3000 hours
mileage (in 10 years)	$150,000$ km
average speed (oer 10 years)	50 km/hour
average number of rides per day (over 10 years, from private up to taxi)	10 times
number of rides (in 10 years)	$50,000$ times
average mileage per ride (over 10 years)	3 km

Table 1.4.6. Estimated number of events per lifetime of some vehicle components.

relevance	example	events per ride	events per life time
safety (rare use)	short circuit, airbag, car alarm	10^{-3}	50
safety	ABS	10	5×10^5
(often use)	indicator	50	2.5×10^6
seasonal dependent	window heating Air conditioning	0.2	10^4
once per starting	door locking, starting relay	1	5×10^4
multiple per ride	lighting, wipers, horn	5	2.5×10^5

1.5 On the contents of this book

The book is divided in three parts. After this introduction, *Part I* gives a short introduction and survey for the various *supervision and fault-diagnosis methods*. After discussion of some terminology basic fault detection methods like limit and trend checking, and plausibility checks are described. Then, advanced signal model and process model-based detection methods like signal analysis, parameter estimation, state observers, state estimators and parity equations are presented in a condensed form. This is followed by approaches for fault diagnosis with classification and inference methods.

Part II is devoted to the *fault diagnosis of internal combustion engines*. Because the diagnosis functions are strongly related to the *engine control functions*, the control structures of gasoline and diesel engines are considered, a subdivision of engine components is provided, the conventional legally required on-board diagnosis (OBD) functions are considered, and a first discussion of model based detection and diagnosis methods for combustion engines is sketched. This is followed by a systematic

treatment of advanced methods for fault detection and *diagnosis of gasoline engines*. The described procedure follows a modular structure for groups of engine components, from the intake air system, fuel supply and ignition system, to the cooling and lubrication system. Several methods were investigated and compared for misfire detection. In all cases the available on-board sensor signals are used and dependent on the kind of faults either signal-analysis methods or process-model-based methods are applied.

Fault detection and *diagnosis of diesel engines* is also based on a modular structure. The considered individual engine components are the intake system, direct fuel injection with distributor pumps, common-rail injection, turbochargers with wastegate or variable geometry, and the exhaust system including particulate filters and catalysts. It is shown for these components how special fault-related features can be generated with periodic signal analysis, parameter estimation and parity equations, which lead to symptoms. These symptoms are then the basis for a fault diagnosis by classification or fuzzy-logic inference methods. The results of these methods are demonstrated by experiments with artificially implemented faults on test benches.

Part III describes results for the *fault detection and diagnosis of electrical drives and actuators* in form of DC, AC and PMSM motors and of electrical, pneumatic and hydraulic actuators, which are used at several places of the engines and the chassis. In the last part *IV* on *fault-tolerant systems* fault-tolerant sensors and actuators are considered. The used terminology for fault detection and diagnosis is summarized in the *Appendix*.

The treated topics in this book enable a progress as follows:

- considerable expansion of engine fault diagnosis compared to conventional OBD methods
- systematic procedures for fault detection by applying different process-model-based and signal-based methods
- modular structure of groups for the diagnosis of engine parts
- fault diagnosis with classification or rule-based inference methods
- validation with implemented faults on test benches
- detection and localization of component, actuator and sensor faults to find faults more quickly and to reduce vehicle downtime
- support for trouble-shooting in service stations and during manufacturing
- basic measurement evaluation and information fusion for
  - online, off-line and remote diagnosis
  - maintenance on demand
- improvement of reliability, availability and lifetime of combustion engines
- design of fault-tolerant actuators, sensors and component parts with regard to safety-related applications
- applications for gasoline and diesel engines in passenger cars, commercial vehicles, trucks, trains, ships and aircraft.

References

ADAC (2002) Pannenstatistik 1999. ADAC e.V., München
ADAC (2005) Pannenstatistik. ADAC e.V., München
ADAC (2011) Pannenstatistik. ADAC e.V., München
Alfes S, Von Querfurth A, Raiser H, Niehaus F (2014) Interactive flexible remote-diagnosis. ATZelektronik worldwide 9(4):46–49
Alt M, Grebe U, Dulzo J, Chang MF (2008) Closed loop combustion control for HCCI. In: 8. Int. Stuttgarter Symposium, Stuttgart
Backhaus R (2008) Gasoline engines with HCCI technique. Motortechnische Zeitschrift – MTZ 69(6):482–485
van Basshuysen R, Schäfer F (2004) Internal combustion engine handbook: basics, components, systems, and perspectives. SAE International, Warrendale
Bauer S, Zhang H, Pfeifer A, Wenzlawski K (2007) PKW-Dieselmotor und EU 6: Ganzheitlicher Systemansatz. In: 28. Int. Wiener Motorensymposium, Wien, Österreich
Berns S (2007) The gasoline challenge. In: Gasoline Systems Symposium, Bosch, Frankfurt, Germany
Berns S, Fehrmann R, Sassen K (2004) Beitrag der Diesel-Elektronik zur Leistungssteigerung und Emissionsreduzierung. In: VDI/VDE-Fachtagung AUTOREG 2004, VDI-Bericht 1828, VDI, Düsseldorf, Wiesloch, Germany
Braess H, Seiffert H (eds) (2005) Handbook of automotive engineering. SAE International, Warrendale
DEKRA Automobil GmbH (2001) Technische Sicherheit und Mobilität. Bericht über Untersuchungen von Fahrzeugen auf unfallrelevenate technische Mängel im Zeitraum 1996–2000. DEKRA Fachzeitschrift (55/01)
Hadler J, Rudolph F, Dorenkamp R, Stehr H, Hilzendeger J, Kranzusch S (2008) Volkswagen's new 2.0 l TDI engine for the most stringent emission standards. MTZ worldwide edition 69(5, 6):12–12, 54–59
Isermann R (2014) Engine Modeling and Control. Springer, Berlin
Köhler E, Flierl R (2012) Verbrennungsmotoren, 6th edn. Vieweg, Wiesbaden
Königstein A, Grebe U, Wu K, Larsson P (2008) Differentiated analysis of downsizing concepts. MTZ worldwide edition 69(6):4–11
Körfer T, Schnorbus T, Miccio M, Schaub J (2014) Emission-based EGR strategies in diesel engines for RDE requirements. MTZ worldwide 75(9):10–17
Marscholik C, Subke P (2011) Datenkommunikation im Automobil: Grundlagen, Bussysteme, Protokolle und Anwendungen, 2nd edn. VDE, Offenbach
Robert Bosch GmbH (ed) (2011) Automotive Handbook, 8th edn. Bentley publishers, Cambridge
Rokosch U (2006) On-Board-Diagnose und moderne Abgasnachbehandlung. Vogel Buchverlag, Würzburg
Samuelsen D, Gerhardt J, Lüders H, Scherer S, Strobel M (2014) With diesel into the age of super ultra low emission vehicles. In: 14. Internationales Stuttgarter Symposium, Springer, pp 1401–1428
Schäffer F (2010) Fahrzeugdiagnose mit OBD, 3rd edn. Elektor-Verlag, Aachen

Schulz T, Kulzer A (2006) Überblick und Bewertung variabler Ventiltriebe. In: VDI/VDE-GMA-Fachausschuss AUTOREG

Steinparzer F, Mattes W, Nefischer P, Steinmayr T (2007) The new BMW four-cylinder diesel engine. MTZ worldwide edition 68(11 and 12):932–943 and 24–27

Storey N (1996) Safety-critical computer systems. Addison Wesley Longman Ltd., Essex

Weingärtner R, Achleitner E, Bächer E (2007) Piezo Direkteinspritzung für Benzinmotoren. In: VDI/VDE-GMA-Fachausschuss AUTOREG

Part I

Supervision, Fault Detection and Diagnosis Methods

2

Supervision, fault-detection and fault-diagnosis methods – a short introduction

Fault detection and fault diagnosis belong to the general area of supervision or condition monitoring of technical processes. The task is to indicate undesired or not permitted states and to take appropriate actions to avoid damage, accidents or environmental pollution. This includes states of the powertrain and the chassis which are dangerous (e.g. spinning vehicle), not economic (e.g. high fuel consumption or high wear and tear) or not ecologic (e.g. too high emissions).

The basic tasks of supervision, fault detection and fault management are described in detail in Isermann (2006) and Isermann (2011). Therefore, only some remarks are given here as an introduction and basis for a systematic approach of automotive diagnosis.

2.1 Basic tasks of supervision

A process (engine or chassis) which operates in open loop is considered, Fig. 2.1.1a). $U(t)$ and $Y(t)$ are input and output signals, respectively. Due to external or internal causes a fault can now appear. Examples for external causes are environmental influences like humidity, dust, chemicals, electromagnetic radiation, high temperature, leading, e.g. to corrosion or pollution. Examples for internal causes are missing lubrication and therefore higher friction or wear, overheating, leaks, and short circuits. These faults $\mathbf{F}(t)$ firstly affect internal process parameters $\boldsymbol{\Theta}$ by $\Delta\boldsymbol{\Theta}(t)$ like changes of resistance, capacitance or stiffness and/or internal state variables $\mathbf{x}(t)$ by $\Delta\mathbf{x}(t)$ like changes of mass flows, currents or temperatures, which are frequently not measurable. According to the dynamic process transfer behavior, these faults $F(t)$ influence the measurable output $Y(t)$ by a change $\Delta Y(t)$. However, it has to be taken into account that also natural process disturbances and noise $N(t)$ and also changes of the manipulated variable $U(t)$ influence the output $Y(t)$.

For a process operating in open loop a remaining fault $f(t)$ (part of $\mathbf{F}(t)$) generally results in a permanent offset $\Delta Y(t)$. In the case of a *closed loop*, Fig. 2.1.1b), the behavior is different. Depending on the time history of parameter changes $\Delta\boldsymbol{\Theta}(t)$ or

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7_2

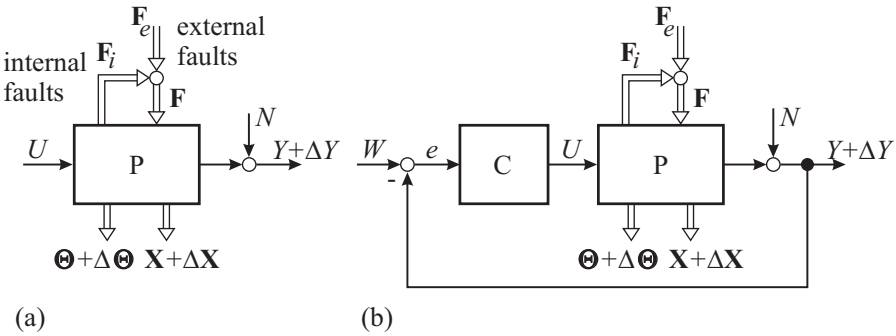


Fig. 2.1.1. Scheme of a process (engine or chassis) P influenced by faults $\mathbf{F}$. **a** process in open loop, **b** process in closed loop.

state-variable changes $\Delta\mathbf{x}(t)$ the output shows only a somewhat shorter and vanishing small deviation $\Delta Y(t)$ if a controller with integral behavior (e.g. a PI-controller) is used. But then the manipulated variable shows a permanent offset $\Delta U(t)$ for proportionally acting processes. If only the output $Y(t)$ is supervised, the fault may not be detected because of the small and short deviation, furthermore corrupted by noise. The reason is that a closed loop is not only able to compensate for disturbances $N(t)$ but also to compensate for parameter changes $\Delta\boldsymbol{\Theta}(t)$ and state changes $\Delta\mathbf{x}(t)$ with regard to the control variable $Y(t)$. This means that faults $\mathbf{F}(t)$ may be compensated by the closed loop. Only if the fault grows in size and causes the manipulated variable to reach a restriction value (saturation) a permanent deviation ΔY may arise. Hence, for processes in closed loop the process input $U(t)$ should be monitored, as well as $Y(t)$.

The *supervision* of technical processes in normal operation or the *quality control* of products in manufacturing is usually performed by *limit checking* or *threshold checking* of some few measurable output variables $Y(t)$, like pressures, forces, liquid levels, temperatures, speeds, and oscillations. This means one checks if the quantities are within a tolerance zone $Y_{\min} < Y(t) < Y_{\max}$. If the tolerance zone is exceeded, an alarm is raised. This is also called *monitoring*.

2.2 Knowledge-based fault detection and diagnosis

As fault detection and fault diagnosis are fundamental for advanced methods of supervision and fault management, these tasks will be considered briefly. Fault detection and diagnosis, in general, are based on *measured variables* by instruments and *observed variables* and states by human technicians. The automatic processing of measured variables for fault detection requires analytical process knowledge and the evaluation of observed variables requires human expert knowledge, which is called *heuristic knowledge*. Therefore, fault detection and diagnosis can be considered within a knowledge-based approach. Figure 2.2.1 shows an overall scheme.

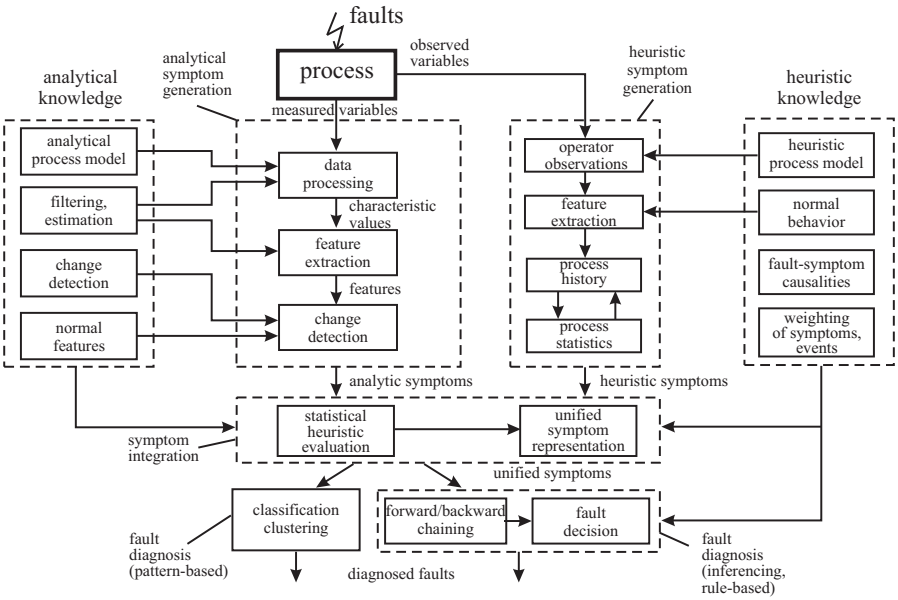


Fig. 2.2.1. Overall scheme of knowledge-based fault detection and diagnosis.

2.2.1 Analytic symptom generation

The analytical knowledge about the process is used to produce quantifiable, analytical information. To do this, data processing based on measured process variables has to be performed to generate first the characteristic values by:

- *limit value checking* of direct, measurable signals. The characteristic values are the violated signal tolerances
- *signal analysis* of directly measurable signals by the use of signal models like correlation functions, frequency spectra, autoregressive moving average (ARMA) and/or the characteristic values, for example, variances, amplitudes, frequencies or model parameters
- *process analysis* by using mathematical process models together with parameter estimation, state estimation and parity equation methods. The characteristic values are parameters, state variables or residuals.

In some cases, special features can then be extracted from these characteristic values, for example, physically defined process coefficients, or special filtered or transformed residuals. These features are then compared with the normal features of the non-faulty process. For this, methods of change detection and classification are applied. The resulting changes (discrepancies) in the mentioned directly measured signals, signal models or process models are considered as *analytic symptoms*. Figure 2.2.2 gives a more detailed survey of analytical fault-detection methods, which are treated in following subchapters.

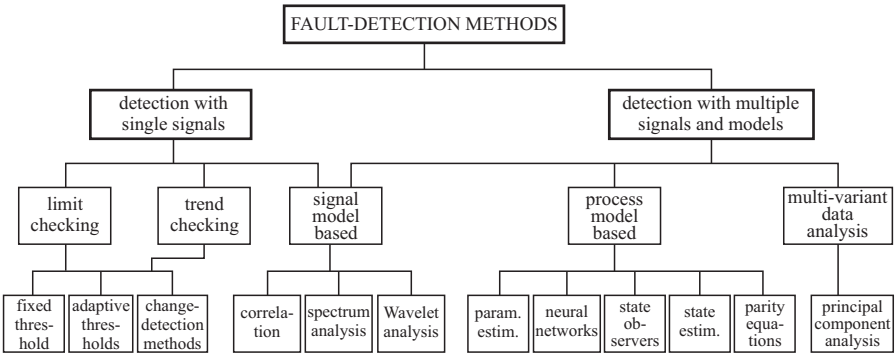


Fig. 2.2.2. Survey of analytical fault-detection methods.

2.2.2 Heuristic symptom generation

In addition to the symptom generation using quantifiable information, heuristic symptoms can be produced by using *qualitative information* from human drivers or workshop personnel. Through human observation and inspection, heuristic characteristic values in the form of special noises, colors, smells, vibration, wear and tear, etc., are obtained. The process history expressed through performed maintenance, repairs, former faults, life-time and load measures, constitutes a further source of heuristic information. Statistical data (e.g. MTTF, fault probabilities) achieved from experience with the same or similar processes can be added. In this way heuristic symptoms are generated, which can be represented as linguistic variables (e.g. small, medium, large) or as vague numbers (e.g. around a certain value), compare Fig. 2.2.1.

2.2.3 Fault diagnosis

The task of fault diagnosis consists in determining the type, size and location of the most possible fault, as well as its time of detection.

Fault-diagnosis procedures use the analytic and heuristic symptoms. Therefore they should be presented in an unified form like confidence numbers, membership functions of fuzzy sets or probability density functions after a statistical evaluation over some time. Then, *classification methods* can be applied if a learned pattern-based procedure is preferred, to determine the faults from symptom patterns or clusters. If, however, more information of fault-symptom relations, e.g. in the form of logic fault-symptom trees or if-then rules are known, *inference methods* (reasoning methods) with forward and backward chaining can be applied, see Isermann (2006). Figure 2.2.3 gives a survey of these methods.

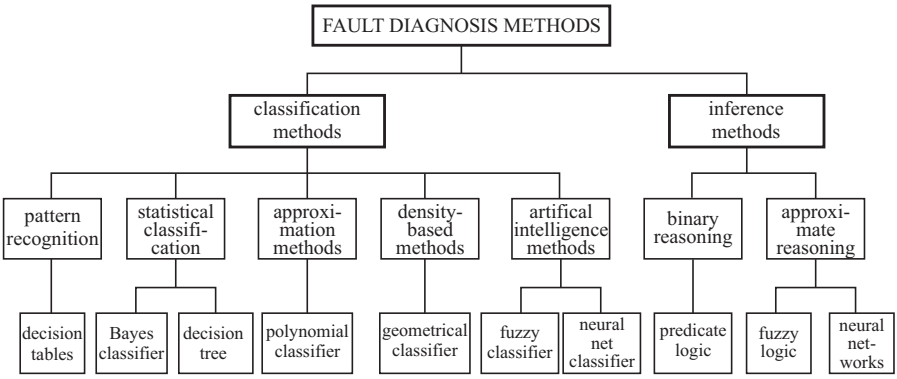


Fig. 2.2.3. Survey of fault-diagnosis methods.

2.3 Signal-based fault-detection methods

A fault detection based on single signal measurements is in simple cases performed with limit checking or trend checking, or, in more complex cases, by operating with special signal models, extraction of special signal features and change-detection methods, see Fig. 2.2.2.

2.3.1 Limit checking of absolute values

Generally, two limit values, called thresholds, are preset, a maximal value $Y_{\max}$ and a minimal value $Y_{\min}$. A normal state is when

$$Y_{\min} < Y(t) < Y_{\max}, \tag{2.3.1}$$

which means that the process is in normal situation if the monitored variable stays within a certain tolerance zone. The exceedance of one of the thresholds then indicates a fault somewhere in the process. This simple method is applied in almost all process automation systems. Examples are the oil pressure (lower limit) or the coolant water (higher limit) of combustion engines or the control error of a control loop. The thresholds are mostly selected based on experience and represent a compromise. On one hand false alarms through normal fluctuations of the variable should be avoided; on the other, faulty deviations should be detected early. Therefore a trade-off between too narrow and too wide thresholds exists.

2.3.2 Trend checking

A further simple possibility is to calculate the first derivative $\dot{Y} = dY(t)/dt$, the trend of the monitored variable and to check if

$$\dot{Y}_{\min} < \dot{Y}(t) < \dot{Y}_{\max}. \tag{2.3.2}$$

If relatively small thresholds are selected, an alarm can be obtained earlier than for limit checking of the absolute value. Trend checking is, for example, applied for oil pressures and vibrations of oil bearings of turbines or for wear measures of machines.

Limit checking of absolute values and trends can also be combined, see Isermann (2006).

2.3.3 Plausibility checks

A rough supervision of measured variables is sometimes performed by checking the plausibility of its indicated values. This means that the measurements are evaluated with regard to credible, convincing values and their compatibility among each other. Therefore, a *single measurement* is examined to find whether the sign is correct and the value is within certain limits. This is also a limit check, however, with usually wide tolerances. If *several measurements* are available for the same process then the measurements can be related to each other with regard to their normal ranges by using logic rules, like

$$\text{IF } [Y_{1\min} < Y_1(t) < Y_{1\max}] \text{ THEN } [Y_{2\min} < Y_2(t) < Y_{2\max}]\,. \tag{2.3.3}$$

For example, one expects for a centrifugal pump with rotating speed n and pressure p

$$\text{IF } [1000 \text{ rpm} < n < 3000 \text{ rpm}] \text{ THEN } [3 \text{ bar } < p < 8 \text{ bar}]\,.$$

The plausibility check can also be made dependent on the operating condition, like

$$\text{IF } [\text{Operating condition 1}] \text{ THEN } [Y_{3\min} < Y_3(t) < Y_{3\max}]\,. \tag{2.3.4}$$

One example is the oil pressure p_{oil} of a combustion engine with speed n and coolant temperature ϑ_{H20}:

$$\text{IF } [n < 800\,\text{rpm}] \text{ AND } [\vartheta_{\text{H20}} < 50^{\circ}\text{C}] \text{ THEN } [3\,\text{bar} < p_{oil} < 5\,\text{bar}]\,. \tag{2.3.5}$$

Hence, plausibility checks may be formulated by using rules with binary logic connections like AND, OR. These rules and ranges of the measurements allow a rough description of the expected behavior of the process under normal conditions. If these rules are not satisfied either the process or the measurements are faulty. Then, one needs further testing to localize the fault and its cause.

These plausibility checks presuppose the ranges of measured process variables under certain operating conditions and represent rough process models. If the ranges of the variables are increasingly made smaller, many rules would be required to describe the process behavior. Then, it may be better to use mathematical process models in the form of equations to detect abnormalities. Therefore, plausibility tests can be seen as a first step towards model-based fault-detection methods.

2.3.4 Signal-analysis methods

Many measured signals of processes and especially engines show oscillations that are either of harmonic or stochastic nature, or both. If changes of these signals are related to faults in the actuators, the process and sensors, signal-model-based fault-detection methods can be applied. Especially for machine vibration, the measurement of position, speed or acceleration allows one to detect, for example, imbalance or bearing faults (turbo chargers), knocking (gasoline engines) and surging (drive train). But also signals from many other engine sensors, like speed, flow and pressure, contain oscillations with a variety of higher frequencies than the engine rotational speed.

The task of fault detection by the analysis of signal models is summarized in Fig. 2.3.1. By assuming special mathematical models for the measured signal, suitable features are calculated, for example, amplitudes, phases, spectrum frequencies and correlation functions for a certain frequency bandwidth $\omega_{min} \leq \omega \leq \omega_{max}$ of the signal. A comparison with the observed features for normal behavior provides changes of the features which then are considered as *analytical symptoms*.

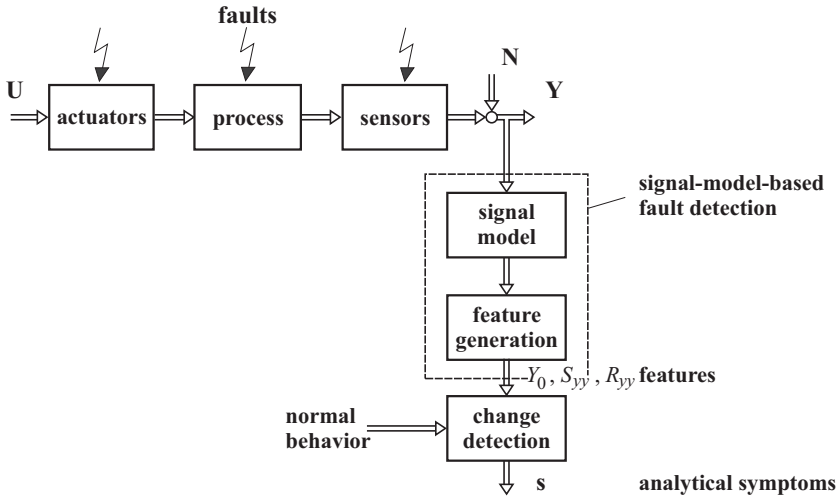


Fig. 2.3.1. Scheme for the fault detection with signal models.

The signal models can be divided into nonparametric models, like frequency spectra or correlation functions, or parametric models, like amplitudes for distinct frequencies or ARMA-type models. Signal-analysis methods exist for harmonic oscillations, stochastic signals and non-stationary signals, compare the scheme of Fig. 2.3.2.

For the analysis of stationary periodic signals band pass filtering or Fourier analysis can be used. Non-stationary periodic signals may be analyzed with, e.g. wavelet transforms. The analysis of stochastic signals is frequently performed by correlation functions, spectrum analysis and signal parameter estimation for ARMA models.

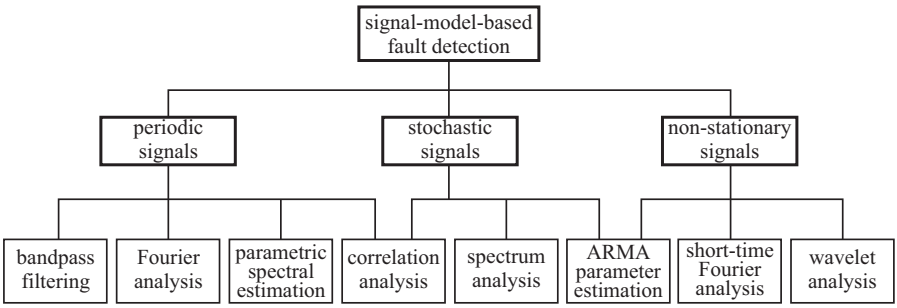


Fig. 2.3.2. Survey of signal-analysis methods for signal-model-based fault detection.

2.4 Process-model-based fault-detection methods

Different approaches for fault detection using mathematical models have been developed in the last decades (see, e.g. Willsky (1976), Himmelblau (1978), Isermann (1984), Isermann and Freyermuth (1991), Isermann (1997), Gertler (1998), Frank (1990), Chen and Patton (1999), Patton et al (2000)), Blanke et al (2006). The task consists of the detection of faults in the processes, actuators and sensors by using the *dependencies between different measurable signals*. These dependencies are expressed by mathematical process models. Figure 2.4.1 shows the basic structure of model-based fault detection. Based on measured input signals $\mathbf{U}$ and output signals $\mathbf{Y}$ the detection methods generate residuals $\mathbf{r}$, parameter estimates $\hat{\boldsymbol{\Theta}}$, or state estimates $\hat{\mathbf{x}}$, which are called features. By comparison with the normal features (nominal values) changes of features are detected, leading to analytical symptoms $\mathbf{s}$.

For the application of model-based fault-detection methods, the process configurations according to Fig. 2.4.2 have to be distinguished. With regard to the inherent dependencies used for fault detection, and the possibilities for distinguishing between different faults, the situation improves greatly from case a) to b) or c) or d), by the availability of some more measurements.

2.4.1 Process models and fault modeling

A fault is defined as an unpermitted deviation of at least one characteristic property of a variable from an acceptable behavior. Therefore, the fault is a state that may lead to a malfunction or failure of the system. The time dependency of faults can be distinguished, see Fig. 2.4.3, as abrupt fault (stepwise), incipient fault (drift-like), or intermittent (sporadic) fault. With regard to the process models, the faults can be further classified. According to Fig. 2.4.4 *additive faults* influence a variable Y by an addition of the fault f, and *multiplicative faults* by the product of another variable U with f. Additive faults appear, e.g., as offsets of sensors, whereas multiplicative faults are parameter changes within a process.

Now, lumped-parameter processes are considered, which operate in open loop. The *static behavior* (steady states) is frequently expressed by a nonlinear characteristic as shown in Table 2.4.1. Changes of parameters β_{i} can be obtained by parameter

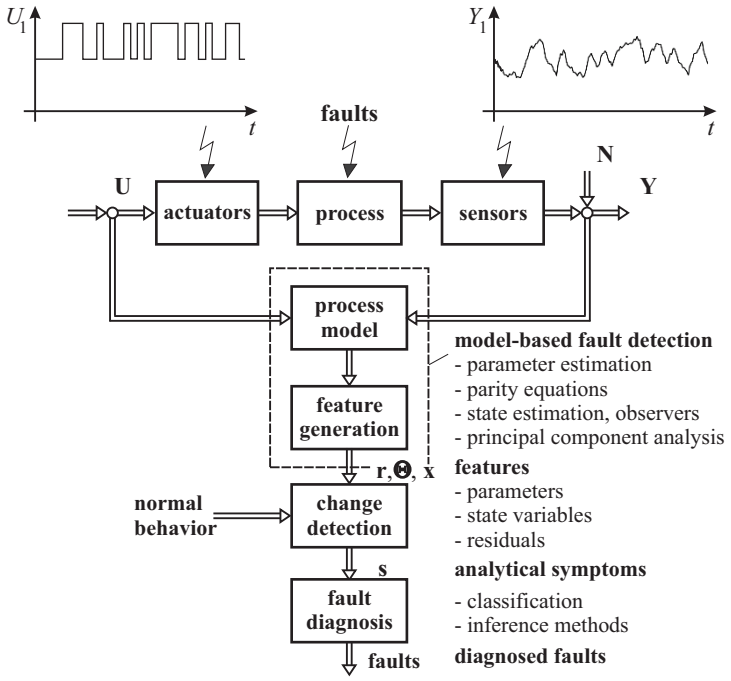


Fig. 2.4.1. General scheme of process-model-based fault detection and diagnosis.

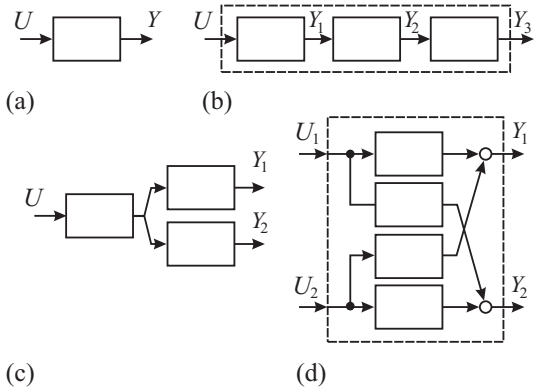


Fig. 2.4.2. Process configuration for model-based fault detection. **a** SISO (single-input single-output); **b** SISO with intermediate measurements; **c** SIMO (single-input multi-output); **d** MIMO (multi-input multi-output).

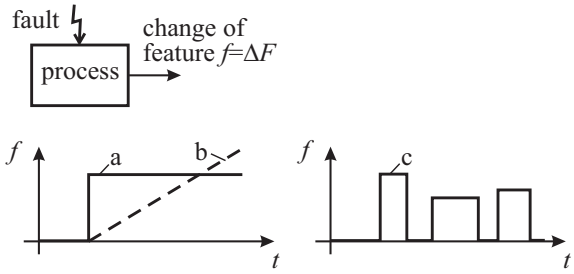


Fig. 2.4.3. Time dependency of faults. **a** abrupt, **b** incipient, **c** intermittent.

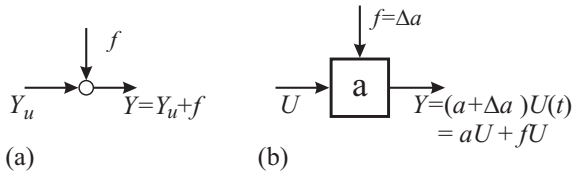
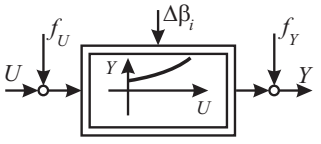


Fig. 2.4.4. Basic models of faults. **a** additive fault, **b** multiplicative faults.

estimation with, e.g., methods of least squares, based on measurements of different input output pairs $[Y_j, U_j]$. This method is applicable for, e.g. valves, pumps, drives, and engines. More information on the process can usually be obtained with dynamic process models. Table 2.4.2 shows the basic input/output models in the form of a differential equation or a state-space model as vector differential equation. Similar representations hold for nonlinear processes and for multi-input multi-output processes, also in discrete time.

Table 2.4.1. Fault detection of a nonlinear static process via parameter estimation for steady states.



Measured signals: $U(t)$, $Y(t)$

Basic equation:

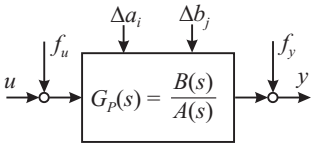
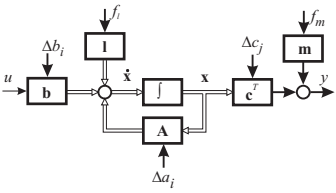
$$Y = \beta_0 + \beta_1 U + \beta_2 U^2 + ... + \beta_q U^q \rightarrow Y = \boldsymbol{\psi}_S^T \boldsymbol{\Theta}_S$$

$$\boldsymbol{\Theta}_S^T = [\beta_0 \beta_1 ... \beta_q] \qquad \boldsymbol{\psi}_S^T = [1\ U\ U^2 ... U^q]$$

Additive faults: f_U input fault; f_Y output fault

Multiplicative faults: $\Delta\beta_i$ parameter faults

Table 2.4.2. Linear dynamic process models and fault modeling.

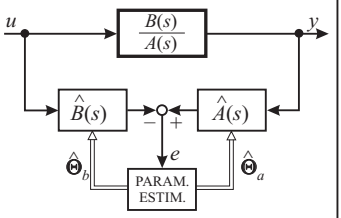
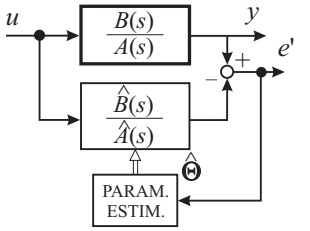
Input/output model	State-space model
 Measured signals: $y(t) = Y(t) - Y_{00}$; $u(t) = U(t) - U_{00}$	
Basic equations: $y(t) + a_1 y^{(1)}(t) + ... + a_n y^{(n)}(t) = b_0 u(t) + b_1 u^{(1)}(t) + ... + b_m u^{(m)}(t)$	$\dot{\mathbf{x}}(t) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{b}u(t)$ $y(t) = \mathbf{c}^T\,\mathbf{x}(t)$
$y(t) = \boldsymbol{\psi}^T(t)\,\boldsymbol{\Theta}$ $\boldsymbol{\Theta}^T = [\, a_1 ... a_n\, b_0 ... b_m \,]$ $\boldsymbol{\psi}^T = [\, -y^{(1)}(t) ... -y^{(n)}(t)\ u(t) ... u^{(m)}(t) \,]$	$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -a_1 \\ 1 & 0 & -a_2 \\ \vdots & \vdots & \vdots \end{bmatrix}$ $\mathbf{b}^T = [\, b_0 \quad b_1 \quad \cdots \,]$ $\mathbf{c}^T = [\, 0 \quad 0 \quad \cdots \quad 1 \,]$
Additive faults: f_u input fault; f_y output fault	f_l input or state-variable fault f_m output fault
Multiplicative faults: Δa_i, Δb_j parameter faults	$\Delta\mathbf{A}$, $\Delta\mathbf{b}$, $\Delta\mathbf{c}$ parameter faults

2.4.2 Fault detection with parameter estimation

Process-model-based methods require the knowledge of a usually dynamic process model in the form of a mathematical structure and parameters. For linear processes in continuous time the models can be impulse responses (weighting functions), differential equations or frequency responses. Corresponding models for discrete-time (after sampling) are impulse responses, difference equations or z-transfer functions. For fault detection in general, differential equations or difference equations are primarily suitable. In many practical cases the process parameters are partially not known or not known at all. Then, they can be determined with parameter estimation methods by measuring input and output signals if the basic model structure is known. Table 2.4.3 shows two approaches by minimization of the equation error and the output error. The first one is linear in the parameters and allows therefore direct estimation of the parameters (least-squares estimates) in non-recursive or recursive form. The second one needs numerical optimization methods and therefore iterative procedures, but may be more precise under the influence of process disturbances. The symptoms are deviations of the process parameters $\Delta\boldsymbol{\Theta}$. As the process parameters $\boldsymbol{\Theta} = f(\mathbf{p})$ depend on physically defined process coefficients $\mathbf{p}$ (like stiffness, damping coefficients, resistance), determination of changes $\Delta\mathbf{p}$ allows usually a deeper insight and

makes fault diagnosis easier, Isermann (1992). Parameter estimation methods operate with adaptive process models, where only the model structure is known. They usually need a dynamic process input excitation and are especially suitable for the detection of multiplicative faults.

Table 2.4.3. Fault detection with parameter estimation methods for dynamic processes.

Minimization of equation error	Minimization of output error
	
Loss function: $V = \Sigma\, e^2(k)$ Method: • non-recursive $\hat{\Theta} = [\Psi^T\Psi]^{-1}\Psi^T\mathbf{y}$ • recursive $\hat{\Theta}(k+1) = \hat{\Theta}(k) + \gamma(k)e(k+1)$	$V = \Sigma\, e'^2(k)$ • nonlinear parameter optimization • recursive form $\hat{\Theta}(\nu+1) = \hat{\Theta}(\nu) + \Gamma(\nu)\frac{\partial V}{\partial \Theta}(\nu)$
Symptoms: • model parameters $\Delta\hat{\Theta}(j) = \hat{\Theta}(j) - \Theta_0$ • process coefficients $\hat{\mathbf{p}} = f^{-1}[\hat{\Theta}]$ $\Delta\mathbf{p}(j) = \hat{\mathbf{p}}(j) - \mathbf{p}_0$	

2.4.3 Fault detection with state observers and state estimation

If the process parameters are known, either state observers or output observers can be applied, Table 2.4.4. Fault modeling is then performed with additive faults $\mathbf{f}_L$ at the input (additive actuator or process faults) and $\mathbf{f}_M$ at the output (sensor offset faults).

a) State observers

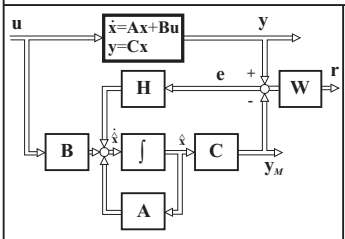
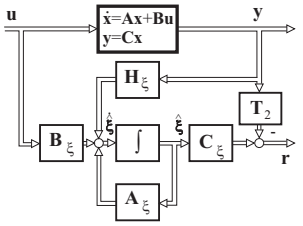
The classical state observer can be applied if the faults can be modeled as state-variable changes $\Delta\mathbf{x}_i$ e.g., for leaks. Special design of the matrix $\mathbf{W}$ allows one to generate structured residuals. In the case of *multi-output processes* special arrangements of observers were proposed:

Dedicated observers for multi-output processes

- Observer, excited by one output: One observer is driven by one sensor output. The other outputs $\mathbf{y}$ are reconstructed and compared with measured outputs $\mathbf{y}$. This allows the detection of single sensor faults, Clark (1990)

- Bank of observers, excited by all outputs: Several state observers are designed for a definite fault signal and detected by a hypothesis test, Willsky (1976)
- Bank of observers, excited by single outputs: Several observers for single sensor outputs are used. The estimated outputs $\mathbf{y}$ are compared with the measured outputs $\mathbf{y}$. This allows the detection of multiple sensor faults, Clark (1990) (dedicated observer scheme)
- Bank of observers, excited by all outputs except one: As before, but each observer is excited by all outputs except one sensor output which is supervised, Frank (1987).

Table 2.4.4. Fault detection with observers for dynamic processes.

State observer	Output observer
Process model: $\dot{\mathbf{x}}(t) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{B}\,\mathbf{u}(t) + \mathbf{F}\,\mathbf{v}(t) + \mathbf{L}\,\mathbf{f}_L(t)$ $\mathbf{y}(t) = \mathbf{C}\,\mathbf{x}(t) + \mathbf{N}\,\mathbf{n}(t) + \mathbf{M}\,\mathbf{f}_M(t)$ $\mathbf{v}(t)$, $\mathbf{n}(t)$: disturbance signals; $\mathbf{f}_L$, $\mathbf{f}_M$: additive fault signals	
	
Observer equations: $\dot{\hat{\mathbf{x}}}(t) = \mathbf{A}\,\hat{\mathbf{x}}(t) + \mathbf{B}\,\mathbf{u}(t) + \mathbf{H}\,\mathbf{e}(t)$ $\mathbf{e}(t) = \mathbf{y}(t) - \mathbf{C}\,\hat{\mathbf{x}}(t)$	$\dot{\hat{\boldsymbol{\xi}}}(t) = \mathbf{A}_\xi\,\hat{\boldsymbol{\xi}}(t) + \mathbf{B}_\xi\,\mathbf{u}(t) + \mathbf{H}_\xi\,\mathbf{y}(t)$ $\boldsymbol{\eta}(t) = \mathbf{C}_\xi\,\hat{\boldsymbol{\xi}}(t)$ $\xi(t) = \mathbf{T}_1\,\mathbf{x}(t)$: transformation
Residuals: • $\Delta\mathbf{x}(t) = \mathbf{x}(t) - \mathbf{x}_0(t)$ • $\mathbf{e}(t)$ • $\mathbf{r}(t) = \mathbf{W}\,\mathbf{e}(t)$ Special observers: – fault-sensitive filters (**H** such that $\mathbf{r}(t)$ defin. direct.) – dedicated observers (for different sensor outputs)	$\boldsymbol{\xi}(t) = \hat{\boldsymbol{\xi}}(t) - \mathbf{T}_1\,\dot{\mathbf{x}}(t)$ $\mathbf{r}(t) = \mathbf{C}_\xi\,\boldsymbol{\xi}(t) - \mathbf{T}_2\mathbf{M}f_M(t)$ – independent of $\mathbf{x}(t)$, $\mathbf{u}(t)$, $\mathbf{v}(t)$ – dependent on $\mathbf{f}_L(t)$, $\mathbf{f}_M(t)$ Design equations: $\mathbf{T}_1\mathbf{A} - \mathbf{A}_\xi\mathbf{T}_1 = \mathbf{H}_\xi\mathbf{C}$ $\mathbf{B}_\xi = \mathbf{T}_1\mathbf{B}$ $\mathbf{T}_1\mathbf{V} = \mathbf{0}$ $\mathbf{C}_\xi\mathbf{T}_1 - \mathbf{T}_2\mathbf{C} = \mathbf{0}$

Fault-detection filters (fault-sensitive filters) for multi-output processes

The feedback $\mathbf{H}$ of the state observer is chosen so that particular fault signals $\mathbf{f}_{\mathrm{L}}(t)$ change in a definite direction and fault signals $\mathbf{f}_{\mathrm{M}}(t)$ in a definite plane, Beard (1971) and Jones (1973).

b) Output observers

Another possibility is the use of output observers (or unknown input observers) if the reconstruction of the state variables $\mathbf{x}(t)$ is not of interest. A linear transformation then leads to new state variables $\xi(t)$. The residuals $\mathbf{r}(t)$ can be designed such that they are independent of the unknown inputs $\mathbf{v}(t)$, and of the state by special determination of the matrices $\mathbf{C}_{\xi}$ and $\mathbf{T}_2$. The residuals then depend only on the additive faults $\mathbf{f}_{\mathrm{L}}(t)$ and $\mathbf{f}_{\mathrm{M}}(t)$. However, all process model matrices must be known precisely. Hence, the observer-based fault-detection methods operate with a *fixed parameter model* and correct the state variables by the feedback of output errors. A comparison with the parity equation approach shows similarities.

c) State estimation

Whereas state observers are designed for deterministic initial states $\mathbf{x}(0)$ and inputs $\mathbf{u}$ and no disturbances, state estimators are optimized filters for stochastic initial states, stochastic state disturbances $\mathbf{v}$ at the input and stochastic disturbances $\mathbf{n}$ at the output with known covariances. In the case of continuous-time signals the Kalman–Bucy filter results and for discrete-time signals the Kalman filter. Table 2.4.5 shows the signal flow and basic equations for the mostly used Kalman filter with discrete-time signals, see also Isermann (2006), and the cited references.

The application of the Kalman filter is similar to that of state observers and should only be applied if considerable stochastic disturbances act on the input and/or the output signals. However, the covariance matrices of both disturbances must be known for determining the filter gain $\bar{\mathbf{K}}$. This needs in many cases some trials to find appropriate values.

2.4.4 Fault detection with parity equations

A straightforward model-based method of fault detection is to take a fixed model G_{M} and run it parallel to the process, thereby forming an output error, see Table 2.4.6:

$$r'(s) = y(s) - y_{\mathrm{M}} = [G_{\mathrm{P}}(s) - G_{\mathrm{M}}(s)]\; u(s). \tag{2.4.1}$$

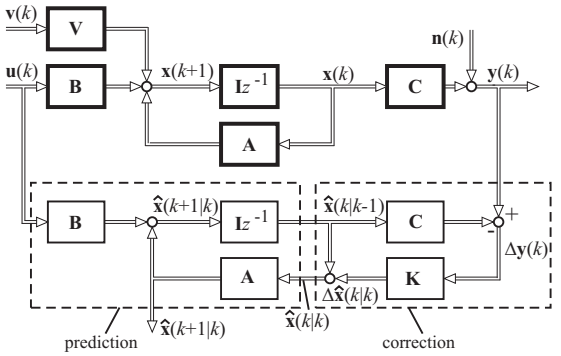
If $G_{\mathrm{P}}(s) = G_{\mathrm{M}}(s)$, the output error for additive input and output faults becomes, Table 2.4.2:

$$r'(s) = G_{\mathrm{P}}(s)\; f_{\mathrm{u}}(s) + f_{\mathrm{y}}(s). \tag{2.4.2}$$

Another possibility is to generate an equation error (polynominal error) or an input error as in Table 2.4.7, Gertler (1998).

Table 2.4.5. Fault detection with state estimation and discrete-time signals (Kalman filter).

State estimation (Kalman filter)



Process models:

$\mathbf{x}(k+1) = \mathbf{A}\,\mathbf{x}(k) + \mathbf{B}\,\mathbf{u}(k) + \mathbf{V}\,\mathbf{v}(k)$

$\mathbf{y}(k) = \mathbf{C}\,\mathbf{x}(k) + \mathbf{n}(k)$

$\mathbf{v}(k)$, $\mathbf{n}(k)$: stochastic disturbances with known covariance matrices $\mathbf{M}$ and $\mathbf{N}$

State estimation equations:

- prediction: $\hat{\mathbf{x}}(k+1|k) = \mathbf{A}\,\hat{\mathbf{x}}(k|k) + \mathbf{B}\,\mathbf{u}(k)$
- correction : $\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k\text{-}1) + \overline{\mathbf{K}}[\mathbf{y}(k) - \mathbf{C}\hat{\mathbf{x}}(k|k-1)]$
- filter gain: $\overline{\mathbf{K}} = \mathbf{P}^{-}\mathbf{C}^{\mathrm{T}}[\mathbf{C}\,\mathbf{P}^{-}\mathbf{C}^{\mathrm{T}} + \mathbf{N}]^{-1}$
- error covariance matrix Riccati equation:

$$\mathbf{P}^{-}(k+1) = \mathbf{A}\mathbf{P}^{-}(k)\,\mathbf{A}^{\mathrm{T}} - \mathbf{A}\mathbf{P}^{-}(k)\,\mathbf{C}^{\mathrm{T}}[\mathbf{C}\mathbf{P}^{-}(k)\,\mathbf{C}^{\mathrm{T}} + \mathbf{N}]^{-1}\mathbf{C}\mathbf{P}^{-}(k)^{\mathrm{T}}\mathbf{A} + \mathbf{V}\mathbf{M}\,\mathbf{V}^{\mathrm{T}}$$

Residuals: $\Delta\hat{\mathbf{x}}(k+1|k)$

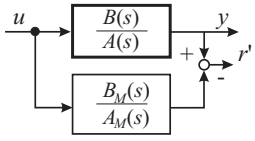
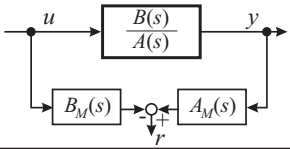
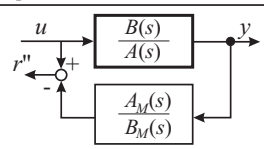
$\Delta\mathbf{y}(k) = \mathbf{e}\,(k)$

In all cases, the residuals only depend on the additive input faults $f_{\mathrm{u}}(t)$ and output faults $f_{\mathrm{y}}(t)$. The same procedure can be applied for multi-variable processes by using a state-space model, see Table 2.4.7.

The derivatives of the signals can be obtained by state-variable filters, Höfling (1996). Corresponding equations exist for discrete time and are easier to implement for the state-space model. The residuals shown in Tables 2.4.6 and 2.4.7 left are direct residuals. If the parity equations are formulated for more than one input and one output, it becomes possible to generate structured residuals such that faults do not influence all residuals. This improves the isolability of faults, Gertler (1998). For example, the components of matrix $\mathbf{W}$ for the state-space model, Table 2.4.7 right, are selected such that, e.g., one measured variable has no impact on a specific residual. Parity equations are suitable for the detection of additive faults. They are simpler to design and to implement than output observer-based approaches and lead

approximately to the same results. A comparison of fault detection with observers, Kalman filter and parity equations is given in Isermann (2006), Sect. 11.4.

Table 2.4.6. Fault detection with different forms of parity equations for linear input/output models.

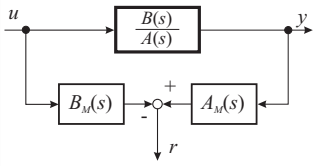
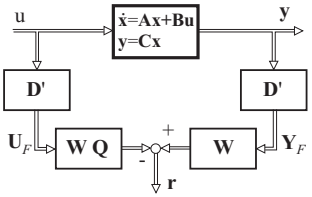
Output error	Equation error	Input error
		
Parity equations: $r'(s) = y(s) - \frac{B_M(s)}{A_M(s)}\, u(s)$ $r'(t) = \boldsymbol{\psi}_r'^T(t)\boldsymbol{\Theta}_{Mr} + \boldsymbol{\psi}_a^T(t)\boldsymbol{\Theta}_{Ma} - \boldsymbol{\psi}_b^T(t)\boldsymbol{\Theta}_{Mb}$	$r(s) = A_M(s)\, y(s) - B_M\, u(s)$ $r(t) = \boldsymbol{\psi}_a^T(t)\, \boldsymbol{\Theta}_{Ma} - \boldsymbol{\psi}_b^T \boldsymbol{\Theta}_{Mb}$	$r''(s) = u(s) - \left(\frac{A_M(s)}{B_M(s)}\right) y(s)$ $r''(t) = \boldsymbol{\psi}_b''^T(t)\boldsymbol{\Theta}_{Mb}'' - \boldsymbol{\psi}_a^T(t)\boldsymbol{\Theta}_{Ma}$
$B_M(s) = b_0 + b_1 s + ... + b_m s^m$ $A_M(s) = 1 + a_1 s + ... + a_n s^n$ $\boldsymbol{\Theta}_{Mr}^T = [\,1\ a_1\ a_2 ...\ a_n\,]$ $\boldsymbol{\psi}_r'^T = [\,y\ y^{(1)}\ y^{(2)} ...\ y^{(n)}\,]$	$\boldsymbol{\Theta}_{Mb}^T = [\,b_0\ b_1\ ...\ b_m\,]$ $\boldsymbol{\Theta}_{Ma}^T = [\,1\ a_1\ ...\ a_n\,]$ $\boldsymbol{\psi}_b^T = [\,u\ u^{(1)}\ u^{(2)} ...\ u^{(m)}\,]$ $\boldsymbol{\psi}_a^T = [\,y\ y^{(1)}\ y^{(2)} ...\ y^{(n)}\,]$	$\boldsymbol{\Theta}_{Mb}''^T = \frac{1}{b_0}[\,1\ b_1\ b_2 ...\ b_m\,]$ $\boldsymbol{\psi}_b''^T = [\,u\ u^{(1)}\ u^{(2)} ...\ u^{(m)}\,]$

2.4.5 Direct reconstruction of non-measurable variables

State observers and Kalman filters reconstruct non-measurable variables contained in the state-vector $\mathbf{x}(t)$ and parameter-estimation methods reconstruct non-measurable parameters $\boldsymbol{\Theta}$ from measured input signals $u(t)$ and output signals $y(t)$. However, process models or parts of it can also be directly used to calculate non-measurable variables from measured variables, for example by using algebraic relationships. A first example is the calculation of the torque M of a DC motor from the current I by using $M(t) = \Psi I(t)$, where Ψ is the flux linkage. A second example is the reconstruction of the volume flow rate $\dot{V}$ of a centrifugal pump transporting a liquid through a pipe from the rotor angular speed ω by $\dot{V}(t) = \kappa\omega(t)$ for steady-state operation, Wolfram (2002). This kind of reconstruction via algebraic relations holds especially for transformers and converters whose behavior is expressed with the power covariables effort $e(t)$ and flow $f(t)$, Isermann (2005).

Figure 2.4.5 summarizes the principles of the most important model based fault detection methods. Parameter estimation usually needs an input excitation. State estimation and parity equations can operate in steady-state conditions without special inputs. they deliver relatively similar results. These and other methods are described in Isermann (2006) and their applications to industrial processes in Isermann (2011).

Table 2.4.7. Fault detection with parity equations and equation errors for dynamic processes.

Input/output model, equation error	State space model
	
Parity equations: $r(s) = A_M(s)y(s) - B_M(s)u(s)$ $r(t) = \boldsymbol{\psi}_a^T(t)\boldsymbol{\Theta}_{Ma} - \boldsymbol{\psi}_b^T(t)\boldsymbol{\Theta}_{Mb}$	$\mathbf{Y}_F(t) = \mathbf{T}\,\mathbf{X}(t) + \mathbf{Q}\,\mathbf{U}_F(t)$ $\mathbf{W}\,\mathbf{Y}_F(t) = \mathbf{W}\,\mathbf{T}\,\mathbf{x}(t) + \mathbf{W}\,\mathbf{Q}\,\mathbf{U}_F(t)$ $\mathbf{W}\,\mathbf{T} = 0$ $\mathbf{r}(t) = \mathbf{W}(\mathbf{Y}_F(t) - \mathbf{Q}\,\mathbf{U}_F(t))$
$B_M(s) = b_0 + b_1 s + \ldots + b_m s^m$ $A_M(s) = 1 + a_1 s + \ldots + a_n s^n$ $\boldsymbol{\Theta}_{Mb}^T = [\, b_0\, b_1 \ldots\, b_m \,]$ $\boldsymbol{\Theta}_{Ma}^T = [\, 1\, a_1\, a_2 \ldots\, a_n \,]$ $\boldsymbol{\psi}_b^T = [\, u\, u^{(1)}\, u^{(2)} \ldots\, u^{(m)} \,]$ $\boldsymbol{\psi}_a^T = [\, y\, y^{(1)} \ldots\, y^{(n)} \,]$	$\mathbf{D'u} = [\, u\, u^{(1)} \ldots\, u^{(m)} \,]^T = \mathbf{U}_F$ $\mathbf{D'y} = [\, \mathbf{y}\, \mathbf{y}^{(1)} \ldots\, \mathbf{y}^{(n)} \,]^T = \mathbf{Y}_F$ $\mathbf{T} = [\, C\ CA\ CA^2 \ldots \,]^T$ $\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 \ldots \\ CB & 0 & 0 \\ CAB & CB & 0 \\ \mathrm{M} & & \end{bmatrix}$

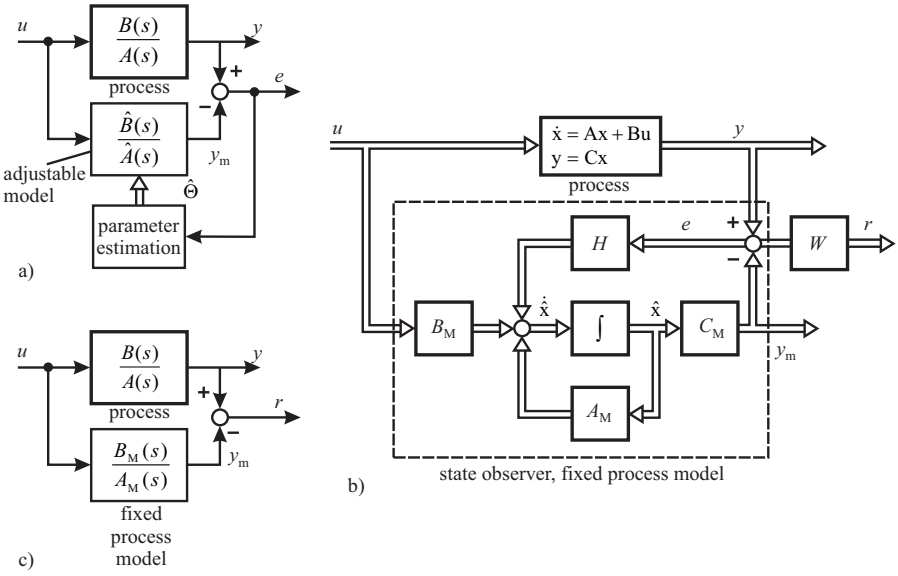


Fig. 2.4.5. Main process-model-based fault-detection methods for dynamic processes with output errors. **a** parameter estimation, **b** state estimation, **c** parity equations.

2.5 Fault-diagnosis methods

The task of fault diagnosis consists of the determination of the type of fault with as many details as possible such as the fault size, location and time of detection. The diagnostic procedure is based on the observed analytical and heuristic symptoms and the heuristic knowledge of the process, see the schemes in Figs. 2.2.1, 2.2.3 and 2.4.1. The inputs to a knowledge-based fault-diagnosis system are all available symptoms as facts and the fault-relevant knowledge about the process, mostly in heuristic form. The symptoms may be presented just as binary values $[0, 1]$ or, e.g. as fuzzy sets to take gradual sizes into account.

2.5.1 Classification methods

If no further knowledge is available for the relations between features and faults, classification or pattern recognition methods can be used, Table 2.5.1. Here, reference vectors $\mathbf{S}_{\mathrm{n}}$ are determined for the normal behavior. Then the corresponding input vectors $\mathbf{S}$ of the symptoms are determined experimentally for certain faults F_{j} applying the fault-detection methods. The relationship between $\mathbf{F}$ and $\mathbf{S}$ is therefore learned (or trained) experimentally and stored, forming an *explicit knowledge base*. By comparison of the observed $\mathbf{S}$ with the normal reference $\mathbf{S}_{\mathrm{n}}$, faults $\mathbf{F}$ can be concluded.

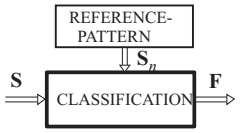
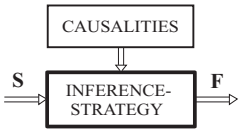
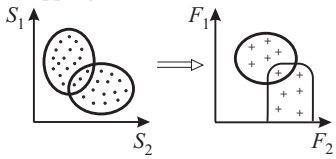
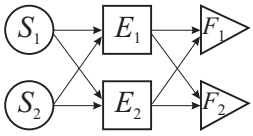
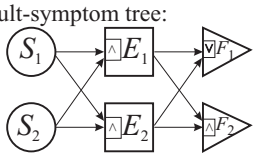
One distinguishes between *statistical* or *geometrical classification methods*, with or without certain probability functions, Tou and Gonzalez (1974). A further possibility is the use of neural networks because of their ability to approximate nonlinear relations and to determine flexible decision regions for $\mathbf{F}$ in continuous or discrete form, Leonhardt (1996). By fuzzy clustering the use of fuzzy separation areas is possible.

2.5.2 Inference methods

For some technical processes, the basic relationships between faults and symptoms are at least partially known. Then this a-priori knowledge can be represented in causal relations: fault $\rightarrow$ events $\rightarrow$ symptoms. Table 2.5.1 shows a simple causal network, with the nodes as states and edges as relations. The establishment of these causalities follows the fault-tree analysis (FTA), proceeding from faults through intermediate events to symptoms (the physical causalities) or the event-tree analysis (ETA), proceeding from the symptoms to the faults (the diagnostic forward-chaining causalities). To perform a diagnosis, this qualitative knowledge can now be expressed in the form of rules: IF $<$ condition $>$ THEN $<$ conclusion $>$. The condition part (premise) contains facts in the form of symptoms S_{i} as inputs, and the conclusion part includes events E_{k} and faults F_{j} as a logical cause of the facts. If several symptoms indicate an event or fault, the facts are associated by AND and OR connectives, leading to rules in the form

$$\begin{aligned} &\text{IF} \;\; < S_1 \;\; \text{AND} \;\; S_2 > \text{THEN} \;\; < E_1 > \\ &\text{IF} \;\; < E_1 \;\; \text{OR} \;\; E_2 > \text{THEN} \;\; < F_1 > . \end{aligned}$$

Table 2.5.1. Methods of fault diagnosis.

Classification methods	Inference methods
	
Without a-priori knowledge on symptom causalities Mapping:  $\mathbf{S}^T = [\, S_1, S_2 \ldots S_n \,]$ $\mathbf{F}^T = [\, F_1, F_2 \ldots F_m \,]$	With a-priori knowledge on symptom causalities Causal network:  Fault-symptom tree: 
Classification: – statistical – geometrical – neural nets – fuzzy clusters	Rules: If $< S_1 \wedge S_2 >$ Then $< E_1 >$ Diagnostic reasoning: – Boolean logic: facts binary – Approximative reasoning: – Probabilistic facts: probability densities – Fuzzy facts: fuzzy sets

For the establishment of this heuristic knowledge several approaches exist, see Frost (1986), Torasso and Console (1989). In the classical fault-tree analysis the symptoms and events are considered as binary variables, and the condition part of the rules can be calculated by Boolean equations for parallel–serial connection, see, e.g., Barlow and Proschan (1975), Freyermuth (1993). However, this procedure has not proved to be successful because of the continuous and gradual nature of faults and symptoms. For the diagnosis of technical processes approximate reasoning is more appropriate. For more details see Isermann (2006).

2.6 Fault detection and diagnosis in closed loop

The main goals for using automatic control loops are precise following of reference variables (setpoints), a faster response than in open loop, compensation of all kind of external disturbances on the controlled variable, stabilization of unstable processes,

reduction of the influence of process parameter changes with regard to the static and dynamic behavior, partial compensation of actuator and process nonlinearities, and, of course, replacement of manual control by humans.

The performance of a SISO control loop with regard to the control error (deviation)

$$e(k) = w(k) - y(k), \tag{2.6.1}$$

i.e. the deviation of the controlled variable $y(k)$ from the reference variable $w(k)$ depends on many facts, compare Fig. 2.6.1, like:

- external disturbance $w(k), u_{\mathrm{v}}(k), v_{\mathrm{i}}(k)$
- structure and parameters of the controller G_c and controller faults f_{c}
- changes of the structure and parameters of the process G_{p} and process faults f_{p}
- changes of the actuator G_{a} and actuator faults f_{a}
- faults f_{s} in the sensor G_{s} and measurement noise n_{s}.

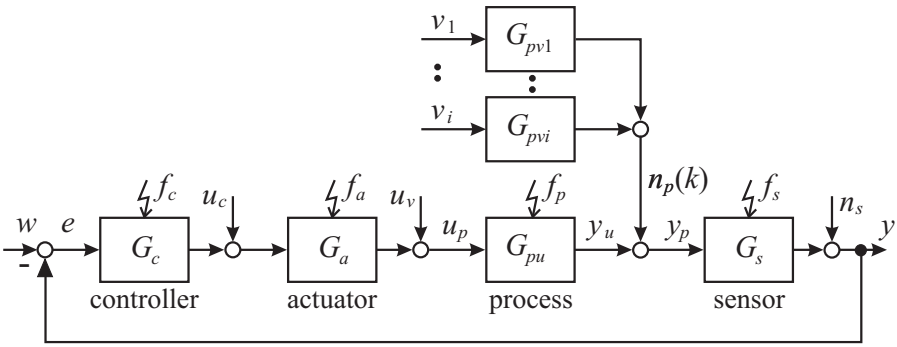


Fig. 2.6.1. Control loop with variables and fault influences.

y	controlled variable	w	reference variable
u_{p}	manipulated variable	e	control deviation
v_{i}	process disturbances	n_{s}	measurement noise
u_{v}	process input disturbances	$f_{\mathrm{c,a,p,s}}$	faults of the controller,
n_{p}	sum of process disturbances		actuator, process and
y_{p}	process output to be controlled		sensor

Hence, many changes and faults influence the performance of closed loops. Usually, only the control deviation e and the control variable y are monitored.

Small faults in the actuator and process, be they additive or multiplicative, will usually be compensated by the feedback controller (with integral action) and they will not be detectable by considering $e(k)$ and $y(k)$ only, as long as the control deviation turns back to approximately zero. Also small sensor offset faults will not be detected. The controller will just make the wrong sensor signal equal to the reference variable. Only by a redundant sensor or other redundant information for the controlled variable the offset fault can usually be detected.

As shown in Isermann (2006), Chap. 12, several *larger faults* have a similar effect on the considered changes of closed-loop behavior such that it is not easy to differentiate them. In addition, some of the behavior is also observed after external disturbances under normal operation.

A first possibility for fault detection in closed loops is to analyze measurable signals like the controlled variable $y(t)$, the manipulated variable $u(k)$, the reference variable $w(k)$ and the control deviation $e(k)$. This is also known as *performance monitoring* of closed loops. For example, the variances of these variables, steady-state deviations, large overshoots or frequency spectra can be monitored. However, it is very difficult to find the reasons for observed changes.

Process-model-based detection methods allow a deeper insight, as they relate the manipulated variable $u(k)$ to the controlled variable $y(k)$. The application of *parameter estimation* in closed loops requires consideration of the identifiability conditions. If no external measurable perturbations can be used, because the closed loop operates with constant reference variable $w(k) = const$ and only compensates for disturbances, special higher-order controller structures are required. With measurable external perturbations, as for servo systems with continuously changing reference variable, parameter estimation methods can be directly applied. Powertrains of vehicles behave like servo systems. Also *parity equations* for the process are directly applicable if the (fixed) process model agrees well with the real process. Especially by combining several detection methods a large portion of faults in the components of a closed loop (actuators, process, sensor, controller) can be detected and isolated.

Concluding remarks

The summary of basic fault-detection and diagnosis methods presented in this chapter was limited to linear processes mainly. Some of the methods can also be directly applied to nonlinear processes, e.g., signal analysis, parity equations and parameter estimations. However, all the methods have to be adapted to the real processes. In this sense the basic methods should be considered as "tools", which have to be combined properly in order to meet the practical requirements for the fault detection of the considered real process.

References

Barlow R, Proschan F (1975) Statistical theory of reliability and life testing. Holt, Rinehart and Winston

Beard R (1971) Failure accommodation in linear systems through self-reorganization. Tech. Rep. MVT-71-1, Man Vehicle Laboratory, Cambridge, MA

Blanke M, Kinnaert M, Lunze J, Staroswiecki M (2006) Diagnosis and fault-tolerant control, 2nd edn. Springer, Berlin

Chen J, Patton R (1999) Robust model-based fault diagnosis for dynamic systems. Kluwer, Boston

Clark R (1990) A simplified instrument detection scheme. IEEE Trans Aerospace Electron Systems 14(3):558–563

Frank P (1987) Advanced fault detection and isolation schemes using nonlinear and robust observers. In: 10th IFAC Congress, München, Germany, vol 3, pp 63–68

Frank P (1990) Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. Automatica 26(3):459–474

Freyermuth B (1993) Wissensbasierte Fehlerdiagnose am Beispiel eines Industrieroboters. Dissertation Technische Hochschule Darmstadt. Fortschr.-Ber. VDI Reihe 8, 315. VDI Verlag, Düsseldorf

Frost R (1986) Introduction to knowledge base systems. Collins, London

Gertler J (1998) Fault detection and diagnosis in engineering systems. Marcel Dekker, New York

Himmelblau D (1978) Fault detection and diagnosis in chemical and petrochemical processes. Elsevier, New York

Höfling T (1996) Methoden zur Fehlererkennung mit Parameterschätzung und Paritätsgleichungen. Dissertation Technische Hochschule Darmstadt. Fortschr.-Ber. VDI Reihe 8, 546. VDI Verlag, Düsseldorf

Isermann R (1984) Process fault detection on modeling and estimation methods - a survey. Automatica 20(4):387–404

Isermann R (1992) Estimation of physical parameters for dynamic processes with application to an industrial robot. International Journal of Control 55(6):1287–1298

Isermann R (1997) Supervision, fault-detection and fault-diagnosis methods – an introduction. Control Engineering Practice – CEP 5(5):639–652

Isermann R (2005) Mechatronic systems – fundamentals, 2nd edn. Springer, London

Isermann R (2006) Fault-diagnosis systems – An introduction from fault detection to fault tolerance. Springer, Heidelberg

Isermann R (2011) Fault-diagnosis applications. Springer, Heidelberg

Isermann R, Freyermuth B (1991) Process fault diagnosis based on process model knowledge. Journal of Dynamic Systems, Measurement and Control 113:Part I, 620–626; Part II, 627–633

Jones H (ed) (1973) Failure detection in linear systems. Dept. of Aeronautics, M.I.T., Cambridge

Leonhardt S (1996) Modellgestützte Fehlererkennung mit neuronalen Netzen - Überwachung von Radaufhängungen und Diesel-Einspritzanlagen. Dissertation Technische Hochschule Darmstadt. Fortschr.-Ber. VDI Reihe 12, 295. VDI Verlag, Düsseldorf

Patton R, Frank P, Clark P (eds) (2000) Issues of fault diagnosis for dynamic systems. Springer, New York

Torasso P, Console L (1989) Diagnostic problem solving. North Oxford Academic, Oxford

Tou J, Gonzalez R (1974) Pattern recognition principles. Addison-Wesley Publishing, Reading, MA

Willsky A (1976) A survey of design methods for failure detection systems. Automatica 12:601–611

Wolfram A (2002) Komponentenbasierte Fehlerdiagnose industrieller Anlagen am Beispiel frequenzumrichtergespeister Asynchronmaschinen und Kreiselpumpen. Fortschr.-Ber. VDI Reihe 8, 967. VDI Verlag, Düsseldorf

Part II

Diagnosis of Internal Combustion Engines

3

On the control and diagnosis of internal combustion engines

As the supervision and diagnosis functions of internal combustion engines are strongly connected with the control functions, this chapter briefly presents the basic control tasks and some control functions. This is followed by a short discussion of present on-board diagnosis (OBD) requirements and realized approaches. Because of the complexity of the sensors, actuators and ECU control functions, the control-oriented structure of the engines from Isermann (2014) is used, which also helps to structure the supervision and diagnosis functions. Finally, an introduction to model-based engine fault detection is given.

3.1 Electronic engine control

The electronic engine control unit (ECU) has the task to control, optimize and supervise all relevant functions of an internal combustion engine (ICE). At first glance this comprises:

- torque generation according to the drivers accelerator pedal
- low fuel consumption
- low exhaust emissions and low noise (compliance with legislation)
- good driving behavior.

With regard to the general assembly of internal combustion engines the *control functions* can be dedicated to following engine subsystems or engine parts:

1. intake system (air system)
2. injection system
3. fuel supply
4. combustion and crankshaft drive, charging unit
5. lubrication
6. exhaust gas system, exhaust gas recirculation
7. cooling system.

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7_3

As the control functions are fundamental for the development of supervision and diagnosis functions, they are briefly described in the following sections, which lean on Isermann (2014). A part of the subsystems is basically controlled mechanically, as for example, the inlet and outlet valve timing and lift through the camshaft for conventional four-stroke engines, the fuel pressure and oil pressure with overpressure relief valves and the coolant temperature with thermostatic expansion valves. Many other important variables are controlled electronically such as the air flow, fuel flow, combustion, torque, speed, exhaust gas recirculation, and exhaust gas treatment. However, also the camshaft can be controlled electronically in the case of variable valve trains (VVT) and the fuel pressure for common rail direct injection systems, as well as the oil pressure. Therefore electrically commanded actuators are state-of-the-art, which operate with electrical, pneumatic or hydraulic auxiliary energy, and sensors with electrical outputs.

The optimization of the thermodynamic, fluid dynamic and mechanical design and construction of internal combustion engines has led to an increasing number of actuators and sensors and to a strong increase of electronic control functions. The *electronic control unit* (ECU) controls the start, idle, warm-up and normal operation. The increase of control functions is demonstrated by the size of programs and data of the digital microcomputer based system. Within the last 15 years the clock frequency of the microprocessors has increased from 12 to 150 MHz, the databus width from 8 bit to 32 bit, program storage from 32 kbyte to 5 Mbyte, the computing time from about 10 to 300 millions instructions per second (MIPS), and the calibration labels from about 2500 to 30000. Within this time window of 15 years these characteristics have increased by following factors: memory: 1:100, computer power (MIPS): 1:30 and calibration parameters 1:10, see Fig. 3.1.1.

This development was mainly influenced by the increase of *variabilities* to optimize the combustion and the exhaust-gas treatment. For gasoline engines this reflected first, for example, in the optimization of valve timing and lift in combination with the injections, depending on load, homogeneous or stratified operation, and λ-control. In the case of diesel engines and partially for gasoline engines this resulted in optimizing multiple injections, valve timing, turbo charging, exhaust-gas recirculation and particulate filters and their regeneration cycles.

Figure 3.1.2 shows a simplified signal flow diagram for *gasoline engines*. The engine control system has to be designed for 6–8 main manipulated variables and 5–8 measured output variables, leading to a complex nonlinear multiple-input multiple-output system. Because several output variables cannot be measured in mass production vehicles, like the torque and some emissions, as they are too costly or may not be robust enough, some output variables of the engine are controlled by *feedforward structures*, also called *open-loop structures*. This requires that the driver‘s pedal position, several influencing variables like the engine speed, air mass flow, air temperature and pressure, oil and coolant temperature have to measured and feedforward control actions on the main manipulated variables have to be specified. In the case of gasoline engines *feedback control* or *closed-loop control* is used for λ-control and knock control. *Diesel engines* use charging pressure, air mass flow rate and overspeed feedback control, see Fig. 3.1.3. Both engine types have idling speed

and coolant fluid temperature feedback control. The feedforward structures have the advantages that they enable fast reactions and do not have stability problems, but they need relatively precise engine models and sensors for the main influencing variables. The closed-loop structures compensate also not measured disturbances, but their controllers must be precisely and robustly adapted to the engine's nonlinear dynamic behavior. All the control functions depend strongly on the load and speed and on the operating state such as starting phase, warm-up, normal state, idling, overrun and shut-down.

The next sections give, as an overview, some details of basic control functions for gasoline and diesel engines.

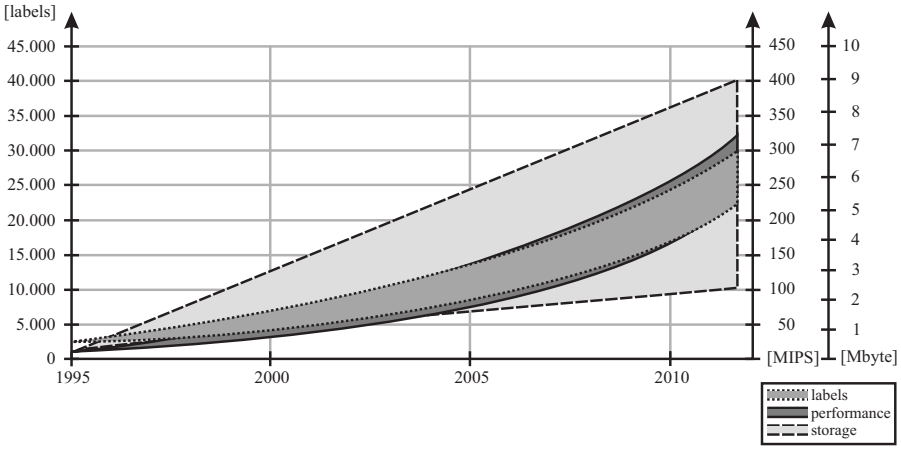


Fig. 3.1.1. Development of ECU characteristics for engine control, ETAS (2010).

3.1.1 On the control of gasoline engines

Figure 3.1.4 illustrates the main components of a direct injection gasoline engine with its actuators and sensors. To generate a certain torque at the crankshaft the gasoline engines require a very specific air-fuel mixture. For an optimal combustion they need a close to stoichiometric mixture of air and fuel in the range $0.8 < \lambda < 1.4$ and because of the mostly applied three-way catalyst the air/fuel ratio must be precisely in the range of $0.98 < \lambda < 1.02$. Therefore the air is usually throttled for part-load by a throttle valve so that the cylinders suck the air from the manifold with subatmospheric pressure. The air mass (air charge) taken in by the cylinders determines together with the fuel the resulting torque. It is measured by an air-mass sensor or determined indirectly by a manifold pressure sensor. However, the conventional load control of the engine by an upstream throttle generates charging losses, in contrast to diesel engines. The control unit (ECU) commands the electrical actuated throttle valve depending on the drivers accelerator pedal position and adjusts the injected

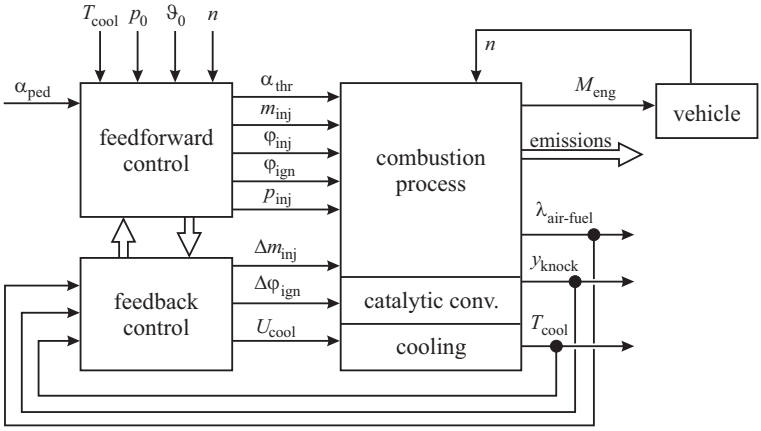


Fig. 3.1.2. Simplified control structure of a gasoline engine with intake fuel injection. 6–8 manipulated variables, 5–8 measured outputs.

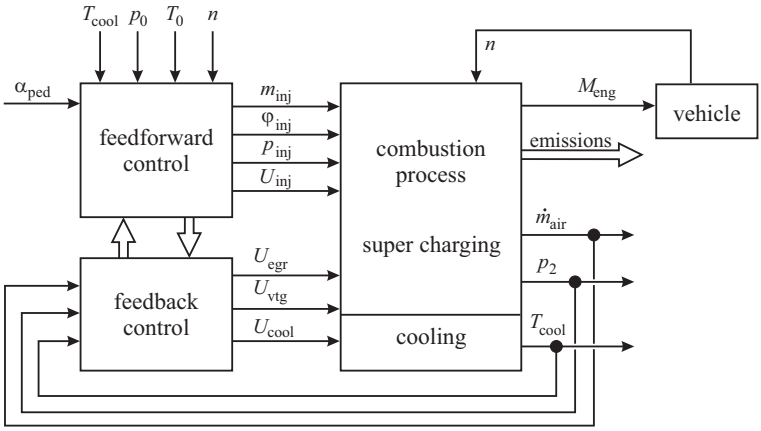


Fig. 3.1.3. Simplified control structure of a diesel engine with turbocharger. 6–8 manipulated variables, 5–8 measured outputs.

fuel quantity with the *fuel injection system* by feedforward control. The fuel is injected either in the intake manifold upstream the inlet valves (port injection) using a low pressure pump (4–6 bar) or directly into the combustion chamber with controlled high pressure fuel pumps (120–200 bar).

The *injection into the intake manifold* during the induction stroke of the cylinders may consist of several injection pulses (multi-point injection) and is applied for a conventional homogeneous combustion. The ECU converts the accelerator pedal position into a desired calculated torque value. By using various correction functions the air charge and the corresponding fuel injection mass per cylinder and an optimal ignition angle relative to the top dead center of the crankshaft are calculated and the manipulated variables throttle valve angle, fuel injection angle and duration and the

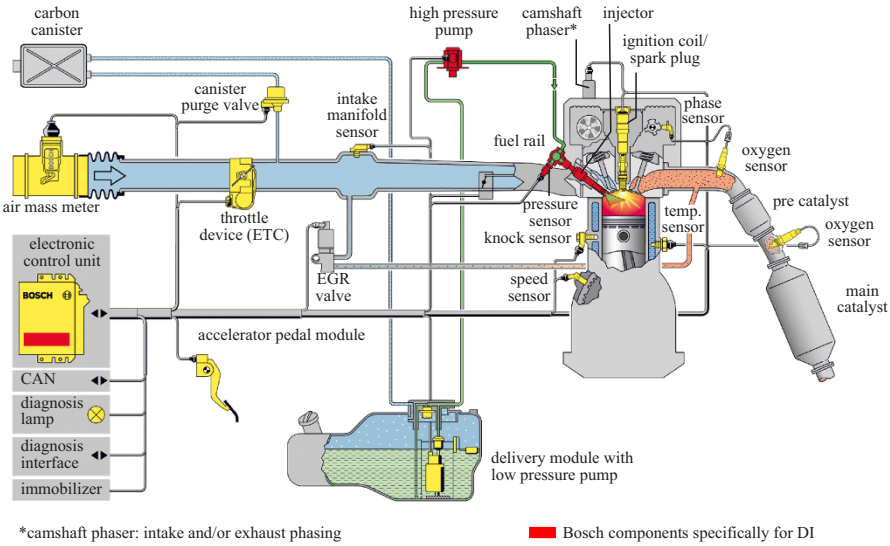


Fig. 3.1.4. Schematic of a direct injection gasoline engine with actuators, sensors and ECU for homogeneous combustion. (Bosch DI Motronic, $\lambda = 1$, by courtesy of Robert Bosch GmbH).

ignition coil activation are set by feedforward control look-up tables. The injected fuel mass is corrected by a (cascaded) lambda feedback controller with measurements of λ in the exhaust system before and after the catalyst in order to reach a stoichiometric combustion for an optimal conversion of CO, HC and NO_x in the three-way catalyst.

The *direct injection* in the cylinders has as additional manipulation the high pressure in the fuel rail and allows different types of combustion. At lower engine speed and torque a stratified cylinder charge can be reached, where in one zone a combustible air/fuel mixture cloud is generated and transferred to the spark plug. The other zone contains excess air and residual exhaust gases. The zones are controlled by a turbulence flap in the manifold. Thus, the throttle valve can be opened, avoiding throttling and charging losses, resulting in better efficiency. The fuel is injected during compression. However, the lean combustion generates larger NO_x concentrations and because of the excess air a three way catalyst cannot be used. Therefore NO_x is reduced by exhaust gas recirculation (EGR), manipulated by an EGR-valve and a NO_x accumulator-type catalytic converter is applied. At high speed and torque the engine has to run in homogeneous mode, because the generation of stratified conditions cannot be realized. Hence, this two-mode (also three-mode) combustion requires precise control functions for several variables and operates by using universal (broad-band) lambda sensors with variable setpoints for the air/fuel ratio. A NO_x sensor may also be used. In addition *knock feedback control* corrects the ignition angle by using one or two knock sensors at the crankcase in order to prevent knocking combustion which may lead to engine damages.

Based on these main control tasks Fig. 3.1.5 shows a control-oriented block structure for a gasoline engine. Altogether seven main actuators manipulate the air flow, the fuel flow and if provided the recirculated exhaust gas, and exhaust gas treatment through different control modules. The control modules are torque control, injection and air/fuel control, ignition control, knock control, air charge and EGR control for normal operation of the warm engine. The speed of the engine results together with the power consuming load, the transmission and the vehicle. In addition, there are special control modules depending on engine operating states, such as for the start, warming-up, shut-off and idling.

3.1.2 On the control of diesel engines

The main components of a diesel engine with turbocharger are shown in Fig. 3.1.6. The air for diesel engines is sucked into the cylinders with slightly lower than atmospheric pressure, if no turbochargers are used and the torque is manipulated by the injected fuel mass. However, most of the modern diesel engines have turbochargers to increase the torque and the power. The air charge then depends on the charging (boost) pressure, which is controlled by the charging pressure controller, e.g. by changing the wastegate position or the position of a variable geometry turbocharger (VTG). At low load, the diesel engine operates with large excess air, i.e. with large air/fuel ratio expressed by the excess-air factor λ. Only at high loads, λ comes closer to 1. If it gets too small, the diesel engine starts smoking. The relative high compression ratio in the cylinders together with the low charging losses because of an inactive throttle valve in the intake and therefore good air supply lead to a higher efficiency compared to gasoline engines, especially at part load. However, the higher the combustion temperature and therefore the higher the thermodynamic efficiency the more NO_x gases are produced because of the oxidation of nitrogen gas contained in the air. In order to lower the combustion temperature and the NO_x concentration, the exhaust gases can be recirculated to the intake on cost of the fuel efficiency.

Modern diesel engines have either a common rail or a unit pump *injection system* with pressures until 2200 bar. The ECU commands the injected fuel mass into the cylinders according to the drivers accelerator pedal position. In order to influence the burning process with regard to optimized fuel consumption, NO_x and soot formation and the generated noise multiple injections are applied, with pre-, main- and after-injection pulses. The air mass flow rate in the *intake system* is indirectly controlled by the EGR valve (passenger car diesels mainly). Because the fresh gas mixture taken in by the cylinders is about constant for a certain operating point, an increasing EGR flow rate $\dot{m}_{\mathrm{egr}}$ results in a reduction of the air flow rate $\dot{m}_{\mathrm{air}}$. This means that the ratio $\dot{m}_{\mathrm{egr}}/\dot{m}_{\mathrm{air}}$ is changed. The control variable of the closed *air mass control loop* is the measured mass flow rate, the manipulated variable is the EGR-valve position, and the reference value is calculated from the desired reference excess air factor λ_{ref}. A further feedback control is the *charging pressure control loop*, which manipulates either a wastegate for fixed geometry turbochargers or the guide vanes of turbochargers with variable geometry (VGT). Both closed control loops are strongly coupled.

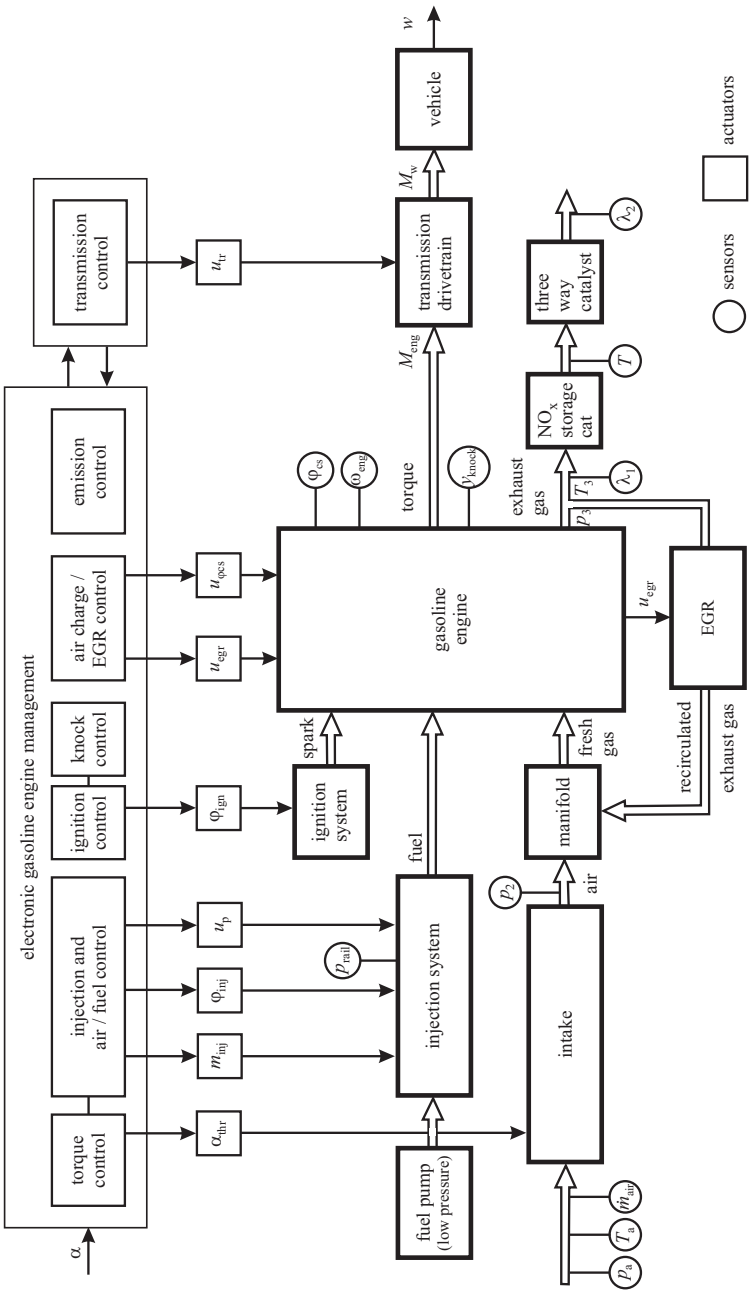


Fig. 3.1.5. Control-oriented block diagram of a gasoline engine with direct injection, actuators, control modules and some sensors.

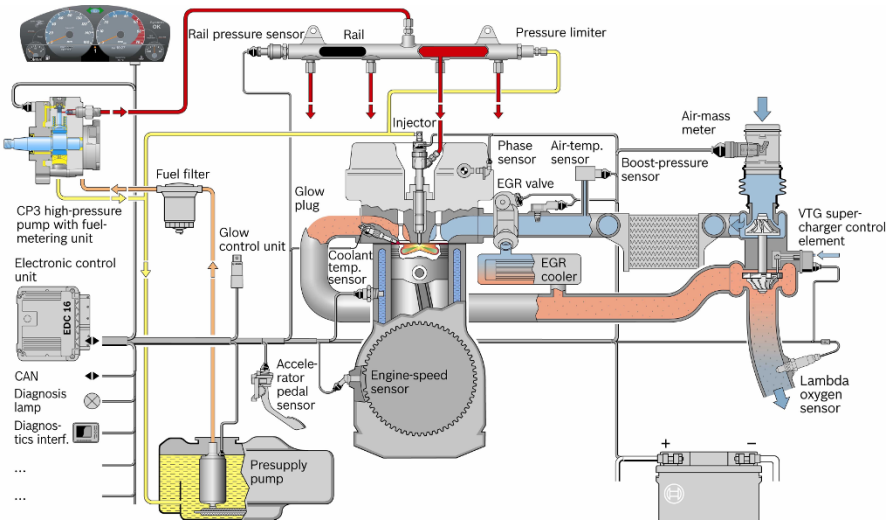


Fig. 3.1.6. Schematic of a direct injection diesel engine with VGT turbocharger, actuators, sensors and ECU. (Bosch EDC 16, by courtesy of Robert Bosch GmbH).

Diesel engines require a *speed controller* for the maximum speed in order to limit the speed by reducing the fuel mass. Otherwise the engine could be damaged by overspeeding. This is because a load control by intake throttling as for gasoline engines is missing. Speed control is also applied for idling, but not during the normal operating range for diesel engines in vehicles.

The *exhaust after-treatment* has the task of reducing HC, CO, NO_x and particulates. In most passenger car diesel engines HC, CO and some soot are removed by an oxidation catalyst. This may be followed by an NO_x *storage catalyst* (NSC), where in the loading phase NO_2 is stored within 30 to 300 s in a lean exhaust gas ($\lambda > 1$). Regeneration and removal of NO_2 take place in a rich exhaust gas ($\lambda < 1$) within 2 to 10 s. This removal or reduction of NO_2 is performed with CO and H_2 in a rich exhaust gas by retarding the injection angle and throttling the intake air. The control of the storage catalyst requires a temperature sensor and NO_x-sensor or λ-sensor. An alternative is a *selective catalytic reduction* (SCR), where NO_x is removed by ammonia NH_3, which stems from a liquid urea. This process operates continuously by injecting an urea/water solution ("add blue") from an extra tank, and is feedback controlled by using NO_x- and NH_3-sensors.

Emitted soot particulates can be removed by *diesel particulate filters* (DPF). They consist of porous ceramics or sintered metal and must be generated by soot burning with the oxygen in the exhaust gas and an increased temperature of minimum 600°C, resulting in CO_2. The regeneration is started when a criterion based on a combination of difference pressure increase and calculated soot storage from a storage model is exceeded. Then the exhaust temperature is increased by a retarded main or extra late injection and intake-air throttling, depending on the operating state

of the engine. This regeneration takes about 10 to 20 minutes and the engine is controlled so that the torque is not remarkably reduced. The control of the DPF requires a difference pressure and a temperature sensor.

Figure 3.1.7 depicts a block-oriented structure of a diesel engine with main actuators and sensors. Seven actuators manipulate the air, the fuel, exhaust gas recirculation and exhaust gas after-treatment. One can distinguish the control modules for torque control, injection control, air charge and EGR control, and emission control for the normal operation phase. These control functions are strongly interconnected. Closed loop control is usually realized for the air flow, common rail pressure, charging pressure and in some cases for exhaust gas after-treatment. This is added by many feedforward control functions. Most feedforward control functions are implemented as grid-based three-dimensional look-up tables (3D-maps) or as two-dimensional characteristics. This is because of the strongly nonlinear static and dynamic behavior of the IC-engines, the good interpretation and the direct programming in microprocessors with fixed point arithmetics. Some of the functions are based on physical models with correction factors, but many of the look-up tables and control algorithms are calibrated after measurements on engine test benches and with vehicles. However, as the complexity and the variants of the engines increase, *engine-model-based identification and control design methods* are of growing importance, see Isermann (2014).

Beside the main control functions several *sublevel controls* are required, such as position control for throttle, injection pumps, camshaft, swirl and tumble flaps, pressure control for fuel and lubrication oil. For special operating conditions or states supplemental control functions are implemented, like knocking control (SI) or smoke limitation control (CI), idling-speed control, cold start-up control and warming-up control.

This concludes a first rough sketch of basic control functions for combustion engines. As an introduction to the the basics of internal combustion engines following books are recommended: Heywood (1988), Urlaub (1995), Cook et al (1996), Jurgen (1999), Stone (1999), Ferguson and Kirkpatrick (2001), Stotsky (2009), van Basshuysen (2009), Pischinger et al (2009), Mollenhauer and Tschöke (2010), Robert Bosch GmbH (2012), Köhler and Flierl (2012), Merker et al (2012), Pucher and Zinner (2012), Spicher (2012), Bargende (2013), Pischinger (2013), Beidl (2014), Wachtmeister (2014).

Internal combustion engine control is treated in the following books: Kiencke and Nielsen (2000), Isermann (2003), Johansson and Rantzer (2003), Robert Bosch GmbH (2006), Guzzella and Onder (2010), Isermann (2010), Eriksson and Nielsen (2014), Isermann (2014).

3.2 On-board and off-board diagnosis of engines

The enlargement of electronic functions in motor vehicles enables not only more control functions but also allows to implement supervision and diagnostic functions. This is necessary because the complexity of the engine, chassis and their components

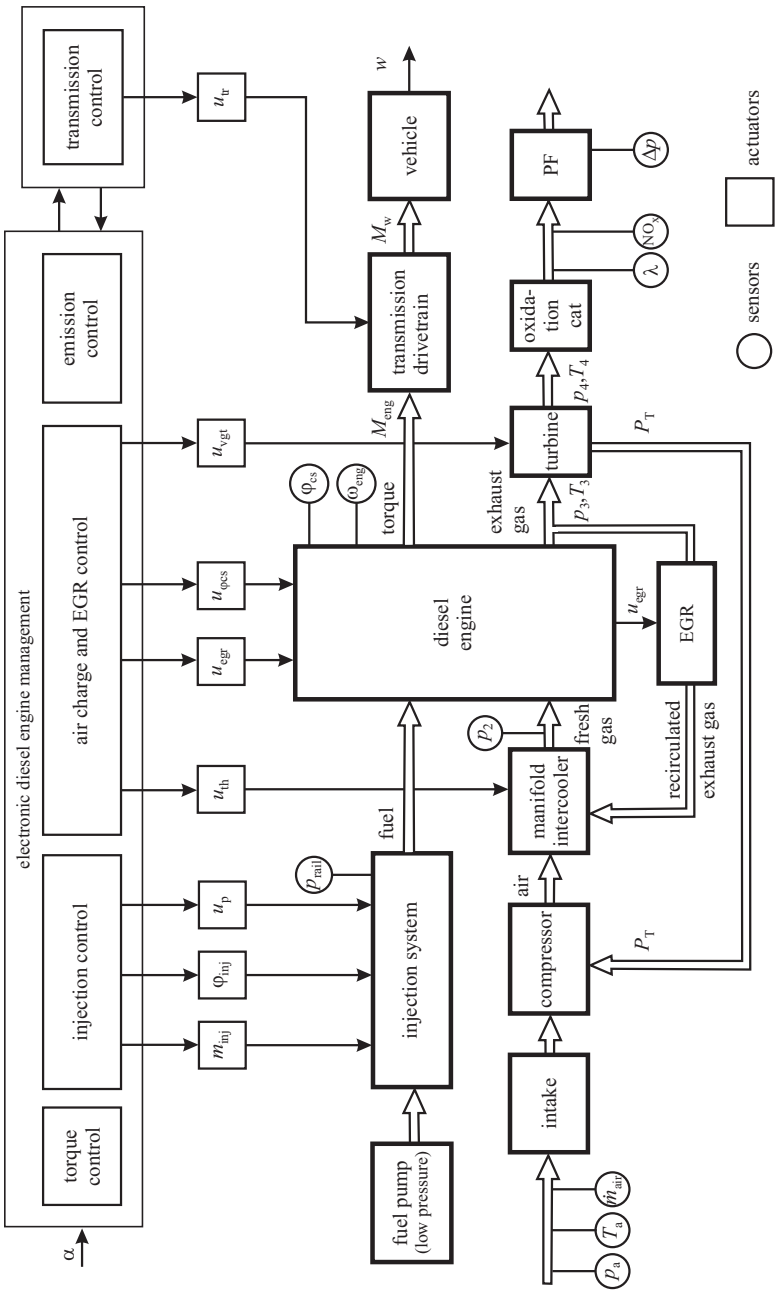


Fig. 3.1.7. Control-oriented block diagram of a diesel engine with common rail injection, actuators, control modules and some sensors.

increased. Therefore, today's vehicles have an on-board diagnostic (OBD) functionality. Its goal is to detect faults and to indicate them to the driver and to change the operation of the engine e.g. by load reduction. The off-board diagnosis is carried out in a workshop by connection with a test equipment, which allows to expand the fault detection coverage. Its intention is to localize the faults down to the smallest possible replaceable unit. Here additional measurements are possible, the inclusion of human observations and use of guided troubleshooting software.

Until about 1979 the *classical supervision* of combustion engines consisted of limit checking the oil pressure, the coolant temperature and board-voltage by alarming the driver through signal lamps. The implementation of microprocessor engine control then allowed not only a more comprehensive engine control, but also much more *on-board diagnosis functions*, see e.g. Robert Bosch GmbH (2012). Monitoring and fault-diagnosis programs check the input and output signals and important control functions during operation. For example, sensor outputs, plug connectors and cables are monitored. This includes the sensor supply voltages, sensor measurement ranges and plausibility checks by comparing different sensor outputs. The ECU output signals to the actuators are monitored with regard to line breaks, short-circuits to battery voltage or vehicle ground and open circuits. Also the actuator positions are monitored with test signals or by plausibility checks. The ECU hardware and software contains many monitoring functions, e.g. for the microcontroller and storages. They may operate after start-up, during normal operation or after switch-off. The communication between ECU's is checked by the detection of transmission errors of the buses like CAN.

After the fault detection *fault-handling procedures* are trigged. If the defective component is detected, replacement values like the last valid value or a substitute value is used allowing a further operation of the engine. The faults are stored in a non-volatile part of the ECU memory by using fault codes and the operating condition at fault occurrence. Warning lamps or text messages indicate severe faults to the driver. Depending on the fault limp-home functions are switched which allow to drive with limited power or speed.

The requirements of the CARB (Californian Air Resources Board) for an *on-board diagnosis* (OBD I) in 1988 extended the monitoring functions. First, only the monitoring of sensors and actuators with the electronic control unit (ECU) was required which influence the emissions. The legal regulations for On-Board Diagnosis II (OBD II) introduced by CARB in 1994 and the European On-Board Diagnosis (EOBD), introduced by the European Union in 1998, forced the development of monitoring methods of all components that cause an increase of exhaust gas emissions because of faults, see e.g. California's OBD II (1996), European Parliament (1998), European Parliament (1999), European Parliament (2002), Achleitner (1997), Stoss and Rupella (2007). EOBD is in operation for passenger cars with gasoline engines since 2000 and with diesel engines since 2003. Statistics have shown that an increase in exhaust gas emissions and decrease in engine performance is in many cases caused by faults in the injection or mixture preparation, by the combustion and by faults in the aftertreatment of exhaust gases, e.g. in the catalytic converter or particulate filter. Faults which pass the OBD limit values by factor 1.5 for USA or

with additively given limits for the EU must be indicated to the driver after the third driving cycle, see e.g. Siemens VDO (2007), DELPHI (2010). The OBD thresholds may be different from the legal emission limits, Almstadt (2007).

The diagnosis functions according to the CARB-OBDII and partially EOBD requirement in 2010 have to include for *gasoline engines*:

- sensors
- actuators
- catalysts
- tank ventilation
- fuel system
- misfire
- exhaust gas recirculation
- NO_x sensors
- secondary air system
- lambda sensors
- coolant thermostat
- cold start emission reduction
- crank case ventilation
- variable valve train
- air conditioning system
- engine cooling system

The on-board diagnosis system for *diesel engines* requires additionally

- oxidation catalyst
- SCR-DeNOx-system
- NO_x-storage catalyst
- particulate filter
- injection mass and timing
- exhaust gas cooler
- charging pressure control
- charging air cooler

The supervision of *sensors* includes a check of the electrical circuit, detecting short circuits to supply and ground, and cable breakage. In addition violation of measurement ranges and plausibility checks are required. The plausibility checks are performed by the comparison with other sensor measurements and signal changes for definite stimulations with test signals. The supervision of *actuators* follows the checks for the electrical circuits as for sensors. Additional functional checks are required as appropriate response to command variables with regard to static precision and reaction time of the actuators.

Some examples for OBD-diagnosis functions are:

- *Catalyst*: A measure for the oxygen storage capacity is obtained by checking the damping of the control generated oscillation amplitudes of the oxygen sensor values before and after the catalyst.
- *Misfire*: the well resolved speed signal shows a decrease, which allows to indicate a misfiring cylinder, e.g. by measuring the time intervals between strokes or parts of them
- *Secondary air*: The oxygen content in the exhaust gases must increase after starting the secondary air pump.
- *λ-sensor*: the continuous range oxygen sensor before the catalyst is checked by a richer fuel/air ratio. The oxygen content then has to decrease.
- *Exhaust gas recirculation*: for increased EGR rate the measured air mass rate has to decrease.
- *Tank leakage*: The air ventilation is closed and a pressure build up by a special diagnosis pump is observed.

Hence, the applied supervision functions operate by special changes of observed signals after stimulation with an appropriate test signal or by comparisons of two

different signals and their signs of signal deviations after stimulation of command variables.

The exceeding of limit values is stored in the on-board-diagnosis system and is indicated to the driver by the MIL-lamp (*malfunction indication lamp*, ISO 2575) after the third driving cycle, if the threshold was passed again. The cars must have a diagnosis plug on the driver side to connect a diagnosis scan tool. This scan tool shows then e.g. readiness (activation of diagnosis functions and diagnosis events), power train data for the fault event, pending and confirmed faults, and various vehicle data. Different protocols are used for the communication, like ISO 9141-2 and ISO 14230-4 in the EU and SAE J 1850 for American passenger cars and SAE J 1708 for U.S. commercial vehicles, Robert Bosch GmbH (2012). The state-of-the-art is e.g. described in Knirsch and Klee (2002), Lücking (2003), Willimowski et al (2005), Pedrelli (2007), SAE (2009), Schäffer (2010), Zimmermann and Schmidgall (2014), Bäker and Unger (2014).

The amount of software for these diagnosis functions presently takes about 50% of the capacity of the engine electronic control units. Because of narrower legal limit values for the tolerated exhaust emissions the efforts for engine diagnosis will increase.

A *software standardization* for the information in vehicle diagnosis systems was introduced by ASAM in 2004 as Open Diagnostic Data Exchange (ODX). It describes the data communication for diagnosis and software programs as a common basis for manufacturers, suppliers, development and after-sales service. ODX is specified with UML (Unified Modeling Language) and uses XML (Extensible Markup Language) for the diagnosis data. Because ASCII-code is used, the data is easily portable. The readable information is also important, see e.g. Marscholik and Subke (2011). According to the ISO 22901-1 the data protocols like UDS (ISO 14229) are standardized.

UDS (Unified Diagnostic Services) standardizes specification of diagnosis service requests and service responses with regard to the ISO/OSI layers of bios systems like CAN. Hence, the UDS data protocols provide data in a standardized format and are the inputs for ODX, which presents the actual measured data with dimensions in a readable form and generates a diagnosis protocol. The standardization of the diagnosis functions is presently in development and is called OTX (Open Text sequence eXchange), Georg et al (2010).

The *off-board diagnosis* is usually applied at service stations. By using special workshop testers the information stored in the fault memory of the ECU is read out over a serial interface. The faults are stored as an error code in the non-volatile area of the ECU data memory.

As the OBD requirements require the supervision of selected functions by limit checking of directly measurable quantities, the expression ”diagnosis”, in the sense of indicating the faulty component is not justified in all cases. Especially the workshops would like to know more details for troubleshooting. A future aspect is the development of *telediagnosis* or *remote-diagnosis* systems, where an onboard-computer or ECU is connected through communication links with capable off-board computers and experts at service stations to support fault-diagnosis and relevant de-

cisions for actions on a remote basis. Therefore more advanced diagnosis methods of powertrain components are considered in the following chapters.

3.3 Control- and diagnosis-oriented subdivision of combustion engines

The development of control and diagnosis functions for internal combustion engines requires a subdivision into signal-flow-oriented parts with physical input and output variables. Table 3.3.1 therefore distinguishes four *engine groups*:

A Basic engine hardware
B Electrical hardware
C Control hardware and software
D Auxiliary components

and subdivides corresponding *engine parts* and shows the dominating physical domains. The engine parts are further detailed in Table 3.3.2 for gasoline engines and in Table 3.3.3 for diesel engines and list the main components, actuators, manipulated variables, sensor variables and auxiliary components for standard engine configurations. The components for the cooling system, electrical system and electronic control system, which are similar for both types of engines are given in Table 3.3.4.

Table 3.3.1. Engine groups and parts.

Engine group	Engine parts	Physical domains
A basic engine hardware	A1 intake system	fluid dynamics, thermodynamics
	A2 injection system	fluid dynamics, mechatronics
	A3 fuel supply	fluid dynamics
	A4 combustion and mechanics	chemistry, thermodynamics, mechanics, materials
	A5 lubrication system	fluid dynamics, heat transfer
	A6 exhaust gas system, exhaust gas recirculation, emission after-treatment	fluid dynamics, thermodynamics, chemistry
	A7 cooling system	fluid dynamics, heat transfer
B electrical hardware	B1 electrical system	low voltage electricity
	B2 ignition system	low & high voltage electricity
C control hardware and software	C1 actuators	mechatronics
	C2 sensors	physics, mechatronics
	C3 electronic control unit	computer technology, control software, algorithms
D auxiliary components	D1 filters	fluid dynamics
	D2 tank ventilation	

Table 3.3.2. Main components of gasoline engines.

Engine part	Components	Actuators	Manipulated variables	Sensors	Auxiliary components
A1 intake system	- intake pipe manifold - blow-by pipe	- el. throttle - intake flaps	- throttle pos. - flap pos.	- air mass flow - manifold temperature - manifold pressure	- air filter
A2 injection system	- high pressure pump (direct injection - injection valves	- metering unit - pressure control valve - injection valve coils or piezo actuat.	- valve pos. - injection duration - injection time	- fuel or rail pressure	
A3 fuel supply	- low pressure fuel pump (intake and direct inj.) - fuel filter	- canister purge valve	- valve pos.	- fuel level - fuel pressure - tank pressure	- tank ventilation - tank diagnosis pump
A4 combustion and mechanics	- combustion chamber - pistons - connecting rods - crankshaft - camshaft - variable valve train - bearings - turbocharger - exhaust gas recirculation	- hydraulic actuators - swirl flaps - tumble flaps - exhaust gas recirculation valve - wastegate actuator	- camshaft position - flap pos. - exhaust gas valve pos. - inlet/outlet valve pos. - wastegate position	- engine speed - camshaft phase - knocking accelerom. - pressures - temperatures	
A5 lubrication	- oil pump - oil channels - mechanics - seals	- oil pump actuator	- oil flow	- oil pressure - oil temp. - oil level	- oil filter
A6 exhaust system	- exhaust pipes - secondary air system - catalytic conv. - muffler		- exhaust gas temperature	- exhaust gas temperature - Lambda - NO_x	
B2 ignition system	- ignition timer - ignition coils - spark plugs - ignition lock	- ignition driver stage	- ignition time/ advance angle	- knock sensor	

Table 3.3.3. Main components of diesel engines.

Engine part	Components	Actuators	Manipulated variables	Sensors	Auxiliary components
A1 intake systems	- intake manifold - blow-by pipe - intercooler	- throttle plate	- throttle plate position	- air mass flow - air temp. - boost pressure	- air filter
A2 injection system	- high pressure pump - glow plugs	- metering unit - pressure control valve - injectors	- valve positions - injection duration - injection time	- fuel pressure	
A3 fuel supply	- low pressure fuel pump - fuel filter			- fuel level - fuel temperature	- tank ventilation
A4 combustion and mechanics	- combustion chamber - pistons - connecting rods - crankshaft - variable valve train - bearings - turbocharger - exhaust gas recirculation cooler	- hydraulic actuators - wastegate - turbine vanes - exhaust recirculation valve - swirl flaps	- exhaust gas flow - valve position - flap position	- engine speed - camshaft phase - camshaft position - boost pressure	
A5 lubrication	- oil pump - oil channels - mechanics - seals	- oil pump actuator	- oil flow	- oil pressure - oil temperature - oil level	- oil filter
A6 exhaust system	- exhaust pipes - catalytic oxygen converter - particulate filter - muffler		- exhaust gas temperature for regeneration	- gas temp. - lambda - NO_x - difference pressure	

Table 3.3.4. Components of cooling, electrical and electronic systems for combustion engines (gasoline and diesel engines)

Engine part	Components	Actuators	Manipulated variables	Sensors
A7 cooling system	- coolant radiator - coolant pump - coolant passages - flexible tubes	- thermostatic expansion valve - air fan - air inlet flaps	- valve position - fan speed - speed of electrical pump	- coolant temperature - coolant level
B1 electrical system	- electrical generator - electrical starter - battery - harness - fuses - CAN bus system	- power transistors - overvoltage protection	- voltages - currents - switches	- voltages - currents
C3 electronic control unit	- ECU		- electrical outputs	- electrical inputs

3.4 Model-based fault detection of combustion engines

Fault detection methods which allow an access to the fault origins require component specific measurements, if possible with input and output variables.

In the case of *actuators* the input variables are known from their command variables of the ECU if not directly measured, e.g. as a position or a current. Some actuators, as the electrical throttle (already investigated in Isermann (2006)), injection pump with pressure measurement, hydraulic camshaft actuators and EGR valves with position measurement have also output measurements. For other actuators no corresponding output variables are measured usually, like flaps in the intake system, injection valves and ignition coils. Then the actuators have to be considered together with the engine components they are build in.

The listed components in Tables 3.3.2 and 3.3.3 indicate that several measurements are available for the intake system, the fuel supply, combustion and mechanics, lubrication circuit, cooling system and exhaust system.

Input signal measurements which can be used for fault detection of the combustion are air mass flow rate, manifold or charging air pressure and temperature, rail pressure, camshaft phase and exhaust gas recirculation valve position. Additionally known inputs are the manipulated variables from the ECU, like injection fuel mass, injection angle, ignition angle (gasoline engines), positions of swirl and tumble actuators. However, their real physical values are not known, as they are usually not directly measured. Only cylinder pressure measurement would allow to reconstruct some of these commanded engine inputs.

Available *output measurements* for the *combustion* and the *mechanical parts* are the angular speed of the flywheel, the excess air ratio λ and, if measured, the exhaust gas temperature, exhaust pressure and NO_x concentration.

Some of the measured variables operate in *closed loops*, like the air flow or manifold pressure, charging pressure, fuel pressure, rail pressure, excess air factor λ and

coolant water temperature. As discussed in Sect. 2.1 the fault detection in closed loops is in general a difficult task.

These discussions show that with regard to the development of *fault detection methods* for combustion engines the various engine parts have to be treated individually. Therefore, some fault detection modules will be defined in the sequel which comprise connected engine parts with regard to the signal flow. Depending on the physical properties, the available measurements, the signal changes and the operating conditions, fault detecting methods like limit checking, parity equations, parameter estimation and signal analysis methods can be designed, see Fig. 3.4.1.

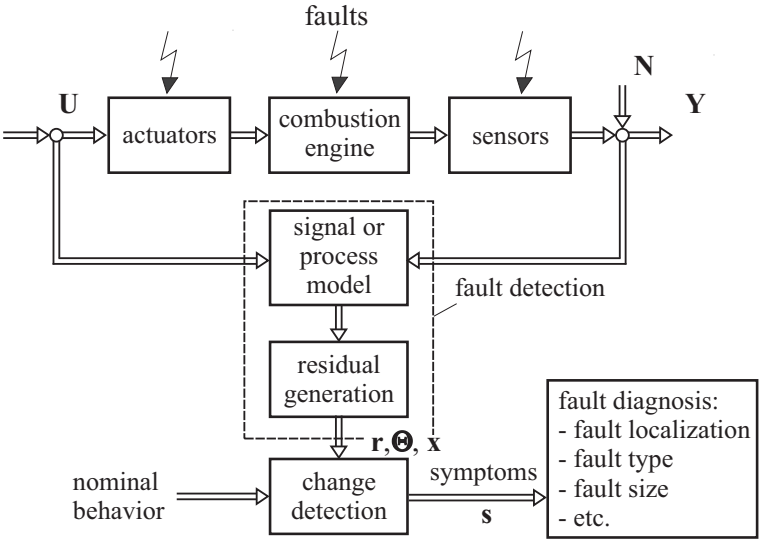


Fig. 3.4.1. Schematic for the fault detection and diagnosis with signal- and process-model-based methods.

For combustion engines linear and nonlinear multi-input single-output (MISO) models are used either based on physical laws or on experiments with identification methods, e.g. in form of polynomial models, special neural networks or look-up-tables for stationary behavior and corresponding types for the dynamic behavior as described in Isermann (2014). For the subsequent *fault-diagnosis* classification or inference methods, including fault-symptom trees, fuzzy-rules etc. are possible choices.

Publications on research activities for fault detection of combustion engines are, e.g. Plapp et al (1990), Führer et al (1993), Nielsen and Nyberg (1993), Ribbens and Rizzoni (1990), Leonhardt (1996), Schüler et al (1996), Gertler (1998), Bidian et al (1995), Nyberg (1999), Willimowski et al (1999), Antory et al (2005), Isermann (2005), Kimmich et al (2005), Beer et al (2006), Franchek et al (2007), Pedrelli (2007), Joshi et al (2009), Eriksson and Nielsen (2014). A survey is given in Mohammadpour et al (2012).

Figure 3.4.2 shows the concept for developed model-based fault-detection and fault-diagnosis systems of a diesel engine, Kimmich et al (2005). The engine is partitioned in several major subsystems as intake air system, injection system, combustion and mechanics, exhaust gas system, lubrication and cooling system. The actuators are commanded by the electronic control unit and act on different components of the subsystems. Based on sensor outputs the fault detection modules detect faults and abnormalities in the components by comparison with a normal status, and after special filtering, and generate symptoms. The symptoms are then combined and processed with diagnosis methods to diagnose, i.e. localize the fault and the corresponding component. This is described in detail in the next two chapters for gasoline and diesel engines. These model-based fault-detection and diagnosis methods then allow to expand the fault-diagnosis coverage of the classical, legislative OBD systems.

References

Achleitner E (1997) Erfahrungen mit den OBDII – Funktionen aus dem Serieneinsatz, Ausblick auf die Weiterentwicklung der Diagnosefunktionen für US-Anwendungen. Wolfsburg

Almstadt K (2007) Trends for OBD threshold limits in Europe and the US. In: Pedrelli O (ed) Onboard-Diagnose II, expert Verlag, Renningen

Antory D, Kruger U, Irwin G, McCullough G (2005) Fault diagnosis in internal combustion engines using non-linear multivariate statistics. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 219(4):243–258

Bäker B, Unger A (eds) (2014) Diagnose in mechatronischen Fahrzeugsystemen VIII. TUDpress, Verlag der Wissenschaften, Dresden

Bargende M (2013) Grundlagen der Verbrennungsmotoren. Lecture notes. Institut für Verbrennungsmotoren und Kraftfahrenwesen. Universität Stuttgart, Stuttgart

van Basshuysen R (2009) Gasoline engine with direct injection. Vieweg+Teubner, Wiesbaden

Beer J, Kainz J, Ketterer A (2006) Diagnosis status of a modern engine management system in view of misfire events. In: AUTOREG 2006, VDI, Wiesloch, Germany, vol VDI-Berichte 1931, pp 243–258

Beidl C (2014) Verbrennungskraftmaschinen I + II. Lecture notes. Technische Universität Darmstadt, Institute for Internal Combustion Engines and Powertrain Systems, Darmstadt

Bidian P, Carlén C, Cascio F, Tatar M, Struss P, Weber R (1995) Vehicle model-based diagnosis (BRITE-EURAM projects). Synthesis report BE95-2128-BRPR-CT95-0138

California's OBD II (1996) Requirements: Title 13, California Code of Regulations (CCR) Section 1968 Resolution 95-34. State of California, Sacramento, CA, USA

Cook J, Grizzle J, Sun J (1996) Automotive control systems. In: Levine W (ed) The Control Handbook, CRC Press, Boca Raton, pp 1261–1274

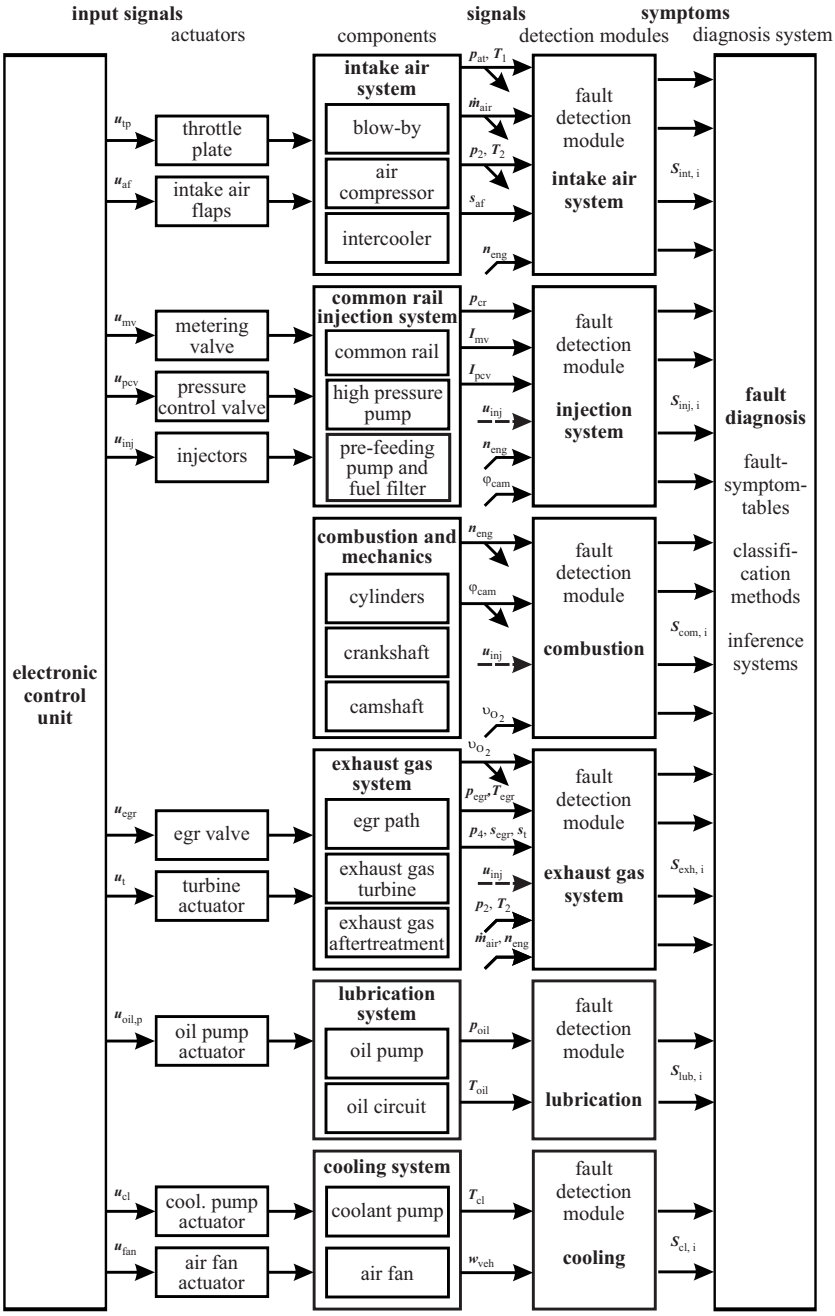


Fig. 3.4.2. Modular structure for the fault detection of combustion engines (for the example of a diesel engine).

DELPHI (2010) Worldwide emission standards. Heavy duty and off-road vehicles. http://delphi.com/pdf/emissions/Delphi-Heavy-Duty-Emissions-Brochure-2010-2011.pdf. DELPHI, Troy, Michigan

Eriksson L, Nielsen L (2014) Modeling and control of engines and drivelines. John Wiley & Sons, Chichester

ETAS (2010) Measurement, electronic control calibration and diagnosis. Brochure. ETAS/COM-Sar, Stuttgart

European Parliament (1998) Directive 98/69/EC of the European Parliament and of the Council of 13 October 1998 relating to measures to be taken against air pollution by emissions from motor vehicles and amending Council Directive 70/220/EEC. European Parliament, Luxembourg

European Parliament (1999) Directive 99/69/EC of the European Parliament. European Parliament, Luxembourg

European Parliament (2002) Directive 02/88/EC of the European Parliament. European Parliament, Luxembourg

Ferguson C, Kirkpatrick A (2001) Internal combustion engines. J. Wiley and Sons, New York

Franchek M, Buehler P, Makki I (2007) Intake air path diagnostics for internal combustion engines. Journal of Dynamic Systems, Measurement, and Control 129(1):32–40

Führer J, Sinsel S, Isermann R (1993) Erkennung von Zündaussetzern aus Drehzahlsignalen mit Hilfe eines Frequenzbereichsverfahrens. In: Proceedings 13. Tagung Elektronik im Kraftfahrzeug, Haus der Technik Essen, Germany

Georg A, Martin F, Buse A (2010) The significance of standards in the diagnostic processs chain. ATZelektronik worldwide 5(3):26–30

Gertler J (1998) Fault detection and diagnosis in engineering systems. Marcel Dekker, New York

Guzzella L, Onder C (2010) Introduction to modeling and control of internal combustion engine systems, 2nd edn. Springer, Berlin

Heywood JB (1988) Internal combustion engine fundamentals. McGraw-Hill, New York

Isermann R (ed) (2003) Modellgestützte Steuerung, Regelung und Diagnose von Verbrennungsmotoren. Springer, Berlin

Isermann R (2005) Mechatronic systems – fundamentals, 2nd edn. Springer, London

Isermann R (2006) Fault-diagnosis systems – An introduction from fault detection to fault tolerance. Springer, Heidelberg

Isermann R (ed) (2010) Elektronisches Management motorischer Fahrzeugantriebe. Vieweg+Teubner, Wiesbaden

Isermann R (2014) Engine Modeling and Control. Springer, Berlin

Johansson R, Rantzer A (eds) (2003) Nonlinear and hybrid systems in automotive control. Springer, London

Joshi A, James S, Meckl P, King G, Jennings K (2009) Assessment of charge-air cooler health in diesel engines using nonlinear time series analysis of intake manifold temperature. Journal of Dynamic Systems, Measurement, and Control 131(4):041,009

Jurgen R (ed) (1999) Automatic electronics handbook, 2nd edn. McGraw-Hill, New York

Kiencke U, Nielsen L (eds) (2000) Automotive control systems. For engine, driveline and vehicle. Springer, Berlin

Kimmich F, Schwarte A, Isermann R (2005) Fault detection for modern diesel engines using signal- and process model-based methods. Control Engineering Practice – CEP 13(2):189–203

Knirsch M, Klee P (2002) On-board diagnosis (OBD) for engine emission control systems. In: Emission control 2002, Dresden, Germany

Köhler E, Flierl R (2012) Verbrennungsmotoren, 6th edn. Vieweg, Wiesbaden

Leonhardt S (1996) Modellgestützte Fehlererkennung mit neuronalen Netzen - Überwachung von Radaufhängungen und Diesel-Einspritzanlagen. Dissertation Technische Hochschule Darmstadt. Fortschr.-Ber. VDI Reihe 12, 295. VDI Verlag, Düsseldorf

Lücking C (ed) (2003) On-Board Diagnose. Konferenz Haus der Technik. Essen, Germany

Marscholik C, Subke P (2011) Datenkommunikation im Automobil: Grundlagen, Bussysteme, Protokolle und Anwendungen, 2nd edn. VDE, Offenbach

Merker G, Schwarz C, Teichmann R (eds) (2012) Combustion engines development. Springer, Berlin

Mohammadpour J, Franchek M, Grigoriadis K (2012) A survey on diagnostics methods for automotive engines. International Journal of Engine Research 13:41–64

Mollenhauer K, Tschöke H (2010) Handbook of diesel engines. Springer, Berlin

Nielsen L, Nyberg M (1993) Model-based diagnosis for the air intake system of si-engines. In: SAE World Congress, Detroit, Michigan, USA.

Nyberg M (1999) Model-based fault diagnosis: Methods, theory, and automotive engine applications. Ph.D. thesis. Linköping University, Linköping

Pedrelli O (ed) (2007) On-board Diagnose II. expert Verlag, Renningen

Pischinger R, Kell M, Sams T (2009) Thermodynamik der Verbrennungskraftmaschinen, 3rd edn. Springer, Wien and New York

Pischinger S (2013) Internal combustion engine fundamentals. Lecture notes. RWTH Aachen, Institute for Combustion Engines, Aachen

Plapp G, Klenk M, Moser W (1990) Methods of on-board misfire detection. In: SAE Technical Paper 900232, Warrendale

Pucher H, Zinner K (2012) Aufladung von Verbrennungsmotoren, 4th edn. Springer, Berlin

Ribbens W, Rizzoni G (1990) Onboard diagnosis of engine misfires. In: Proc. SAE 90, Warrendale, PA, SAE 901768

Robert Bosch GmbH (ed) (2006) Gasoline-Engine Management, 3rd edn. Bentley publishers, Cambridge

Robert Bosch GmbH (ed) (2012) Automotive Handbook, 12th edn. Bentley publishers, Cambridge

SAE (2009) On-Board Diagnostics for Light and Medium Duty Vehicles Standards Manual – 2010 Edition. SAE International, Warrendale, Pennsylvania

Schäffer F (2010) Fahrzeugdiagnose mit OBD, 3rd edn. Elektor-Verlag, Aachen

Schüler M, Leonhardt S, Ludwig C, Ayoubi M, Isermann R (1996) Model-based and signal-based fault detection of diesel engines with turbocharger. In: 13th IFAC World Congress, San Francisco, USA

Siemens VDO (2007) Worldwide emission standards and related regulations. www.vdo.com. Siemens VDO, Regensburg

Spicher U (ed) (2012) Direkteinspritzung im Ottomotor VIII. expert Verlag, Renningen

Stone R (1999) Introduction to internal combustion engines, 3rd edn. SAE International and Maxmillan Press, Warrendale

Stoss P, Rupella A (2007) Technische und legislative Randbedingungen für die Implementierung der Software zum Testen der OBD- und EOBD-Funktionen. In: Pedrelli O (ed) Onboard-Diagnose II, expert Verlag, Renningen

Stotsky A (2009) Automotive engines. Springer, Berlin

Urlaub A (1995) Verbrennungsmotoren. Grundlagen - Verfahrenstheorie - Konstruktion, 2nd edn. Springer, Berlin

Wachtmeister G (2014) Verbrennungsmotoren. Lecture notes. Technische Universität München, Institute of Internal Combustion Engines, München

Willimowski M, Füssel D, Isermann R (1999) Misfire detetion for spark-ignition engines by exhaust gas pressure analysis. MTZ worldwide pp 8–12

Willimowski M, Klee P, Küsell M, Knirsch M (2005) Zukünftige Zukünftige Diagnose in Motorsteuerungen. In: 5. Symposium Steuerungssysteme für den Antriebsstrang von Kraftfahrzeugen, IAV GmbH, Berlin, Germany

Zimmermann W, Schmidgall R (2014) Bussysteme in der Fahrzeugtechnik, 5th edn. Springer Fachmedien, Wiesbaden

4

Diagnosis of gasoline engines

The model-based approach for the fault diagnosis of gasoline engines follows a modular structure. The engine is divided in engine parts and their components, actuators, standard sensors and additional sensors. Following fault-detection modules are defined, compare Tables 3.3.2 and 3.3.3:

A1 Intake system
A2 Fuel injection system
A3 Fuel supply
A4 Combustion system and mechanics
A5 Lubrication system
A6 Exhaust gas system, emission aftertreatment
A7 Cooling system
B1 Electrical system
B2 Ignition system
C3 Electronic control.

The following sections describe a selection of advanced fault-detection and diagnosis methods. These methods have to be considered in addition to the established OBD functions, sketched in Sect. 3.2. As well signal-model as process-model-based fault-detection methods are developed. The process models used for model-based fault detection are based on physical laws or on experimental models, or on both, called semi-physical models.

4.1 Intake system (air path manifold)

4.1.1 Fault diagnosis of the intake system with physical models

In the following, the detection of sensor faults, leakage and clogging in the intake system is considered.

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7_4

Air charge determination
As the air charge into the cylinders m_{air} cannot be measured directly it has to be reconstructed based on the measurements of other variables, like air flow rate $\dot{m}_1$ before the throttle, manifold pressure p_2 or throttle position α_{th}, see also Müller (2003a), Guzzella and Onder (2004), Isermann (2014).

Air flow measurement
The most direct method is to measure the air flow rate $\dot{m}_1$ before the throttle and after the air filter, e.g. by a hot-film sensor HFM. In contrast to other flow meters, the HFM sensor directly measures the mass flow rate, possesses a wide measurement range of $1:50$ and has a small time constant. In stationary condition it holds for the air flow rate into the cylinders

$$\dot{m}_{2,\mathrm{air}}(t) = \dot{m}_{\mathrm{HFM}}(t) = \dot{m}_1(t) \tag{4.1.1}$$

if no gases enter the manifold, i.e. for $\dot{m}_{\mathrm{add}} = 0$. In dynamic operation, however, $\dot{m}_{2,\mathrm{air}}$ into the cylinders is delayed by the intake manifold pressure dynamics. This can be taken into account by, Isermann (2014)

$$^{1}\tilde{\dot{m}}_{2,\mathrm{air}}(t) = \dot{m}_1(t) - \frac{V_{\mathrm{int}}}{RT_2}\frac{dp_2(t)}{dt}, \tag{4.1.2}$$

where V_{int} is the intake volume. Correcting the measured air flow rate by the first derivative of the manipulated pressure allows to determine the air charge in dynamic operation. Because of the included noise and periodicity in $p_2(t)$ this signal has to be low-pass-filtered before the derivation, see Schwarte et al (2002). Further required measurements are p_2, T_2. If additional gases $\dot{m}_{\mathrm{add}}$ enter the manifold intake, the manifold pressure p_2 does not only depend on $\dot{m}_{\mathrm{air,in}}$ but also on $\dot{m}_{\mathrm{add}}$, see Isermann (2014). Then, $p_2(t)$ and dp_2/dt changes. However, the dynamic correction of (4.1.2) is approximately applicable, if e.g. $\dot{m}_{\mathrm{egr}}$ is not very large compared to $\dot{m}_1(t)$.

Manifold pressure measurement
In order to save the air flow sensor, the charge air can be determined by the manifold pressure p_2. The gas flow rate $\dot{m}_{\mathrm{gas,cyl}}$ into the cylinders follows, taking the additional gases with flow rate $\dot{m}_{\mathrm{add}}$ into account

$$^{2}\tilde{\dot{m}}_{2,\mathrm{air}}(t) = \frac{\eta_{\mathrm{v}}\,(n,\rho_2)\;V_{\mathrm{D}}\,n}{2\,R\,T_2}p_2(t) - \dot{m}_{\mathrm{add}}(t). \tag{4.1.3}$$

As the sucked gas directly depends on the dynamically measured $p_2(t)$ the dynamics of the intake manifold are already included. This air charge determination needs the measurement of p_2 and T_2, the calibrated volumetric efficiency $\eta_{\mathrm{v}}(n,\rho_2)$, due to (5.1.2) and determination of $\dot{m}_{\mathrm{add}}$.

Measurement of throttle position
Another possibility to circumvent an air flow sensor is to base the air charge determination on the known throttle angle α_{th}. Then, the flow equation of a throttle can be used

$$^{3}\tilde{\dot{m}}_{2,\mathrm{air}}(t) = c_{\mathrm{th}}\left(\alpha_{\mathrm{th}}, n\right) A\left(\alpha_{\mathrm{th}}\right) p_{\mathrm{a}} \sqrt{\frac{2}{RT_{\mathrm{a}}}} \Psi\left(\frac{p_2(t)}{p_{\mathrm{a}}}\right) - \frac{V_{\mathrm{int}}}{RT_2} \frac{dp_2(t)}{dt}, \quad (4.1.4)$$

where $A(\alpha_{\mathrm{th}})$ is the effective cross section. This method requires measurement of $\alpha_{\mathrm{th}}, p_{\mathrm{a}}, T_{\mathrm{a}}, p_2, T_2$ and calibration of the loss factor $c_{\mathrm{th}}\left(\alpha_{\mathrm{th}}, n\right)$. The final air charge mass for one cylinder follows according to

$$m_{\mathrm{air,cyl}} = \tilde{\dot{m}}_{2,\mathrm{air}} \frac{2}{i_{\mathrm{c}} n}. \quad (4.1.5)$$

Table 4.1.1 shows a comparison for the required sensors of the three methods. In all cases p_2, T_2 are required. Using the manifold pressure p_2 is dynamically fast but requires precise calibration of the volumetric efficiency and determination of additional gas flows $\dot{m}_{\mathrm{add}}$. The application of the throttle angle needs also ambient pressure p_{a} and temperature T_{a} and precise calibration of the throttle loss factor c_{th}. The required measurements for stationary operating conditions only are summarized in Table 4.1.2. Because the dynamic compensation is then not needed, p_2 or T_2 must not be measured for some cases.

A comparison of the three ways of air flow determination was made by Müller (2003a). Most suited is the direct air flow measurement with dynamic compensation. The accuracy of the other methods depends much on a precise calibration of the correction factors. Average deviations of about 2.5% could be reached.

Table 4.1.1. Methods for the determination of the air flow rate into the cylinders with different sensors and dynamic compensation of the intake manifold delay.

Basic sensor		Additional sensors						Calibration
		p_{a}	T_{a}	p_2	T_2	n	$\dot{m}_{\mathrm{add}}$	of
air mass flow	$\dot{m}_1$	-	-	✓	✓	-	✓	V_{int}
manifold pressure	p_2	-	-	-	✓	✓	✓	η_{v}
throttle angle $p_2/p_{\mathrm{a}} \geq 0.528$	α_{th}	✓	✓	✓	✓	✓	-	c_{th},V_{int}
throttle angle $p_2/p_{\mathrm{a}} < 0.528$	α_{th}	✓	✓	-	✓	✓	-	c_{th},V_{int}

Sensor faults

If the intake manifold is equipped with sensors for air flow rate, pressure, temperature and throttle angle, then there exists some analytical redundancy. The three methods for air flow determination treated in Isermann (2014) can be used to formulate output residuals of parity equations. This becomes straight forward for the case of stationary behavior, $\dot{m}_{\mathrm{egr}} = 0$, overcritical p_2/p_{a} and negligible other mass flows $\dot{m}_{\mathrm{add}}$, since then different sensors are used. Three residuals can be determined:

$$r_{\mathrm{mair1}} = \dot{m}_1 - {}^{1}\tilde{\dot{m}}_{2,\mathrm{air}}\left(p_2, T_2\right) \quad (4.1.6)$$

$$r_{\mathrm{mair2}} = {}^{2}\tilde{\dot{m}}_{2,\mathrm{air}}\left(p_2, T_2\right) - {}^{3}\tilde{\dot{m}}_{2,\mathrm{air}}\left(p_{\mathrm{a}}, T_{\mathrm{a}}\right)_{\mathrm{overcrit.}} \quad (4.1.7)$$

$$r_{\mathrm{mair3}} = \dot{m}_1 - {}^{3}\tilde{\dot{m}}_{2,\mathrm{air}}\left(p_{\mathrm{a}}, T_{\mathrm{a}}\right)_{\mathrm{overcrit.}} \quad (4.1.8)$$

Table 4.1.2. Methods for determination of the air flow-rate into the cylinders for stationary operation.

Basic sensor		Additional sensors						Calibration
		p_a	T_a	p_2	T_2	n	$\dot{m}_\mathrm{add}$	of
air mass flow	$\dot{m}_1$	-	-	-	-	-	✓	-
manifold pressure	p_2	-	-	-	✓	✓	✓	η_v
throttle angle $p_2/p_\mathrm{a} \geq 0.528$	α_th	✓	✓	✓	-	✓	-	c_th
throttle angle $p_2/p_\mathrm{a} < 0.528$	α_th	✓	✓	-	-	✓	-	c_th

The used measurements are depicted in Fig. 4.1.1. These residuals are becoming symptoms of the air system (resp. intake system) S_int1, S_int2, S_int3 if they exceed certain thresholds and are valid for the respective engine operating point. Table 4.1.3 lists the resulting changes of these symptoms for offset faults of six sensors. The resulting patterns are different and therefore the faults are isolable. However, if one of the three basic sensors $\dot{m}_1$, p_2 and α_th for air flow determination is missing, only one residual can be calculated and a unique determination of the faulty sensor is not possible.

Table 4.1.3. Expected fault-symptom table for *positive offset sensor faults*, stationary engine operation, overcritical flow through throttle, $\dot{m}_\mathrm{egr} = 0$ and no additional gas flows $\dot{m}_\mathrm{add} = 0$.

Sensor faults		Symptoms		
		S_int1	S_int2	S_int3
		r_mair1	r_mair2	r_mair3
air mass flow	$\Delta\dot{m}_1$	$+$	0	$+$
manifold pressure	Δp_2	$-$	$+$	0
manifold temperature	ΔT_2	$-$	$-$	0
ambient pressure	Δp_a	0	$+$	0
ambient temperature	ΔT_a	0	$-$	$+$
throttle angle	$\Delta\alpha_\mathrm{th}$	0	$-$	$-$

Leakages and clogging

As *leaks* are additive faults to the gas flows the parity equations (4.1.6) to (4.1.8) can be used, as for offset sensor faults. Table 4.1.4 shows the resulting symptoms.

The residual r_mair3 expresses differences between the measured air flow and the calculated air flow through the throttle. Its increase indicates a leak between the air flow sensor and the throttle. A leak between the throttle and the cylinder input increases p_2 and reduces $\dot{m}_1$. Therefore, residuals r_air1 and r_air3 become negative. (A residual for the manifold pressure can also be formed. However, its information is included in r_air2 as it depends directly on p_2 due to (4.1.3)).

A further symptom for leakage can be obtained from the λ-controller output. Additional air flow with regard to the measured $\dot{m}_1$, which determines the injected

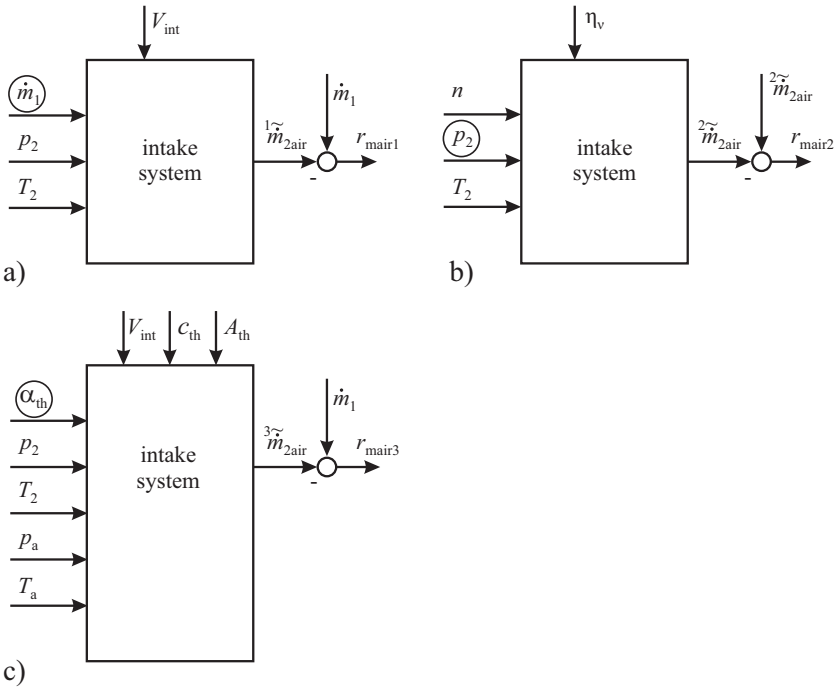


Fig. 4.1.1. Used measurements for the fault detection of the intake system with parity equations by using the air mass flow rate. Basic measurements are: **a** mass flow rate. **b** manifold pressure. **c** throttle position.

fuel mass, leads to a larger correcting factor $c_\lambda > 1$ within the ECU. Therefore, the residual

$$r_{\text{exh1}} = r_\lambda = c_\lambda - 1.0 \tag{4.1.9}$$

is an indication for a leakage after the air flow measurement, leading to symptom S_{exh1}.

Partial *clogging* or pollution with deposits is typical for the air filter in front of the throttle, at the throttle cross sectional area, in the manifold after the exhaust gas recirculation connection and at swirl or tumble flaps. The main effect is generally an increase of flow resistance which shows up in a change of the manifold pressure and/or gas flow and reduces the volumetric efficiency. For a clogged air filter the residual r_{mair2} and r_{mair3} decrease because of the decreased pressure before the throttle and for clogging within the manifold r_{mair2} decreases because of reduced volumetric efficiency η_v.

The resulting patterns for the symptoms S_{int2} and S_{int3} in Table 4.1.4 are different and therefore the leakages and the location of clogging can be diagnosed. A comparison with Table 4.1.3 shows, that for positive and negative sensor offsets the patterns of the symptoms S_{int1},S_{int2} and S_{int3} (with one exception) are different and therefore isolable. The overcritical pressure ratio over the throttle was assumed, be-

cause then the air flow equation for the throttle (4.1.4) is less dependent on calibrated parameters.

Summarizing, the measurement of the signals shown in Fig. 4.1.1 allow to detect and isolate sensor faults, leakages and partial clogging. However, this is no more possible if one of the sensors air flow, throttle angle and manifold pressure is missing.

Table 4.1.4. Expected fault-symptom table for the intake system with *physical models* of the air mass flow for leaks and clogging with overcritical flow through throttle and no additional gas flows, $\dot{m}_{\mathrm{add}} = 0$.

Faults	Symptoms and residuals			
	S_{int1}	S_{int2}	S_{int3}	S_{exh1}
	r_{mair1}	r_{mair2}	r_{mair3}	r_{λ}
leakage before throttle	$-$	0	$-$	0
leakage after throttle	$-$	$+$	$-$	$+$
clogging of air filter	$-$	0	0	0
clogging within manifold	$-$	$-$	0	0

4.1.2 Fault diagnosis of the intake system with experimentally identified models

a) Case 1: fuel stratified direct injection gasoline engine

As treated in the last section the physically based models for the behavior of the intake manifold system require several experimentally determined parameters, correcting factors in the form of at least two dimensional look-up tables. An alternative is then to apply directly nonlinear models which are developed by identification methods. This was investigated by Hartmanshenn and Isermann (2005) and Leykauf and Isermann (2008).

A direct-injection gasoline engine VW FSI 1.6 l (max. power 81 kW, max. torque 155 Nm) on a dynamic test rig was used for the experimental investigations. The engine combustion changes, after the warm-up phase depending on the engine load, between different operating modes: homogeneous ($\lambda = 1$), stratified-homogeneous ($\lambda = 1.5$) and stratified mode ($\lambda > 1$). Figure 4.1.2 depicts the standard and some additional sensors of the investigated engine. A scheme of the modular structure for the diagnosis system is presented in Fig. 4.1.3.

Modeling with local linear net models

The intake system is experimentally modeled with local linear net models and a multiple-input single-output (MISO) structure. The used output signals for the intake system are air flow $\dot{m}_{\mathrm{air}}$ and manifold pressure p_2 with the input signals shown in Fig. 4.1.4. The identification and parameter estimation was performed with the LOLIMOT method, Hartmanshenn and Isermann (2005), Isermann and Münchhof

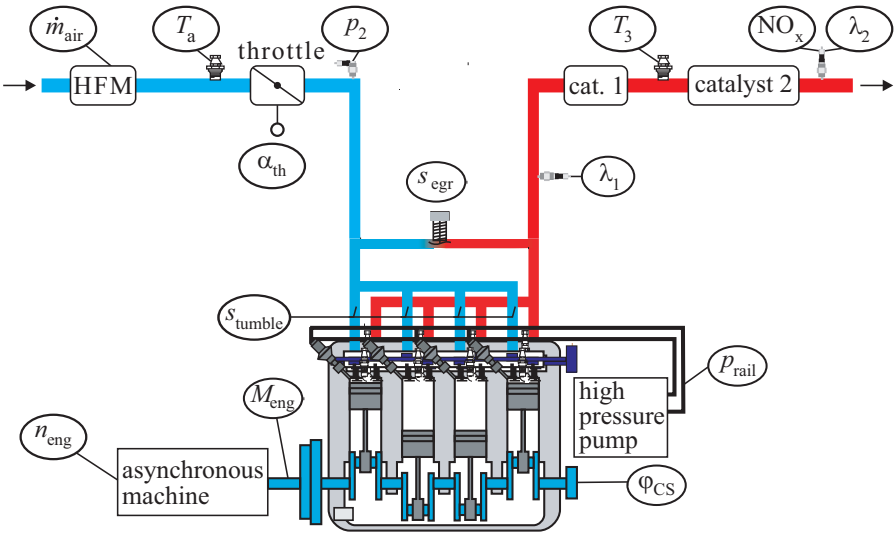


Fig. 4.1.2. Scheme of the DI gasoline engine with sensors. VW FSI 1.6 l, 81 kW, 155 Nm.

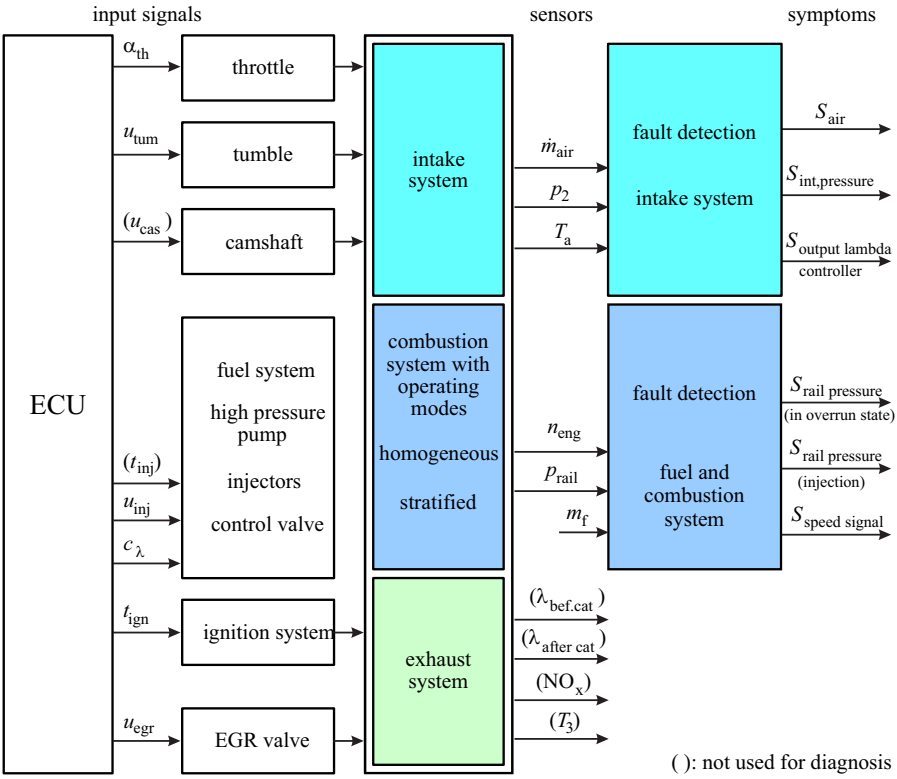


Fig. 4.1.3. Modular structure of the diagnosis system for the investigated DI gasoline engine.

(2011). For each operating mode of the engine separate LOLIMOT models were generated.

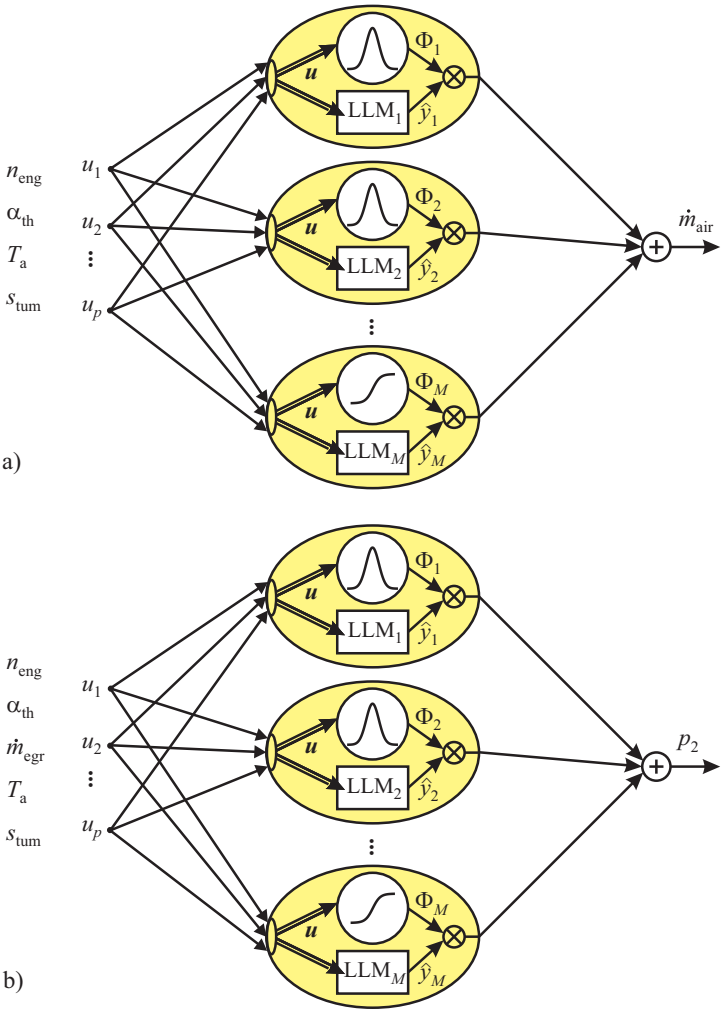


Fig. 4.1.4. Nonlinear MISO-LOLIMOT models for **a** air flow, **b** manifold pressure.

With regard to a theoretical intake manifold model in Isermann (2014), Chap. 4, the model a) corresponds to the throttle and model b) to the throttle and intake volume. The temperature T_2 in the manifold is not applied. The used input test signals for engine speed and pedal value for the measurements in the homogeneous mode are shown in Fig. 4.1.5, and cover the speed-load-range $n_{eng} = 1200 \ldots 3700\,\text{rpm}$ and $M_{eng} = -10 \ldots 110\,\text{Nm}$. They consist of a sequence of step and ramp func-

tions. In stratified mode the operating area was limited to $n_{\mathrm{eng}} = 1200 \ldots 3000\,\mathrm{rpm}$ and $M_{\mathrm{eng}} = -10 \ldots 60\,\mathrm{Nm}$.

The mean error of the identified LOLIMOT models between measured and simulated outputs is maximum 3 percent of the maximum value for extra validation data.

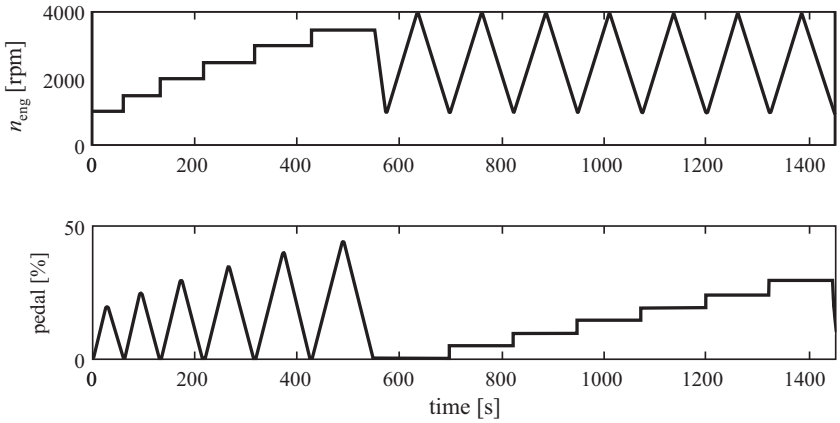


Fig. 4.1.5. Test signals for the input signals engine speed and pedal position for the identification in the fault free case.

Generation of residuals and symptoms with experimental models
Output residuals for the mass flow $\dot{m}_1$, the manifold pressure p_2 and the λ-controller output u_λ are used to establish parity equations, compare Fig. 4.1.6.

$$\begin{aligned} r_{\mathrm{int1}} &= r_{\mathrm{m_{air}}} = \dot{m}_1 - \dot{m}_{1,\mathrm{mod}} \\ r_{\mathrm{int2}} &= r_{\mathrm{p_2}} = p_2 - p_{2,\mathrm{mod}} \\ r_\lambda &= r_4 = c_\lambda - 1. \end{aligned} \tag{4.1.10}$$

The residual $r_{\mathrm{p_2}}$ is positive in case of a leakage after the throttle valve, as due to the low pressure in the manifold air flows in and results in a higher manifold pressure. Dependent on the residual a post-processing is applied. It consists of a low-pass filter and masking if the residual is not valid in a certain operation area. This filtered residual is then one input in a map for the thresholds with the additional inputs for the operating mode, engine speed and torque, resulting in a symptom, Fig. 4.1.7. In total, three symptoms were generated for the fault detection of the intake and exhaust system: S_{int1}: air mass flow; S_{int2}: manifold pressure; S_{exh1}: output of the λ-controller in homogenous operation mode.

Operation-point-dependent fault detection
As the output signals p_2, $\dot{m}_1$ and λ-controller output c_λ depend on the operation point $[M_i, n_j]$ of the engine, the complete operation area was measured with different fault cases. Measurements with small and large leaks before and after the throttle

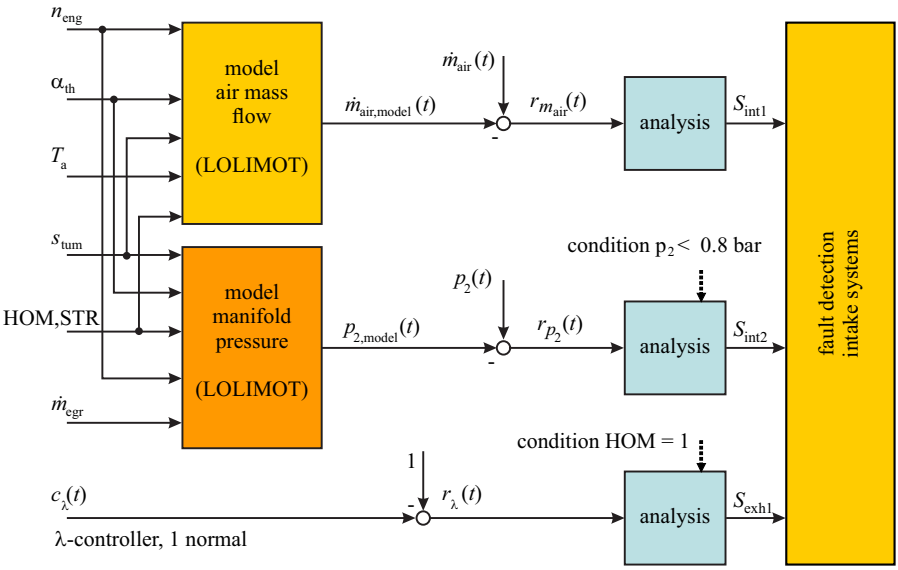


Fig. 4.1.6. Signal-flow diagram for the fault detection with parity equations of the intake system (DI gasoline engine).

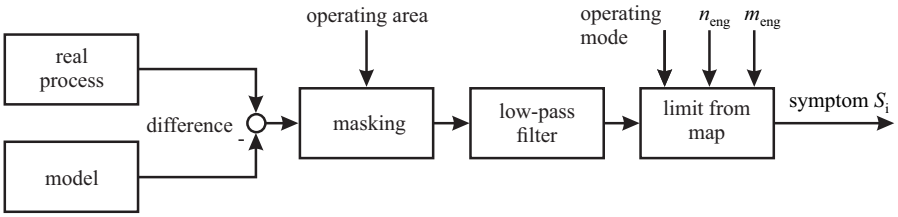


Fig. 4.1.7. Generation of residuals, their filtering and masking and resulting symptoms.

valve and different restrictions before the first catalyst to simulate a higher EGR rate were performed. Each measurement results in one static map for air flow and manifold pressure, generated with LOLIMOT identification. For example the calculated map of the manifold pressure in the fault free case in homogeneous mode is shown in Fig. 4.1.8. The differences between the maps without and with faults are the residuals in the complete speed-load-range of the engine. These maps were generated in homogeneous and stratified mode. Figure 4.1.9 shows the size of the residuals of the manifold pressure and the air flow caused by different leak sizes ($d = 1 \ldots 3\,\text{mm}$) after the throttle valve in homogenous mode. Hence, the fault detection depends strongly on the operating point.

Using a logic AND-combination of the two corresponding residual maps for the same fault one obtains areas where the detection of the faults, depending on the sizes of the residuals are possible. This will be later used for the final fault diagnosis procedure. To increase the robustness the areas with residual deflection of minimum

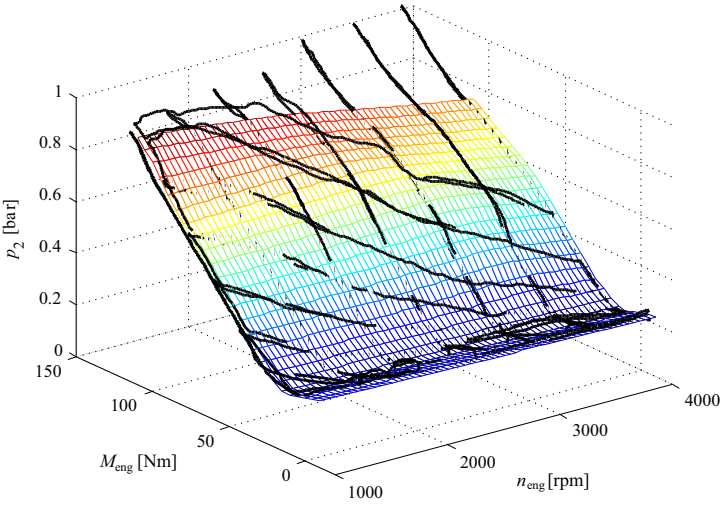


Fig. 4.1.8. Map of the manifold pressure p_2 in the complete speed-load-range for the fault-free case in homogeneous mode (bold lines: measured data).

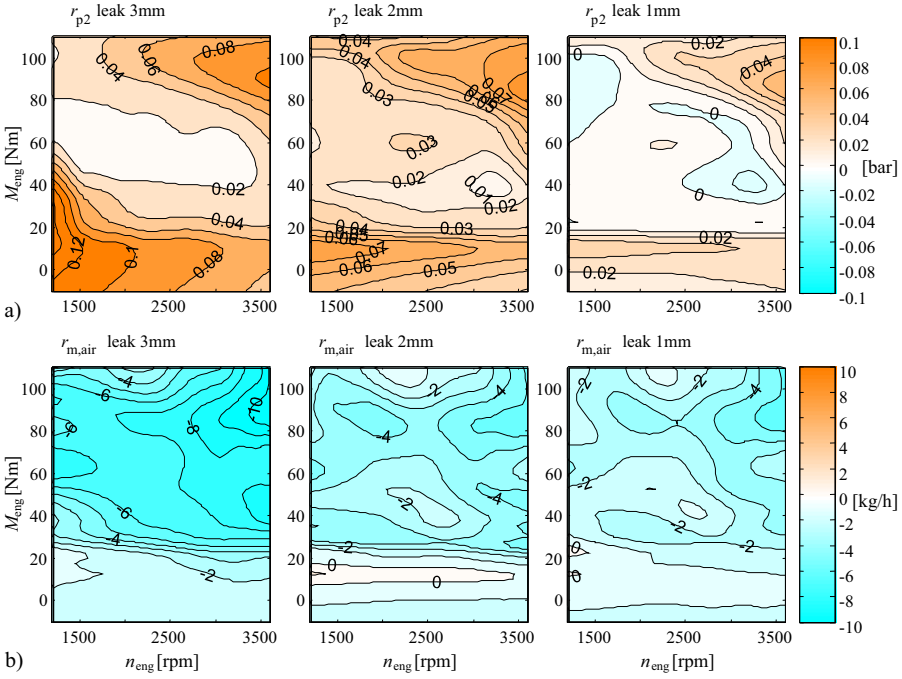


Fig. 4.1.9. Map of the two residuals caused by different leakage sizes ($d = 1 \ldots 3$ mm) after the throttle valve in the complete speed-load-range in homogeneous mode. **a** residuals of manifold pressure. **b** residuals of air flow.

10 percent compared to the fault free case were marked. Figure 4.1.10 depicts the results for leakages after and before the throttle valve. By this procedure limit values for the residuals are generated which depend on the operating point $[M_i, n_j]$ of the engine.

Best conditions to detect leakages after the throttle valve are in the low load area in homogeneous mode. The detection of leakages before the throttle valve is better possible in the higher load area. Generally, of course larger leaks are better detectable than small leaks, see Fig. 4.1.9.

The residual maps from Fig. 4.1.9 can also be used to generate new limits depending on the speed-load-range and the operation mode of the engine. This increases the possible detection area in comparison to Hartmanshenn and Isermann (2005). Figure 4.1.11 shows the residual of the manifold pressure for different stationary speed-load-range points in the MVEG driving cycle in stratified and homogeneous mode for a leakage with $3\,mm$-diameter after the throttle valve. The fault was present during the complete measurement. In stratified mode the residual is only in a short time higher as the limit, but in homogeneous mode the residual exceeds the operating point dependent limit values significantly at low torques.

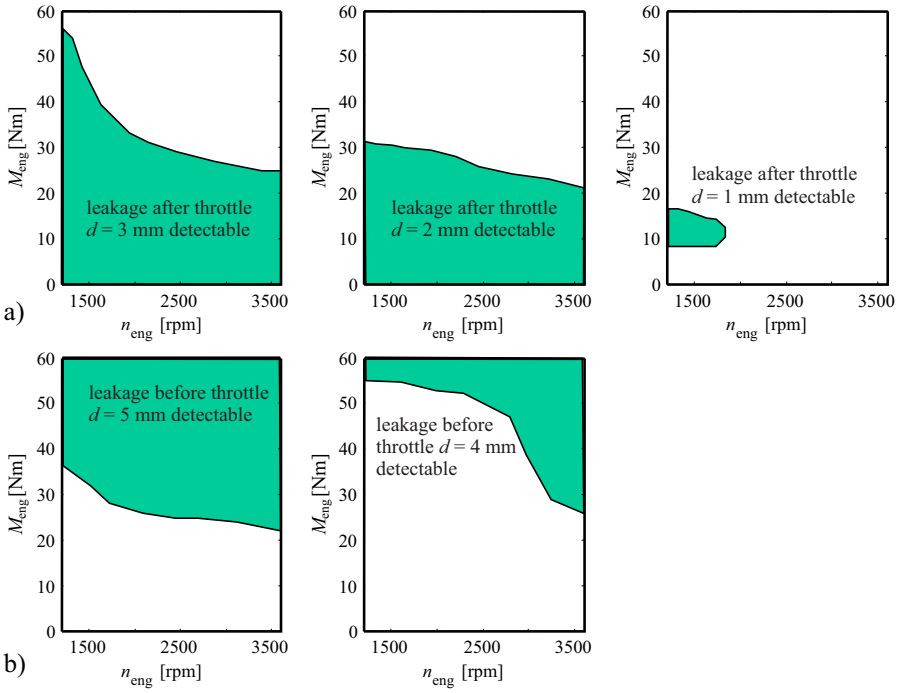


Fig. 4.1.10. Operating point dependent area for the detection of leakages with different diameters before and after the throttle valve in homogeneous operation mode. **a** leakage after throttle. **b** leakage before throttle.

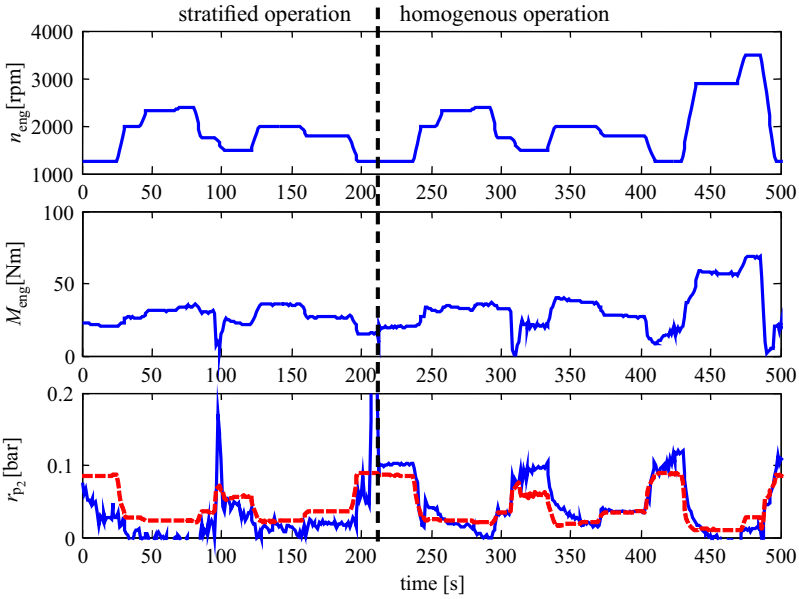


Fig. 4.1.11. Residuals of the manifold intake pressure for a leak of 3 mm diameter after the throttle valve (MVEG driving cycle)

Diagnosis system for the intake and exhaust system

The fault diagnosis resulting in the kind and size of fault is based on all determined symptoms and may use binary logic or approximate reasoning by fuzzy logic rules, Isermann (2006). In both cases IF-THEN rules are applied. Binary logic in the simplest cases uses only the signs of the symptoms (e.g. residuals exceeding a sharp threshold). If fuzzy logic is applied, the symptoms are assigned to membership functions $\mu(S_i)$. Figure 4.1.12 shows an example for the diagnosis of a leakage after throttle valve. The three residual sizes $r_{\mathrm{m_{air}}}$, $r_{\mathrm{p_2}}$ and r_λ activate membership functions and a UNION operator (minimum function) for the accumulation to determine the kind of a certain fault. For a maximum change of one symptom value 80 percent of the expected residual change is assumed to assign the membership function "strong positive/negative", based on the results of Fig. 4.1.9. Maps with the limits depending on the actual operation point of the engine were calculated from the residual maps as shown in Fig. 4.1.11.

The following rules were determined for the intake and exhaust system, based on the results of Figs. 4.1.9, 4.1.10 and 4.1.11.

- If S_{int1} negative AND S_{int2} neutral AND S_{exh1} positive (hom. mode), THEN leakage before throttle valve
- If S_{int1} negative AND S_{int2} positive AND S_{exh1} positive (hom. mode), THEN leakage after throttle valve ($d = 2$ mm)
- If S_{int1} negative AND S_{int2} strong positive AND S_{exh1} positive (hom. mode), THEN leakage after throttle valve ($d = 3$ mm)

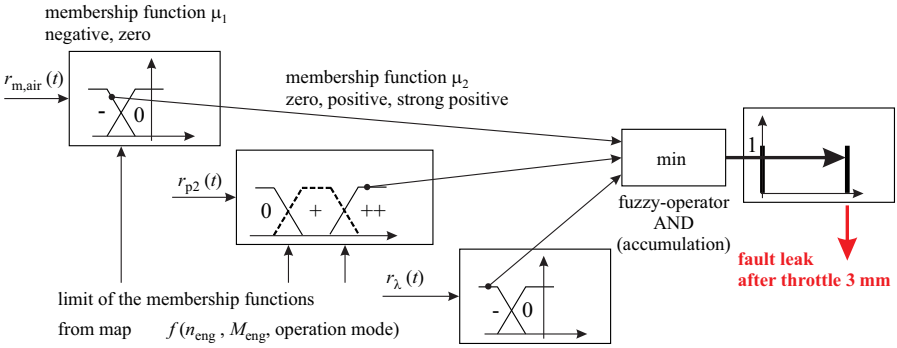


Fig. 4.1.12. Example of the rule-based fuzzy logic reasoning for a "leakage after throttle".

- If S_{int1} negative AND S_{int2} positive, THEN increased EGR rate.

These different fault-symptom-patterns enable the diagnosis of leakages before and after the throttle valve and an increased EGR rate because of a restriction in the exhaust system.

The symptoms and the resulting fuzzy-logic fault possibility are depicted in Fig. 4.1.13 for the case of a leakage after throttle. If the singleton output exceeds a value of 0.8, a leak is diagnosed. This is possible for stratified as well as for homogeneous operation of the engine. A fault-symptom table for this engine is given in Sect. 4.4 together with other faults, which uses for simplification a binary logic reasoning.

b) Case 2: homogeneous combustion gasoline engine

The fault detection in the intake system will now be considered for a conventional gasoline engine with (only) homogeneous near stoichiometric combustion. The investigated engine is a three-cylinder engine 1.0 l, Opel ECO TEC, 40 kW, 82 Nm with intake-manifold-injection and electrical throttle, Hartmanshenn and Isermann (2005). The instrumentation corresponds to Fig. 4.1.2. The modeling is performed as depicted in Fig. 4.1.4 with nonlinear MISO-LOLIMOT static local linear models for air mass flow and manifold pressure as outputs, however, without tumble flap input. These identified models were obtained with appropriate input excitation as described in the last section in the operating range $1000 - 4000\,rpm$ and $5 - 75\%$ relative load. The resulting mean squared output error between measurements and the model is about 3%, Hartmanshenn and Isermann (2005).

Based on these models for the normal behavior output parity equations are generated as (4.1.10), leading to the residuals $r_{\mathrm{int1}} = r_{\mathrm{m_{air}}}$ and $r_{\mathrm{int2}} = r_{\mathrm{p_2}}$. Additionally, deviations of the λ-controller correcting factor c_λ yield the residual $r_{\mathrm{exh1}} = r_\lambda$, as in (4.1.10). Therefore, three symptoms can be generated: S_{int1}: air mass flow, S_{int2}: manifold pressure and S_{exh1}: correction λ-controller.

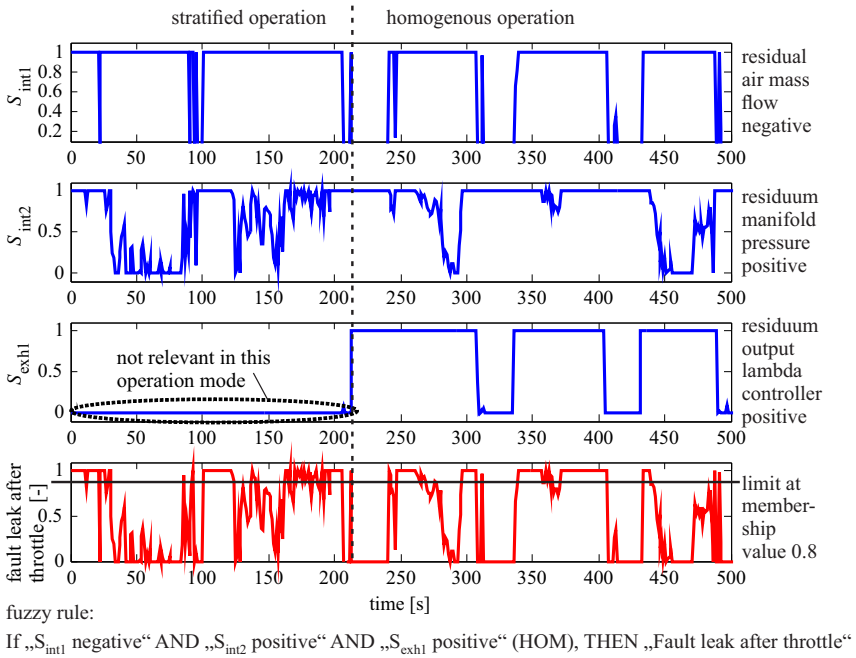


Fig. 4.1.13. Generation of the fuzzy logic possibility measure for a “leakage after throttle”, using one if-then-rule for three symptoms for stratified and homogeneous operations.

Figures 4.1.14 to 4.1.16 depict the course of the three residuals for leaks of different sizes in front of the throttle, after the throttle and for a disconnected blow-by tube and a leak in the exhaust pipe.

The manifold pressure residuals indicate large changes for leaks after the throttle, especially near idling where the manifold pressure is small. Largest changes for the air mass residual can be seen for leaks in front of the throttle and for part and high load. The influence of the leak in the exhaust pipe is small. The residual for the λ-correction factor is large for leaks after the throttle and in the exhaust pipe, especially for low loads. Hence, the three residuals react differently for the faults and therefore they are well suited for the fault diagnosis. The results show also, that leaks of 0.5–$1\,mm$ diameter are detectable after the throttle, those with $2\,mm$ before the throttle and leaks with about $5\,mm$ diameter in the exhaust pipe.

The resulting symptoms are summarized in Table 4.1.5. Also changes of the residuals for other investigated faults are listed. Hence, based on the identified air flow and manifold pressure model and the correcting factor of the λ-controller several faults can be detected by using standard sensors. Five of them are isolable, because of different patterns of the deviation signs. If same or rather similar patterns appear, as for leakage after the throttle and disconnected blow-by tube, these faults are not isolable. But then groups of faults can be given for further investigation, e.g. by inspection.

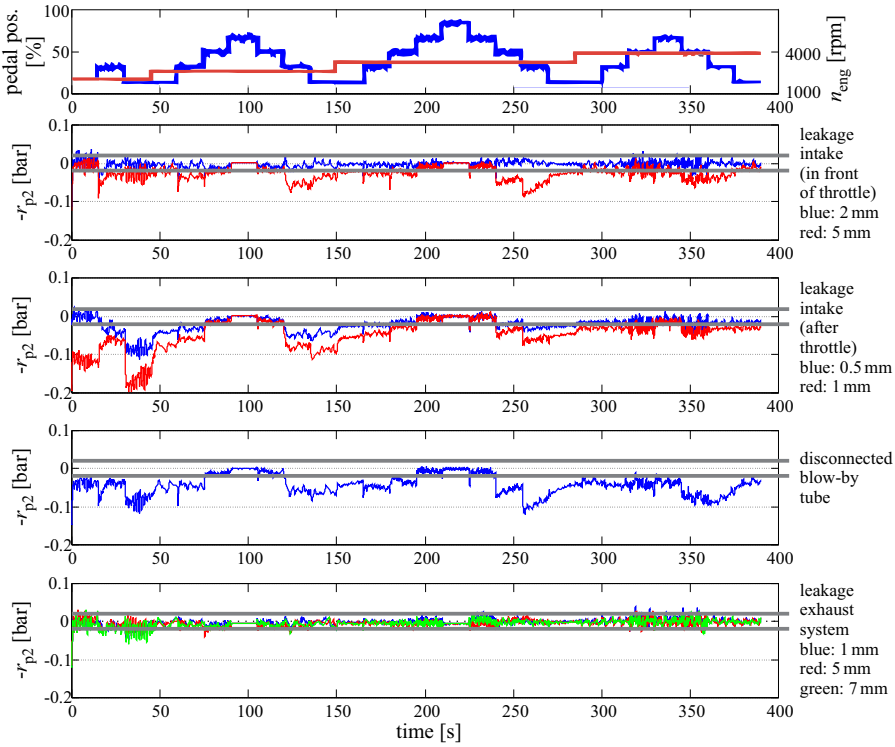


Fig. 4.1.14. Residuals for the manifold pressure p_2 with parity equations for different leakages in the air system.

Table 4.1.5. Fault-symptom table for the homogeneous combustion gasoline engine with *experimental models*

Faults	Symptoms and residuals			Isolability	
	S_{int1}	S_{int2}	S_{exh1}		
	$r_{m_{air}}$	r_{p_2}	r_λ	isolable	not isolable
leakage before throttle	−	0	+		✓
leakage after throttle	−	++	++	✓	
disconnected blow-by tube	−	+	++		✓
leakage in exhaust pipe before catalyst	0	0	++		✓
clogging of air−filter	−	−	0	✓	
EGR−valve stuck open	−	+	0	✓	
EGR−valve stuck close	+	−	0	✓	
injection mass too high (1 cyl.)	0	0	−	✓	
injection mass too low (1 cyl.)	0	0	+		✓

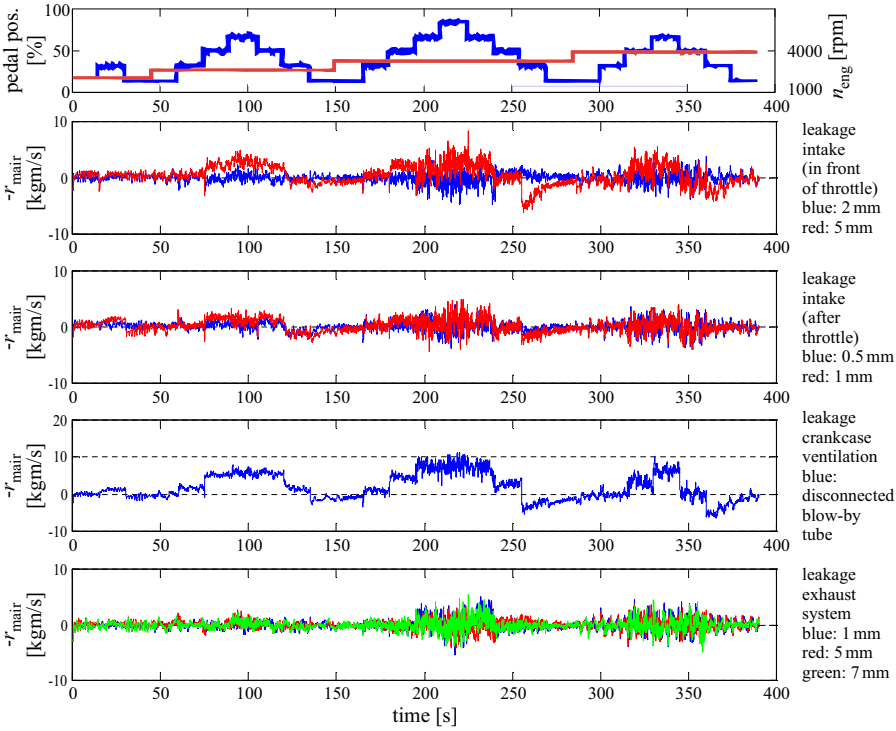


Fig. 4.1.15. Residuals for the air mass flow $\dot{m}_{\mathrm{air}}$ with parity equations for different leakages in the air system.

4.2 Misfire detection

As the measurement of the cylinder pressure is not standard in most gasoline engines, a signal which follows immediately the burning process is the crankshaft speed. Due to the acceleration and deceleration during the strokes the angular speed shows oscillations, Isermann (2014), Sect. 4.3.3. Their frequencies and amplitudes indicate irregularities of the burning cycles, see Mauer and Watts (1989), Ribbens and Rizzoni (1990), Plapp et al (1990), Führer et al (1993), Fehrenbach and Quante (1987), Förster et al (1997). Therefore, the engine-speed analysis is an important basis for fault detection of fuel injections, ignition, burning and mechanics. But also the analysis of the ion-current or the exhaust pressure signal are suitable ways. Some methods are described in the sequel.

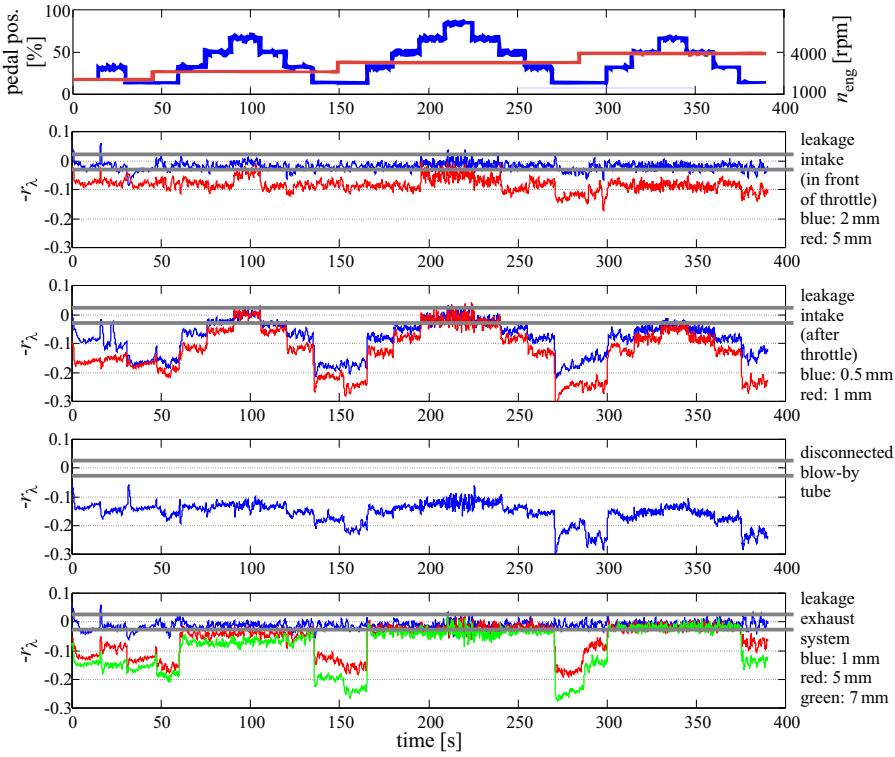


Fig. 4.1.16. Correction factor c_λ of the λ-controller for different leakages in the air system, see (4.1.9).

4.2.1 Engine speed analysis

Of special importance for combustion engines is misfire detection. When a cylinder misfires, e.g. due to faults in the mixture preparation or ignition system with the effect that no combustion or incomplete combustion occurs, unburned fuel enters the exhaust system, which then burns in the hot catalytic converter. The released heat may damage or destroy the catalytic converter by thermal overloading. If a given misfire ratio is exceeded, the fuel supply for the misfiring cylinders can be cut off in order to protect the catalytic converter from damage and to avoid exceeding the emission standard. One way to detect misfiring cylinders is to evaluate the engine speed signal at the engine flywheel.

The signal characteristics of a combustion engine are determined by the batch behavior of the combustion, which depends on the crankshaft angle φ_{CS}. Each cylinder of a four-stroke engine fires every $720°\mathrm{CS}$. This corresponds to one working cycle and specifies the engine base period. All relevant signal components are multiples of this base frequency. During a working cycle, each cylinder fires one time so that for a four-cylinder engine a combustion results every $180°\mathrm{CS}$. If the engine angular speed, measured at the flywheel, is denoted by ω_{eng}, the frequency of this oscillation

corresponds to the ignition frequency f_{I}

$$f_{\mathrm{I}} = \frac{\omega_{\mathrm{eng}}}{4\pi}\, i_{\mathrm{c}}, \quad i_{\mathrm{c}} \ : \ \text{number of cylinders} \tag{4.2.1}$$

In Fig. 4.2.1 a typical engine speed signal of a spark-ignition (SI) engine measured at idle speed without misfiring is depicted showing speed oscillations with the ignition frequency around the engine speed mean value (approx. 800 rpm).

If misfires or faults in the injection mass occur, the engine speed decreases significantly. Figure 4.2.1 shows the measured engine speed of a four-cylinder engine in the case of continuous misfiring of one cylinder. Then, additional low-frequency oscillations arise, as can be clearly seen from the low-pass filtered engine speed signal. The appearing frequency components are harmonics of the engine base frequency. Depending on the misfiring cylinders, different frequency patterns result. In the past, methods have been investigated using the Fourier and the fast Fourier analysis to evaluate these frequency components, see Ribbens and Rizzoni (1990). Figure 4.2.3 shows the Fourier transforms of engine speed signals with no misfire and with misfire in cylinder 1. Without misfire, the ignition frequency means that only the fourth engine harmonic appears in the spectrum. In the case of misfires, additional frequency components arise. Evaluating these frequency components means that not only misfires, but also the misfiring cylinder, can be detected and located.

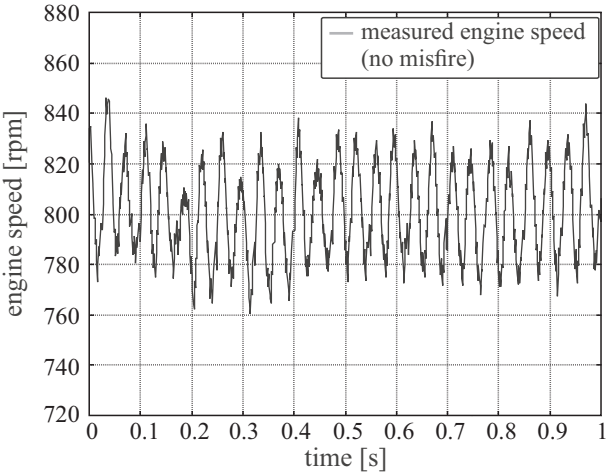


Fig. 4.2.1. Measured engine speed signal at idle speed of an SI engine (no misfires).

Another method to be considered, Führer et al (1993), uses the real and imaginary components of the discrete Fourier transformation (DFT) applied to the engine speed signal. Thereby, a four-stroke four-cylinder engine shall be considered, whereas the principle of the method was also successfully implemented in a six-cylinder spark-ignition engine up to 6000 rpm and for loads higher than 20%.

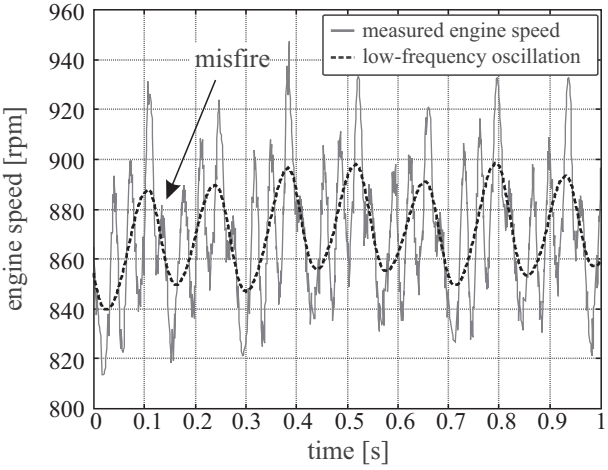


Fig. 4.2.2. Measured engine speed signal and low-pass filtered signal at idle speed with misfires in cylinder 1.

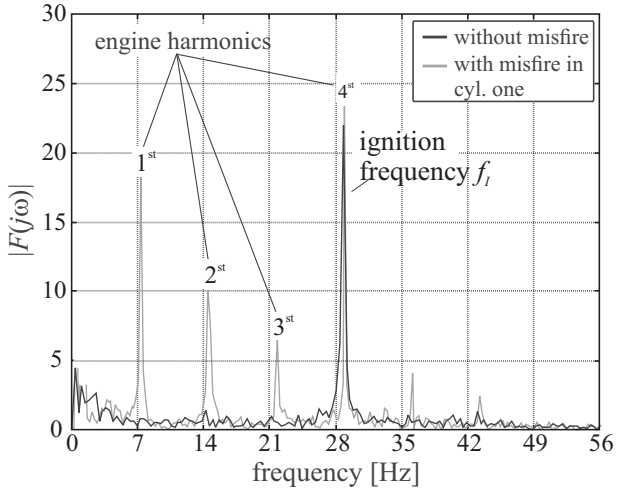


Fig. 4.2.3. Fourier transform of the engine speed signal without and with misfires in cylinder 1.

Since the engine speed varies with time, the data acquisition is performed crank angle synchronously so that no sampling time adaptation is necessary. For calculation of the DFT, the data is sampled all at 90°CS. This corresponds to the double ignition frequency, such satisfying the Shannon sampling theorem. The resulting speed-dependent sampling time for a four-stroke engine then follows from the ignition frequency:

$$T_0 = \frac{1}{2\, f_{\mathrm{I}}} = \frac{\pi}{2\, \omega_{\mathrm{eng}}} \tag{4.2.2}$$

The DFT evaluation can now be determined by using a few sampling points $N \geq 8$ (minimum one working cycle), which is an easy real-time application. To compute the DFT, the amplitudes and the phase angle can be calculated as follows: compare (8.17), (8.18) in Isermann (2006) and (5.3.1) to (5.3.8)

$$A_m = \sqrt{\left(\sum_{i=0}^{N-1} \omega_i \cos\left(\frac{2\pi mi}{N}\right)\right)^2 + \left(\sum_{i=0}^{N-1} \omega_i \sin\left(\frac{2\pi mi}{N}\right)\right)^2} \tag{4.2.3}$$

$$\varphi_m = \arctan\frac{\sum_{i=0}^{N-1} \omega_i \sin\left(\frac{2\pi mi}{N}\right)}{\sum_{i=0}^{N-1} \omega_i \cos\left(\frac{2\pi mi}{N}\right)} \tag{4.2.4}$$

whereas $\omega_i = \omega_{\text{eng}} = 2\pi n_{\text{eng}}$ is the speed measured and m denotes the order of the harmonics. Because of the usually non-cyclic combustion variations, an average value for several working cycles can be calculated from the measured data.

Faults to be taken into consideration are misfires or combustion differences in one or two cylinders. To locate the misfiring cylinders, only the first and second engine harmonics ($m = 1$ and $m = 2$) have to be evaluated, see also Fig. 4.2.4. Representing the real and imaginary components of these two frequencies, values equal to zero for no misfires and unequal to zero for misfires appear. For pattern recognition and misfire detection respectively, comparisons of the amplitude values and the real and imaginary components have to be performed. Six different patterns P have to be distinguished for the relative location of the misfiring cylinders to each other:

P0: no fault
P1: one cylinder oversupplies
P2: one cylinder undersupplies
P3: two subsequent cylinders undersupply
P4: two oppositely cylinders undersupply
PX: undetectable.

Also, two thresholds T1, T2, which are dependent on engine speed and load, have to be determined. The flowchart in Fig. 4.2.4 shows the signal flow of monitoring and analyzing the possible fault patterns. Depending on the fault case, different vector patterns arise, with which the defective cylinders can be detected. Thus, the fault diagnosis is executed by a pattern-recognition method of the amplitudes and phases of the DFT.

The performance of the proposed method is on one hand limited by faults in the data acquisition (for example, error in measurements) and on the other by overlaid disturbances on the measured signal. It can be used for misfire detection as well as for monitoring smooth engine operation for the whole operation area, except for too low loads and high speeds.

The speed analysis method for misfire detection with six-cylinder engines is described in Führer et al (1993).

For higher angular speeds the rotational oscillations of the vibrating system formed by crankshaft, connecting rods and pistons influence the measured speed.

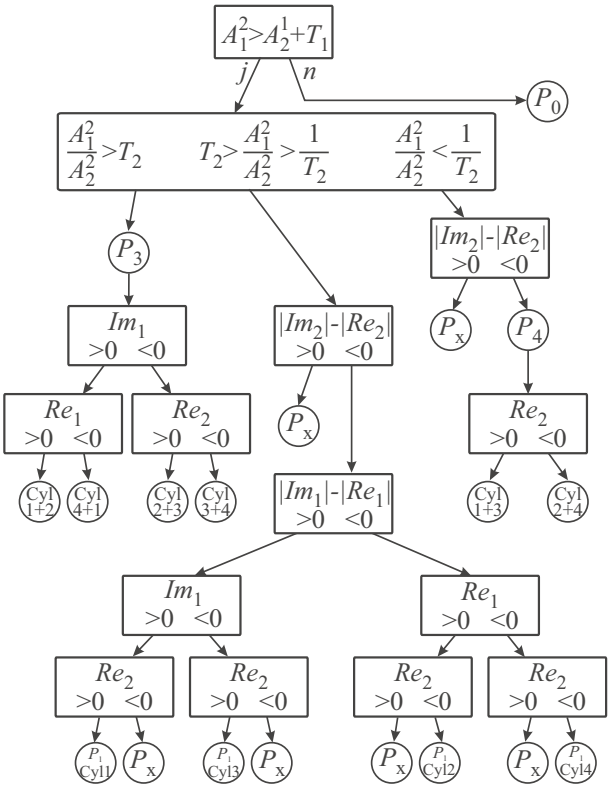


Fig. 4.2.4. Scheme for detection of misfires and diagnosis of the faulty cylinders for a four-cylinder engine by using the amplitudes, real and imaginary components of the discrete Fourier transform (DFT).

However, compensation of the free moments of inertia allows to reconstruct the gas forces responsible for torque generation see Kimmich (2004).

As the amplitudes of the crankshaft speed becomes the smaller the larger the number of cylinders, there is a limit of the speed analysis from about eight cylinders upwards, because the cylinder strokes then overlap resulting in a smooth running. Then ion-current analysis for each cylinder is an (expensive) alternative, see Sect. 4.2.2. If multi-cylinder engines are arranged in two banks to three, four or five cylinders exhaust pressure measurement, Willimowski (2003) can be applied, as described in Sect. 4.2.3, or oxygen-content measurement at both sides, see Kimmich (2004). Disturbing factors for the speed analysis are generally the influences from the powertrain (more for manually shifted gears, less for automatic hydraulic transmissions) and the torsional behavior of the crankshaft. These effects may need special filtering.

The angular speed oscillation can also be used to calculate the indicated engine torque, see Isermann (2014), Sect. 4.3.3.

4.2.2 Ion-current analysis

Because of the chemical reactions and the high temperatures during the burning process ions and electrons are generated which increase the electrical conductivity of the burning gas in the cylinders. Figure 4.2.5a) depicts a measurement configuration, which uses the spark plug electrodes to measure an ion current. During the current flow for the ignition through the secondary coil the capacitor C1 is charged, until the Zener-diode DZ_1 reaches the breakthrough voltage of about 82 V. After the ignition the capacitor C1 supplies the spark plug with this voltage and generates an electrical field over the electrodes of the spark plug. The resulting ion current is measured as voltage drop at the measurement resistance R_m. As shown in Fig. 4.2.6 the ion current is characterized by a first peak through the chemical ionization around 0 to $10°\mathrm{CS}$ and a second peak through the thermal ionization around 15 to $20°\mathrm{CS}$. The ion current shortly after the ignition shows strong oscillations, which cannot be used for extracting useful information. The second ion-current peak is mainly due to the high temperature of the burning gas, which increases the mobility of ions and electrons. It correlates well with the measured cylinder pressure, see Fig. 4.2.6.

The evaluation of the ion current follows the block diagram depicted in Fig. 4.2.7, which is the result of measurements with different multi-cylinder engines, Willimowski et al (1999), Willimowski (2003). First, bursts and high-frequent noise directly after ignition are removed by median filtering. Then, the offset is eliminated within a crank angle window $\varphi_{\mathrm{ref},1} \ldots \varphi_{\mathrm{ref},2}$, where no burning appears. The resulting zero mean ion-current signal is then integrated within two windows $\varphi_1 \ldots \varphi_2$ and $\varphi_3 \ldots \varphi_4$ leading to the features

$$I_1 = \int_{\varphi_1}^{\varphi_2} I_{\mathrm{cyl}}(\varphi)\, d\varphi \qquad \text{misfire detection} \qquad (4.2.5)$$

$$I_2 = \int_{\varphi_3}^{\varphi_4} I_{\mathrm{cyl}}(\varphi)\, d\varphi \qquad \text{late burning indicator 1} \qquad (4.2.6)$$

and the difference angle

$$\Delta\varphi_6 = \varphi_5 - \varphi_{\mathrm{ign}} \qquad \text{late burning indicator 2} \qquad (4.2.7)$$

The window angles are calibrated in dependence on speed and torque such, that the values of the integrals become maximal values. If I_1 does not exceed a calibrated threshold no burning took place, hence a misfire appeared, and if I_2 exceeds a threshold a too late or partial burning happened. $\Delta\varphi_6$ additionally indicates a late burning. The magnitude of the ion current I_{cyl} very much depends on speed and torque and the signal is of course weak for low speed and low load, e.g. for idling. However, the described ion-current analysis methods allows to detect misfires, partial or late burnings and knocking in the whole drivable operating range for gasoline engines and is especially suitable for multi-cylinder engines where misfire detection from speed analysis is not reliable enough.

The information gained from the ion-current analysis can also be used to detect knocking and to optimize the ignition angle and the peak pressure location from

the thermal ion-current peak position, Willimowski (2003), see also Eriksson et al (1996).

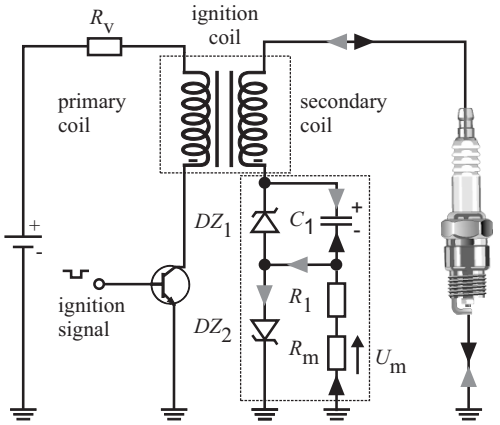


Fig. 4.2.5. Ion-current measurement: Measurement system from Delphi, Willimowski (2003).

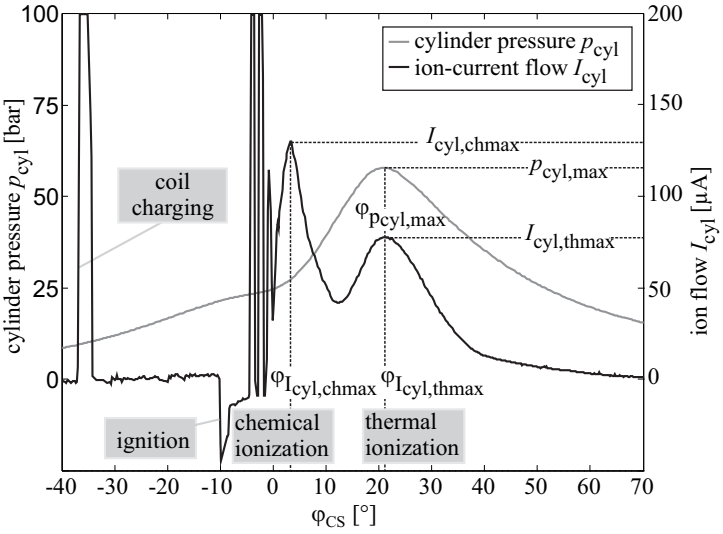


Fig. 4.2.6. Measured ion current and cylinder pressure.

4.2.3 Exhaust gas pressure analysis

A further alternative for misfire detection is the analysis of the exhaust pressure oscillations. For each combustion cycle, each exhaust valve opens once in the gas ex-

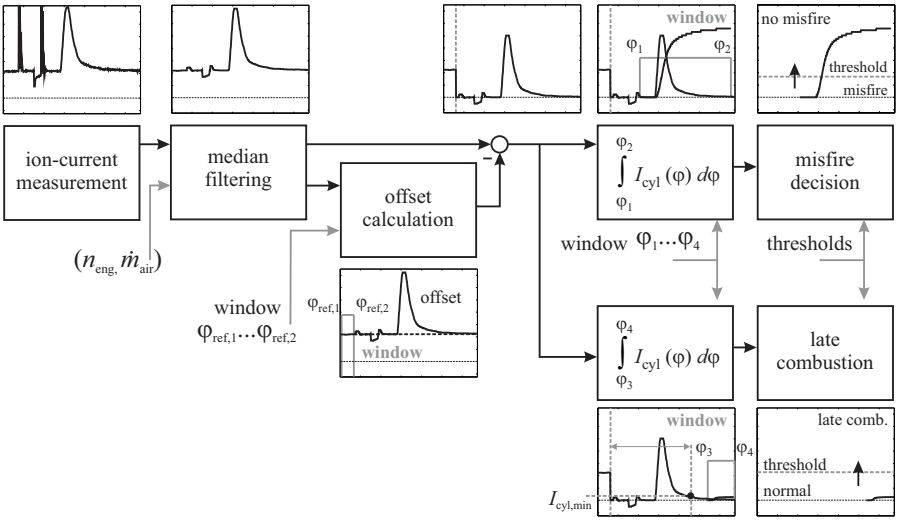


Fig. 4.2.7. Signal flow for the misfire detection with ion-current analysis.

change phase of the engine. The pressure waves in the exhaust manifolds are excited by the sudden relaxation of the combustion gases and the piston movement as the exhaust valves open. The exhaust gas pulsations depend on the combustion process and therefore on the cylinder pressure when the exhaust valve opens. These exhaust gas pulsations are measured dynamically. For the interpretation of the exhaust gas pressure signals, both the multiplicity of phenomena occurring in coupled exhaust pipes with mufflers and combustion-related influences have to be examined. The resulting frequency component amplitudes of the exhaust gas pulsations are influenced by the formation of standing waves caused by reflections and the propagation and attenuation mechanisms.

The investigations were carried out on a BMW 750i test car with a V-12 engine. This vehicle is equipped with a dual exhaust system, with one catalyst for each six-cylinder bank. The exhaust gas pressure is measured by one pressure transducer per exhaust strand. Thus, the primary oscillation frequency of the exhaust gas pressure in each exhaust strand corresponds to the ignition frequency of a six-cylinder engine (half of the ignition frequency of a 12-cylinder engine). This frequency corresponds to the sixfold camshaft rotational speed (6th engine harmonic, abbreviated: 6th EH), Fig. 4.2.8 shows the frequency spectrum of the exhaust gas pressure in the catalyst mixing tube of one exhaust strand. It can be seen that especially the 6th EH and multiples of it with a lower amplitude are dominant.

If a cylinder misfires, the exhaust gas pressure drops significantly, due to the missing combustion. Compared to regular combustion without misfires, additional frequency components of high intensity below the 6th EH arise, as shown by the frequency analysis in Fig. 4.2.8b). The frequency components caused by misfires are always multiples of the camshaft frequency and can be calculated from the angular

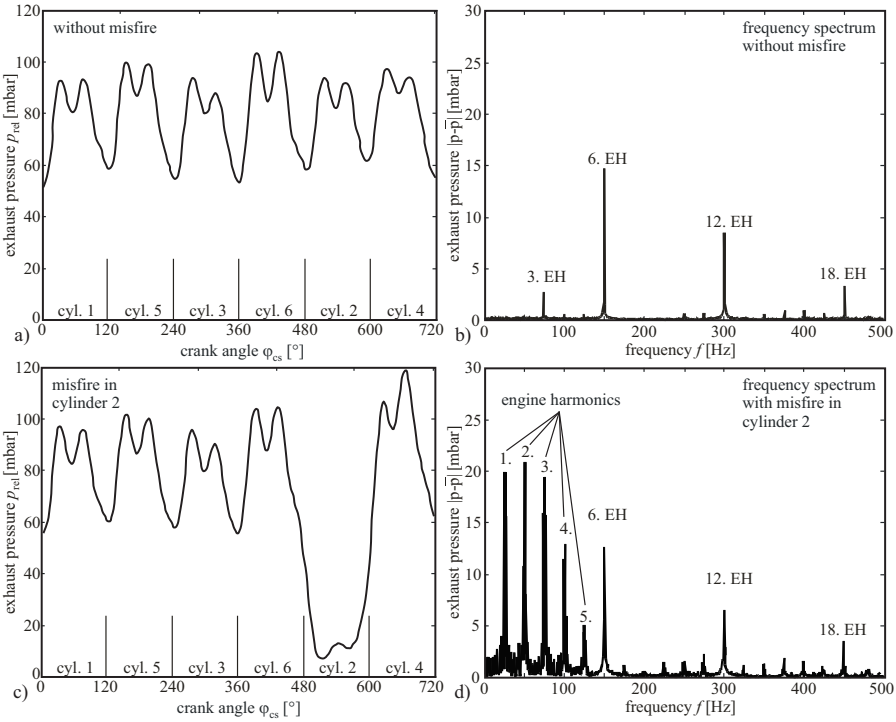


Fig. 4.2.8. Measured exhaust pressure signal and Fourier analysis. **a**, **b** without misfire. **c**, **d** with misfire in cylinder 2. (EH: engine harmonics).

speed ω_{eng} of the crankshaft for four-stroke engines

$$f_j = \frac{\omega_{\mathrm{eng}}}{4\pi}\, i_c\, j \tag{4.2.8}$$

and with $\omega_{\mathrm{eng}} = 2\,\pi\,n$

$$f_j = \frac{n}{2}\, i_c\, j \tag{4.2.9}$$

where i_c is the number of cylinders, j the number of harmonics, n the engine speed in rounds per second. The ignition frequency is obtained for $j = 1$ and the first engine harmonic is defined as the ignition frequency for a one cylinder engine

$$f_1 = \frac{\omega_{\mathrm{eng}}}{4\pi} = \frac{n}{2} \tag{4.2.10}$$

The intensity of these engine harmonics essentially depends on the combination of misfiring cylinders as well as on the current engine operating point.

The pressure measurement is performed in a short pipe coupled to the exhaust system by a T-joint. Compared to a wall-mounted sensor, this offers the advantage

that no complex cooling mechanism is needed. Thus, low price pressure transducers can be used, which is especially important with regard to use in mass production.

The most suitable location for exhaust gas pressure measurement was determined to be the catalyst mixing tube. The length of the T-joint pipe was chosen as $l_{\mathrm{st}} = 165\,\mathrm{mm}$. In this way, a maximum temperature at the diaphragm of the transducer of approximately $85°\mathrm{C}$ was achieved.

One possibility to detect misfires from the time behavior of the pressure signal is to apply the discrete time Fourier transform, Willimowski et al (1999). However, as a misfire first shows a significant decrease and then an increase of the pressure an analysis procedure which is sensitive to sharp signal changes is more appropriate. This is the case for the wavelet analysis, Willimowski (2003). Therefore, the measured pressure signal is transformed by

$$W(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} p(t)\, \Psi\left(\frac{t-\tau}{a}\right) d\tau \tag{4.2.11}$$

in the time domain where Ψ is a mother-wavelet time-scaled (dilatation) by a and time-shifted (translation) by τ, see e.g. Strang and Nguyen (1996), Best et al (1997), Isermann (2006). It turned out that using the Haar-wavelet with $a = 240°\mathrm{CS}$ is a suitable choice, as it is easy to compute and damps well the ignition frequency and its higher harmonics, leading to $W \approx 0$ for normal operation. The pressure signal $p(\varphi)$ is sampled at $\varphi_0 = 6°\mathrm{CS}$ and segmented over $120°\mathrm{CS}$ with regard to the cylinder, taking into account corresponding pressure dead times from outlet valves to the pressure sensor.

The evaluation of the wavelet coefficient $W(a,\varphi)$ in dependence on the crank angle φ makes it independent on the engine speed. Thus the parameters of prefiltering with FIR (finite impulse response) and of the wavelet coefficient have to be calibrated only once and hold for the whole operation range of the engine. Figure 4.2.9 depicts the measured pressure signal $p(\varphi)$ over 10 strokes with several misfires and the wavelet coefficient – $W(240,\varphi)$ (negative for decreasing pressure). If this wavelet coefficient passes a negative and then a positive threshold, a misfire is detected. The corresponding $120°\mathrm{CS}$ segment and the local extremal values of W are stored to localize the misfiring cylinder. A more detailed description and a signal-flow diagram of the misfire detection and cylinder localization is described in Willimowski (2003). For very low speed and low torque better results are obtained with a Haar-mother wavelet over $60°\mathrm{CS}$. This misfire-detection method with exhaust pressure oscillation analysis has turned out to operate reliably over the whole operating range of the 12 cylinder V-engine (two banks of six-cylinder engine) from $600 \ldots 5500$ rpm and low load until maximum driveable load with about 95 to 100% detection rate, also for simultaneous misfires in several cylinders.

4.3 Fuel supply and injection system

The low-pressure fuel supply system and its modeling is described in Isermann (2014), Sect. 4.2. Now, model-based fault-detection methods based on available mea-

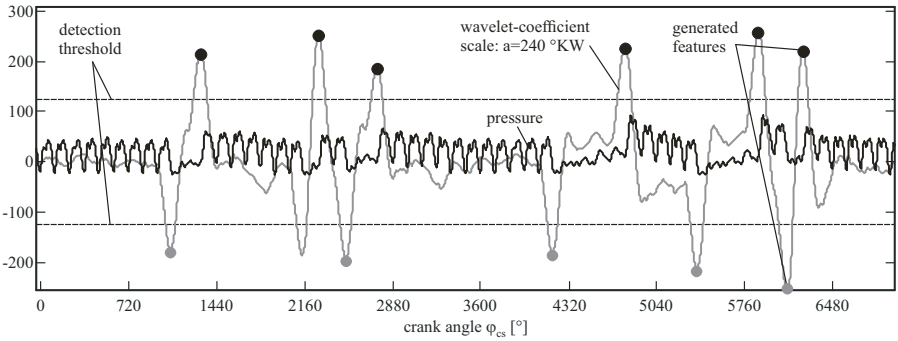


Fig. 4.2.9. Exhaust pressure signal with stochastically generated misfires and corresponding wavelet analysis for one 6 cylinder strand. $n_{\mathrm{mot}} = 2000$ rpm, fuel injection time period $TL = 3$ ms/round (weak torque). Pressure signal and wavelet coefficient $-W(240, \varphi)$.

surements are discussed. Figure 4.3.1 illustrates two fuel supply systems with and without fuel return.

4.3.1 Low-pressure supply system

Some faults which influence a correct fuel pressure or flow rate to the injectors are:

- Filters: contaminated
- DC motor: commutator fault, windings shortcut, mechanical wear, electrical connections
- Valves: blocked non-return valve, blocked pressure relief valve
- Pump: low pressure, high friction, wear, vapor bubbles
- Pressure controller: mechanical defect, contaminated return line, leakage
- Tubes: leakages, contaminations, vapor bubbles
- Rail-pressure sensor: offset fault, gain fault.

In the case of the fuel supply system with fuel return of Fig. 4.3.1 a) the only available electrical signal in the ECU is the rail pressure, if the voltage or current to the fuel pump meter is not measured. Most of the possible faults lead to a pressure drop of the rail pressure. If the mechanical rail pressure regulator is able to compensate this by lowering the excess return fuel flow, smaller faults are covered and are not detectable. Only if larger deviations of the rail pressure arise and given thresholds are exceeded in the ECU, faults become obvious, but cannot be diagnosed with regard to their origin. If the rail pressure can be measured after shut-off of the engine, a decreasing pressure indicates a leakage or a defective non-return valve. The situation improves with an electronic controlled rail pressure for a demand controlled pump, as now also the applied voltage to the DC motor is known, see Fig. 4.3.1 b) Then relations between the pressure reference value and the motor voltage can be used. If e.g. a too high voltage is required, several faults (contamination, leakages, wear, etc.) can be

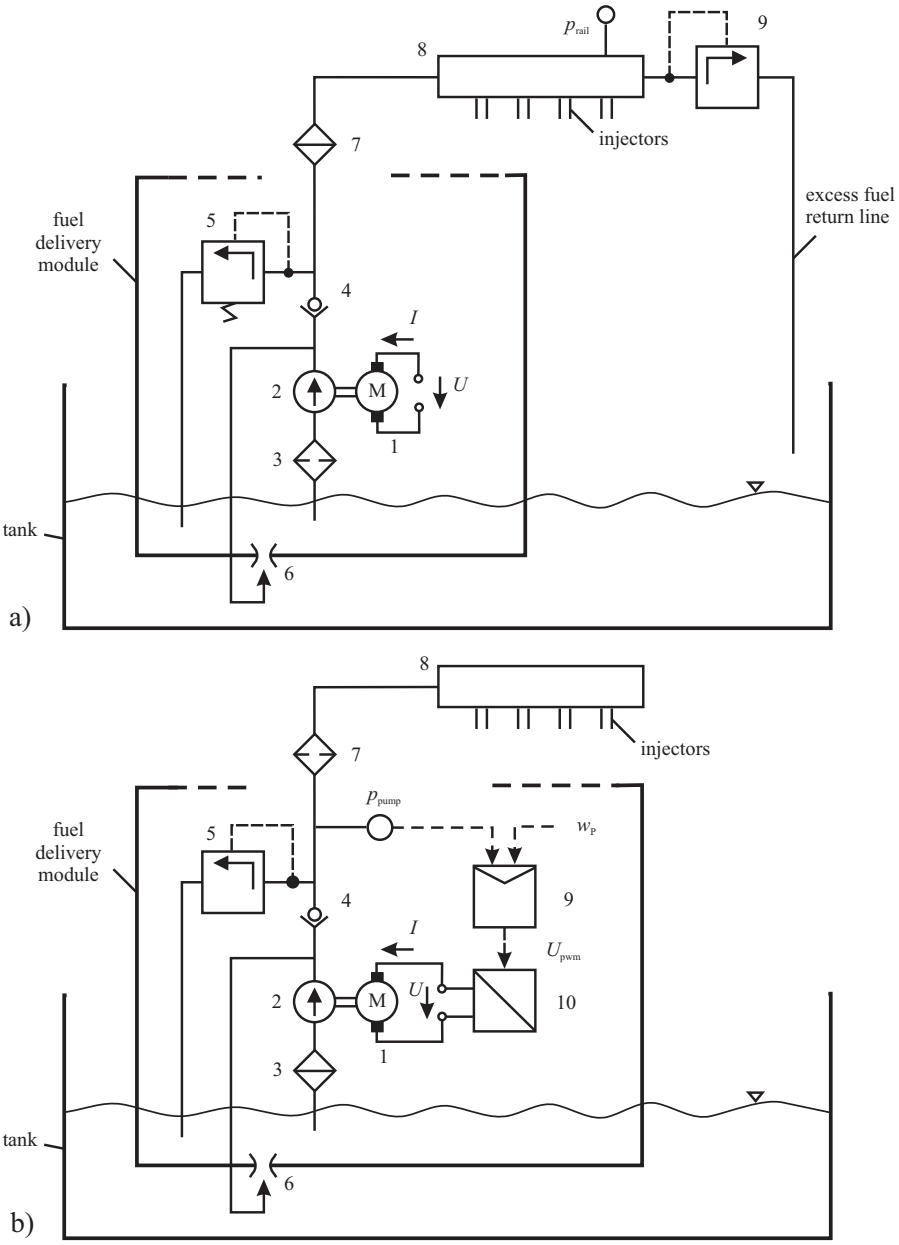


Fig. 4.3.1. a Low-pressure fuel supply system for intake manifold injection, with fuel return. 1 DC motor; 2 pump; 3 filter (sieve); 4 non-return valve; 5 pressure relief valve; 6 jet pump; 7 filter; 8 fuel rail; 9 pressure regulator. **b** Fuel supply for intake manifold injection, with demand (pressure) control. 1 DC motor; 2 pump; 3 filter (sieve); 4 non-return valve; 5 pressure relief valve; 6 jet pump; 7 filter; 8 fuel rail (without return line); 9 pressure controller (ECU); 10 pulse width modulator for pump voltage.

detected, but not diagnosed. Fault detection in a closed loop with few measurements is relatively limited, see Isermann (2006).

A more detailed fault detection and fault diagnosis becomes possible if more measurements are available for the motor-pump-tube system. As shown in Wolfram (2002), Füssel (2002) and Isermann (2011) parameter estimation methods or parity equations can be applied for fault detection of pumps. Figure 4.3.2 depicts a signal-flow scheme for parity equations. Following residuals between measured values and model outputs are possible:

$$\left.\begin{aligned} r_1 &= \Delta p - \Delta\hat{p}\,(\omega) \\ r_2 &= \dot{V} - \hat{\dot{V}}\,(\Delta p) \\ r_3 &= \dot{V} - \hat{\dot{V}}\,(\Delta\hat{p}) \\ r_4 &= M_{\mathrm{el}} - \hat{M}_{\mathrm{el}}\,(\omega) \end{aligned}\right\} \tag{4.3.1}$$

If the pump pressure $\Delta p = p_{\mathrm{pump}}$ and the pump speed ω are measured, only r_1 can be determined indicating differences to the modeled pump pressure. The additional measurement of the motor current I allows to have a measure for the motor torque M_{el} and to calculate a further residual r_4. Then, e.g., motor faults, bearing faults, impeller faults, increased flow resistances (filters, valves) or sensor faults can be detected, but not diagnosed separately. A further improvement results if a measure for the volume flow can be used, based on the injected fuel quantities and assuming they are correctly available from the ECU. Then, the residuals r_2 and r_3 can be determined and more faults are detectable and diagnosable, as shown in Füssel (2002), Isermann and Leykauf (2009) and Isermann (2011).

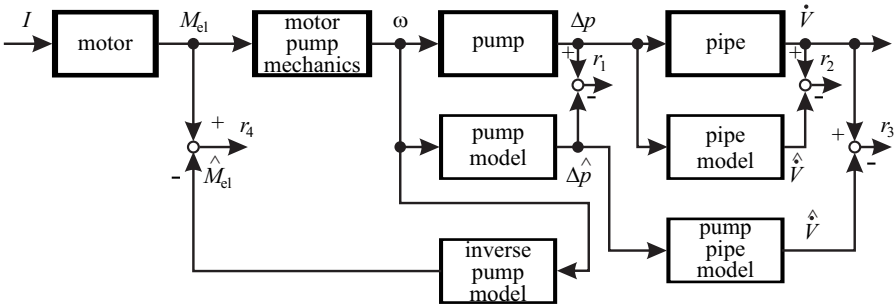


Fig. 4.3.2. Residual generation with parity equations for a motor-pump-tube system.

4.3.2 High-pressure fuel supply and injection system

A high-pressure fuel supply system according to Fig. 4.3.3 is considered. It consists of the low-pressure part (with fuel pump and filter) and the high-pressure part with the components high-pressure triple-plunger pump, fuel rail, injectors, rail pressure

sensor and control valve, see Fig. 4.3.3 a). The only measurement of the injection system available for the ECU is the rail pressure p_{rail}, which was measured at a test rig with $1^{\circ}\mathrm{CS}$ resolution. The goal is now to investigate which kind of faults can be detected in the high-pressure injection system, like injection or pump faults, by using the rail pressure and engine speed signals with high resolution.

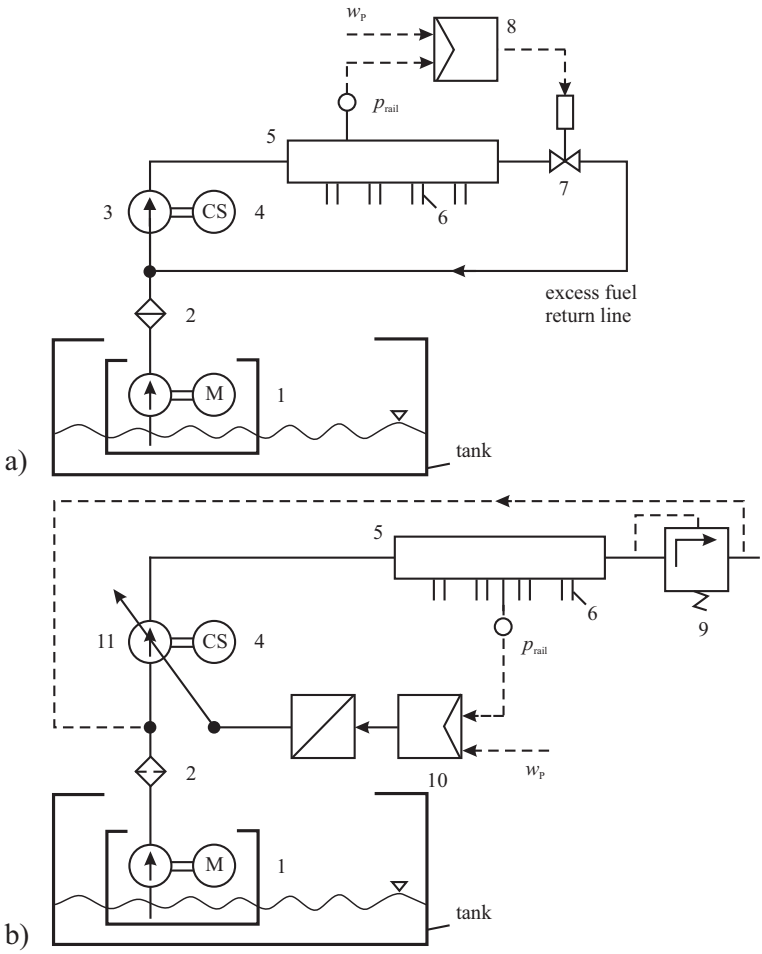


Fig. 4.3.3. a High-pressure fuel supply system with pressure control valve and fuel return. **b** High-pressure fuel supply system with rail pressure control by an adjustable pump (control demand). 1 fuel delivery module with low-pressure pump; 2 filter; 3 high-pressure pump; 4 camshaft; 5 high-pressure fuel rail; 6 high-pressure injectors; 7 pressure control valve and relief valve; 8 rail pressure controller; 9 rail pressure relief valve; 10 rail pressure controller; 11 adjustable high-pressure pump.

a) Wavelet analysis of the rail pressure signal

The rail pressure is a non-stationary periodic signal, containing several frequencies, see Fig. 4.3.4. For an analysis of non-stationary periodic signals the Short-Time-Fourier-Transformation or the Wavelet analysis can be applied, Isermann (2006), Kammeyer and Kroschel (2009). Wavelet transformation uses basis functions – called mother wavelets – which can be scaled in time (dilatation) with the factor a and shifted in time (translation) with τ. The continuous time Wavelet transformation with the mother-wavelet function Ψ is defined as:

$$W(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} y(t)\, \Psi\left(\frac{t-\tau}{a}\right) dt \tag{4.3.2}$$

The rail pressure signal was analyzed for the detection of less fuel injection in one cylinder with a simple Haar wavelet (double rectangular pulse) and a Morlet wavelet, given by

$$\Psi_{\text{Morlet}}(t) = Ce^{-t^2/2} \cos(t)\,. \tag{4.3.3}$$

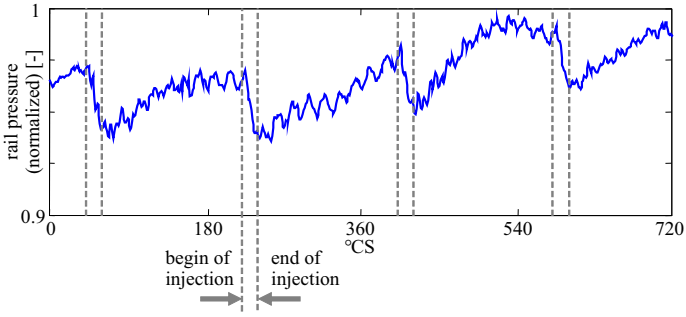


Fig. 4.3.4. Normalized rail pressure signal for $n = 1500$ rpm, without fault, VW 1.6 l FSI.

Figure 4.3.5 a) and b) shows that differences between the high-pressure signal in the fault free case and with a fault of 30% less fuel injection in cylinder 1 are hardly visible, Isermann and Leykauf (2009). But the wavelet analysis can detect these small changes. The wavelet coefficients $W(a,\tau)$ can be seen in Fig. 4.3.5 c) and d) for different scales a and time shift τ. The scale a corresponds to a frequency by $1/f$. In the case of an injection fault different coefficients appear for a frequency $f = 50$ Hz which is the injection frequency f_{inj}. At some operation points also wavelet coefficients at $f = 12.5$ Hz could be observed, which is a quarter of the injection frequency and corresponds directly to one cylinder with less fuel injection, see Fig. 4.3.5 d). Because the injection frequency is known from the engine speed, only two specific coefficients $W(f_{\text{inj}},\tau)$ and $W(f_{\text{inj}}/4,\tau)$ have to be calculated. Positive results could be found with this analysis method in the engine-speed-range $n_{\text{eng}} = 1200 \ldots 3000$ rpm and engine loads $M_{\text{eng}} > 80$ Nm, and a fault of 20 –

30% less fuel injection in one cylinder, Leykauf and Isermann (2006), Leykauf and Isermann (2008).

For fault detection the following residual is calculated by the comparison of the actual calculated wavelet coefficient for the injection frequency with the coefficient in fault-free case stored in a map:

$$r_{\text{inj1}} = W\left(f_{\text{inj}}, \tau\right) - W\left(f_{\text{inj}}, \tau\right)_{\text{norm}} \tag{4.3.4}$$

This residual is then used to calculate a symptom S_{inj1} for less fuel injection in one cylinder for engine loads $M_{\text{eng}} > 80\,\text{Nm}$. S_{inj1} is set to $+1$ if the residual is $r_{\text{inj1}} \geq 0.1$.

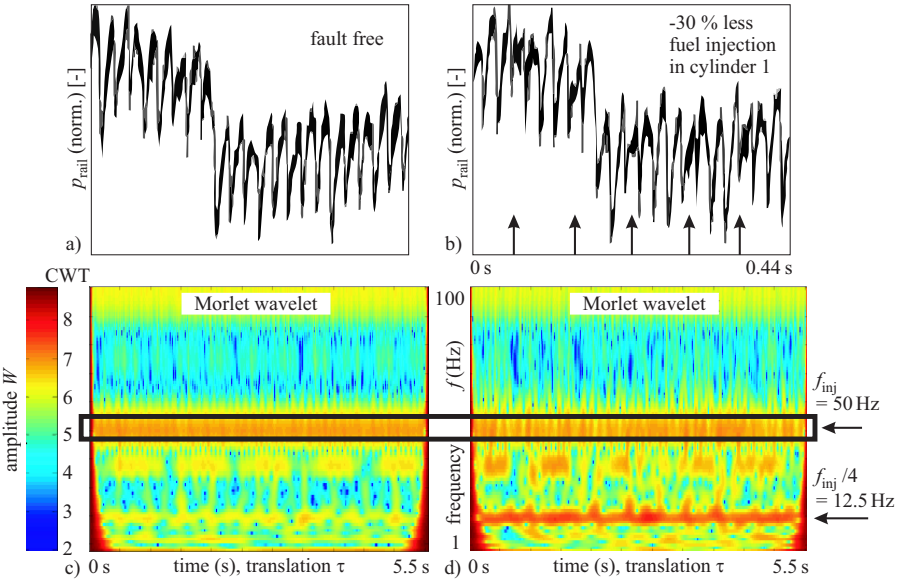


Fig. 4.3.5. Wavelet analysis of the rail pressure for $n = 1500\,\text{rpm}$, $M = 90\,Nm$. **a** fault free. **b** 30% less fuel injection in cylinder 1. **c** wavelet coefficients in fault free case. **d** wavelet coefficients for injection fault (The arrows show the instants of less injection).

b) Analysis of the engine speed signal

Another method to detect faults in the injection system is the analysis of the engine speed as shown in Sect. 4.2.1. If investigations are made on a test bench, the coupling between the engine and the asynchronous machine of the test bench makes the measurement not as sensitive as a measurement in a vehicle, where the speed is not controlled by an electrical motor and, therefore, shows larger amplitudes. Figure 4.3.6 shows the speed signal over 720°CS with 15% less fuel injection in cylinder 1. It is

obvious that the speed amplitude of the first cylinder is smaller than the amplitude of the cylinders 2–4. The amplitude is calculated with

$$\Delta n_{\mathrm{Cyl-i}} = max\left[\Delta n_{\mathrm{Cyl-i}}\right] - min\left[\Delta n_{\mathrm{Cyl-i}}\right](i = 1 \ldots 4)\,. \tag{4.3.5}$$

Figure 4.3.7 depicts the calculated amplitudes of the speed signal over 50 working cycles with the same fault in cylinder 1. The amplitude of cylinder 1 is significantly smaller than the amplitudes of the other three cylinders. Also a variation of the size of the amplitude is obvious. Therefore, the fault detection compares the actual amplitude with this amplitude in fault free case stored in a map for five working cycles:

$$r_{\Delta n_{\mathrm{Cyl-i}}} = \frac{1}{5}\sum_{k=0}^{4}\left[\Delta n_{\mathrm{Cyl-i}}(k) - \Delta n_{\mathrm{Cyl-i,map}}\right](i = 1 \ldots 4)\,. \tag{4.3.6}$$

A symptom S_{mec1} is then set to -1 if the change of the amplitude in a fault case is larger than 15 rpm. This symptom is very fast to calculate and relative sensitive for cylinder specific injection mass changes, especially in a vehicle, see Kimmich (2004), Isermann (2005).

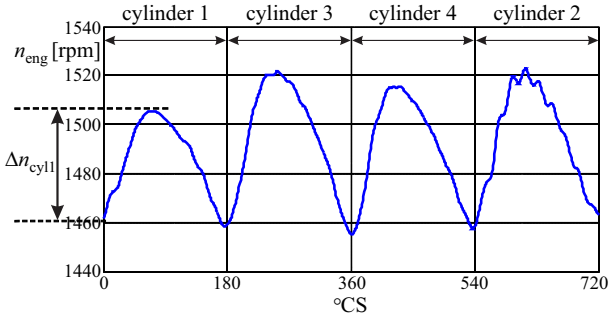


Fig. 4.3.6. High-resolution speed signal for one working cycle of $n_{\mathrm{eng}} = 1500$ rpm and $M_{\mathrm{eng}} = 70$ Nm with 15 % less fuel injection in cylinder 1 at engine test bench with VW 1.6 l FSI.

c) Fault detection and diagnosis in the rail pressure system

Based on the wavelet analysis of the rail pressure and the speed analysis three symptoms can be generated:

- $S_{\mathrm{inj1}} = 1$ if $r_{\mathrm{inj1}} > 0.1$: detection of less fuel injection in a specific cylinder ($S_{\mathrm{mec1}} = 0$ else) – wavelet analysis of rail pressure, detectable changes: 20 – 30 % less injection
- S_{inj2} from rail pressure: no injection during overrun state, as an information on the high-pressure pump. Then $S_{\mathrm{inj2}} = 1$ if p_{rail} too small, see Leykauf and Isermann (2006)

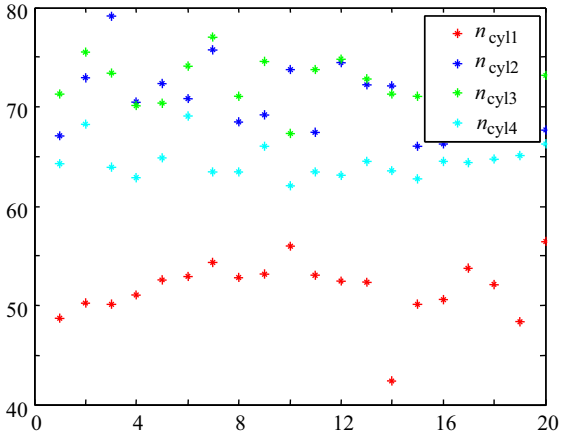


Fig. 4.3.7. Engine speed amplitudes of cylinders 1-4, $n_{\mathrm{eng}} = 1500\,\mathrm{rpm}$ and $M_{\mathrm{eng}} = 70\,\mathrm{Nm}$ with 15 % less fuel injection in cylinder 1 (at engine test bench).

- $S_{\mathrm{mec1}} = -1$ if $r_{\Delta \mathrm{n-cyl-i}} < -15\,\mathrm{rpm}$ engine speed amplitude ($S_{\mathrm{mec1}} = 0$ else), from high-resolution measurement of the speed signal. Detectable changes: $10\,\%$ less injection.

The detection uses stored data measured with $1^{\circ}\mathrm{CS}$ resolution and evaluates the signal over several cycles. The wavelet analysis uses 200 working cycles. Diagnosis is possible if the injected fuel mass is minimum 20–30 % less than normal. For the detection with the engine speed signal only a few working cycles are necessary (five were adequate for n = const.). The diagnosis works with binary logic. A fault-symptom table is included in Table 4.4.1.

4.3.3 Tank leak diagnosis

For vehicles with gasoline engines, emission reduction legislation requires an evaporative emission control system with the goal to avoid that fuel vapor from the tank leaves the tank. The fuel in the tank evaporates if the temperature increases or the tank pressure increases relative to the ambient pressure. The fuel in the tank is warmed up internally because of the energy losses of the fuel pump and fuel return from the injection system. Figure 4.3.8 shows a scheme for a fuel evaporation-emission control system, Robert Bosch GmbH (2011).

A fuel tank vent line is connected with a canister containing activated coal. If fuel vapor leaves the tank, the activated coal absorbs it such that only air leaves the canister. A regeneration of the activated coal is started by opening the canister purge solenoid valve. The vacuum in the intake manifold forces fresh air through the carbon to the intake after the throttle. The air then takes the absorbed fuel and transports it to the cylinders. Opening and closing of the purge-valve according to a PWM signal is controlled by the ECU depending on the operation point and time, such that the activated coal is able to absorb fuel vapor. Because of the higher boiling temperature

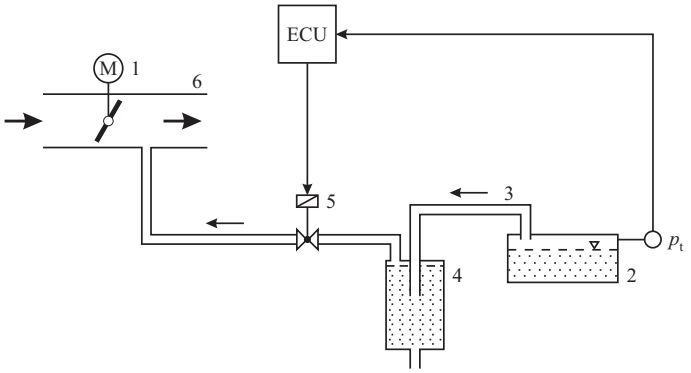


Fig. 4.3.8. Fuel evaporative emissions control system: p_t tank pressure, 1 electrical throttle, 2 fuel tank, 3 fuel tank vent line, 4 carbon canister, 5 canister-purge valve, 6 intake manifold.

of diesel fuel (180°C), evaporative-emission control is not required for diesel engine vehicles.

The EOBD limits the supervision of the evaporation-emission control system in checking the electrical circuit of the tank pressure sensor and the canister-purge valve. The OBD II in US however must detect leaks. Since 2000, CARB regulations require that leaks of 0.5 mm diameter have to be detected. Two methods for leak detection are applied.

The *low-pressure method* monitors the tank pressure during the tank ventilation with the carbon-canister purge valve. In case of a leak to atmosphere the pressure gradient is smaller or zero.

The *over-pressure method* uses an electrical vane pump to increase the tank pressure. In the case of a well sealed tank the required electrical current of the pump is large. Thus the size of the measured current is a measure for the leak size (or for electrical motor defects, like high friction).

A low-pressure leak-detection method is described in Krysander and Frisk (2009). The carbon canister possesses a diagnosis valve to atmosphere and a purge control valve to the intake manifold. For leak detection the diagnosis valve is closed and the purge valve is opened. The tank pressure then drops by about 0.02 bar within a few seconds. The purge valve is closed and the tank pressure observed. In the case of no leak, it will increase slowly because the fuel evaporates until it reaches saturation pressure, which is, however, temperature-dependent. A leakage leads to a larger pressure increase. Assuming a constant fuel vapor volume in the tank, applying the gas law and a leak mass flow through a hole then leads to a first order differential equation for the tank pressure. Parameter estimation of an evaporation related pressure gradient then is used for leak detection of holes in the size of 1 to 5 mm in less than 10 s.

4.4 Ignition system

A further source of faults in the combustion is the ignition system. For an experimental analysis, the ignition energy can be reduced simultaneously for all four cylinders using the ECU for manipulation the charging-time of the ignition coil. Of special interest is if there is a transitional area between a normal ignition and a measurable less effective ignition. Measurements were performed for different speed-load-points in the stratified operation mode. For the detection the engine speed amplitudes were used, because they are a measure of the produced power of the relevant cylinder. The engine speed amplitude was analyzed over 300 working cycles. For the operation point $n = 1500\,\text{rpm}$ and $M = 55\,\text{Nm}$ in stratified mode the amplitudes in fault free case (charging time for the ignition coil $t_{\text{char}} = 2.6\,\text{ms}$) and with 80% reduced charging time ($t_{\text{char}} = 0.5\,\text{ms}$) can be observed in Fig. 4.4.1. In the fault case the engine speed amplitudes of all four cylinders are irregularly reduced from an average value of 65 rpm to less than 30 rpm with only a small transitional area of the reduced charging time indicating misfires.

Using the speed amplitudes the coefficient of variance COV is determined by

$$COV_{\Delta n_{\text{Cyl-i}}} = \frac{\sigma\left(\Delta n_{\text{Cyl-i}}\right)}{\Delta n_{\text{Cyl-i,mean}}} \cdot 100\%\,(i = 1 \ldots 4)\,. \tag{4.4.1}$$

Figure 4.4.2 depicts how COV depends on reducing the charging time. No change by reducing the charging until about 70% could be observed. However, after reducing it more than about 75% the COV increases strongly due to misfires.

Hence, a larger reduction of the ignition energy can be detected by the symptom S_{mec1} for engine speed amplitudes, as for injection faults. However, different to about 30% less fuel injection a large reduced ignition energy results in a drastic decrease of the speed signal amplitude.

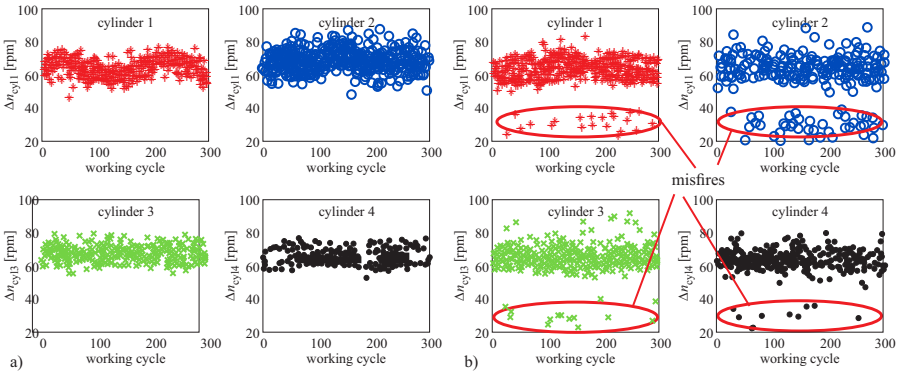


Fig. 4.4.1. Engine speed amplitudes of cylinders 1–4 $n_{\text{eng}} = 1500\,\text{rpm}$ and $M_{\text{eng}} = 55\,\text{Nm}$ in stratified operation mode for VW 1.6 l FSI. **a** fault free case: charging time of the ignition coils $t_{\text{char}} = 2.6\,ms$. **b** faulty case: charging time $t_{\text{char}} = 0.5\,\text{ms}$.

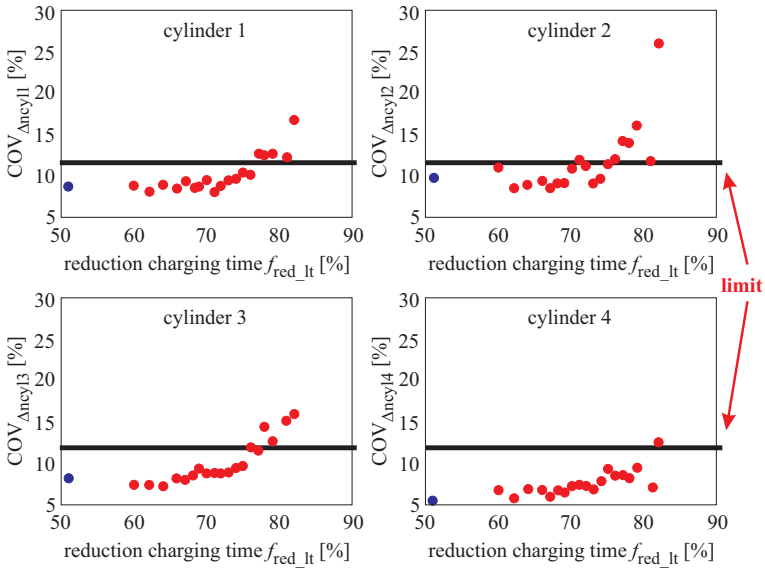


Fig. 4.4.2. Coefficients of variance for engine speed amplitudes of cylinders 1–4, $n_{eng} = 1500\,\text{rpm}$, $M_{eng} = 55\,\text{Nm}$ in stratified operation mode for different charging times of the ignition coils.

Summary for the DI gasoline engine:
The developed fault-diagnosis methods for the considered direct injection gasoline engine in Sects. 4.1.2a), 4.3.2 and 4.4 concentrated on detecting faults in the intake manifold, the fuel injection system and the ignition system. It was shown how leaks before and after the throttle valve can be detected with the measurement of air flow, manifold pressure and output of the λ-controller in homogeneous and stratified operation. The detection of different leak sizes depends on the operating speed-load range of the engine. It is also possible to detect too high exhaust gas recirculation rates because of, for example, contaminated catalysts or silencers or stuck EGR valves.

The frequency analysis of the rail pressure signal with wavelet analysis and the amplitude determination of the speed oscillation allows to detect less fuel injection of one cylinder. Furthermore, it was shown that too low ignition energy is detectable by the speed analysis of the engine.

Table 4.4.1 summarizes the generated symptoms for the different investigated faults. The patterns show that the investigated faults are detectable and isolable. Therefore they can be diagnosed, partially also in size. Reduced injection mass and misfires due to ignition faults can be separated by evaluating both, the engine speed and the rail pressure, which is a considerable progress.

The fault diagnosis can be based on a fuzzy logic treatment of the symptoms. However, to simplify the evaluation, binary logic was used, also to demonstrate the ability to isolate the faults. It is noted that only typical series production sensors are

required for the fault diagnosis, however, with a high resolution of the rail pressure and engine speed signal.

Hence, a comprehensive fault diagnosis system of a combustion engine can be realized by an appropriate combination of signal-model and process-model-based fault-detection methods. Different operating points have to be used to gather enough and well sized symptoms.

Table 4.4.1. Fault-symptom table for the intake system and the combustion of a DI gasoline engine VW 1.6 l FSI (+: positive symptom, −: negative symptom, 0: no symptom change, d: don't care, ✓: applicable, ×: non-applicable).

Symptom:	S_{int1}	S_{int2}	S_{exh1}	S_{inj1}	S_{inj2}	S_{mec1}	Operating mode	
fault type:	Air mass flow	Intake pressure	Output Lambda-controller	Rail pressure (injection)	Rail pressure (in overrun state)	Speed signal amplitude	homogeneous	stratified
leakage before throttle valve	−	0	+				✓	✓
leakage after throttle valve (2mm)	−	+	+				✓	✓
leakage after throttle valve (3mm)	−	++	++				✓	✓
increased exhaust gas recirculation (with restriction)	−	+	d				✓	✓
less injection mass in one cylinder				+	d	−	✓	×
less fuel supply				0	−	0	(✓)	(✓)
reduced ignition energy				0	d	− −	✓	✓

4.5 Combustion pressure analysis

The crank-angle-dependent analysis of the combustion pressure in the individual cylinder yields of course a direct measure of the combustion process. The indicated combustion pressure is a standard procedure for developing combustion processes, at least with special sensors on test benches. However, only recently combustion pressure measurement is considered for series production engines, especially in connection with new combustion processes like homogeneous charge compression ignition (HCCI), both for gasoline and diesel engines. Then simple combustion pressure features like "center of gravity" or "peak difference pressure", can be used for fault detection, which is described in Isermann (2014), Sect. 4.2. An alternative is a heat release analysis. For example, the mass-fraction-burned MFB can be calculated for different crank angles $\varphi = \varphi_1, \varphi_2, \varphi_3, \ldots$ and differences $\Delta x_{\mathrm{MFB}}(\varphi)$ then indicate differences to the normal combustion process, see e.g. Müller (2003b) Isermann (2014).

4.6 Exhaust system

Faults in the exhaust system have their origin in leaks or congestions in the tail pipe or in the functions of the catalysts.

4.6.1 Leaks and congestions

Usually, only one or two oxygen sensors and sometimes the exhaust gas temperature are available to obtain a direct information on the exhaust path. Hartmanshenn and Isermann (2005) observed that in the case of a leak the amplitude of the lambda sensor before the catalyst increases both the average value and the amplitude of oscillations. Hence, residuals

$$r_{\mathrm{exh2}} = \overline{\lambda(t)} - \lambda_{\mathrm{norm}} \tag{4.6.1}$$

$$r_{\mathrm{exh3}} = E\left\{\lambda_{\max}(t) - \lambda_{\min}(t)\right\} \tag{4.6.2}$$

can be determined. In addition the correction factor c_λ of the λ-control increases

$$r_{\mathrm{exh1}}(t) = c_\lambda(t) - 1, \tag{4.6.3}$$

which was also used for leak detection in the air path, see Sect. 4.1.

An increased restriction in the intake path leads to a higher exhaust pressure and therefore to an increased exhaust gas recirculation. As usually the exhaust gas pressure is not measured, the intake manifold pressure can be used for symptom generation, because it increases with increasing exhaust gas recirculation. The corresponding residual $r_{\mathrm{p_2}} = r_{\mathrm{int2}}$ was already described in (4.1.10) for leak detection in the air path.

4.6.2 Catalyst diagnosis

The exhaust-gas aftertreatment for gasoline engines is based on catalytic converters. Here, the three-way catalyst is dominating which is capable to reduce the main pollutants CO, HC and $\mathrm{NO_x}$. CO and HC are converted by oxidation to CO_2 and H_2O. $\mathrm{NO_x}$is converted by CO to N_2, CO_2 and O_2. The oxygen O_2 which is required for the oxidation of the main pollutants is taken from residual oxygen of the exhaust gases or from the $\mathrm{NO_x}$.

However, the optimal conversion is only possible for stoichiometric combustion, i.e. for the excess-air factor $\lambda = 1$, and for temperatures above $250^\circ\mathrm{C}$. As the range around $\lambda = 1$ (two-step lambda sensor, LSF, HEGO: heated exhaust gas oxygen sensor) is relatively narrow for an effective reduction of all three pollutants, a closed-loop lambda-control is required, such that a stoichiometric air-fuel ratio is reached.

The closed-loop control may operate with one or two lambda sensors. As the catalytic converter must operate in a very narrow range of $\lambda = 1 \pm 0.005$ the sensor is mounted before the converter (upstream), if only one sensor is applied. Applying the zirconium-dioxide sensor with a jump characteristic at $\lambda = 1$ (two-step lambda

sensor, LSF or HEGO) the required control performance can be reached by a two-step control with a ramp function of the injected fuel mass as manipulated variable. The manipulation variable then changes the correcting factor $u_\lambda = c_\lambda$ of the excess-air factor resulting in small amplitudes of about 3%.

An improvement of the control performance is obtained by using a second sensor after the converter. Then, a steady characteristic sensor (abbreviated by LSU, broad-band lambda sensor or UEGO: universal exhaust gas oxygen sensor) is placed upstream, and used for an inner control loop and the jump characteristic sensor downstream, correcting the setpoint of the inner loop with a slow integral type controller (cascaded control). This results in smaller control oscillations and small corrective controls from the downstream sensor, see e.g. the summary in Isermann (2014), Chap. 7.2.

With two lambda sensors it is possible to detect faults as well in the sensors as in the catalyst. A good functioning catalyst is able to store oxygen (because of a cerium oxide layer) in the lean phase ($\lambda > 1$) and to consume it in the rich phase ($\lambda < 1$) of the control oscillations. This damps the amplitudes of the λ-value from the upstream sensor through the catalyst to the downstream sensor. The stored oxygen mass is in the range of 100 mg to 1 g, Robert Bosch GmbH (2011). For ageing catalysts the storage capacity reduces and therefore the downstream amplitudes increase.

The *fault detection of the catalyst* is usually based on monitoring the oxygen storage capacity of the catalyst. Therefore, a first method is to start in a stationary operation stage by changing the setpoint of the inner controller, e.g. to $\lambda = 0.95$ and to 1.05 in order to generate defined oscillations by this test signal. The amplitudes of the downstream sensor then must be considerably smaller than those of the upstream sensor, otherwise the catalyst is defect due to ageing or poisoning e.g. by led and sulfur, Knirsch and Klee (2002). This method can be used in a normal stationary operation phase. If only one setpoint change is commanded, the stored oxygen mass is calculated from rich to lean and is calculated for one cycle between the two saturation limits, Willimowski et al (2005)

$$S_{\mathrm{cat1}} = \Delta m_{\mathrm{O2}} = 0.23 \int_{t_1}^{t_2} [\lambda(t) - 1]\ \dot{m}_{\mathrm{air}}(t)\, dt \tag{4.6.4}$$

with $\dot{m}_{\mathrm{air}}$ air-mass flow rate. (mass of O_2 is 0.23 multiplied with the mass of air). In order to avoid the use of an extra set point test signal to estimate the oxygen storage capacity, motoring operations of the engine can be used, Louen et al (2015). Because then no fuel injection and no combustion occurs, pure air with 21 % O_2 flows to the catalyst. This leads to a storage of oxygen and the downstream two-step λ-sensor indicates an air excess factor $\lambda \gg 1$. If the storage is filled, the oxygen content after the catalyst increases, and the λ-sensor jumps to $\lambda < 1$, indicating the end of the storage time Δt_{in}. After the end of the motoring phase, combustion begins again and a phase with decreasing storage with duration Δt_{out} follows. However, this duration depends on the exhaust gas mass flow $\dot{m}_{\mathrm{exh}}$ and other variables. Monitoring of the oxygen storage capacity can be based on features like Δt_{in}, Δt_{out}, $\bar{\dot{m}}_{\mathrm{exh}}$ and by training with new and aged catalysts multi-dimensional thresholds for limit-checking can be found.

For further details and methods it is referred to Auckenthaler (2005), Peyton Jones and Muske (2004), Peyton Jones et al (2006), Feßler (2011), Louen et al (2015), Odendall (2015), and Odendall (2016).

4.7 Cooling system

Since about one third of the heat provided by the fuel has to be removed to the environment, a precise functioning control and supervision of the cooling system is an important requirement. Figure 4.7.1 depicts a scheme of a typical, basic cooling system. The coolant pump is either driven by a V-belt and therefore the speed is given through the engine speed or by an electric motor, which allows independent speed control. After cold start the thermostatic valve opens to the bypass only in order to reach fast the operation coolant temperature. The thermostatic controller then opens the way to the radiator such that external cooling begins. The valve is a double-acting disk which is actuated by a temperature-dependent expansion valve, in which both sides are opened allowing a mixture of uncooled and cooled coolant flow to the engine such that according to the setpoint of the thermostat a certain operating temperature at the entrance to the engine is reached. The setpoint of an electronic map controlled thermostat may be changed through a heating resistor to raise the operating temperature at part-load and to reduce it at higher load. This allows to save fuel, to reduce wear and improves the heating of the vehicle interior. In the case of electrical driven pumps the speed is also changed and optimized with regard to operating requirements.

As the temperature time behavior of the coolant system is much slower than, for example, the torque generation, the coolant circuit is frequently in a dynamic state during driving with different load.

Modeling of the cooling system is treated e.g. in Corbel (1987), Chang et al (1991), Salah et al (2010), Müller et al (2016), and Nahim et al (2016). It was shown in Isermann (2014) that the dynamic behavior of the coolant temperature ϑ_c in dependence on the manipulated variable u_1 of the expansion valve can be approximated by a nonlinear second order model or after linearization by a first order model with dead time and varying parameters.

4.7.1 Fault detectionof the cooling system with mechanical driven pumps

The supervision of the cooling system is usually based on the measurements of the coolant temperature at the thermostatic expansion valve and a coolant level measurement e.g. in the coolant expansion tank. Limit value checking is applied for too high temperature and too low coolant level and violation is signalled in the cockpit. Faults in the coolant control can be based on control deviations between the coolant reference value and its measured value

$$e_\vartheta(t) = \vartheta_c(t) - \vartheta_{c,\mathrm{ref}}(t). \tag{4.7.1}$$

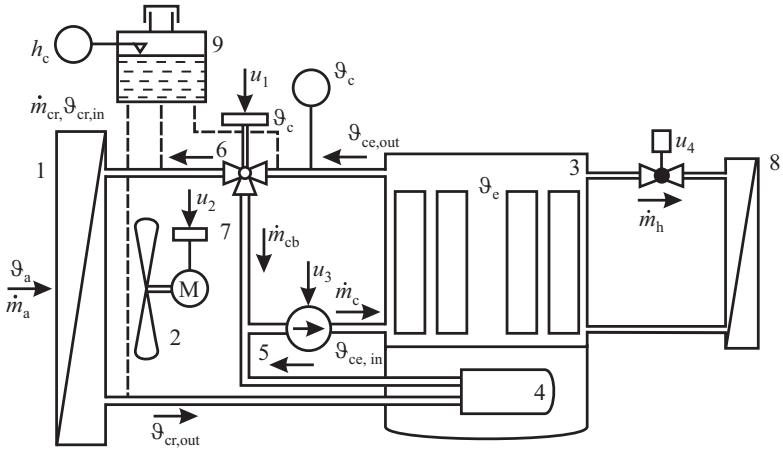


Fig. 4.7.1. Schematic of an engine cooling system: 1 coolant radiator 2 fan 3 combustion engine 4 oil cooler 5 coolant pump 6 thermostatic valve 7 bypass 8 cabin heat exchanger 9 coolant expansion tank. u_1 manipulated variable for thermostatic valve, u_2 manipulated variable for fan speed, u_3 manipulated variable for pump speed, u_4 manipulated variable for cabin heating, ϑ_c coolant temperature, h_c coolant level.

This is directly possible for electronic map-controlled thermostats, where the reference value is determined by the ECU and the cooling temperature is measured electrically and is also available in the ECU. However, the reasons for remaining control deviations can be manifold, as e.g. defective thermostat, defective coolant pump, contaminated radiator or defective fan. OBD requires to detect emission related faults of the cooling system. As too low coolant temperature may increase emissions, the measured coolant temperature is compared with a minimum temperature, that has to be reached.

To monitor the temperature sensor plausibility checks are provided as the engine cools down. This may indicate a sticking sensor.

In Nahim et al (2016) model-based fault detection for a marine diesel engine with sea water cooling is considered using temperature measurements. Simulated faults are wear and leak of the pump, locked thermostat and leaks and fouling of the heat exchanger.

4.7.2 Fault detection with electrical driven coolant pumps

A more advanced supervision of the cooling system becomes possible if electrical cooling pumps are used. Their speed can be controlled independently of engine speed. Measurement of the voltage, current and speed then enables to determine the coolant mass flow based on a mathematical coolant pump model, Schäfer et al (2006), Schäfer (2012). Therefore coolant pump models are considered.

The steady state behavior of a centrifugal pump is usually described by the generated pressure difference $\Delta p_\mathrm{p} = p_\mathrm{out} - p_\mathrm{in}$ or delivery head $H = \Delta p_\mathrm{p}/\rho\, g$ in

dependence of the volume flow $\dot{V}_\mathrm{p}$ and speed ω_p

$$\Delta p_\mathrm{p} = f(\dot{V}_\mathrm{p}, \omega_\mathrm{p}). \tag{4.7.2}$$

Taking into account a finite number of blades, blade and tube friction losses, impact losses due to non-tangential flow of the blade entrance, the basic equation for the delivery head becomes, Dixon (1966), Pfleiderer and Petermann (2005), Isermann (2011)

$$\Delta p_\mathrm{p} = \rho g H = \rho g \left(h_\mathrm{nn} \omega_\mathrm{p}^2 - h_\mathrm{nv} \omega \dot{V}_\mathrm{p} - h_\mathrm{vv} \dot{V}_\mathrm{p}^2 \right). \tag{4.7.3}$$

In order to obtain dimensionless characteristic quantities and to be more independent on the geometrical size of the pump, following variables are introduced, which originate from fluid mechanical laws for turbomachines:

delivery number

$$\varphi = \frac{\dot{V}_\mathrm{p}}{A_\mathrm{p} c_\mathrm{u}} = \frac{4\dot{V}_\mathrm{p}}{\pi^2 \omega_\mathrm{p} D_2^3}, \tag{4.7.4}$$

pressure number

$$\psi = \frac{2gH}{c_\mathrm{u}^2} = \frac{2gH}{(\pi \omega_\mathrm{p} D_2)^2}, \tag{4.7.5}$$

power number

$$\lambda = \frac{\psi \varphi}{\eta_\mathrm{p}}. \tag{4.7.6}$$

Herewith are:

rotor area	$A = \pi D_2^2/4$,
rotor outer velocity	$c_\mathrm{u} = 2\pi D_2 \omega_\mathrm{p}/2$,
rotor outer diameter	D_2,
pump efficiency	η_p.

Introducing (4.7.4) and (4.7.5) in (4.7.3) yields

$$\psi = k_1 + k_2 \varphi + k_3 \varphi^2 \tag{4.7.7}$$

and instead of a family of characteristics with a parameter ω_p one obtains only one characteristic.

(4.7.4) leads to

$$\dot{V}_\mathrm{p} = D_2^3 \pi^2 \omega_\mathrm{p} \varphi(\lambda)/4 \tag{4.7.8}$$

and the volume flow can be determined if

$$\varphi(\lambda) = \frac{\lambda}{\psi} \eta_\mathrm{p} \tag{4.7.9}$$

is known.

If the pump is driven by a DC motor, which takes the electrical power $P_\mathrm{el} = UI$ from the supply net, then the overall efficiency of the motor-pump system is

$$\eta = \eta_{\mathrm{mot}}\,\eta_{\mathrm{p}} = \frac{P_{\mathrm{out}}}{P_{\mathrm{in}}} = \frac{p_{\mathrm{p}}\dot{V}}{UI}. \tag{4.7.10}$$

Hence, if the motor efficiency η_{mot}

$$\eta_{\mathrm{mot}} = \frac{M\,\omega}{UI} = f(P_{\mathrm{in}}, \omega_{\mathrm{mot}}) \tag{4.7.11}$$

is known (e.g. as a look-up table), the pump efficiency

$$\eta_{\mathrm{p}} = \frac{\eta}{\eta_{\mathrm{mot}}} \tag{4.7.12}$$

can be determined. For the power number holds with (4.7.10), (4.7.9), (4.7.8)

$$\lambda = \frac{2gH}{(\pi\omega_{\mathrm{p}}D_2)^2} \cdot \frac{4\dot{V}_{\mathrm{p}}}{\pi^2\omega_{\mathrm{p}}D_2^3} \cdot \frac{UI\eta_{\mathrm{mot}}}{\rho g H \dot{V}_{\mathrm{p}}} = \frac{8\,UI\eta_{\mathrm{mot}}}{\pi^4 D_2^5 \rho\,\omega_{\mathrm{p}}^3}. \tag{4.7.13}$$

Hence, λ can be determined if U and I are measured and η_{mot} is known from experiments. Further, if the characteristic $\lambda = f(\varphi)$ for the pump is known by measurements its inversion yields $\varphi = f^{-1}(\lambda)$ and with (4.7.8) one obtains the volume flow based on measurements of U, I and ω_{p}, Schäfer (2012).

For a coolant pump with rated power $P_{\mathrm{el}} = 800\mathrm{W}$, $\dot{V}_{\mathrm{c}} = 250\mathrm{l/min}$, $H = 14\mathrm{m}$, $n_{\mathrm{p}} = 5000\mathrm{rpm}$ the volume flow could be determined for $P_{\mathrm{el}} > 150\mathrm{W}$, $n_{\mathrm{p}} < 3000\mathrm{rpm}$, $\dot{V}_{\mathrm{c}} > 90\mathrm{l/min}$ with a mean standard deviation of about 4%. For smaller power and speed the motor efficiency becomes small, so that the reconstruction of the pump power is not accurate enough.

Based on the estimated coolant mass flow $\dot{V}_{\mathrm{p}}$ and thermal engine cooling models as described in Isermann (2014), model-based fault-detection methods can be developed, which allow to improve the diagnosis of faults in the cooling system.

The flow dynamic behavior of the cooling system is treated in Isermann (2014). For changes of the pump speed and linearization around an operation point, it leads to a first order differential equation

$$T_{\mathrm{c}}\frac{d\dot{V}_{\mathrm{p}}}{dt} + \Delta\dot{V}_{\mathrm{p}}(t) = K_{\mathrm{c}}\,\Delta\omega_{\mathrm{p}}(t). \tag{4.7.14}$$

For example, a parameter estimation of the gain K_{c} allows to detect faults like internal flow restrictions, which change the volume flow after a pump speed variation.

4.8 Lubrication system

The lubrication system serves to lubricate and cool the pairings of the crankshaft drive, the valve train and other moving components and to remove local contaminants, wear particles and combustion residues. The most frequently used forced-feed lubrication is based on a positive displacement oil pump (mostly gear pump or vane pump), which delivers a defined volumetric oil flow rate from the oil sump to an

oil cooler and and oil filter, see Fig. 4.8.1. The oil volume flow rate is about 30 to 35 l/h per kW engine power, Eberan-Eberhorst (2010). As the oil pressure increases with the speed of the pump, a pressure relief or pressure control valve limits the oil pressure to a maximum of, for example $p_{p,lim} = 5$ bar, to prevent damage of the oil cooler, oil filter and seals. The oil cooler is cooled by the engine coolant or by air. The oil filter can be circumvented by a pressure released bypass valve in the case of strong contamination, if e.g. the pressure difference exceeds 0.8 ... 2.5 bar. The main oil filter is a full-flow filter which filters the complete oil flow of the engine and holds back particles in the range of 0.5 ... 500 μm which could cause damage or wear. This filter is regularly exchanged as part of the oil service. Additionally, bypass filters can be provided, which hold back superfine particles, such as abrasive and soot particles. These bypass filters are mainly used for diesel engines and are limited to an oil flow of maximal 10%.

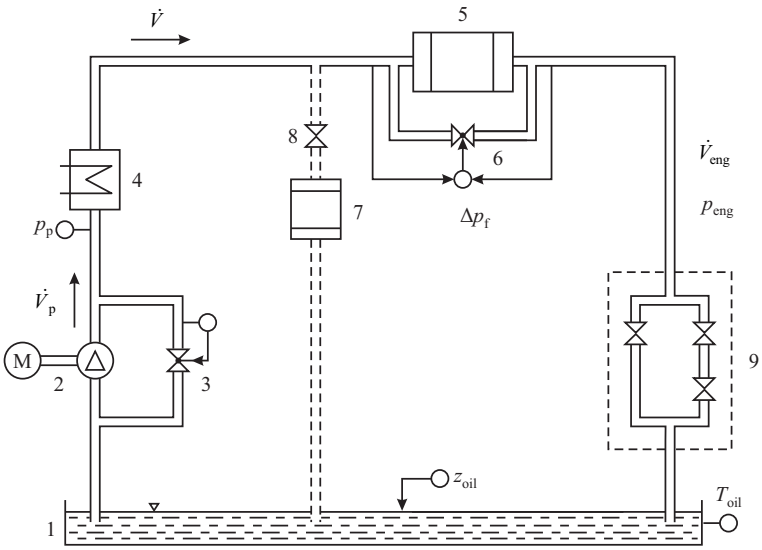


Fig. 4.8.1. Scheme of a lubrication oil circuit: 1 oil pan, 2 oil pump, 3 pressure control valve, 4 oil cooler, 5 full-flow filter, 6 bypass valve, 7 bypass filter, 8 throttle, 9 engine oil flow passages.

The oil first flows to an oil gallery and then through many oil ducts to the crankshaft, connecting rods, piston pins, camshaft, valves, turbocharger etc. and then flows back unpressurized to the oil sump in the oil pan underneath the crank case, see for example Mollenhauer and Tschöke (2010), Robert Bosch GmbH (2011). An alternative to this wet-sump lubrication is the dry-sump lubrication, when the oil is gathered in an external oil reservoir. This system gives more freedom for the design of the engine and functions also for high lateral accelerations.

For steady-state conditions, the oil volume flow rate and the oil pressures depend on the flow characteristics of the pump, the filters and the engine. The delivered flow of the pump $\dot{V}_\mathrm{p}$ is mainly determined by the pump speed ω_p and the flow absorbed by the engine $\dot{V}_\mathrm{eng}$, which depends on the oil pressure p_eng, the engine speed ω_eng and the oil temperature T_oil. In the following the behavior of the pressures and flows is considered with simplifying assumptions. Because the oil pump sucks the oil from the oil pan with approximately atmospheric pressure, it holds for the pump difference pressure $\Delta p_\mathrm{p} = p_\mathrm{p}$. The design and basic equations for gear and vane pumps are described in Ivantysyn and Ivantysynova (1993), Reinhardt (1992), Eberan-Eberhorst (2010), Berg (2009), Mollenhauer and Tschöke (2010), Affenzeller and Gläser (1996). Contributions on modeling the oil circuit can be found e.g. in Neu et al (1977), Reulein (1998), Haas et al (1991), and Chun (2003).

4.8.1 Models of a lubrication circuit

The delivered theoretical volume flow $\dot{V}_\mathrm{p}$ of a positive displacement pump, as for example, a gear pump is

$$\dot{V}_\mathrm{p,th} = V_\mathrm{p,th}\,\omega_\mathrm{p} \tag{4.8.1}$$

where $V_\mathrm{p,th}$ is the theoretical (ideal) displacement volume per turn. The required driving power becomes

$$P_\mathrm{p,th} = \dot{V}_\mathrm{p,th}\,p_\mathrm{p} = V_\mathrm{p,th}\,\omega_\mathrm{p}\,p_\mathrm{p}. \tag{4.8.2}$$

The real, effective volume flow is lowered by internal leak flows $\dot{V}_\mathrm{p,l}$, see Fig. 4.8.2 a), which depend on the generated pressure p_p and speed ω_p

$$\dot{V}_\mathrm{p,eff}\,(p_\mathrm{p},\omega_\mathrm{p}) = V_\mathrm{p,th}\,\omega_\mathrm{p} - \dot{V}_\mathrm{p,l}\,(p_\mathrm{p},\omega_\mathrm{p}) = V_\mathrm{p,th}\,\omega_\mathrm{p}\,\eta_\mathrm{vol} \tag{4.8.3}$$

and can be taken into account by a volumetric efficiency η_vol, see Isermann (2005). This efficiency may also comprise charging losses, which reduce $V_\mathrm{p,th}$, for example, for higher pump speeds.

The corresponding power loss through the internal leakages is

$$P_\mathrm{p,l} = \dot{V}_\mathrm{p,l}\,p_\mathrm{p} \tag{4.8.4}$$

leading to a reduced delivery power

$$\begin{aligned} P_\mathrm{p,vol} &= P_\mathrm{p,th} - P_\mathrm{p,l} = \left(V_\mathrm{p,th}\,\omega_\mathrm{p} - \dot{V}_\mathrm{p,l}\right) p_\mathrm{p} \\ &= V_\mathrm{p,th}\,\omega_\mathrm{p}\,p_\mathrm{p}\,\eta_\mathrm{vol} = P_\mathrm{p,th}\,\eta_\mathrm{vol}. \end{aligned} \tag{4.8.5}$$

Pump internal flow losses yield a pressure drop $\Delta p_\mathrm{p,l}\,(\omega_\mathrm{p})$ such that the effective pump difference pressure becomes

$$p_\mathrm{p,eff} = p_\mathrm{p}(\omega_\mathrm{p}) - \Delta p_\mathrm{p,l}(\omega_\mathrm{p}). \tag{4.8.6}$$

Figure 4.8.2 b) shows qualitatively a characteristic of the effective pump pressure, where the pump operates against a constant throttle resistance. This characteristic may be approximated by

$$p_{\mathrm{p,eff}} = p_{\mathrm{p0}} + c_{\mathrm{p1}}\omega_{\mathrm{p}} + c_{\mathrm{p2}}\omega_{\mathrm{p}}^2. \tag{4.8.7}$$

The corresponding pump delivery power to the fluid is then reduced

$$P_{\mathrm{p,vol,l}} = P_{\mathrm{p,vol}} - \dot{V}_{\mathrm{p,th}} \Delta p_{\mathrm{p,l}} \tag{4.8.8}$$

and can be taken into account by a hydraulic loss efficiency η_{h}

$$P_{\mathrm{p,vol,l}} = P_{\mathrm{p,vol}}\,\eta_{\mathrm{h}} = P_{\mathrm{p,th}}\,\eta_{\mathrm{vol}}\,\eta_{\mathrm{h}}. \tag{4.8.9}$$

The bearing and gear friction losses are covered by a mechanical efficiency η_{m} such that the effective pump delivery power to the pumped fluid becomes

$$P_{\mathrm{p,eff}} = P_{\mathrm{p,vol,l}}\,\eta_{\mathrm{m}} = P_{\mathrm{p,th}}\,\eta_{\mathrm{vol}}\,\eta_{\mathrm{h}}\,\eta_{\mathrm{m}} = P_{\mathrm{p,th}}\,\eta_{\mathrm{tot}}. \tag{4.8.10}$$

The overall efficiency η_{tot} covers all losses. The required shaft driving power is therefore

$$\begin{aligned} P_{\mathrm{p,in}} &= \frac{1}{\eta_{\mathrm{tot}}} P_{\mathrm{p,eff}} = \frac{1}{\eta_{\mathrm{tot}}} \dot{V}_{\mathrm{p,eff}}(p_{\mathrm{p}}, \omega_{\mathrm{p}}) p_{\mathrm{p,eff}}(\omega_{\mathrm{p}}) \\ &= \frac{1}{\eta_{\mathrm{vol}}\,\eta_{\mathrm{h}}\,\eta_{\mathrm{m}}} V_{\mathrm{p,th}}\omega_{\mathrm{p}} \left(p_{\mathrm{p0}} + c_{\mathrm{p1}}\omega_{\mathrm{p}} + c_{\mathrm{p2}}\omega_{\mathrm{p}}^2\right). \end{aligned} \tag{4.8.11}$$

It increases linearly as well as mainly quadratic with the pump speed.

The oil flow absorbed by the engine is assumed to be determined by a turbulent flow which approximately follows the flow law through a contraction

$$\dot{V}_{\mathrm{eng}}\left(p_{\mathrm{eng}}, \omega_{\mathrm{eng}}\right) = c_{\mathrm{eng}}\left(\omega_{\mathrm{eng}}, T_{\mathrm{oil}}\right) \sqrt{\frac{2}{\rho_{\mathrm{oil}}}} \sqrt{p_{\mathrm{eng}}}. \tag{4.8.12}$$

Since the flow also depends on the engine speed, the coefficient c_{eng} changes with ω_{eng}. As the oil temperature has a large influence on the oil viscosity, it has a considerable effect on c_{eng}, and the absorbed oil flow through the engine behaves as depicted in Fig. 4.8.2 d).

p_{eng} is the acting pressure at the engine entrance (gallery) and is reduced by pressure drops in the filter, oil cooler and channels

$$p_{\mathrm{eng}} = p_{\mathrm{p}} - \Delta p_{\mathrm{f}}\left(\dot{V}_{\mathrm{p,eff}}\right). \tag{4.8.13}$$

The volume flow delivered by the pump $\dot{V}_{\mathrm{p}}$ may be reduced by the flow rate $\dot{V}_{\mathrm{cv}}$ of an active pressure control valve

$$\dot{V}_{\mathrm{p}}(t) = \dot{V}_{\mathrm{p,eff}}(t) - \dot{V}_{\mathrm{cv}}(u_1, t) \tag{4.8.14}$$

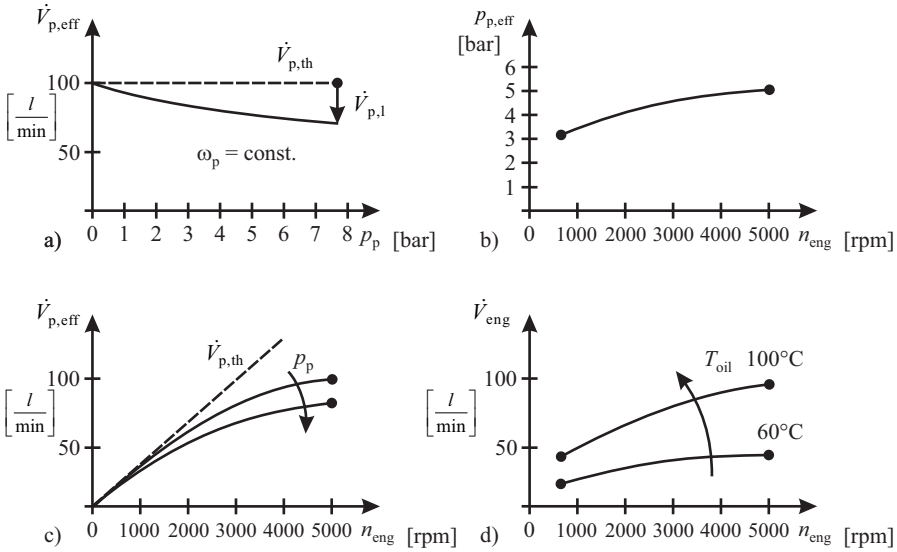


Fig. 4.8.2. Characteristics of lubrication circuit components (qualitative). **a** pump flow $\dot{V}_p(p)$. **b** pump pressure $p_p(n_{eng})$. **c** pump flow $\dot{V}_p(n_{eng})$. **d** engine absorption flow $\dot{V}_{eng}(n_{eng})$.

with

$$\dot{V}_{cv}(u_1) = \frac{1}{\sqrt{c_v(u_1)}}\sqrt{p_{p,eff}}, \tag{4.8.15}$$

see Fig. 4.8.3, where $c_v(u_1)$ is a resistance of the control valves.

The dynamic behavior of the *pressures* in the oil circuit can be determined by considering the mass balance equation taking a volume storage into account. The dynamics of the *flows* follow from the balance equation for the moved oil mass. Both dynamics were treated in Isermann (2013), Chapter 4.9. Here only the pressure behavior is considered, as the volume flows usually cannot be measured.

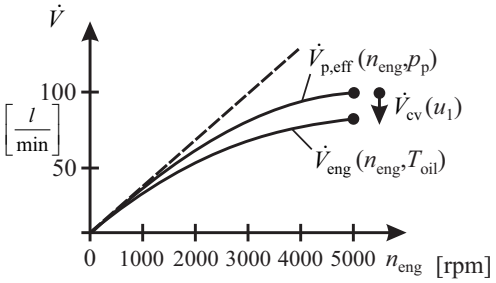


Fig. 4.8.3. Flow characteristic for the pump delivery and the engine absorption.

The dynamic behavior of the pressures in the oil circuit follows from the interaction of the mass flows. Therefore the *mass balance equation* between the delivered mass flow of the pump $\dot{m}_{\mathrm{p}}(t) = \rho_{\mathrm{oil}}\dot{V}_{\mathrm{p}}(t)$ and the absorbed mass flow of the engine $\dot{m}_{\mathrm{eng}}(t) = \rho_{\mathrm{oil}}\dot{V}_{\mathrm{eng}}(t)$ is considered, taking into account a mass storage $m_{\mathrm{pe}} = \rho_{\mathrm{oil}}V_{\mathrm{pe}}$ between the pump, and the distribution ducts to the lubrication points. This means that a mass storage in the flow line including oil cooler, oil filter and main gallery is considered. Assuming now $\rho_{\mathrm{oil}} = \text{const.}$ a volume balance equation results

$$\frac{d}{dt}V_{\mathrm{pe}}(t) = \dot{V}_{\mathrm{p}}(t) - \dot{V}_{\mathrm{eng}}(t). \tag{4.8.16}$$

The storage volume of this volume balance equation follows assuming a compressible fluid with bulk modulus

$$\beta = -V_{\mathrm{pe}}(\frac{\partial p}{\partial V}). \tag{4.8.17}$$

This results in a relation between pressure change and volume change $\Delta p = \Delta V\,\beta/V$ or

$$\frac{dp}{dt} = \frac{\beta}{V_{\mathrm{pe}}}\Delta\dot{V}. \tag{4.8.18}$$

A flow difference between pump and engine thus becomes for closed pressure control valve with $\dot{V}_{\mathrm{cv}} = 0$ (can be assumed if $p_{\mathrm{p}} < p_{\mathrm{p,lim}}$)

$$\frac{dp_{\mathrm{eng}}(t)}{dt} = \frac{\beta}{V_{\mathrm{pe}}}\left(\dot{V}_{\mathrm{p,eff}}(\omega_{\mathrm{p}},t) - \dot{V}_{\mathrm{eng}}(\omega_{\mathrm{eng}},p_{\mathrm{eng}},t)\right). \tag{4.8.19}$$

Insertion of (4.8.3) and (4.8.12) in (4.8.19) results in

$$\frac{V_{\mathrm{pe}}}{\beta}\frac{dp_{\mathrm{eng}}(t)}{dt} = \left(V_{\mathrm{p,th}}\,\omega_{\mathrm{p}}(t)\,\eta_{\mathrm{vol}} - c_{\mathrm{eng}}(\omega_{\mathrm{eng}})\sqrt{\frac{2}{\rho_{\mathrm{oil}}}}\sqrt{p_{\mathrm{eng}}(t)}\right). \tag{4.8.20}$$

Linearization around an operating point for pump speed changes yields, assuming c_{eng} as constant,

$$T_{\mathrm{p}}\frac{dp_{\mathrm{eng}}(t)}{dt} + \Delta p_{\mathrm{eng}}(t) = K_{\mathrm{p}}\Delta\omega_{\mathrm{p}}(t) \tag{4.8.21}$$

with gain and time constant

$$\begin{aligned} K_{\mathrm{p}} &= \frac{V_{\mathrm{p,th}}\eta_{\mathrm{vol}}}{a_1}, \\ T_{\mathrm{p}} &= \frac{V_{\mathrm{pe}}}{\beta\,a_1}, \\ a_1 &= \frac{c_{\mathrm{eng}}(\omega_{\mathrm{eng}},T_{\mathrm{oil}})}{\sqrt{2\rho_{\mathrm{oil}}\,\bar{p}_{\mathrm{eng}}}}. \end{aligned} \tag{4.8.22}$$

Thus, the pressure at the engine entrance behaves, after pump speed changes and with simplifying assumptions, as a linear first order system with a small time constant T_{p}. However, as the parameters depend on ω_{eng} and $\omega_{\mathrm{p}} = i_{\mathrm{p}}\,\omega_{\mathrm{eng}}$, with i_{p} the pump gear

ratio, the dynamic equation is nonlinear in reality. The time constant is the larger, the larger the storage volume, the smaller the bulk modulus and the smaller the flow resistance coefficient of the engine.

If a variable pump with manipulated displacement volume $V_{\mathrm{th}}(up)$ is applied (as for a gear pump with axial shifted gear wheels), it follows from (4.8.20) after linearization for constant speed ω_{eng}

$$T_{\mathrm{p}} \frac{dp_{\mathrm{eng}}(t)}{dt} + \Delta p_{\mathrm{eng}}(t) = K_{\mathrm{pV}} \Delta V_{\mathrm{p,th}} \tag{4.8.23}$$

with

$$K_{\mathrm{pV}} = \frac{\overline{\omega}_{\mathrm{p}} \eta_{\mathrm{vol}}}{a_1}. \tag{4.8.24}$$

4.8.2 Model-based fault detection of a lubrication circuit

As Fig. 4.8.1 shows the usually available measurements in lubrication circuits are the pressure p_{p} after the pump, the pressure difference Δp_{f} over the full-flow filter, the oil temperature $T_{\mathrm{oil,eng}}$ usually in the oil sump of the engine, where the different oil flows are mixed, and the oil level z_{oil} in the oil pan. The pump speed $\omega_{\mathrm{p}} = i_{\mathrm{p}} \omega_{\mathrm{eng}}$ is also known from the engine speed. Hence, various variables in the lubrication system like flows are not measured and therefore the possibilities of oil circuit diagnosis are rather limited.

The *conventional supervision* of the oil circuit is performed by limit value checking of the oil pressure of the pump, the oil temperature in the engine and the oil level in the oil pan. Alarms are given for

$$p_{\mathrm{p}} \leq p_{\mathrm{p,min}} \tag{4.8.25}$$

$$T_{\mathrm{oil,eng}} \geq T_{\mathrm{oil,eng,max}} \tag{4.8.26}$$

$$z_{\mathrm{oil}} \leq z_{\mathrm{oil,min}}. \tag{4.8.27}$$

Protection functions for a too high pump pressure become active, if the pressure control return valve opens or for a too high pressure drop of a contaminated full-flow filter a bypass valve opens, see Fig.4.8.1.

The electronic supervision of the oil level is usually based on simple float switches for measurements at stand still phases or on continuously operating sensors, which function also during driving. Known electronic oil level sensors use thermal, capacitive or ultrasonic measurement principles. As the oil level sloshes during driving useful level measurement is obtained by appropriate filtering respective averaging.

For the lubrication and cooling of the engine parts the oil pressure p_{eng} at the entrance to the engine, e.g. in the gallery, where the oil flows are distributed, is an important variable. If this oil pressure is measured, then (4.8.13) can be used to determine the pressure drops in the oil filter, oil cooler and channels

$$\Delta p_{\mathrm{f}} = p_{\mathrm{p}} - p_{\mathrm{eng}} \tag{4.8.28}$$

A comparison with the stored characteristic for the normal behavior by calculation of the residual

$$r_{\mathrm{pf}}(t) = \Delta p_{\mathrm{f}}(t) - \Delta p_{\mathrm{f,nom}}(\dot{V}_{\mathrm{p,eff}}, T_{\mathrm{oil}}). \tag{4.8.29}$$

then gives hints for contaminations or leaks, where $\dot{V}_{\mathrm{p,eff}}$ follows from (4.8.3).

A further possibility is to supervise the pump pressure increase in dependence on the pump speed. Based on the measured pressure $p_{\mathrm{p,eff}}(\omega_{\mathrm{p}})$ according to (4.8.6) and the characteristic shown in Fig. 4.8.2 b) with the approximation (4.8.7) the parameters

$$\underline{\hat{\theta}}_{\mathrm{p}}^{T} = [\hat{p}_{\mathrm{p0}} \;\; \hat{c}_{\mathrm{p1}} \;\; \hat{c}_{\mathrm{p2}}] \tag{4.8.30}$$

can be determined with least squares parameter estimation. Deviations of the parameters then can give hints for pump defects.

The volume flow balance between the pump delivered flow $\dot{V}_{\mathrm{p}}$ and the engine absorbed flow $\dot{V}_{\mathrm{eng}}$ yields according to (4.8.16) and (4.8.20) for the stationary behavior

$$p_{\mathrm{eng}} = \frac{V_{\mathrm{p,th}}^2 \eta_{\mathrm{vol}}^2 \rho_{\mathrm{oil}}}{2 c_{\mathrm{eng}}^2} \omega_{\mathrm{p}}^2. \tag{4.8.31}$$

Hence, if the oil pressure in the oil gallery can be measured, a lower volumetric pump efficiency due to pump leaks or a higher overall resistance coefficient c_{eng} of the oil passages through the engine lead to smaller pressure values $p_{\mathrm{eng}}(\omega_{\mathrm{p}}^2)$. However, as c_{eng} depends much on the oil temperature and engine speed, relatively wide tolerances of Δp_{eng} must be allowed in determining the thresholds for faulty behavior.

Instead of (4.8.31) the gain $K_{\mathrm{p}}\,(\omega_{\mathrm{p}}) = \Delta p_{\mathrm{eng}}/\Delta\omega_{\mathrm{p}}$ due to (4.8.22) can be used as feature. Some results for model-based fault detection in marine diesel lubrication system based on temperature and pressure measurements are shown in Nahim et al (2016).

The possibilities for fault detection in the lubrication circuit increase considerably if *electrical driven oil pumps* are used. Then voltage and current measurements allow to determine the driving power $P_{\mathrm{el,in}}(U, I)$. The electrical power delivered to the pump shaft is then lowered by an electrical efficiency $P_{\mathrm{p,el}} = \eta_{\mathrm{el}} P_{\mathrm{el,in}}$. Then it holds with (4.8.11)

$$P_{\mathrm{p,eff}} = \dot{V}_{\mathrm{p,eff}} p_{\mathrm{p,eff}} = \eta_{\mathrm{tot}} \eta_{\mathrm{el}} P_{\mathrm{el,in}}. \tag{4.8.32}$$

If the efficiencies are known, then the effective pump flow $\dot{V}_{\mathrm{p,eff}}$ can be determined. This can for example be used to determine flow resistance coefficients like c_{eng} in (4.8.12). Compare the determination of volume flow and the fault detection with an electrical driven centrifugal coolant pump in Sect. 4.7.2.

4.9 Overall gasoline engine fault diagnosis

To obtain an overall presentation of model-based fault-detection methods for gasoline engines the results of the fault-detection modules for interconnected parts, as

the intake system, the fuel and combustion system and the exhaust system are now joined together. These signal- and model-based fault-detection methods have to be seen as an addition to the classical OBD methods.

The fault diagnosis based on the observed symptoms can be performed with fault-symptom tables or with rule-oriented approaches as described in Sect. 4.1.2 and represented in Fig. 4.1.12. Figure 4.9.1 illustrates an overall model-based diagnosis system as an expansion of Fig. 4.1.3.

The resulting overall fault-symptom table is depicted in Table 4.9.1 for the considered three-engine process parts. It is obtained by combining Tables 4.1.5, 4.4.1 and the results of Sect. 4.6. It shows that most symptoms are related to the respective process parts but some symptoms contribute also from other engine parts. The patterns of the symptoms indicate that they are in most cases different and therefore the faults can be isolated, except in some cases, where the effect on the symptoms is identical as e.g. for "leakage after throttle" and "disconnected blow-by tube". For the detection of some sensor faults in the intake system, Table 4.1.3 can be applied, where three ways of determining the air mass flow are used.

Hence, a recommendation is to develop the fault-detection and -diagnosis methods for the respective engine process parts with their actuators and sensors and then to see if also symptoms from other engine parts enlarge the basics and therefore the coverage for the diagnosis.

References

Affenzeller J, Gläser H (1996) Lagerung und Schmierung von Verbrennungsmotoren. Springer, Wien

Auckenthaler T (2005) Modelling and control of three-way catalytic converters. Doctoral thesis. No. 16018. Swiss Federal Institute of Technology, Zürich

Berg M (ed) (2009) Ölkreislauf im Verbrennungsmotor III. expert Verlag, Renningen

Best R, Boddeke H, Stainhauser D, Wyler K (1997) Time-frequency analysis and digital image processing. part 4. Technisches Messen 64(2):73–82

Chang X, Chiang EC, Johnson JH (1991) The theoretical development of vehicle engine cooling airflow models using incompressible flow methods. In: SAE Technical paper Series, Warrendale, PA, 910644

Chun S (2003) Network analysis of an engine lubrication system. Tribology international 36(8):609–617

Corbel JC (1987) An original simulation method for car engine cooling systems: a modular system. In: SAE Technical paper Series, Warrendale, PA, 870713

Dixon S (1966) Fluid mechanics, thermodynamics of turbomachinery. Pergamon Press, Oxford

Eberan-Eberhorst C (ed) (2010) Schmierung von Verbrennungsmotoren, 3rd edn. expert Verlag, Renningen

Eriksson L, Nielsen L, Nytomt J (1996) Ignition control by ionization current interpretation. In: SAE World Congress. Paper 960045, Detroit, Michigan, USA

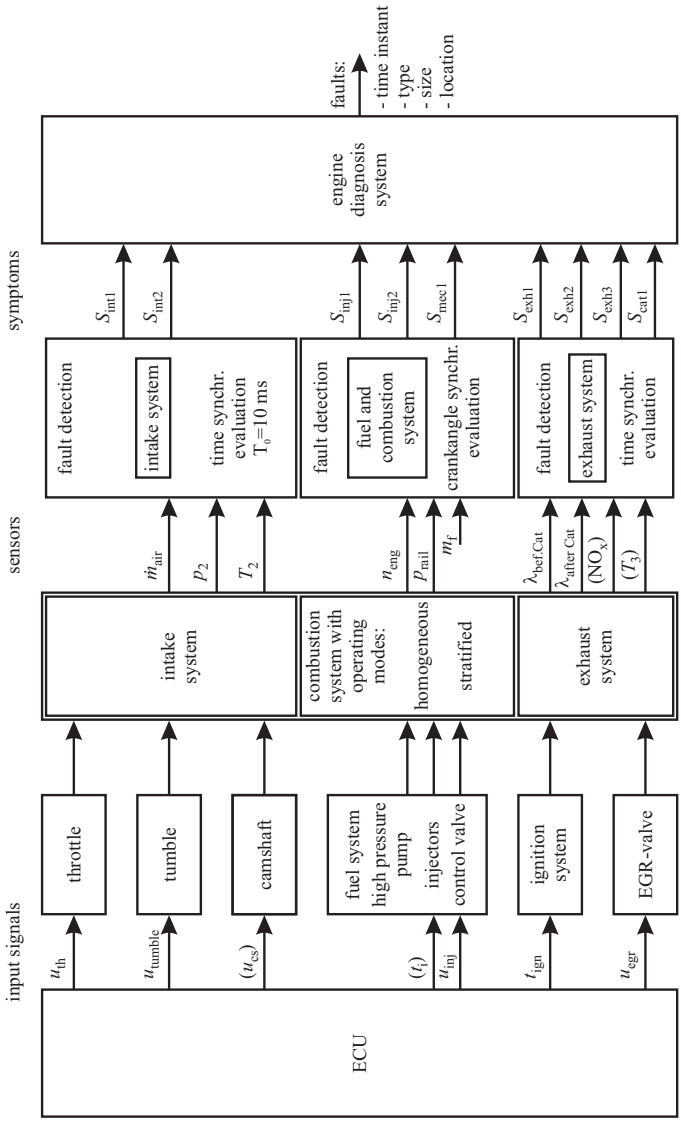


Fig. 4.9.1. Overall model-based diagnosis system for three engine process parts of a DI gasoline engine.

Table 4.9.1. Overall model-based fault-symptom table for the intake, fuel and combustion, and exhaust system of a DI gasoline engine (+: positive symptom, −: negative symptom, 0: no symptom change, d: don't care, ✓: applicable, ×: non-applicable).

Engine process part		A1: intake system		A2/A3: fuel & combustion			A6: exhaust system			
	symptoms	S_{int1}	S_{int2}	S_{inj1}	S_{inj2}	S_{mec1}	S_{exh1}	S_{exh2}	S_{exh3}	S_{cat1}
	residuals	r_{int1}	r_{int2}	r_{inj1}	r_{inj2}	r_{mec1}	r_{exh1}	r_{exh2}	r_{exh3}	Δm_{02}
	faults	air mass flow	intake pressure	rail pressure (injection)	rail pressure (overrun)	speed amplitude	output lambda controller	average lambda	amplitude lambda	oxygen storage
Intake system faults	leakage before throttle	−	0				+			
	leakage after throttle	−	+				+			
	disconnected blow-by tube	−	+				+			
	clogging of air filter	−	−				0			
	clogging within manifold	−	−				0			
	increased EGR	0	+				d			
Injection & ignition faults	less injection in one cylinder	0	0	+	d	−	+			
	less fuel supply			0	−	0				
	ignition energy reduced			0	d	−				
Exhaust system faults	leakage in exhaust pipe						+	+	+	
	clogging in exhaust pipe	−	+							
	aged catalyst									−

Fehrenbach H, Quante F (1987) Diagnose von Verbrennungsmotoren durch Analyse des Winkelgeschwindigkeitsverlaufs der Kurbelwelle. FhG-Berichte 4 87

Feßler D (2011) Modellbasierte On-Board-Diagnoseverfahren für Drei-Wege-Katalysatoren. Dissertation TU Karlsruhe, vol 8. KIT Scientific Publishing, Karlsruhe

Förster J, Lohnmann A, Mezger M, Ries-Müller K (1997) Advanced engine misfire detection for SI-engines. In: SAE Technical paper Series, Warrendale, PA, 970855

Führer J, Sinsel S, Isermann R (1993) Erkennung von Zündaussetzern aus Drehzahlsignalen mit Hilfe eines Frequenzbereichsverfahrens. In: Proceedings 13. Tagung Elektronik im Kraftfahrzeug, Haus der Technik Essen, Germany

Füssel D (2002) Fault diagnosis with tree-structured neuro-fuzzy systems. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, 957. VDI Verlag, Düsseldorf

Guzzella L, Onder C (2004) Introduction to modeling and control of internal combustion engine systems. Springer, Berlin

Haas A, Esch T, Fahl E, Kreuter P, Pischinger F (1991) An original simulation method for car engine cooling systems: a modular system. In: SAE Technical paper Series, Warrendale, PA

Hartmanshenn E, Isermann R (2005) Diagnosemethoden Ottomotor – Modellgestützte präventive Diagnosemethoden (Fehlerfrüherkennung für Ottomotoren. Abschlussbericht. Heft 806-2005. FVV, Frankfurt

Isermann R (2005) Mechatronic systems – fundamentals, 2nd edn. Springer, London

Isermann R (2006) Fault-diagnosis systems – An introduction from fault detection to fault tolerance. Springer, Heidelberg

Isermann R (2011) Fault-diagnosis applications. Springer, Heidelberg

Isermann R (2013) Model-based fault diagnosis of combustion engines. ASME Dynamic Systems and Control (DSC) Magazine (December):6–16

Isermann R (2014) Engine Modeling and Control. Springer, Berlin

Isermann R, Leykauf M (2009) Model-based fault diagnosis of a direct injection gasoline engine with homogeneous and stratified operation. In: 7th IFAC International Symposium on Fault Detection, Supervision and Safety of Technical Processes. SAFEPROCESS 2009, Barcelona, Spain, pp 899–904

Isermann R, Münchhof M (2011) Identification of dynamic systems. Springer, Berlin, Heidelberg

Ivantysyn J, Ivantysynova M (1993) Hydrostatische Pumpen und Motoren, Konstruktion und Berechnung. Vogel, Würzburg

Kammeyer K, Kroschel K (2009) Digitale Signalverarbeitung: Filterung und Spektralanalyse, 7th edn. Vieweg+Teubner, Wiesbaden

Kimmich F (2004) Modellbasierte Fehlererkennung und Diagnose der Einspritzung und Verbrennung von Dieselmotoren. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 12, 549. VDI Verlag, Düsseldorf

Knirsch M, Klee P (2002) On-board diagnosis (OBD) for engine emission control systems. In: Emission control 2002, Dresden, Germany

Krysander M, Frisk E (2009) Leakage detection in a fuel evaporative system. Control Engineering Practice – CEP 17(11):1273–1279

Leykauf M, Isermann R (2006) Diagnosemethoden GDI-Motor – Modellgestützte präventive Diagnosemethoden (Fehlerfrüerkennung für Ottomotoren. Abschlussbericht. Heft 828-2006. FVV, Frankfurt

Leykauf M, Isermann R (2008) Model-based fault diagnosis of a direct injection gasoline engine with homogeneous and stratified operation. In: 8. Internationales Stuttgarter Symposium Automobil- und Motorentechnik, Stuttgart

Louen C, Ding S, Pietsch I, Zwinzscher S (2015) On-Board-Diagnose von Drei-Wege-Katalysatoren mit Hilfe von SVM im Schubbetrieb. In: Liebl J, Beidl C (eds) Internationaler Motorenkongress 2015, Baden-Baden, Springer Fachmedien, Wiesbaden, pp 515–528

Mauer G, Watts R (1989) On-line cylinder diagnostics on combustion engines by noncontact torque and speed measurements. In: SAE Technical paper Series, Warrendale, PA, 890485

Mollenhauer K, Tschöke H (2010) Handbook of diesel engines. Springer, Berlin

Müller N (2003a) Adaptive engine control for gasoline engines using combustion pressure sensors (in German). Fortschr.-Ber. VDI Reihe 12, 545. VDI Verlag, Düsseldorf

Müller N (2003b) Adaptive Motorregelung beim Ottomotor unter Verwendung von Brennraumdruck-Sensoren. Dissertation Technische Universität Darmstadt. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 12, 545. VDI Verlag, Düsseldorf

Müller T, Reichenbach M, Selinger M (2016) Thermal management in vehicle integration. In: 3. Internationaler Motorenkongress 2016, Baden-Baden, Germany, pp 307–323

Nahim H, Younes R, Shraim H, Ouladsine M (2016) Modeling with fault integration of the cooling and the lubricating systems in marine diesel engine: Experimental validation. In: 8th IFAC Symposium Advances in Automotive Control, Norrköping, Sweden, pp 570–575

Neu E, Wade J, Chu A (1977) Simulating the lubrication system of a diesel engine. In: SAE Technical paper Series, Warrendale, PA, 770032

Odendall B (2015) Softwareentwicklung mittels Simulationsumgebung. In: Liebl J, Beidl C (eds) Internationaler Motorenkongress 2015, Baden-Baden, Springer Fachmedien, Wiesbaden, pp 475–497

Odendall B (2016) Catalytic converter properties for diagnosability. MTZ worldwide 77(1):26–33

Peyton Jones J, Muske K (2004) Model-based obd for three-way catalyst systems. In: SAE World Congress. Paper 2004-01-0639

Peyton Jones J, Makki I, Muske K (2006) Catalyst diagnostics using adaptive control system parameters. In: SAE World Congress. Paper 2006-01-1070

Pfleiderer C, Petermann H (2005) Strömungsmaschinen, 7th edn. Springer, Berlin

Plapp G, Klenk M, Moser W (1990) Methods of on-board misfire detection. In: SAE Technical Paper 900232, Warrendale

Reinhardt G (1992) Schmierung von Verbrennungskraftmaschinen. expert Verlag, Renningen

Reulein C (1998) Simulation des instationären Warmlaufverhaltens von Verbrennungsmotoren. Doctoral Thesis Technische Universität München. Hieronymus-Print, München

Ribbens W, Rizzoni G (1990) Onboard diagnosis of engine misfires. In: SAE Technical paper Series, Warrendale, PA, 901768

Robert Bosch GmbH (ed) (2011) Automotive Handbook, 8th edn. Bentley publishers, Cambridge

Salah M, Mitchell T, Wagner J, Dawson D (2010) A smart multiple-loop automotive cooling system-model, control, and experimental study. IEEE/ASME Transactions On Mechatronics 15(1):117–124

Schäfer S (2012) Modellbasierte Steuerung des Kühlkreislaufes einer Brennstoffzelle mit automatisiertem Test der Software. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, Nr. 1219 VDI Verlag, Düsseldorf

Schäfer S, Willimowski P, Maier O, Weispfenning T, Isermann R (2006) Modellbasierte Volumenstrombestimmung zum Betrieb eines Brennstoffzellen-Kühlsystems. In: Fachtagung "Steuerung und Regelung von Kraftfahrzeugen und Verbrennungsmotoren" AUTOREG 2006, Wiesloch, Germany, pp 587–596

Schwarte A, Kimmich F, Isermann R (2002) Model-based fault detection and diagnosis for diesel engines. MTZ worldwide 63(7–8)

Strang G, Nguyen T (1996) Wavelets and filter banks. Wellesley-Cambridge Press, Wellesley, MA, USA

Willimowski M, Füssel D, Isermann R (1999) Misfire detetion for spark-ignition engines by exhaust gas pressure analysis. MTZ worldwide pp 8–12

Willimowski M (2003) Verbrennungsdiagnose von Ottomotoren mittels Abgasdruck und Ionenstrom. Dissertation Technische Universität Darmstadt. Shaker Verlag, Aachen

Willimowski M, Klee P, Küsell M, Knirsch M (2005) Zukünftige Zukünftige Diagnose in Motorsteuerungen. In: 5. Symposium Steuerungssysteme für den Antriebsstrang von Kraftfahrzeugen, IAV GmbH, Berlin, Germany

Wolfram A (2002) Komponentenbasierte Fehlerdiagnose industrieller Anlagen am Beispiel frequenzumrichtergespeister Asynchronmaschinen und Kreiselpumpen. Fortschr.-Ber. VDI Reihe 8, 967. VDI Verlag, Düsseldorf

5

Diagnosis of diesel engines

In order to develop advanced fault-detection and diagnosis systems for diesel engines a modular structure is used, which is based on the available input and output signals from the ECU and comprises the actuators and engine components and their sensors, compare Tables 3.3.1 to 3.3.4 and Fig. 3.1.7. Then, so called *detection modules* are defined which allow the fault detection of the interacting components by using their signal readings. This leads to the detection modules for the intake system, injection, combustion, and exhaust system, as shown in Fig. 5.0.1. The detection modules generate certain symptoms as deviations of features from the normal (fault-free) behavior. These symptoms are then the basis of the fault diagnosis for the overall engine, using fault-symptom relationships for the components of the engine and their interrelations indicating the type and location of the faults.

An improvement of engine diagnosis systems over the implemented OBD functions is obtained by using the inherent properties of measured signals and relations between the signals. Therefore, model-based methods are described in the following sections which use both, *signal models* and *process models*. These models must be able to express the influence between faultless and faulty behavior. Because only relatively few measured signals are available in mass-production engines a grouping into the mentioned detection modules will be made.

Several signals are periodic because of the repetitive cylinder charging and combustion operations. Hence, methods of harmonic signal analysis, like Fourier or Wavelet analysis, can be applied. The process models, relating input and output signals, use as a basis physical models, which have to be simplified, because of real-time computational demands. However, as physically based models of the engine include nonlinear fluidic and thermodynamic processes many parameters have to be estimated experimentally, because they are not known in advance and vary strongly. Then it may be more straightforward to use process models based on identification and parameter-estimation methods. If the structure of these models is based on simplified physical models, this results in semi-physical models, Töpfer et al (2002). Because of the strongly nonlinear behavior of combustion engines, also neural-network approaches may be used. Here, a practical feasible type is the use of local linear models, based on parameter-estimation methods and weighting with radial basis func-

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7_5

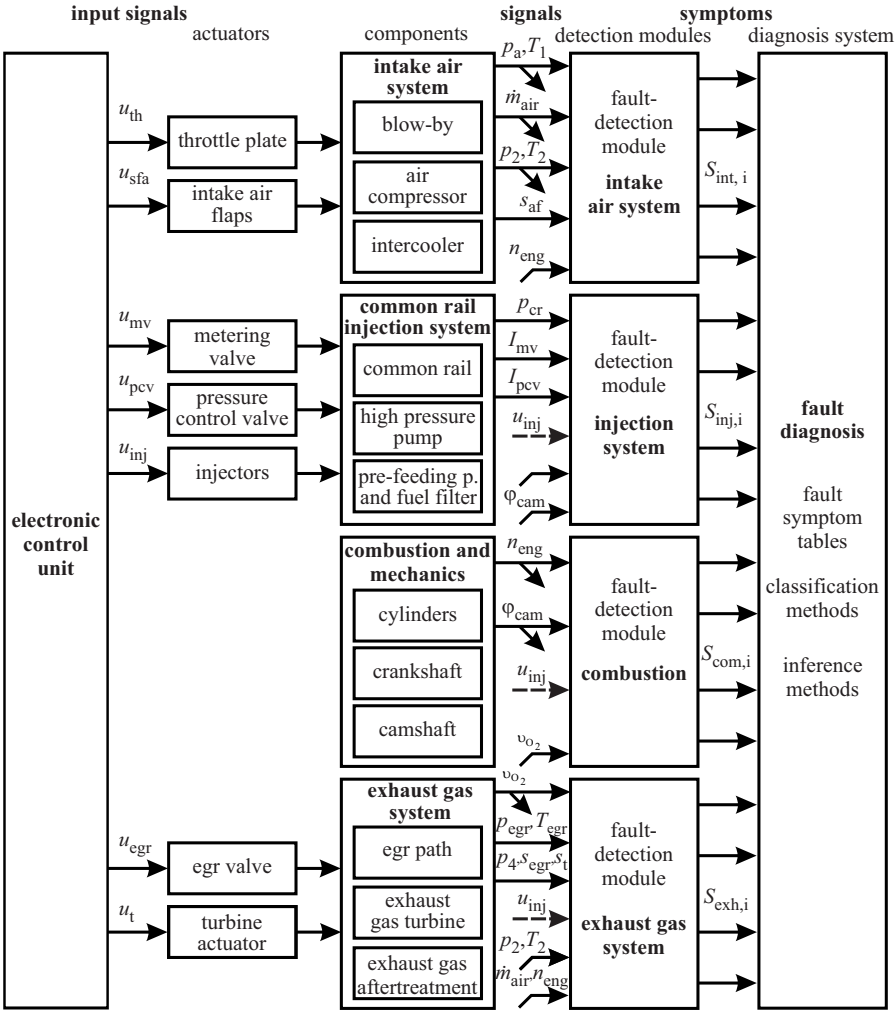


Fig. 5.0.1. Modular structure of a diesel-engine fault-diagnosis system with four detection modules.

tions (e.g. LOLIMOT method). The procedure described in the next sections was mainly developed by Schwarte et al (2004), Kimmich (2004), Kimmich et al (2005), Clever (2011), Eck et al (2011), and Sidorow (2014).

The available sensor signals depend on the type of the diesel engine. A first difference comes from the kind of fuel injection. *Cam-driven distributor injection pumps* with axial or radial pistons for all cylinders or *unit injector pumps* for each cylinder possess solenoid valves for determining pilot injection, main injection and post injection. The electronic control is either integrated in the pump device or part of the engine ECU and uses (internally) in addition to the begin of injection and the fuel

quantity command signal, sensor values like cam rotation angle, pump speed and engine temperature. However, the injection pressure is usually not measured and is therefore not available for fault detection.

Common-rail injection systems separate pressure generation and fuel injection by an accumulator volume in the rail. The pressure is generated by an in-line fuel injection pump or by a radial-piston pump. The rail pressure is controlled by a pressure control valve on the high pressure side with freely varying reference value. It is now the injector which opens and closes based on triggered signals from the engine ECU and such determines beginning of injection and injected fuel quantity, including multiple injections. Solenoid valve injectors open with a hydraulic high pressure force-boost design. Piezo injectors are faster and smaller and allow more and smaller injections. As the high-pressure pump is designed for high delivery quantities, an excess of compressed fuel in idle and part load is returned to the fuel tank via the pressure control valve. An alternative is the low-pressure side control, compressing only as much fuel as injected through a metering unit. Then, the rail has either a pressure relief or control valve with smaller fuel return. In both cases the rail pressure is measured and can be used with advantage for fault detection of the pump and injection system.

Two types of fuel-injection systems, the distributor injection and the common-rail injection system are treated in the Sects. 5.2 and 5.3 with regard to methods of fault detection by using the given measurements. The intake system shows a more common basic construction for diesel engines, Sect. 5.1, as turbochargers with wastegate control and variable geometry, Sect. 5.4. However, the exhaust aftertreatment varies, depending on applied catalysts and particulate filters or selective catalytic reduction (SCR). In Sect. 5.5 some selected examples are described.

Finally, the generated symptoms from different detection modules are combined to perform a fault diagnosis of the overall engine, Sect. 5.6.

5.1 Intake system

The considered fault detection for the intake system is developed to detect leakages, congestions and faults in the EGR valve and swirl flaps. The used intake models describe the air mass flow as an average value and oscillations of the air mass flow and the manifold pressure and are semi-physical models with identification through local linear network models.

The investigated engine is a 2 liter, 4 cylinder, 16 valve diesel engine with a power of 74 kW and a torque of 230 Nm. It is charged by a wastegate turbocharger and possesses exhaust gas recirculation, variable swirl flaps for the inlet gas, see Fig. 5.1.1. As shown in Fig. 5.1.2, the air flows through the air filter passes an air mass flow sensor and then flows through the compressor, intercooler and inlet manifold. The blow-by of the crankshaft casing is led back between air mass flow sensor and compressor wheel. Recirculated exhaust gas is mixed to the air by the exhaust gas recirculation (EGR) valve. To reduce emissions, each cylinder is filled by a swirl port and a filling port. By throttling the filling port of each cylinder with the swirl

flaps actuator (SFA) the swirl can be adjusted. Further components of the intake system are the pneumatic membrane actuator to manipulate the swirl flaps, the EGR valve and the electro-pneumatic converters which convert a pulse-width-modulated (PWM) signal of the electronic engine control unit to the pressure for the pneumatic actuators. By cutting off the intake system from the surrounding subsystems, namely environment, engine block, exhaust system and electronic control unit, input and output variables emerge, see Fig. 5.1.3, only some variables are measured or given by the electronic control unit: engine speed n_{eng}, pulse-width-modulated signals for EGR and SFA control u_{egr}, u_{sfa}, atmosphere pressure and temperature p_{a}, T_{a}, intake manifold pressure and temperature $p_{2\mathrm{i}}$, $T_{2\mathrm{i}}$, and the Hot Film Mass air flow rate $\dot{m}_{\mathrm{air,HFM}}$. Other variables like turbine power P_{turb} , compressor power P_{comp}, turbocharger speed n_{tc}, EGR mass flow $\dot{m}_{\mathrm{egr}}$, air mass flow into the engine $\dot{m}_{\mathrm{air,e}}$, and the positions of the actuators are not measured.

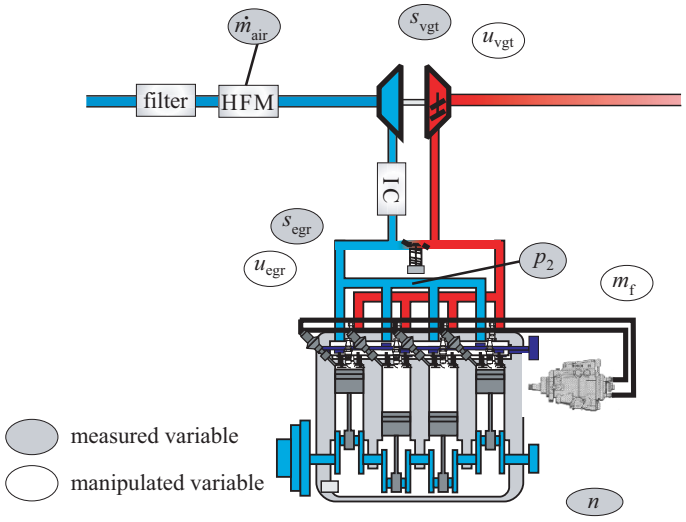


Fig. 5.1.1. Scheme of the investigated diesel engine with distributor pump direct injection, wastegate turbocharger and exhaust gas recirculation. Opel, 2 liter, 16 valves, 74 kW, 230 Nm. Used variables and measurements.

5.1.1 Modeling of the intake system with semi-physical nonlinear models

The engine pumping, describing the air mass flow into the engine, was modeled with a semi-physical and local linear network model. It is a mean value model of one working cycle neglecting the periodic working principle. The physical model part describes the engine pumping corresponding to an ideal positive-displacement pump, which is the so-called theoretical air mass flow into the engine

$$\dot{m}_{\mathrm{air,th}} = \frac{1}{2}\, n_{\mathrm{eng}}\, V_{\mathrm{D}}\, \frac{p_{2,\mathrm{i}}}{R\, T_{2,\mathrm{i}}} \tag{5.1.1}$$

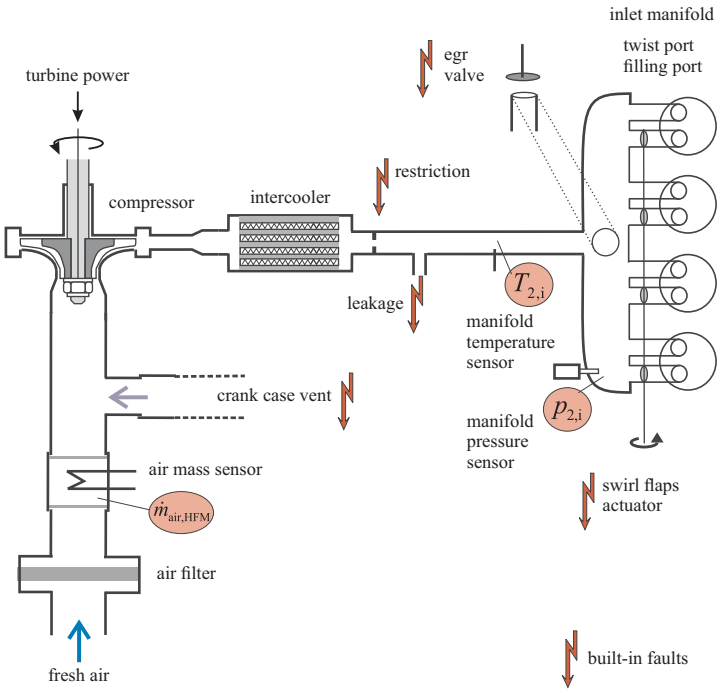


Fig. 5.1.2. Air path of the intake system with sensors and considered faults.

where V_D is the engine displacement volume and R is the specific gas constant.

All other nonlinear induction phenomena are combined and described with the operating-point-dependent *volumetric efficiency coefficient*, Heywood (1988). This coefficient is the ratio of real air mass flow and theoretical air mass flow:

$$\eta_V = \frac{\dot{m}_{air,e}}{\dot{m}_{air,th}} = \frac{\dot{m}_{air,e}}{\frac{1}{2}\, n_{eng}\, V_D\, \frac{p_{2,i}}{R\, T_{2,i}}}. \tag{5.1.2}$$

The nonlinear phenomena like charge heating, backflow, flow friction, ram effect, etc. are very hard to model physically and therefore the real air mass flow is modeled and identified with a neural net model in dependence on engine speed and manifold air density $\rho_{2,i} = p_{2,i}/RT_{2,i}$. The stationary modeled air mass flow then results:

$$\dot{m}_{air,e} \approx f_{\eta_V}\,(n_{eng}, \rho_{2,i})\;\frac{1}{2}\, n_{eng}\, V_D\, \frac{p_{2,i}}{R\, T_{2,i}} = f_{\eta v}\,(n_{eng}, \rho_{2,i}) \cdot \rho_{2,i}. \tag{5.1.3}$$

The *air mass flow oscillation* is caused by the periodic flow into each cylinder. In the time domain the frequencies of the air mass flow oscillation are proportional to the engine speed. But in the angle domain these frequencies are constant. The main (angle-)frequency of the air mass flow oscillation has a constant 180°CS (crank angle) period for four-cylinder four-stroke engines. Measurements have shown that an approximation with a mean value and one harmonic describes the real mass

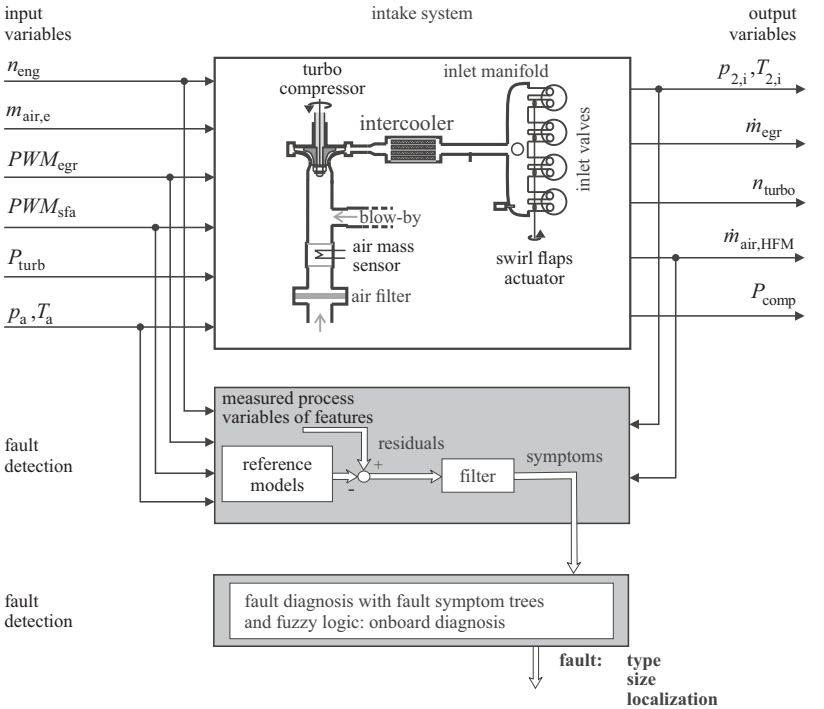


Fig. 5.1.3. Fault-diagnosis structure of the intake system.

flow sufficiently precise. Therefore, a signal model was set up with operating-point-dependent amplitude and phase of the mass flow oscillation:

$$\dot{m}_{\mathrm{air,HFM}}(\alpha) = \bar{\dot{m}}_{\mathrm{air,HFM}} + A_{\dot{m}_{\mathrm{air,HFM}}} \cos\left(2\pi \frac{\alpha}{180^\circ} - \varphi_{\dot{m}_{\mathrm{air,HFM}}}\right), \quad (5.1.4)$$

$$A_{\dot{m}_{\mathrm{air,HFM}}} \approx f_{\mathrm{A_{\dot{m}}}}(n_{\mathrm{eng}}, \rho_{2,\mathrm{i}}), \quad (5.1.5)$$

$$\varphi_{\dot{m}_{\mathrm{air,HFM}}} \approx f_{\varphi_{\dot{\mathrm{m}}}}(n_{\mathrm{eng}}, \rho_{2,\mathrm{i}}). \quad (5.1.6)$$

Figure 5.1.4 shows the air mass flow for one operation point with a first harmonic approximation. Like the air mass flow oscillation the *charging pressure oscillation* is also stimulated by the periodic flow into each cylinder and its signal model is similar to the one before:

$$p_{2,\mathrm{i}}(\alpha) = \bar{p}_{2,\mathrm{i}} + A_{p_{2,\mathrm{i}}} \cos\left(2\pi \frac{\alpha}{180^\circ} - \varphi_{\mathrm{p_{2,i}}}\right), \quad (5.1.7)$$

$$A_{p_{2,\mathrm{i}}} \approx f_{A_{\mathrm{p}}}(n_{\mathrm{eng}}, \rho_{2,\mathrm{i}}), \quad (5.1.8)$$

$$\varphi_{p_{2,\mathrm{i}}} \approx f_{\varphi_{\mathrm{p}}}(n_{\mathrm{eng}}, \rho_{2,\mathrm{i}}). \quad (5.1.9)$$

The estimation of the amplitude and phase of the massflow and the pressure oscillation can be performed by a Fourier analysis in the time-domain, Clever (2011). Then, the Fourier coefficients

$$a\left(\nu\omega_0\right) = \frac{2}{L_\mathrm{p}} \sum_{k=0}^{L_\mathrm{p}-1} y\left(kT_0\right) \cos\left(\nu\omega_0 kT_0\right), \tag{5.1.10}$$

$$b\left(\nu\omega_0\right) = \frac{2}{L_\mathrm{p}} \sum_{k=0}^{L_\mathrm{p}-1} y\left(kT_0\right) \sin\left(\nu\omega_0 kT_0\right) \tag{5.1.11}$$

are calculated with $\nu = 1$ and $L_\mathrm{P} = T_\mathrm{P}/T_0$ for the first harmonic frequency $f_0 = 2\omega_\mathrm{eng}$ respectively $\omega_0 = 2\pi f_0 = 4\pi n_\mathrm{eng}$. The signal y is either $\dot{m}_\mathrm{air}$ or p_2i and T_0 is the sampling time. T_p needs to be a multiple of the period $T_\mathrm{p} = \kappa/f_0$, $\kappa = 1, 2, \ldots$. The amplitude and phase then follow by

$$A_\mathrm{y}\left(\nu\omega_0\right) = \sqrt{a_\nu^2 + b_\nu^2}, \tag{5.1.12}$$

$$\varphi_\mathrm{y}\left(\nu\omega_0\right) = \arctan\left(b\left(\nu\omega_0\right)/a\left(\nu\omega_0\right)\right). \tag{5.1.13}$$

Local linear networks are employed for modeling parts of the intake system with little required knowledge of the inner physical model structure. The special net model LOLIMOT (local linear model tree), Isermann and Münchhof (2011), uses weighted local linear models, see Sect. 4.1.2. The local linear models are identified with an orthogonal divided input space and define the placement of radial basis functions, see Fig. 5.1.5. The radial basis functions are normalized and define the validity of each model. The validity of a local model is almost 100% in its center and decreases towards its neighbor models, so that the superposition of local model validity at each point is 100%. Because of this there is a smooth transition between the local models and the overall model is steadily differentiable. The approach with local linear models leads to fast training properties, because the local model parameters are computed by a least squares method in one step. LOLIMOT comprises both automatic model structure generation and the identification of model parameters. The model structure is generated in an iterative way, starting with one local model and adapting to the variable complexity, resp. nonlinearity of the identified system. The input space of the worst performing local model is divided in two local models, resulting in a growing number of local models. This procedure is stopped when a given number of local models is reached or the training fault falls below a given limit. As an example Fig. 5.1.5 shows the input space and the local linear model structure of the neural net model which models the volumetric efficiency of the engine pumping with 11 local models, resp. 11 iterations.

For the fault-free description of the intake system five static *reference models* were identified, which describe the volumetric efficiency, the amplitude of air mass flow oscillation, the phase of air mass flow oscillation, the amplitude of charge pressure oscillation and the phase of charge pressure oscillation depending on engine speed and manifold density. The reference models were identified for a closed EGR valve and opened swirl flaps with a quasi-stationary identification cycle, Fig. 5.1.6. The identification cycle stimulates only very low frequencies and evenly distributes data points over the complete input space. The distribution of data points is shown in Fig. 5.1.7. As a high boost pressure cannot be achieved at low engine speed the input space region with low engine speed and medium to high manifold density is empty.

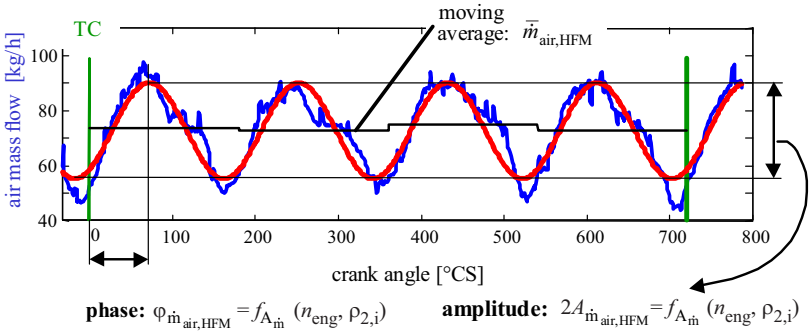


Fig. 5.1.4. Example of the air mass oscillation with a first harmonic approximation at n_{eng} = 1200 rpm.

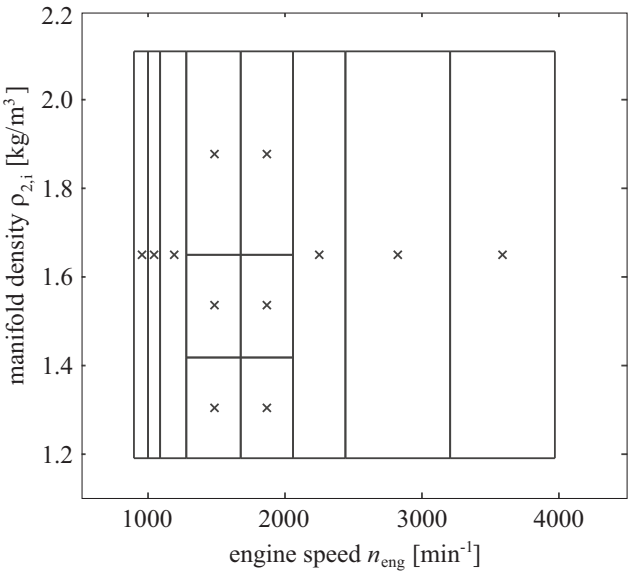


Fig. 5.1.5. Placement of local linear models with their centers 'x' and range of dominant validity for the volumetric efficiency with linear local models LOLIMOT.

Because local models are extrapolated in the empty data region they are corrected manually to a constant value without affecting the validity of the model. Figure 5.1.8 shows the identified reference model for the volumetric efficiency compared with measured data points.

LOLIMOT identifies the model with a continuous mathematical function. However, for online application look-up tables, calculated from the LOLIMOT model, are employed for the reference models in order to reduce the computational effort and to be compatible to the general representation of nonlinear models in the engine ECU. This approach of the fast neural net model LOLIMOT to obtain look-up tables

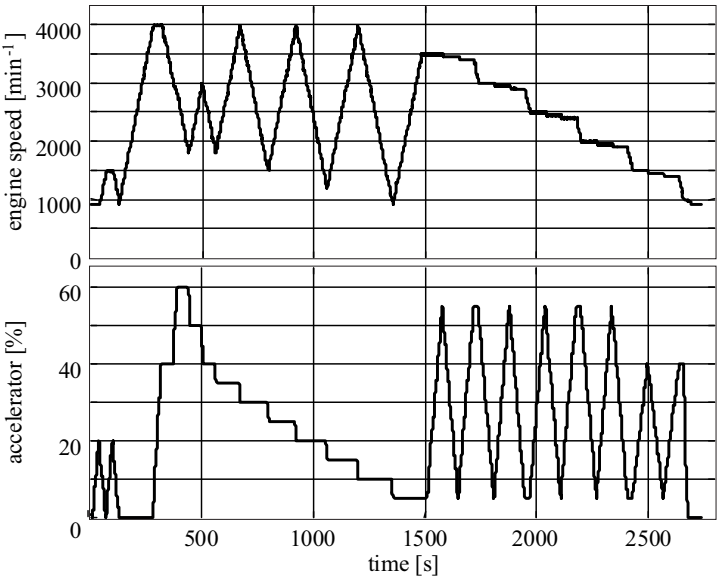


Fig. 5.1.6. Identification cycle for obtaining the intake models.

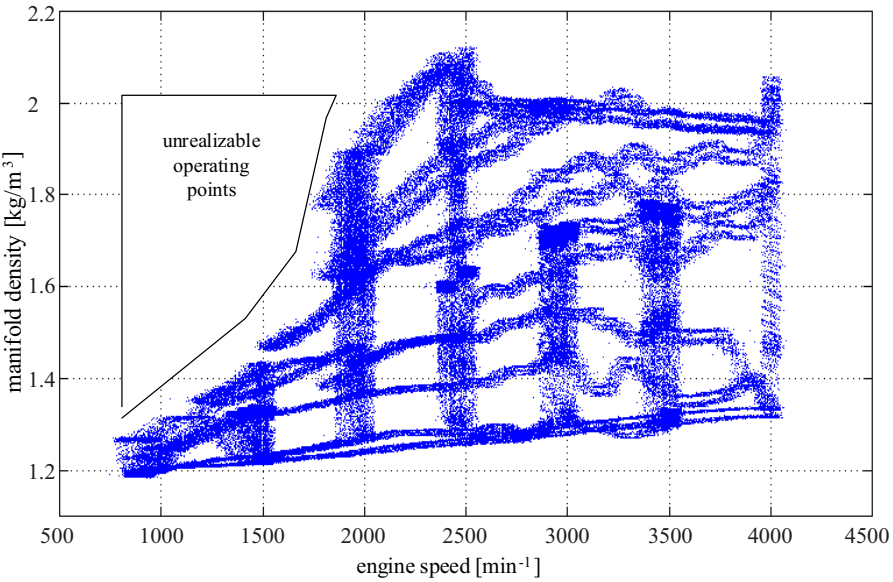


Fig. 5.1.7. Input space for the identification with the distribution of measured data points.

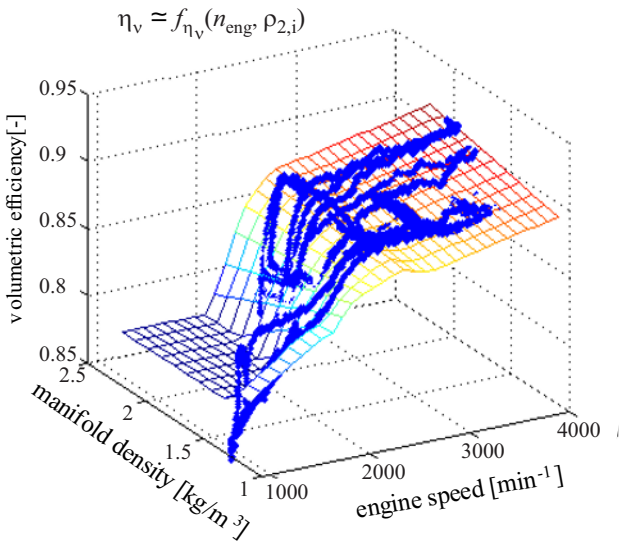


Fig. 5.1.8. Identified stationary reference model of the volumetric efficiency with LOLIMOT compared to measured data points.

provides a great advantage compared to the general used grid measurement regarding flexibility, complete measurements and the effort to measure exactly certain operating points. With an identified neural net model any desired resolution of the look-up table in the electronic engine control can be extracted afterwards. Figure 5.1.9 shows the identified reference model for the amplitude of the air mass flow oscillation. At engine speed of about 2000 rpm the largest amplitude results indicating the resonance frequency of the air path pipe. Figure 5.1.10 then shows the corresponding phase of the air mass flow oscillation. The phase is mainly a linear function of the engine speed. The identification of the charging pressure oscillation with models for amplitude and phase shows similar results as the air mass flow oscillation and is not shown here, see Schwarte et al (2002), Kimmich et al (2005). An application to another common-rail diesel engine, see Fig. 5.4.1, has shown similar results, Clever (2011).

5.1.2 Fault detection with nonlinear parity equations and diagnosis

The five identified reference models calculating special features of the intake system are used to set up five independent *parity equations* yielding the residuals:

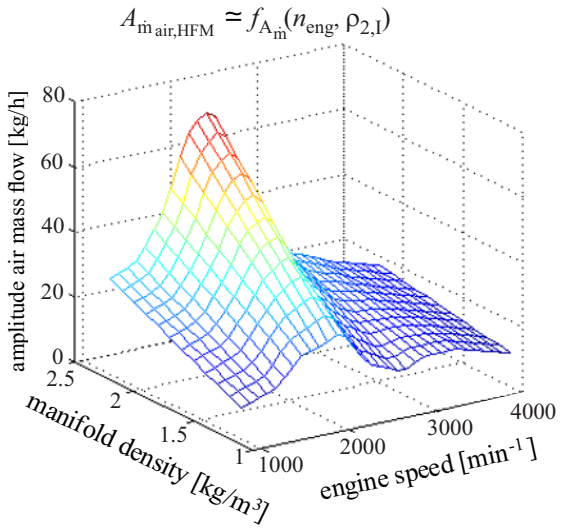


Fig. 5.1.9. Identified reference model for the amplitude of air mass flow oscillation.

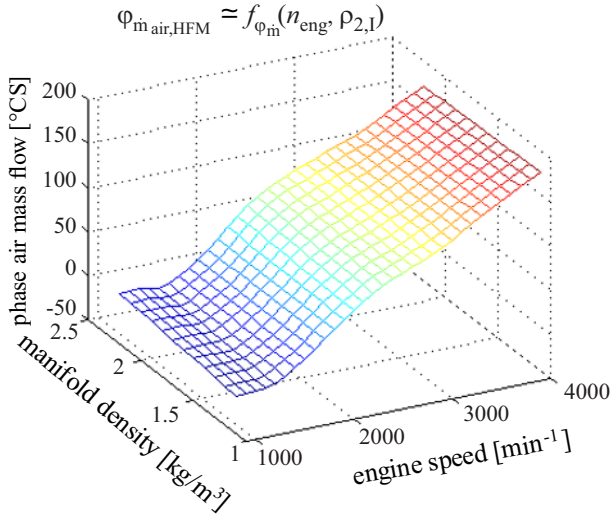


Fig. 5.1.10. Identified reference model for the phase of air mass flow oscillation.

$$r_{\eta_\nu} = f_{\eta_\nu}\left(n_{\mathrm{eng}}, \rho_{2\mathrm{i}}\right) - \eta_\nu, \tag{5.1.14}$$

$$r_{\mathrm{A}_{\dot{\mathrm{m}}}} = A_{\dot{\mathrm{m}}_{\mathrm{air,HFM}}} - f_{\mathrm{A}_{\dot{\mathrm{m}}}}\left(n_{\mathrm{eng}}, \rho_{2\mathrm{i}}\right), \tag{5.1.15}$$

$$r_{\varphi_{\dot{\mathrm{m}}}} = \varphi_{\dot{\mathrm{m}}_{\mathrm{air,HFM}}} - f_{\varphi_{\dot{\mathrm{m}}}}\left(n_{\mathrm{eng}}, \rho_{2\mathrm{i}}\right), \tag{5.1.16}$$

$$r_{\mathrm{A_p}} = A_{\mathrm{p_{2i}}} - f_{\mathrm{A_p}}\left(n_{\mathrm{eng}}, \rho_{2\mathrm{i}}\right), \tag{5.1.17}$$

$$r_{\varphi_{\mathrm{p}}} = \varphi_{\mathrm{p_{2i}}} - f_{\varphi_{\mathrm{p}}}\left(n_{\mathrm{eng}}, \rho_{2\mathrm{i}}\right). \tag{5.1.18}$$

The signal processing and the model-based fault-detection algorithms are implemented on a Rapid Control Prototyping System with MATLAB/SIMULINK. In Fig. 5.1.11 the realized model for online fault detection is shown with its explanation of the corresponding processing steps. The measured signals are crank-angle synchronously preprocessed for a better signal to noise ratio. Then, firing cycle synchronous algorithms calculate the characteristic features. These features are compared to the output of the reference models yielding the residuals.

The residuals are further processed leading to symptoms. Finally, the symptoms indicate faults by deviating from zero and exceeding a threshold. The results of real-time fault detection are presented in Fig. 5.1.12 for an exemplary operating point. Several faults were temporarily built in at the intake system. The residual thresholds are marked by dashed lines. In the fault-free case the residuals are almost zero. The reference models for the volumetric efficiency, amplitude air mass flow oscillation and amplitude charging pressure oscillation show the expected behavior. The reference models for phase air flow and phase boost pressure have wider stochastic deviations in the fault-free case of approximately $4°\mathrm{CS}$. Therefore, the first three symptoms are most suitable for fault detection. The first fault example in Fig. 5.1.12 shows the case of undesired closed swirl filling ports. The residuals amplitude charging pressure oscillation and amplitude air mass flow oscillation exceed the thresholds and with the corresponding symptoms this fault is clearly detected (without position sensor at the swirl actuator). The second fault example is an undesired opened EGR valve. The residual volumetric efficiency responses intensively and it is obvious that smaller EGR faults are detected as well. Additionally, the residuals amplitude charging pressure oscillation and amplitude air flow oscillation show a strong deflection, too. The third example shows different sizes of leakages in the air path between intercooler and engine. The bigger the leakage diameter, the stronger is the deflection of the residual volumetric efficiency. The leakages with 4 mm, 5 mm and 7 mm in diameter are well detected.

As a further fault example the crank case vent pipe was removed. This is equivalent to a leakage between air mass flow sensor and compressor. Additional air which is not measured is sucked into the air path. With the symptom volumetric efficiency this fault is clearly detected as well. Note that the symptom volumetric efficiency is different, dependent on the leakage location before or after the compressor, because air is lost or additionally sucked in. The last fault example presented depicts a short-time restriction (higher air resistance) between intercooler and engine. This fault is detected by the symptoms amplitude boost pressure oscillation and amplitude air mass flow oscillation. The demonstrated faults were detected very fast in just a few 100 ms. Table 5.1.1 summarizes the fault symptom relations. It is differentiated

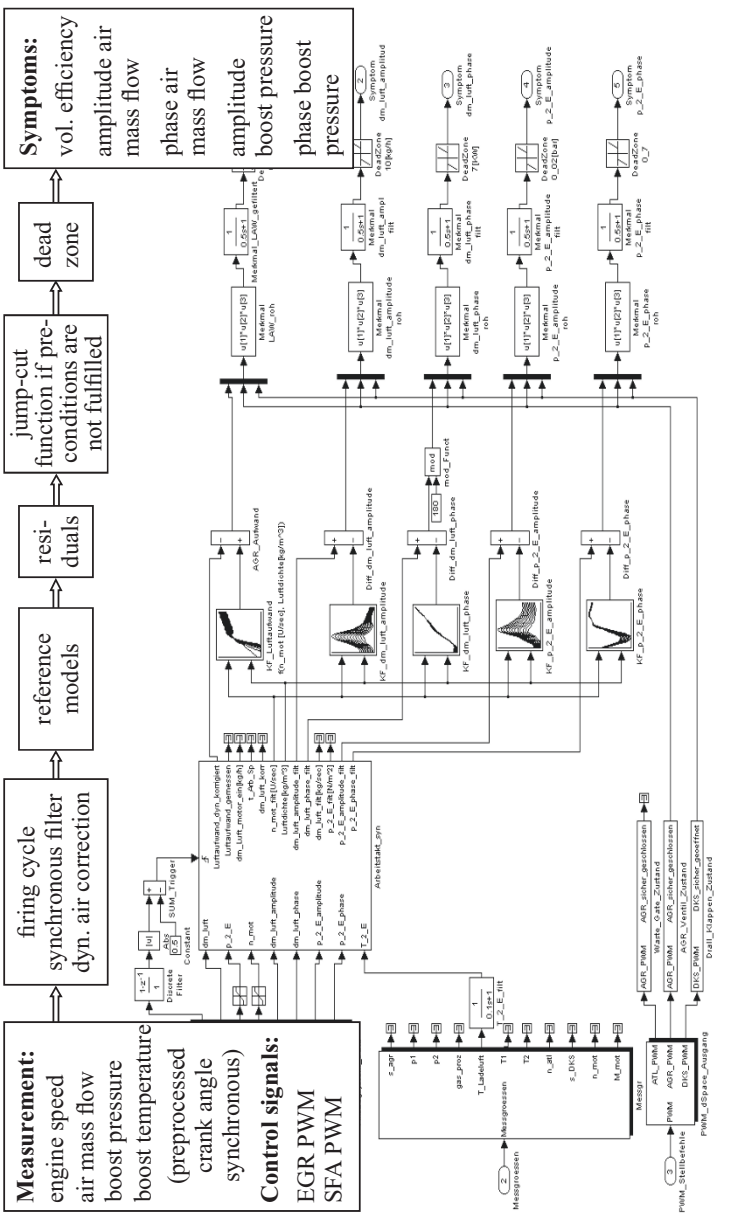


Fig. 5.1.11. Implementation of real-time model-based fault detection of the intake system on a dSPACE Rapid Control Prototyping hardware with MATLAB/SIMULINK.

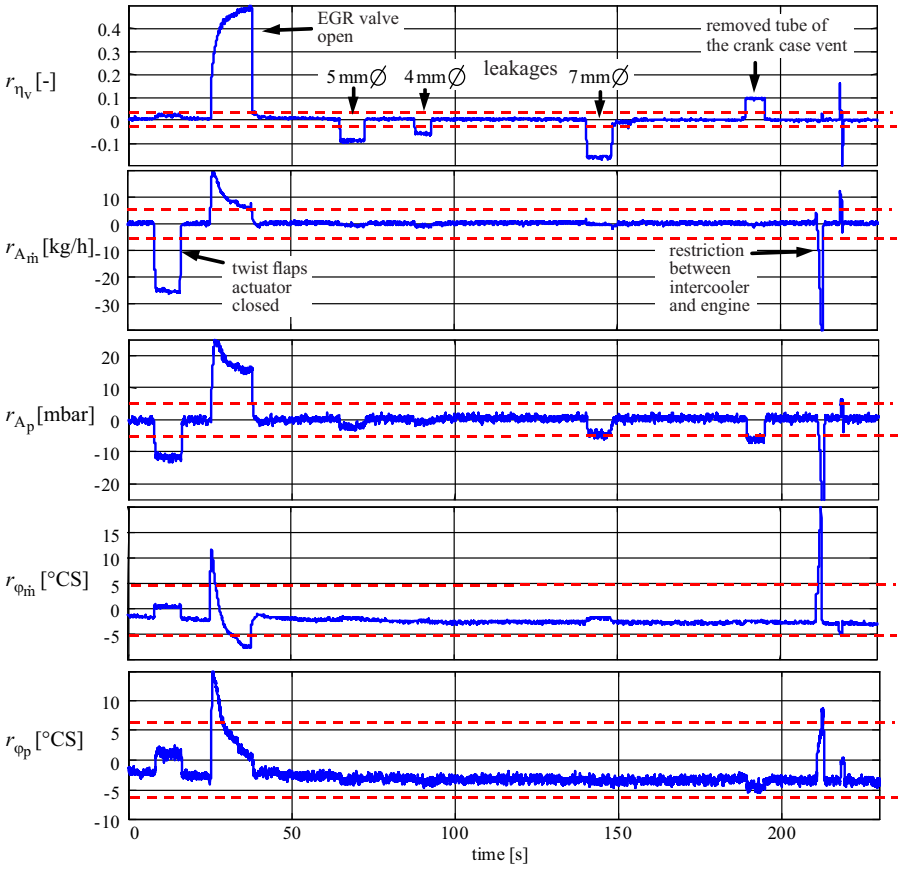


Fig. 5.1.12. Residual deflections in dependence on faults (online), 2000 rpm, 130 Nm, p_{2i}=1.5 bar, air flow 165 kg/h.

between intense positive, positive, negative, and no response. Taking the signs of the deviations into account, different symptom patterns result for each fault, so that all faults except two different leaks are isolated and therefore can be diagnosed.

The results show that the application of net models for model-based fault detection with parity equations offers advantages. The identification of nonlinear models is faster, more flexible and covers more data points than the usual identification with grid measurements. The advantage will even be larger if one is forced to identify nonlinear models with higher dimensions. The special network LOLIMOT can also easily be extended to dynamic nonlinear models. In final applications look-up tables can be generated. The computational effort for real-time fault detection is relatively low and can be realized in future engine control units. An expansion of this fault-diagnosis method to a diesel engine with low pressure EGR is performed by Eck et al (2011). An application of principal component analysis (PCA) to the partial blockage of the airflow to the intercooler is shown in Antory et al (2005).

Table 5.1.1. Fault-symptom table of the intake system.

Faults	Symptoms				
	$S_{\eta \mathrm{v}}$	$S_{\mathrm{A}\dot{\mathrm{m}}}$	S_{Ap}	$S_{\varphi \dot{\mathrm{m}}}$	$S_{\varphi \mathrm{p}}$
Removed tube of the crank case vent	$+$	0	0	0	0
Leakage between intercooler and engine	$-$	0	0	0	0
Restriction between intercooler and engine	0	$-$	$-$	$+$	$+$
Swirl flaps actuator, filling port is closed	0	$-$	$-$	0	0
EGR valve: stuck at open	$++$	$+$	$+$	0	0
Leaky EGR valve	$+$	0	0	0	0

legend:

$++$ symptom responds intense positive
$+$ symptom responds positive
$-$ symptom responds negative
0 symptom does not respond

symptoms:

$S_{\eta \mathrm{v}}$ volumetric efficiency
$S_{A\dot{m}}$ amplitude air mass flow oscillation
S_{Ap} amplitude boost pressure oscillation
$S_{\varphi \dot{\mathrm{m}}}$ air mass flow oscillation
$S_{\varphi p}$ phase boost pressure oscillation

5.2 Direct injection system with distributor pump and combustion

Figure 5.2.1 shows the engine components for the fault-detection-module injection and combustion, which are integrated here and shortly called combustion module. After passing the intake system the aspirated air mass flows into the cylinders. The exhaust gas recirculation (EGR) mixes exhaust gas to the aspirated air mass in order to keep nitrogen oxide emissions low. A high pressure injection pump (Bosch, VP44) provides the injection mass and the right injection timing. Fuel pressures up to 1500 bar can be reached at the injection nozzle. The engine speed is measured at the open end of the crankshaft by means of an optical incremental sensor with a resolution of one degree crank angle ($^\circ$CS). Behind the turbine of the turbocharger a broadband oxygen-sensor (Bosch, LSU 4.2) is installed which is used for measuring the oxygen concentration in the exhaust gas. This sensor serves as an additional source of information about the combustion.

Figure 5.2.2 depicts the basic structure of the fault detection and diagnosis system with essential input and output variables of the engine process, see also Kimmich et al (2001), Kimmich and Isermann (2002), Schwarte et al (2002). For the fault detection following variables are used: the mean and high resolved (1°CS) measured engine speed, $\bar{n}_{\mathrm{eng}}$ and n_{eng}, the oxygen concentration ν_{O2} in the exhaust gas, the aspirated air mass flow $\dot{m}_{\mathrm{air}}$, the manifold pressure $p_{2\mathrm{i}}$, the engine temperature T_{eng}, resulting from the mean value of oil and coolant temperature, T_{oil} and T_{cool}, and the nominal injection mass $m_{\mathrm{f,nom}}$ for one cylinder from the electronic diesel control (Bosch EDC 15). The measured signals are processed by means of signal and process models to generate features. The comparison of the current process behavior with the nominal one yields the output residuals.

Based on engine speed and oxygen evaluation following residuals of output parity equations can be generated:

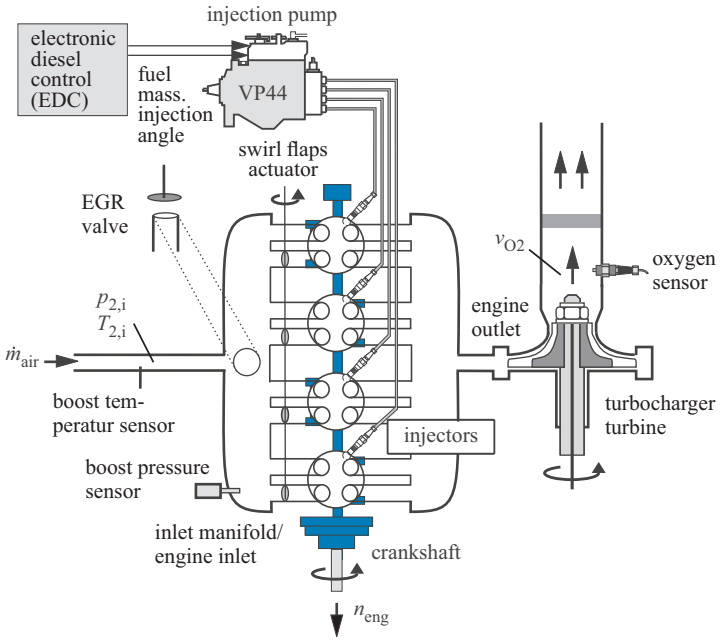


Fig. 5.2.1. Engine components for the fault-detection module: distributor pump injection, combustion and crankshaft drive.

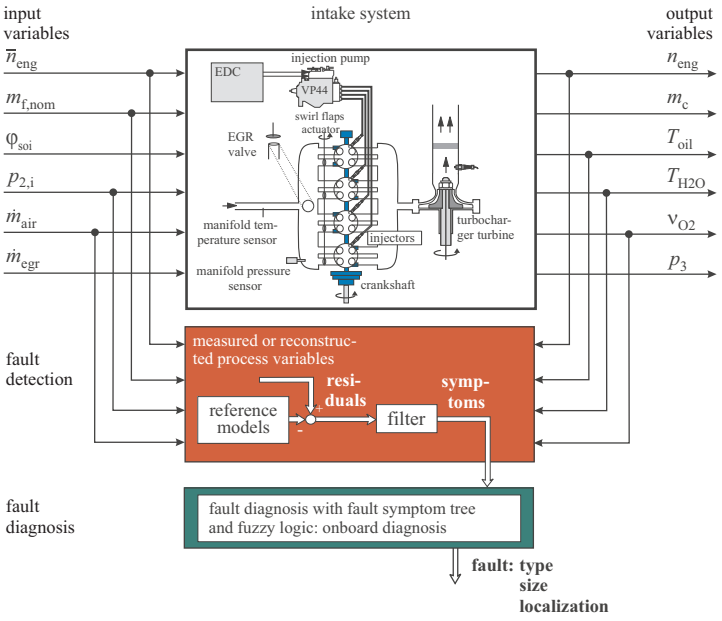


Fig. 5.2.2. Fault-diagnosis structure of the combustion module.

1. Residuals from engine speed
   a) mean effective engine torque r_{MME}
   b) engine roughness r_{ER}
   c) engine roughness cylinder individual $r_{\mathrm{ERC1...4}}$
2. Residuals from oxygen content
   a) injection mass r_{MF}
   b) injection mass fluctuations $r_{\Delta\mathrm{MF}}$
   c) cylinder selective injection mass $r_{\mathrm{MFC1...4}}$.

By filtering the residuals, for instance with a dead zone (threshold function), the corresponding symptoms S_i result, see Fig. 5.2.3. Thereby the size of the threshold value is a compromise between robust fault detection against disturbances and the detection of small faults. In a next step the generated symptoms serve as input variables for the fault diagnosis.

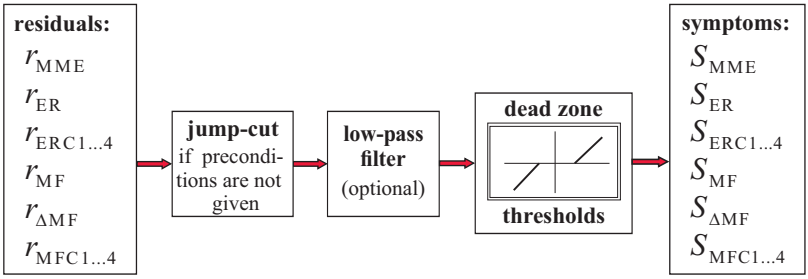


Fig. 5.2.3. Symptom generation from the observed residuals for the combustion module.

5.2.1 Fault detection with combustion features and speed measurement

To generate a residual for the mean effective engine torque r_{MME} the engine torque calculated by a process model is compared with the engine torque reconstructed from the engine speed oscillation, Kimmich and Isermann (2002). The torque model, which represents the nominal (fault free) engine, is based on the calculation of the mean effective cylinder pressure p_{me}, which can be split up into the mean indicated pressure of the low pressure part $p_{\mathrm{mi,lp}}$, the mean effective pressure of the high pressure part $p_{\mathrm{mi,hp}}$ and the mean friction pressure p_{mf}. For each of the components a model is established in the following.

The model of the *low pressure part*, the mean indicated cylinder pressure $p_{\mathrm{mi,lp\_mod}}$ is based on the difference between manifold and exhaust gas pressure using a linear experimentally determined function:

$$p_{\mathrm{mi,lp\_mod}} = 1.47 \cdot (p_{2,\mathrm{i}} - p_{3,\mathrm{mod}}\,(\bar{n}_{\mathrm{eng}}, m_{\mathrm{f,nom}})) - 0.14. \tag{5.2.1}$$

Since the exhaust gas pressure $p_{3,\mathrm{mod}}$ in (5.2.1) is not a measured variable, it has to be calculated by a model using the mean engine speed $\bar{n}_{\mathrm{eng}}$ and the nominal injection

mass $m_{\mathrm{f,nom}}$ as input variables. The model of the exhaust gas pressure is based on the local linear neural net model LOLIMOT, which was identified by measured data.

The model of the mean indicated cylinder pressure $p_{\mathrm{mi,hp\_mod}}$ of the *high pressure part* is also based on a LOLIMOT model using the mean engine speed and the nominal injection mass of the ECU as input variables:

$$p_{\mathrm{mi,hp\_mod}} = f(\bar{n}_{\mathrm{eng}}, m_{\mathrm{f,nom}}). \tag{5.2.2}$$

According to Fischer (2000) the *mean effective friction pressure* $p_{\mathrm{mf,mod}}$ is calculated dependent on the engine temperature T_{eng} and the mean engine speed using an empirical model approach:

$$p_{\mathrm{mf,mod}} = C_0 + C_1 \cdot \left[A_0\left(T_{\mathrm{eng}}\right) + A_1\left(T_{\mathrm{eng}}\right) \cdot \bar{n}_{\mathrm{eng}} + A_2\left(T_{\mathrm{eng}}\right) \cdot \bar{n}_{\mathrm{eng}}^2\right] \tag{5.2.3}$$

with the parameters

$$A_i = f(T_{\mathrm{eng}}, T_{\mathrm{eng}}^2).$$

The exact function f to calculate the parameters A_i can be found in Fischer (2000). The model is adapted with the parameters C_0 and C_1. This is done experimentally by means of measured friction pressures at two reference operating points.

The mean effective cylinder pressure results from the sum of the three pressure components. The desired *mean effective engine torque* can then be calculated for a four-cylinder engine using the displacement volume V_{D}:

$$M_{\mathrm{me,mod}} = \underbrace{\left(p_{\mathrm{mi,hp}} + p_{\mathrm{mi,lp}} - p_{\mathrm{mf}}\right)}_{\mathrm{p_{me,mod}}} \cdot \frac{V_{\mathrm{D}}}{4\pi}. \tag{5.2.4}$$

This value is used as a *reference value* for the faultless engine. On the other side the *real released engine torque* can be determined from the oscillation of the engine speed signal, see Isermann (2014). For an appropriate determination of the speed amplitude and the engine torque respectively, all superimposed disturbances have to be eliminated. These disturbances are on one hand the impact of the mass torques caused by the oscillation of the pistons and the connecting rods and on the other side the torsional oscillations of the crankshaft. The mass torque components in the engine speed signal can be modeled physically and therefore be eliminated directly. The high frequency signal components of the torsional oscillations are suppressed by low pass filtering. After disturbance rejection the *engine speed amplitude* is determined from the difference of speed minimum and maximum for each cylinder interval (180°CS). Thereby, the *delivered engine torque* of one cylinder is proportional to the engine speed amplitude and to the mean engine speed, see Führer et al (1993),

$$M_{\Delta\mathrm{n}} = \frac{2}{\Delta\varphi} J_{\mathrm{eng}} \bar{\omega}_{\mathrm{eng}}\, \Delta\omega_{\mathrm{eng}} = \frac{8\pi^2}{3600} \frac{J_{\mathrm{eng}}}{\Delta\varphi}\, \bar{n}_{\mathrm{eng}}\, \Delta n_{\mathrm{eng}} \tag{5.2.5}$$

which is called the "speed oscillation torque", where $\Delta\varphi$ denotes the angle difference between engine speed minimum and maximum, J_{eng} the engine inertia and Δn_{eng} the engine speed amplitude. The mean effective engine torque M_{me} can then be

determined using an experimentally generated look-up table, see Fig. 5.2.4. Input variables of the look-up table are the mean engine speed $\bar{n}_{\mathrm{eng}}$ and the averaged speed oscillation torque

$$M_{\Delta \mathrm{n},720} = \frac{1}{4} \sum_{i=1}^{4} M_{\Delta n,i} \tag{5.2.6}$$

of the speed oscillation over one working cycle (720 °CS), since $M_{\Delta n}$ is only a measure for the torque output of one cylinder.

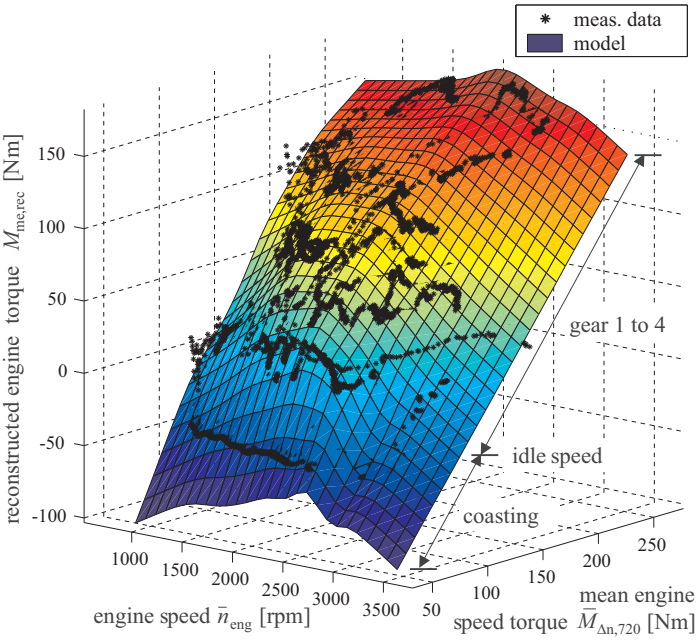


Fig. 5.2.4. Mean effective engine torque $M_{\mathrm{me,rec}}$ in dependence on mean engine speed $\bar{n}_{\mathrm{eng}}$, mean speed oscillation torque $\bar{M}_{\Delta \mathrm{n},720}$ from engine speed amplitude (the look-up table was generated from measurements in the vehicle for coasting, idle speed and for stationary driving conditions with different gear ratios).

With the torque of the reference model (5.2.4) and the reconstructed torque (5.2.6) of the engine speed oscillation the *residual mean effective engine torque* r_{MME} results as

$$r_{\mathrm{MME}} = M_{\mathrm{me,rec}} - M_{\mathrm{me,mod}}, \tag{5.2.7}$$

where $M_{\mathrm{me,rec}} = M_{\Delta \mathrm{n},720}$ denotes the *reconstructed real engine torque* and $M_{\mathrm{me,mod}}$ the reference torque. After executing a threshold according to Fig. 5.2.3 the corresponding symptom S_{MME} results.

For supervision of *engine roughness* the mean acceleration of the crankshaft between successive combustions can be utilized according to Schmidt et al (2000). The

method is based on the fact that the mean engine acceleration over 180°CS (for four-cylinder engines this is the distance between two successive combustions) indicates if the torque output of one cylinder is larger or less than the average torque output of all cylinders. In crank angle domain the mean engine speed acceleration can be calculated by means of a Crank Angle Synchronous Moving Average (CASMA) filter:

$$\frac{1}{\pi}\int_{\alpha}^{\alpha+\pi}\dot{\omega}_{\mathrm{eng}}(\varphi)\,d\varphi \cong \frac{1}{N_{180}}\sum_{i=0}^{N_{180}-1}\dot{\omega}_{\mathrm{eng}}((k-i)\varphi_0) \tag{5.2.8}$$

where π denotes the considered interval of 180°CS in radian and $k\varphi_0$ is the considered crank angle. The number of sampling points N_{180} results from the sample angle φ_0 and the window interval φ_{w} (180°CS):

$$N_{180} = \frac{\varphi_{\mathrm{w}}}{\varphi_{\mathrm{s}}}. \tag{5.2.9}$$

Because in production vehicles only the engine angular speed ω_{eng} is available, the angular acceleration in equation (5.2.8) is determined from the first derivative of the engine speed signal:

$$\hat{\dot{\omega}}_{\mathrm{eng},180}(k\varphi_0) \cong \frac{1}{N_{180}}\sum_{i=0}^{N_{180}-1}\frac{\omega_{\mathrm{eng}}\left((k-i)\varphi_0\right)-\omega_{\mathrm{eng}}\left((k-i-L)\varphi_0\right)}{L\varphi_0}\,\omega_{\mathrm{eng}}\left((k-i)\varphi_0\right) \tag{5.2.10}$$

where L denotes the number of sample points over which the derivative is calculated ($L\varphi_0 \approx 90°\mathrm{CS}$), see Kimmich (2004).

For stationary operating points the mean angular acceleration is zero if no combustion differences occur. Faults like misfires or injection mass deviations lead to a deflection of the angular acceleration accordingly. By elimination of the mean engine acceleration over one working cycle the method can also be used during dynamic driving operations.

For residual generation, first the rms (root of mean squared) value of the mean acceleration is calculated to supervise the *overall engine roughness* for all four cylinders:

$$r_{\mathrm{ER}} = \sqrt{\frac{1}{N_{720}}\sum_{i=1}^{N_{720}-1}\hat{\dot{\omega}}^2_{\mathrm{eng},180}\left((k-i)\varphi_0\right)}. \tag{5.2.11}$$

According to (5.2.8) N_{720} indicates the number of measured data points over one working cycle. The further evaluation of the residual leads to the overall engine roughness symptom S_{ER}.

The torque output of each cylinder compared to the torque output of the whole working cycle over 720°CS can be determined from the cylinder individual evaluation. For this, the mean angular acceleration is evaluated around the TDC (top dead center) of each cylinder i in the interval

$$\varphi_{\mathrm{TDC}(i)} + 70\,\varphi_0 \leq \varphi \leq \varphi_{\mathrm{TDC}(i)} + 110\,\varphi_0 \quad (\varphi_0 = 1^\circ\mathrm{CS})$$

according to (5.2.10). Thereby, the TDC of each cylinder has to be corrected first by the time delay of the filter. Four residuals $r_{\mathrm{ERC1...4}}$ can be derived by averaging the mean angular acceleration in the interval defined above. The symptoms $S_{\mathrm{ERC1...4}}$ for cylinder individual roughness can then be generated according to Fig. 5.2.3. Figure 5.2.5 visualizes the evaluation process. According to the inserted fault it can be seen, that the torque output of cylinder one is less than the average torque output of the engine. The delivered torques of the cylinders 2, 3 and 4 are each larger than the average torque output.

5.2.2 Fault detection with combustion features and excess air measurement

Oxygen evaluation in the exhaust gas allows also to detect faults in the injection and combustion system.

A first residual injection mass r_{MF} results from the difference of the nominal injection mass $m_{\mathrm{f,norm}}$ given by the ECU and the reconstructed injection mass $m_{\mathrm{f,rec}}$, calculated from the measured oxygen concentration ν_{O2} in the exhaust gas and the aspirated air mass m_{air} over one working cycle. Disturbances in the measured oxygen concentration and air mass are suppressed before by low pass filtering. For the calculation of the reconstructed injection mass a physical/chemical approach according to Pischinger and Heywood (1989) is used, assuming a normal combustion

$$m_{\mathrm{f,rec}} = \frac{m_{\mathrm{air}}\,(1 - 4.8\,\nu_{\mathrm{O2}})}{14.5 + 4.6\,\nu_{\mathrm{O2}}}. \tag{5.2.12}$$

The model accuracy obtained with this equation is shown in Fig. 5.2.6 for the fault-free case. The absolute error $e = m_{\mathrm{f,rec}} - m_{\mathrm{f,nom}}$ between the reconstruction and reference setting for engine speeds of 1000 to 4000 rpm and injection masses of 10 to 40 mg/stroke is illustrated. As it can be seen, the error varies from -3.5 to 2 mg/stroke over the operating range. To minimize the deviations, a correction look-up table was generated based on the measured data to correct the nominal fuel mass. Herewith an approximation of the preset injection mass can be achieved with an accuracy of ± 1.5 mg/stroke over the whole operating range. With the corrected fuel mass $m_{\mathrm{f,nom\_corr}}$, the *residual injection mass for all cylinders* r_{MF} results:

$$r_{\mathrm{MF}} = m_{\mathrm{f,rec}} - m_{\mathrm{f,nom\_corr}}. \tag{5.2.13}$$

The residual injection mass deviation $r_{\Delta\mathrm{MF}}$ is generated by a signal-based method. If there are *injection mass deviations between the cylinders*, the measured oxygen concentration starts to oscillate with an oscillation period of one working cycle. Thereby the standard deviation of the oxygen signal can be used as residual

$$r_{\Delta\mathrm{MF}} = \sigma_{\nu_{\mathrm{O2}}} = \sqrt{\frac{1}{N_{720} - 1} \sum_{i=0}^{N_{720}-1} \left(\nu_{\mathrm{O2}}\left((k-i)\varphi_0\right) - \bar{\nu}_{\mathrm{O2}}(k\varphi_0)\right)^2}. \tag{5.2.14}$$

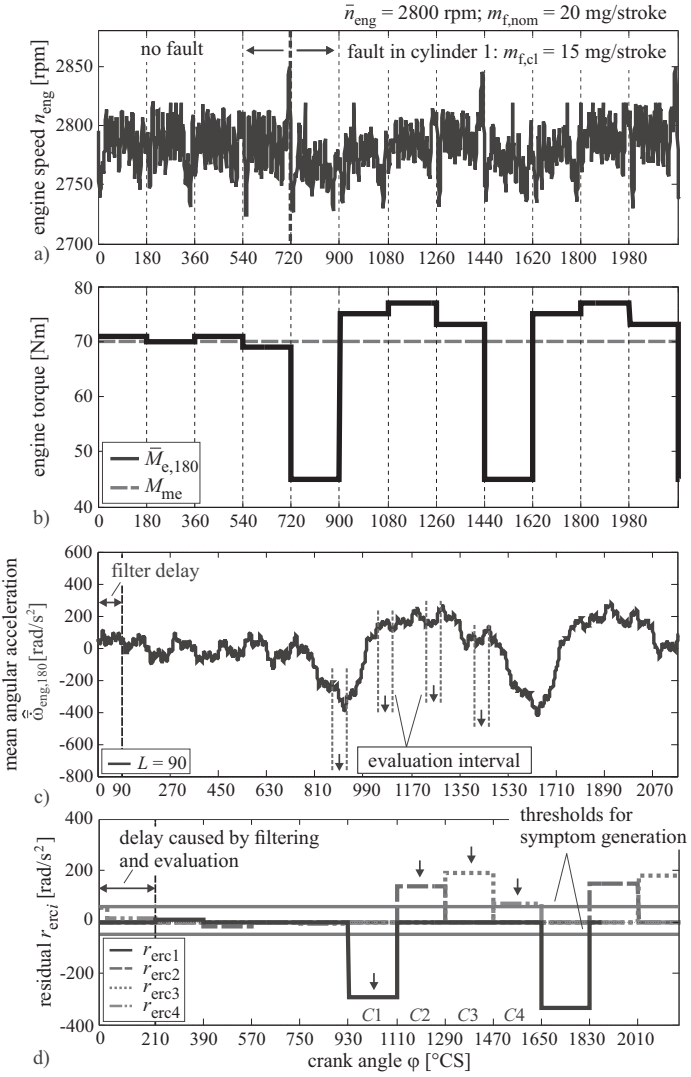


Fig. 5.2.5. Cylinder individual residual and symptom generation for the supervision of combustion: **a** Measured engine speed n_{eng} (resolution 1°CS). **b** Measured mean effective engine torque (at test rig) over 180°CS ($\bar{M}_{\mathrm{eng},180}$) and 720°CS ($M_{\mathrm{me}}$). **c** Calculated mean angular acceleration $\hat{\bar{\dot{\omega}}}_{\mathrm{eng},180}$ according to equation (5.2.10) **d** Cylinder individual residuals for engine roughness $r_{\mathrm{ERC1\ldots4}}$.

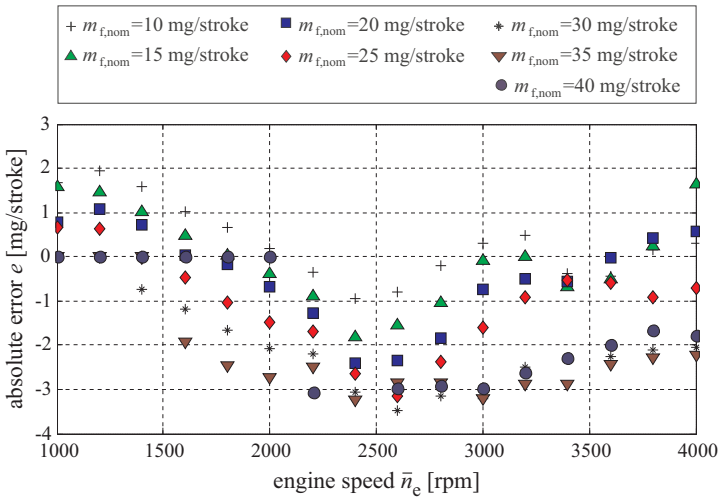


Fig. 5.2.6. Absolute error between reconstructed and nominal injection mass.

In the fault-free case the standard deviation and its residual, respectively, are zero. In the fault case the residual deflects and if a predefined threshold is exceeded, the corresponding symptom *injection mass fluctuation* $S_{\Delta \mathrm{MF}}$ results indicating differences in the injection masses between the cylinders.

For cylinder individual injection mass calculation an inverse model of the oxygen sensor dynamics is used, Kimmich and Isermann (2002). Experiments have shown that the sensor dynamics can be described by a first-order system. Therefore, time constant and gain have to be identified. Since the oxygen concentration is sampled crank angle synchronously, the inverse sensor model becomes:

$$\frac{\nu_{\mathrm{O2,rec}}(\zeta)}{\nu_{\mathrm{O2,m}}(\zeta)} = G^{-1}(\varsigma) = \frac{1 + p_1\, \zeta^{-1}}{p_2} \tag{5.2.15}$$

where $\nu_{\mathrm{O2,m}}$ denotes the measured and $\nu_{\mathrm{O2,rec}}$ the reconstructed and dynamic corrected oxygen concentration. The ζ-transfer function for crank-angle synchronous systems corresponds to the z-transfer function for discrete time systems, Schmidt et al (2000), Isermann (2014). The parameters p_1 and p_2 of the crank-angle synchronous process are connected with the time constant and gain in time domain via the engine angular speed $\bar{\omega}_{\mathrm{eng}}$ and the sample angle φ_0:

$$T = \frac{\varphi_0}{(p_1 + 1)\, \bar{\omega}_{\mathrm{eng}}}\,, \quad K = \frac{T\, \bar{\omega}_{\mathrm{eng}}\, p_2}{\varphi_0}. \tag{5.2.16}$$

The model parameters were identified by means of a least squares parameter estimation for different data sets. Figure 5.2.7 shows the results of the parameter estimation, where the values are plotted over the aspirated air mass. As it can be seen, the time constant and the gain varies depending on the aspirated air mass, so that both variables have to be adapted corresponding to the engine operating point.

According to Fig. 5.2.7 the time constants of the oxygen sensor are in a range of 60 to 100 ms, which is much larger than the time constants of the combustion. Therefore, the sensor is too slow for cylinder-individual evaluation of oxygen concentration directly and a dynamic correction of the lagged sensor signal is necessary. Since the dynamic behavior of the sensor is known, the inverse sensor model according to (5.2.15) can be used to compensate the sensor dynamics approximately. (5.2.12) can then be used to calculate the injected fuel mass $m_{\mathrm{f,rec}}$. Figure 5.2.8 shows the different calculation steps.

For cylinder assignment the runtime $T_{\mathrm{egr,C}i}$ of the exhaust gas from the cylinder outlet to the measuring point has to be determined. This can be done using thermodynamic models and the geometric data of the exhaust system, which is shown in Fig. 5.2.9. As is can be seen, the calculated exhaust gas runtimes vary between 10 and 70 ms in dependence on the mean engine speed and the exhaust gas mass flow (sum of aspirated air mass and injected fuel mass).

With the inverse model of the sensor dynamics and the runtime of the exhaust gas it is possible to evaluate the oxygen signal cylinder individually and to assign the reconstructed fuel mass to the corresponding cylinders. By an appropriate evaluation of the signal in the individual cylinder segments, for each cylinder a *cylinder individual injection mass residual* $r_{\mathrm{MFC1\ldots4}}$ results from the difference of nominal and reconstructed injection mass:

$$r_{\mathrm{MFC1\ldots4}} = m_{\mathrm{f,rec1\ldots4}} - m_{\mathrm{f,nom}} \tag{5.2.17}$$

which forms the symptoms $S_{\mathrm{MFC1\ldots4}}$ after passing a threshold.

5.2.3 Combined diagnosis for injection and combustion

In the following, experimental results for the residuals engine torque, injection mass, injection mass deviations, and cylinder individual injection mass are presented. The thresholds for symptom generation were experimentally determined at the test bench. They are a compromise of robust fault detection against disturbances and a fast and early detection of small faults.

First, the engine torque supervision is considered. Figure 5.2.10 shows the result of an injection mass reduction on the amplitude of the engine speed oscillation.

Starting from normal conditions (injection mass $m_{\mathrm{f,nom}} = 31$ mg/stroke, engine torque $M_{\mathrm{me}} = 145$ Nm) an injection mass fault is produced by incrementally reducing the injected fuel mass for all cylinders in a bypass mode of the RCP system. Since the other input variables are kept constant the engine torque calculated from the torque reference model (5.2.4) remains almost unchanged. Opposite to this, the delivered engine torque (5.2.6) and the engine-speed amplitude, respectively, decreases with decreasing injection mass. Thus, the residual (5.2.7) differs significantly from zero, which can be seen in Fig. 5.2.11. If the residual exceeds the depicted thresholds, the corresponding symptom deflects and indicates a fault.

For an individual injection mass fault in cylinder 3 at an engine speed of 1400 rpm and a nominal injection mass of 10 mg/stroke Fig. 5.2.12 depicts the oxygen concentration and the residuals from oxygen evaluation. Thereby, the injection mass in

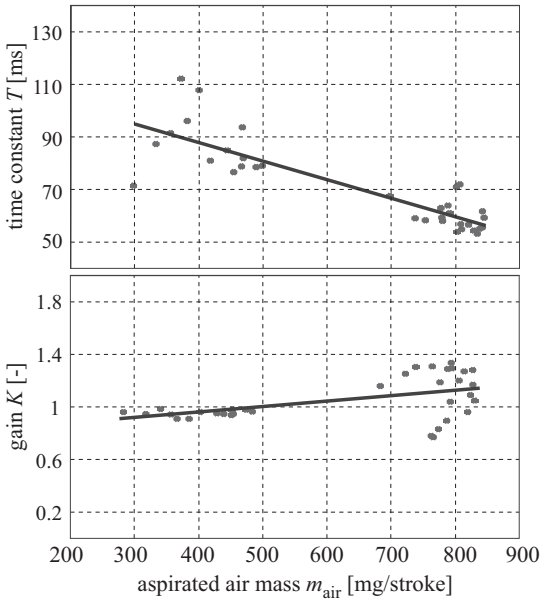


Fig. 5.2.7. Identified time constant and gain of the oxygen sensor.

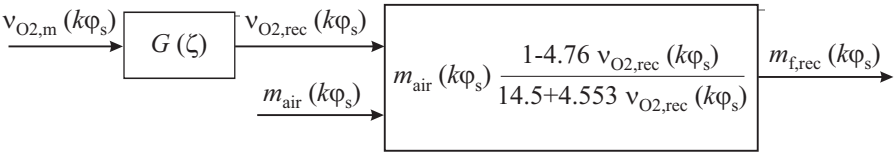


Fig. 5.2.8. Compensation of the oxygen sensor dynamics and calculation of the injected fuel mass.

cylinder 3 was incrementally reduced with 1 mg/stroke increments from the nominal injection mass to 5 mg/stroke. As it can be seen each of the residuals shows a deflection from zero, but only the residuals $r_{\Delta \mathrm{MF}}$ and r_{MFC3} exceed the thresholds (the thresholds of the residual r_{MF} lie outside of the axes scaling). Especially the cylinder selective residual r_{MFC3} of cylinder 3 shows an increasing deflection with increasing fault. If the thresholds are exceeded, the individual cylinder fault can be detected because of the symptom deflections.

The generated symptoms based on the engine speed and oxygen concentration enable the detection of different faults in the injection system and the mechanical components of the cylinders. Additionally, they can be used to supervise or adapt engine control functions, for example for torque or injection control. Table 5.2.1 presents different symptom patterns for possible faults in the injection system. The symptom deflections show different patterns for the indicated faults of the injection pump, the injectors and engine mechanics. Therefore, these faults can be isolated

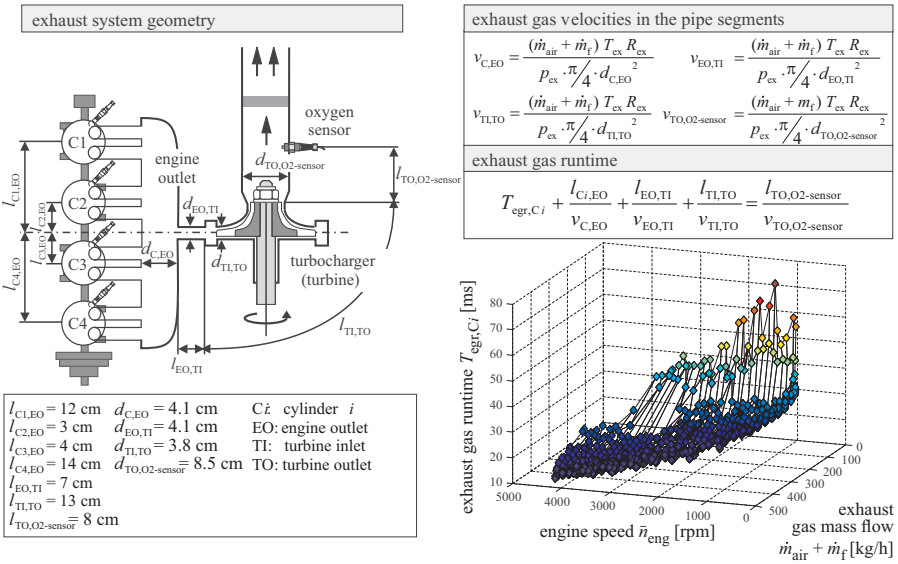


Fig. 5.2.9. Calculation of the exhaust gas runtimes from the cylinder outlet to the oxygen sensor using thermodynamic models and the geometric data of the exhaust system.

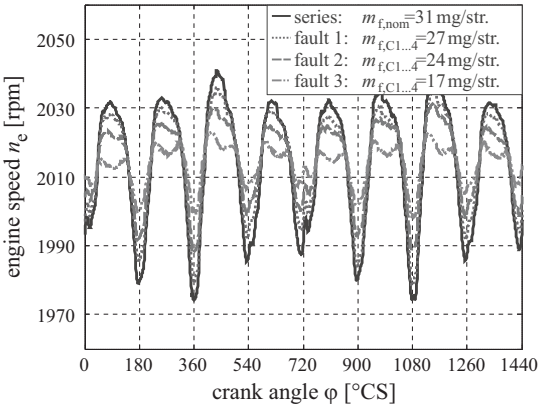


Fig. 5.2.10. Engine speed oscillation for common different injection mass deviations at all cylinders for a four-cylinder diesel engine.

and diagnosed. This table is then used as a basis for fault diagnosis, for instance with fuzzy if-then-rules.

5.2.4 Combustion pressure measurement analysis

The combustion pressure analysis in Isermann (2014), Sect. 4.3.3a) has shown how relatively simple features of the cylinder pressure course $p_{cyl}(\varphi)$ in dependence on

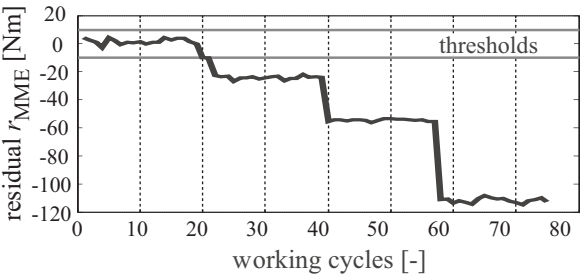


Fig. 5.2.11. Mean effective engine torque residual r_{MME} for a injection mass reductions of all cylinders.

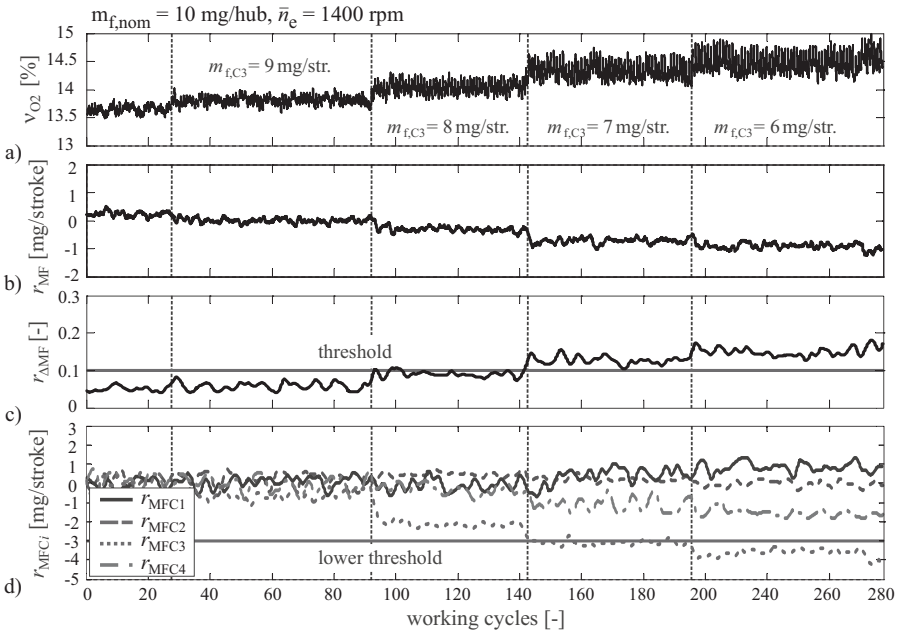


Fig. 5.2.12. Residuals from oxygen evaluation for a reduction of injection mass in cylinder 3: **a** measured oxygen concentration ν_{O2} (resolution 1°CS), **b** residual injection mass r_{MF} according to equation (5.2.13), **c** residual injection mass fluctuations $r_{\Delta\mathrm{MF}}$ according to equation (5.2.14), **d** residual cylinder selective injection mass $r_{\mathrm{MFC}i}$ according to equation (5.2.17).

the crank angle can be extracted. Herewith the information on the combustion lies in the difference pressure $\Delta p_{\mathrm{cyl}}(\varphi)$ between the combustion pressure $p_{\mathrm{cyl}}(\varphi)$ and the motored (towed) pressure $p_{\mathrm{cyl,m}}(\varphi)$. Suitable features for fault detection are the coordinates π_{c} and φ_{c} of the center of gravity and the peak difference pressure coordinates Δp_{peak} and φ_{peak}, as shown by Leonhardt (1996). They depend directly on the injected fuel mass m_{f}, begin of the injection φ_{soi} and speed n_{eng}. Therefore, the injected fuel mass and begin of injection can be reconstructed by the nonlinear

Table 5.2.1. Fault-symptom table for the distributor pump injection and combustion system.

	Faults	Symptoms					
		S_{MME}	S_{ER}	S_{ERCi}	S_{MF}	$S_{\Delta\mathrm{MF}}$	S_{MFCi}
Faults effecting all cylinders	pump pressure low	$-$	0	0	$-$	0	0
	chip tuning	$+$	0	0	$+$	0	0
Faults effecting a single cylinder i	injectors: clogging of the nozzles	0	$+$	$-$	0	$+$	$-$
	nozzle needle stuck at open	0	$+$	$-$	0	$+$	$+$
	nozzle needle stuck at closed	0	$++$	$--$	0	$++$	$--$
	cylinder mechanics (pistons, valves)	0	$+$	$-$	0	0	0

legend:

$++$ symptom responds intense positive
$+$ symptom responds positive
0 symptom does not respond
$-$ symptom responds negative
$--$ symptom responds intense negative

symptoms:

S_{MME} mean effective torque
S_{ER} overall engine roughness
S_{ERCi} cylinder individual engine roughness
S_{MF} injection mass all cylinders
$S_{\Delta\mathrm{MF}}$ injection mass fluctuation between cylinders
S_{MFCi} cylinder individual injection mass

approximation

$$m_f = f_m\left[\pi_c\,\varphi_c\,\Delta p_{\mathrm{peak}}\,\varphi_{\mathrm{peak}}\,n_{\mathrm{eng}}\right], \tag{5.2.18}$$

$$\varphi_{\mathrm{soi}} = f_\varphi\left[\pi_c\,\varphi_c\,\Delta p_{\mathrm{peak}}\,\varphi_{\mathrm{peak}}\,n_{\mathrm{eng}}\right]. \tag{5.2.19}$$

Experimental investigation with a turbocharged four-cylinder diesel engine VW 1.2l, 126 Nm, 51 kW, swirl chamber injection, showed, that an approximation with radial basis function (RBF) neural networks led to a relatively good performance with training data $m_f = 5, 10, 15, 20, 25\mathrm{mg}$, $\varphi_{\mathrm{soi}} = -4, -7, -10, -13, -16\,^\circ\mathrm{CS}$ and $n_{\mathrm{eng}} = 1200, 1500, 1800, ..., 3600\mathrm{rpm}$. The standard deviation of the two reconstructed values are 2-3% for the training data ($\pm 1.5\mathrm{mg}$ resp. $\pm 1^\circ\mathrm{CS}$) and for other validation data about 10% for stationary operation. The trained static model can also be used for instationary driving. The reconstructed values for an acceleration with manual gear shifting from the 1st to the 5th gear with a simulated vehicle at the test bench are shown in Leonhardt (1996), Figures 5.75 and 5.76. Both values show stochastic fluctuations, but follow in principle the reference values from the ECU with moderate accuracy if the zero fuel injection during gear shifting is excluded. Based on the averaged values of reconstructed fuel injection and begin of injection over certain operation points, the calculated differences indicate faults in the injection system, assuming that the combustion is normal.

A further possibility based on the measured combustion pressure is to calculate features like the brake mean effective pressure p_{mean} (IMEP) or the crank angle φ_{Q50} for the 50 % mass fraction burned, see Isermann (2014), Section 4.2. Some examples for control and monitoring of a dual fuel gas diesel engine are shown in Barta and Hampson (2016).

5.3 Common-rail injection system

High pressure injection systems play a central role in any direct injection internal combustion engine. Many faults, malfunctions and failures of the engines origin in the highly stressed injection pump, common-rail pressure control and injectors. Therefore, it is of considerable interest to detect and diagnose early incipient faults before they lead to intermittent and permanent drastic faults and even failures of the injection followed by failures of the combustion process.

In the frame of the OBD diagnosis the injection system is already supervised with regard to some emission-related faults. Hence, larger faults of the sensors, the rail pressure control deviations and wire connections are already checked and examined for plausibility. However, a detailed fault detection and diagnosis is usually not provided.

In the following a model-based fault-detection module for an electronically controlled common-rail injection system for diesel engines is described. It is the result of an in-depth case study, Clever and Isermann (2008), Clever (2010), Clever and Isermann (2010), Clever (2011). The investigated injection system is the type CP1H from Bosch. A corresponding scheme is depicted in Fig. 5.3.1. Measurements are made on a test bench with a four-cylinder diesel engine, Opel/GM, 1.9l, 110 kW, 315 Nm, as depicted in Fig. 5.4.1a).

The dynamic models of the components of a high pressure common-rail system were already established in Isermann (2014), Sect. 5.5.3, resulting in the signal flow chart, Fig. 5.3.2, for the common-rail pressure.

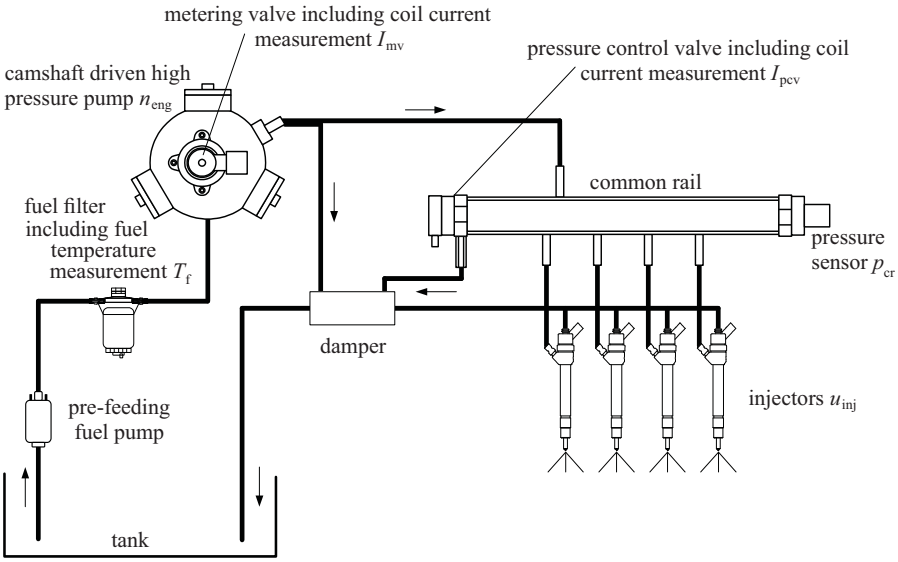


Fig. 5.3.1. Scheme of a common-rail injection system.

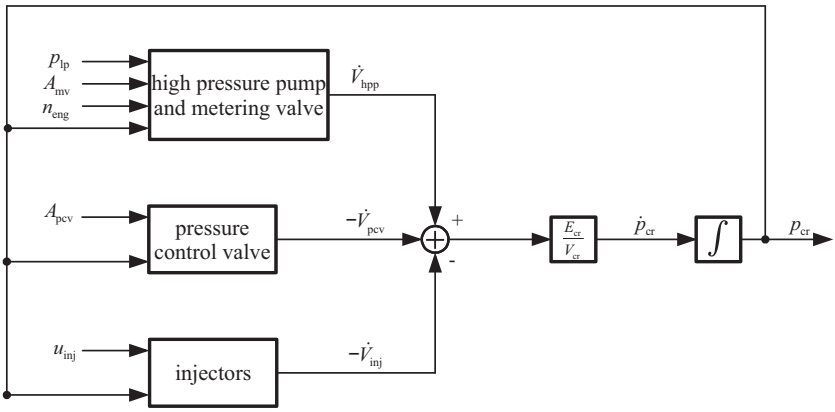


Fig. 5.3.2. Signal flow chart for the common-rail pressure, Isermann (2014).

5.3.1 Analysis of the rail pressure signal

The common-rail pressure consists of a mean value according to the reference variable of the pressure control loop and periodic oscillations through the discontinuous delivery of the radial piston pump and periodic opening and closing of the injectors. The mean value as well as the oscillation are subject to changes if faults in the injection system arise. Before corresponding fault-detection features are stated, first the signal characteristics of the oscillations are analyzed.

Because these oscillations are composed of periodic signals with different frequencies, they are modeled by a Fourier series

$$y(t) = \frac{a_0}{2} + \sum_{\nu=1}^{\infty} a_\nu \cos(\nu\omega_0 t) + \sum_{\nu=1}^{\infty} b_\nu \sin(\nu\omega_0 t) \tag{5.3.1}$$

with

a_ν, b_ν	Fourier coefficients
$\omega_0 = 2\pi f_0$	characteristic angular frequency
f_0	fundamental frequency.

The Fourier coefficients are defined as

$$a_\nu(\nu\omega_0) = \frac{2}{T_\mathrm{p}} \int_0^{T_\mathrm{p}} y(t) \cos(\nu\omega_0 t) dt, \tag{5.3.2}$$

$$b_\nu(\nu\omega_0) = \frac{2}{T_\mathrm{p}} \int_0^{T_\mathrm{p}} y(t) \sin(\nu\omega_0 t) dt. \tag{5.3.3}$$

The time period T_p needs to be a multiple of the fundamental period

$$T_\mathrm{p} = r \frac{1}{f_0}, \quad r = 1, 2, \ldots \tag{5.3.4}$$

Given the Fourier coefficients, the amplitude of the oscillation for a characteristic angular frequency can be calculated by

$$A_y(\nu\omega_0) = \sqrt{a_\nu^2 + b_\nu^2}. \tag{5.3.5}$$

In order to get the coefficients from sampled data, (5.3.1) has to be discretized:

$$a_\nu(\nu\omega_0) \approx \frac{2}{L_\mathrm{p}} \sum_{k=0}^{L_\mathrm{p}-1} y(kT_0) \cos(\nu\omega_0 k T_0), \tag{5.3.6}$$

$$b_\nu(\nu\omega_0) \approx \frac{2}{L_\mathrm{p}} \sum_{k=0}^{L_\mathrm{p}-1} y(kT_0) \sin(\nu\omega_0 k T_0), \tag{5.3.7}$$

$$L_\mathrm{p} = \frac{T_\mathrm{p}}{T_0}, \tag{5.3.8}$$

where T_0 is the sampling time and N the signal length.

Figure 5.3.3 shows the amplitude spectra of a common-rail pressure sensor signal while the engine was in a stationary operation point in overrun (no fuel injections) and the common-rail pressure was in open loop. The values of the main variables for the operation condition are also shown. The shown measurement is mainly characterized by an oscillation with the angle period

$$\text{main pump cycle:} \qquad \tau_\mathrm{pist} = 180^\circ\mathrm{CS}. \tag{5.3.9}$$

This oscillation is forced by the high pressure pump. Because the radial piston pump consists of three pistons, and is driven by the camshaft belt at a ratio of 2:3 relating to the crankshaft, each $180^\circ\mathrm{CS}$ one pump element delivers fuel leading to the main pump cycle.

Usually, the three pump elements deliver the same amount of fuel in a stationary operation point. If the delivery quantity of at least one element differs, an additional oscillation appears (see Fig. 5.3.4). Since each of the three pump elements delivers fuel once in $540^\circ\mathrm{CS}$, the cycle angle of this oscillation is

$$\text{second pump cycle:} \qquad \tau_\mathrm{hpp} = 540^\circ\mathrm{CS}. \tag{5.3.10}$$

If fuel is injected, the considered four-cylinder engines opens the injectors each $180^\circ\mathrm{CS}$, yielding the

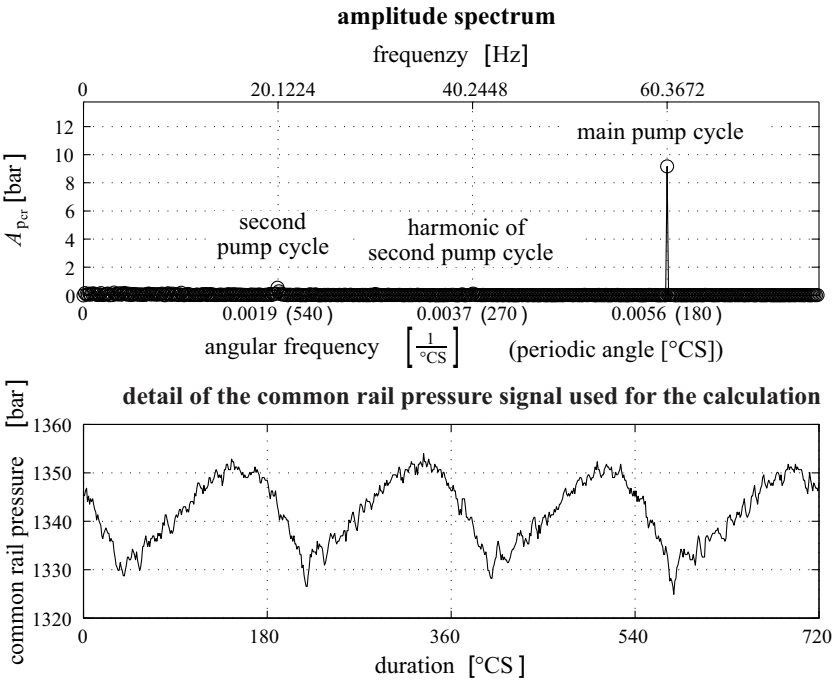
$$\text{main injection cycle:} \qquad \tau_\mathrm{inj} = 180^\circ\mathrm{CS} \tag{5.3.11}$$

which has the same cycle angle as the main pump cycle, see Fig. 5.3.5.

If the volume flow to at least one injector differs due to different recycle flows or injection quantities, the cycle angle of additional oscillations becomes

$$\text{second injection cycle:} \qquad \tau_\mathrm{bank} = 720^\circ\mathrm{CS}. \tag{5.3.12}$$

This second injection cycle now differs from the second pump cycle.



operation point

common rail pressure controlled in open loop

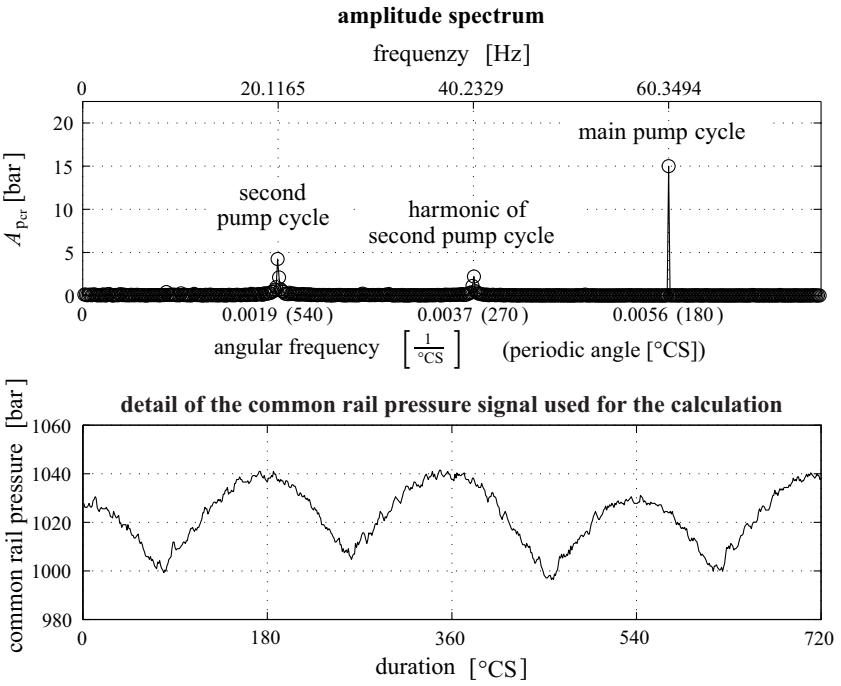
essential variables	n_{eng}	$\dot{V}_{inj}$	I_{pcv}	I_{mv}
values	≈1800 rpm	$0\frac{mm^3}{cyl}$	1.05A	0.87A

Fig. 5.3.3. Measured common-rail pressure signal and its amplitude spectrum in overrun. The basic oscillation is forced by the three pump elements of the camshaft belt driven radial piston pressure pump.

If the delivery quantity of at least one of the three pump elements is different from the others and at the same time the volume flow to at least one of the four injectors differs from the others, all of the above discussed oscillations are present in the signal but have different amplitudes. This leads to several combinations of different fuel delivery quantities on the one hand and different volume flows to the injectors on the other. These combinations repeat after the period

$$\text{superimposed cycle: } \tau_{sup} = \tau_{hpp,bank} = 3 \cdot 4 \cdot 180°\text{CS} = 2160°\text{CS}. \quad (5.3.13)$$

Because the amplitudes of the common-rail pressure oscillations forced by the high pressure pump and the injections are small compared to the mean common-rail pressure, it can be assumed that the volume flows through the pressure control valve and through the leakages are constant for a stationary operating point:



operation point

common rail pressure controlled in open loop

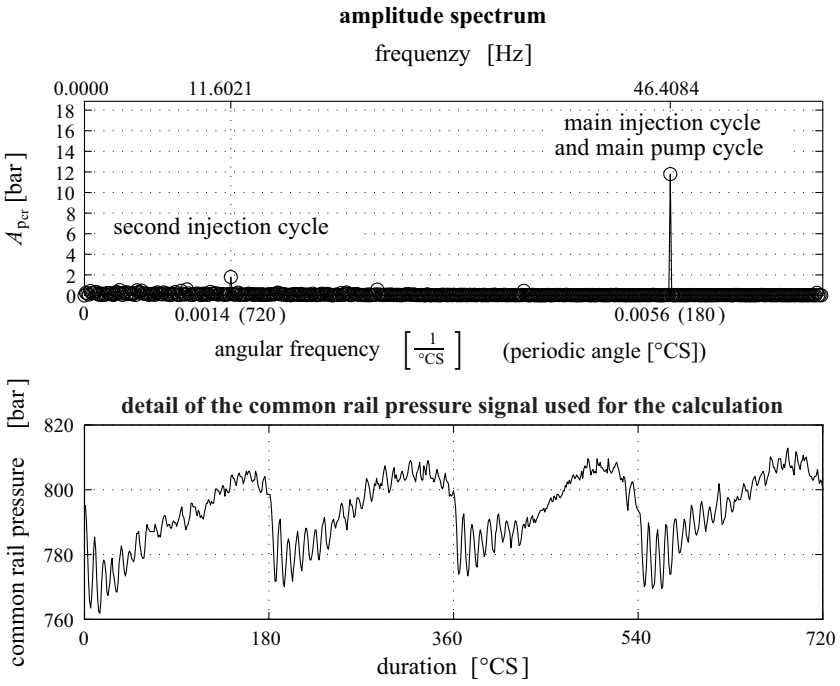
essential variables	n_{eng}	$\dot{V}_{\mathrm{inj}}$	I_{pcv}	I_{mv}
values	≈1800 rpm	$0\frac{\mathrm{mm}^3}{\mathrm{cyl}}$	1.15A	0.72A

Fig. 5.3.4. Measured common-rail pressure signal and its amplitude spectrum in overrun. One pump piston delivers less fuel, leading to a second pump cycle with period $540°\mathrm{CS}$.

$$V_{\mathrm{pcv}} + V_{\mathrm{leak}} \approx \text{constant for a stationary operation point.} \tag{5.3.14}$$

Thus, these volume flows do not cause additional oscillations of the common-rail pressure. Figure 5.3.6 illustrates the discussed oscillations with the cycle angles τ_{inj}, τ_{bank}, τ_{pist}, τ_{hpp} and τ_{sup}.

The amplitude of the main pump oscillation at $180°\mathrm{CS}$ depends on the delivered fuel and angular speed. The delivered fuel to the high pressure pump from the low pressure pump is feedforward controlled by the metering valve. This valve determines the fuel flow rate to the pump in order to save pumping energy and unnecessary heating-up of the fuel, Robert Bosch GmbH (2005). Therefore, the high pressure pump has some non-delivery rotation angles $0 \leq \chi \leq 180°\mathrm{CS}$. As shown in Clever (2011), the amplitudes of the common-rail pressure change in dependence on the engine speed and metering valve position respectively valve current



operation point

common rail pressure controlled in open loop

essential variables	n_{eng}	$\dot{V}_{\text{inj}}$	I_{pcv}	I_{mv}
values	≈1300 rpm	≈20.02 $\frac{\text{mm}^3}{\text{cyl}}$	1.05A	0.80A

Fig. 5.3.5. Measured common-rail-pressure signal and its amplitude spectrum for active injectors. One injector injects more fuel than the others, leading to a second injection cycle with period $540°\text{CS}$.

$4\,\text{bar} \leq A_{\text{pcr}}(180°\text{CS}) \leq 15\,\text{bar}$ for overrun and can be represented as normal values in a look-up table.

The measured oscillations of the common-rail pressure finally are a superposition of the individual induced oscillations by the high pressure pump and the injectors. The superposition leads to following measurable periods:

$$\begin{aligned} &\text{first sum period} && \tau_{\text{pist,inj}} = 180°\text{CS} \\ &\text{second pump period} && \tau_{\text{hpp}} = 540°\text{CS} \\ &\text{second injector period} && \tau_{\text{bank}} = 720°\text{CS} \\ &\text{second sum period} && \tau_{\text{hpp,bank}} = 2160°\text{CS} \end{aligned} \tag{5.3.15}$$

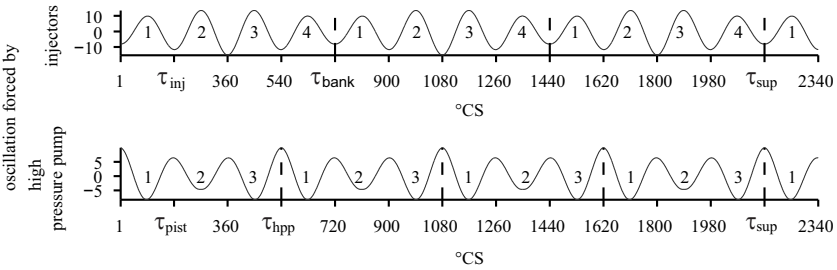


Fig. 5.3.6. Illustration of the oscillations forced by the volume flows through the high pressure pump and the injectors. (Each number indicates the injector which is currently active and the pump element which currently delivers fuel, respectively. In order to illustrate the oscillations caused by differences of the volume flows to the injectors on the one hand and differences of the delivery quantities of the three pump elements on the other, different amplitudes of the forced oscillations are assumed. For the injector flows it is assumed that the volume flows to the injectors 1 and 4 are smaller than the flows to the injectors 2 and 3. For the delivery quantities of the pump elements it is assumed that the delivery quantity of pump element 2 is smaller than the delivery quantities of the other pump elements).

where it holds for the resulting frequencies with $f = n_{\mathrm{eng}}\tau/360$. For $n_{\mathrm{eng}} = 1800\mathrm{rpm} = 30\mathrm{rps}$ one obtains

$$\begin{aligned} &\text{first sum frequency} && f_{\mathrm{pist,inj}} = 60\mathrm{Hz} \\ &\text{second pump frequency} && f_{\mathrm{hpp}} = 20\mathrm{Hz} \\ &\text{second injector frequency} && f_{\mathrm{bank}} = 15\mathrm{Hz} \\ &\text{second sum frequency} && f_{\mathrm{hpp,bank}} = 5\mathrm{Hz} \end{aligned} \tag{5.3.16}$$

5.3.2 Model-based fault diagnosis

A failure mode and effect analysis (FMEA) with faults of the fuel filter, metering valve, pump mechanics, pressure control valve, injector including seals and possible failure effects on the engine is reported in Clever (2011). Based on this analysis, artificially introduced faults have been selected for experiments.

The fault detection is based on output residuals between measured or calculated quantities and their normal values, see Fig. 5.3.7.

a) Mean common-rail pressure

Figure 5.3.2 shows the used signal flow diagram of the pressure build-up in the common-rail. In order to remove the oscillations of the common-rail signal, it is low-pass filtered.

Based on physical modeling leading to the model structure, a common-rail pressure model can be identified from measured data. But not all of the inputs in Fig. 5.3.2 are available as measurements. For this reason the model structure, depicted in

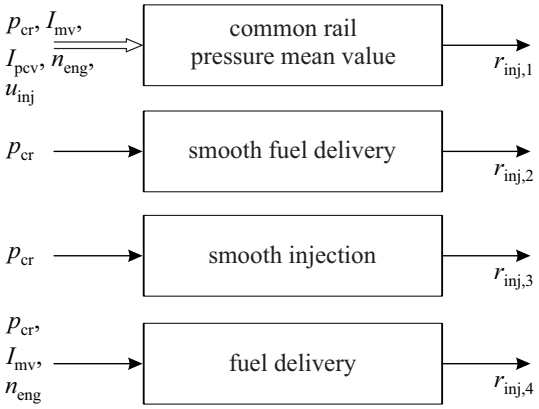


Fig. 5.3.7. Measured input variables and residuals for the model-based fault detection of the common-rail injection system.

Fig. 5.3.8, is used. It is assumed that the pressure before the metering valve is constant. Thus, it is not used as an input for the model anymore. Further it is assumed that the opening cross section of the metering valve and the pressure control valve are approximately proportional to the measured current through the electromagnetic actuators. Thus, the measured currents through the two valves are used as model inputs. The hysteresis of the position of the pressure control valve is taken into account by offset values for the maximum hysteresis width.

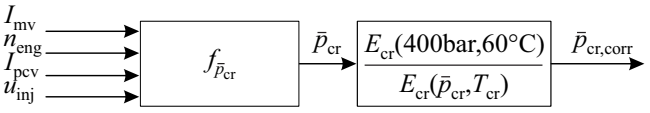


Fig. 5.3.8. Signal flow scheme for the common-rail pressure with used measured input signals.

By means of a model for the bulk modulus, the measured common-rail pressure is related to a bulk modulus at a particular pressure (e.g. 400 bar) and at a certain temperature (e.g. 60°C). Also the fuel temperature is calculated based on a model.

The data-based modeling and identification of the common-rail pressure is conducted using the net model LOLIMOT with the inputs:

I_{mv}	supply current of the metering valve
I_{pcv}	supply current of the pressure control valve
n_{eng}	engine speed
u_{inj}	desired injection quantity

On basis of the estimated model following parity equation for the “mean common-rail pressure” can be formulated:

$$r_{\mathrm{inj},1} = \overline{p}_{\mathrm{cr,corr}} - f_{\overline{\mathrm{p}}_{\mathrm{cr,corr}}}(I_{\mathrm{mv}}, n_{\mathrm{eng}}, I_{\mathrm{pcv}}, u_{\mathrm{inj}}) \tag{5.3.17}$$

with

$\overline{p}_{\mathrm{cr,corr}}$	large signal of the common-rail pressure signal corrected with respect to a condition at 400 bar and 60°C
T_{cr}	calculated temperature of the fuel inside the common-rail

b) Uniformity

The residuals "smooth fuel delivery" and "smooth injections" are calculated on the basis of the characteristic periods of the common-rail pressure oscillations, stated in (5.3.15). In order to distinguish between non-uniform fuel delivery and non-uniform injections, it is sufficient to monitor the difference of the rail pressure signals at crankshaft φ and its value for a phase shift of $720°\mathrm{CS}$ and $540°\mathrm{CS}$ by calculating:

- non-uniform fuel delivery by the pump:

$$r_{\mathrm{inj},2} = \frac{1}{2160°\mathrm{CS}} \int_{\mathrm{t}(\varphi - 2160°\mathrm{CS})}^{t(\varphi)} |(p_{\mathrm{cr}}(\varphi) - p_{\mathrm{cr}}(\varphi + 720°\mathrm{CS}))|\, dt, \tag{5.3.18}$$

- non-uniform injections:

$$r_{\mathrm{inj},3} = \frac{1}{2160°\mathrm{CS}} \int_{\mathrm{t}(\varphi - 2160°\mathrm{CS})}^{t(\varphi)} |(p_{\mathrm{cr}}(\varphi) - p_{\mathrm{cr}}(\varphi + 540°\mathrm{CS}))|\, dt. \tag{5.3.19}$$

c) Fuel delivery

A further residual "fuel delivery" evaluates the amplitude of the $180°\mathrm{CS}$ synchronous periodic oscillation during the *overrun state*. As long as no fuel is injected into the cylinders, the amplitude of this oscillation is a measure for the fuel mass flow through the high pressure pump. The fuel mass flow through the pressure control valve is assumed to be constant. The mean amplitude of this oscillation calculated over one pump cycle ($540°\mathrm{CS}$) – as measure for the fuel mass flow through the high pressure pump – mainly depends on the pressure before the metering valve, the opening cross section of the metering valve, the engine speed and finally on the bulk modulus of the fuel in the common-rail. Under the same assumptions, that were already made for the residual "mean common-rail pressure", it is feasible to identify the amplitude of the signal from measurable variables. The dependence of the common-rail pressure on the bulk modulus E_{cr} of the fuel, which again depends on the fuel temperature, is incorporated by correcting the measured pressure as shown in Fig. 5.3.9.

The mean amplitude over $540°\mathrm{CS}$ is calculated by a Fourier analysis according to:

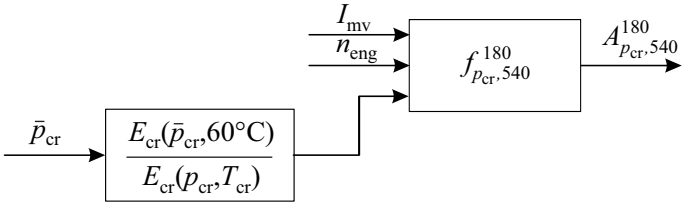


Fig. 5.3.9. Signal flow scheme for the amplitude of the $180°\mathrm{CS}$ rail pressure oscillation in overrun. E_{cr} [N/m^2] bulk modulus of the fuel.

$$A^{180}_{\mathrm{p_{cr},540}}(t(\varphi)) = \left(\left(\frac{2}{540°\mathrm{CS}} \int_{\varphi-540°\mathrm{CS}}^{\varphi} p_{\mathrm{cr}}(\varphi) \cos(\omega_{\tau_{\mathrm{pist}}} \cdot \varphi)\, d\varphi \right)^2 + \left(\frac{2}{540°\mathrm{CS}} \int_{\varphi-540°\mathrm{CS}}^{\varphi} p_{\mathrm{cr}}(\varphi) \sin(\omega_{\tau_{\mathrm{pist}}} \cdot \varphi)\, d\varphi \right)^2 \right)^{\frac{1}{2}} \tag{5.3.20}$$

with

$A^{180}_{\mathrm{p_{cr},540}}$ mean amplitude of the oscillations with cycle period $180°\mathrm{CS}$ over $540°\mathrm{CS}$

$\omega_{\tau_{\mathrm{pist}}} = \frac{2\pi}{\tau_{\mathrm{pist}}}$

The identification of the model for the pressure amplitudes is performed by means of local-polynomial models on basis of the LOLIMOT-partition-algorithm. The resulting parity equation for the residual "mean common-rail pressure amplitude in overrun state" is finally given by:

$$r_{\mathrm{inj},4} = A^{180}_{p_{\mathrm{cr}},540} - f^{180}_{p_{\mathrm{cr}},540}(I_{\mathrm{mv}}, n_{\mathrm{eng}}, \overline{p}_{\mathrm{cr,corr,A}}) \tag{5.3.21}$$

with

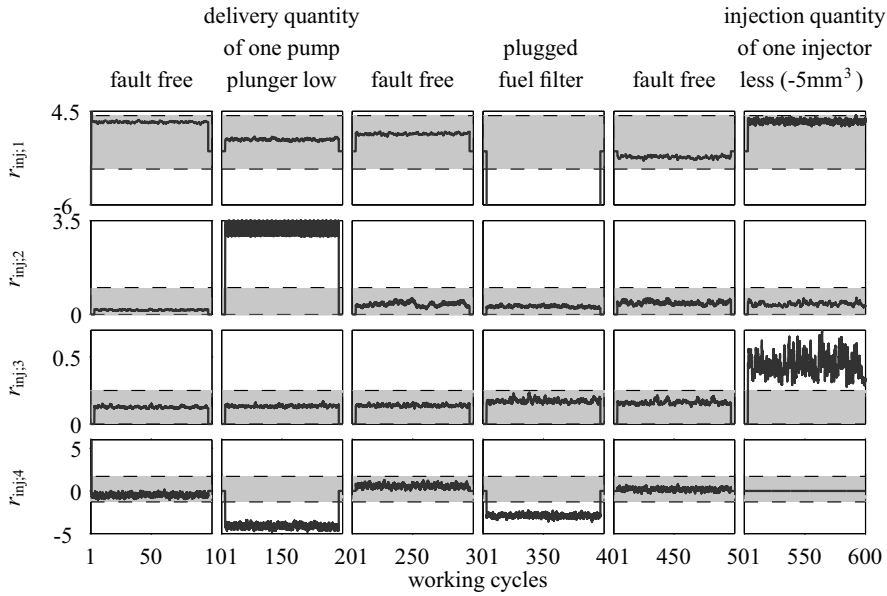
$f^{180}_{p_{\mathrm{cr}},540}$ model of the rail pressure amplitude for normal state

$\overline{p}_{\mathrm{cr,corr,A}}$ large signal of the common-rail pressure signal with respect to the actual mean common-rail pressure and $60°\mathrm{C}$

d) Experimental results

The fault detection method has been tested at the engine testbench with a four-cylinder common-rail diesel engine. Three faults are introduced:

- A reduced fuel delivery quantity of one of the pump elements by manipulating the return spring of the corresponding piston
- A plugged fuel filter by throttling the fuel pipe.
- Non-uniform fuel mass flows to the injectors by a modification of the setpoint for the injection quantity of one cylinder in the ECU.



operation points

essential variables	n_{eng}	u_{inj}	I_{pcv}	I_{mv}
working cycles 1 to 100	$\approx$ 1010 rpm	$0\frac{mm^3}{cyl}$	0.54 A	1.16 A
working cycles 101 to 200	$\approx$ 1010 rpm	$0\frac{mm^3}{cyl}$	0.55 A	1.17 A
working cycles 201 to 300	$\approx$ 1410 rpm	$0\frac{mm^3}{cyl}$	0.64 A	1.13 A
working cycles 301 to 400	$\approx$ 2600 rpm	$0\frac{mm^3}{cyl}$	0.8 A	0.96 A
working cycles 401 to 500	$\approx$ 1810 rpm	$0\frac{mm^3}{cyl}$	0.74 A	1.08 A
working cycles 501 to 600	$\approx$ 1980 rpm	$\approx 10\frac{mm^3}{cyl}$	0.87 A	1.07 A

Fig. 5.3.10. Residual deflections for three faults at different operating points.

The results of the fault detection, with respect to different operation points, are shown in Fig. 5.3.10

The faults were inserted sequentially and for each fault a measurement was conducted. Then, the engine was stopped. In the next step the fault was corrected and the next fault was introduced. Afterwards the engine was started again and the next measurement was conducted. The measured signals were measured with a sampling angle of 1°CS. Different operation points have been chosen to demonstrate the algorithms' performance in the entire operation range. The gray areas mark the values of the residual, while the system is assumed to be in a fault-free condition. The thresholds have been set under the objective to enable the detection of even small faults on the one hand and to prevent false alarms on the other.

Figure 5.3.10 shows that all simulated faults are detected. While the delivery rate of one pump element is reduced, the residuals "smooth fuel delivery" $r_{\mathrm{inj},2}$ and "fuel

delivery" $r_{\mathrm{inj},4}$ exceed the thresholds. The plugged fuel filter can be detected by the residuals "mean common-rail pressure" $r_{\mathrm{inj},1}$ and "mean common-rail pressure amplitude" $r_{\mathrm{inj},4}$. Reduced injection quantities can be detected by means of the residual "smooth injections" $r_{\mathrm{inj},3}$.

Note, that the deflection of the residuals "smooth injections" does not depend on the actual injection quantity level, but on the injection quantity difference between two injectors. Thus, the injection quantity fault is detectable regardless to the actually desired injection quantity.

Since the models that are applied in the parity equations are based on a number of assumptions (use of electric current signals instead of the opening cross section of the valves, constant pressure in front of the metering valve) plus the fact, that they are identified from measured data, the oscillations of the residual "mean common-rail pressure" and "fuel delivery" are larger than the oscillations of the uniformity residuals.

The presented fault-detection methods have been tested in the operation range shown in Table 5.3.1.

The fault-symptom table, Table 5.3.2, gives a qualitative overview of the type and number of faults, that can be detected and diagnosed by means of the introduced fault-detection module. The symptoms $S_{\mathrm{inj},1}$ to $S_{\mathrm{inj},3}$ are gained from the presented residuals as follows:

Table 5.3.1. Operation ranges for testing the fault-detection algorithms.

Varied variable	Minimum	Maximum
n_{eng}	$1000\,\mathrm{rpm}$	$3000\,\mathrm{rpm}$
u_{inj}	$0\,\frac{\mathrm{mm}^3}{\mathrm{cyl}}$	$60\,\frac{\mathrm{mm}^3}{\mathrm{cyl}}$
p_{cr}	$300\,\mathrm{bar}$	$1800\,\mathrm{bar}$
I_{mv}	$0.7\,\mathrm{A}$	$1.5\,\mathrm{A}$
I_{pcv}	$0.2\,\mathrm{A}$	$1.2\,\mathrm{A}$

$$S_{\mathrm{inj,i}} = \begin{cases} +1, & \text{if } \; r_{\mathrm{inj,i}} > r^{+}_{\mathrm{inj,i}} \\ -1, & \text{if } \; r_{\mathrm{inj,i}} < r^{-}_{\mathrm{inj,i}} \\ \mathbf{0}, & \text{otherwise} \end{cases} \tag{5.3.22}$$

with

$r^{+}_{\mathrm{inj},i}, r^{-}_{\mathrm{inj},i}$ threshold, that defines the value in which the process is assumed to be fault-free.

The symptoms $S_{\mathrm{inj},4}$ to $S_{\mathrm{inj},6}$ are all gained from residual $r_{\mathrm{inj},4}$ using (5.3.21). However, symptom $S_{\mathrm{inj},4}$ is derived from the absolute value of $r_{\mathrm{inj},4}$, symptom $S_{\mathrm{inj},5}$ is only calculated for a widely opened metering valve and $S_{\mathrm{inj},6}$ for an almost closed one. This is because the deflection of the residual $r_{\mathrm{inj},4}$ on particular faults depends on the operating point. As can be seen from the table, the symptoms $S_{\mathrm{inj},4}$ and $S_{\mathrm{inj},6}$

Table 5.3.2. Fault-symptom table for the common-rail system. $0, +, -$: symptom is zero, positive, ✓, ×: yes, no.

		$S_{\mathrm{inj},1}$	$S_{\mathrm{inj},2}$	$S_{\mathrm{inj},3}$	$S_{\mathrm{inj},4}$	$S_{\mathrm{inj},5}$	$S_{\mathrm{inj},6}$	isolable
$F_{\mathrm{inj},1}$	low delivery quantity of one pump piston	0	+	0	+	−	+	✓
$F_{\mathrm{inj},2}$	reduced injection quantity of one injector	0	0	+	0	0	0	✓
$F_{\mathrm{inj},3}$	pressure loss in front of high pressure pump (e.g. a plugged fuel filter)	−	0	0	+	−	+	✓
$F_{\mathrm{inj},4}$	pressure in front of high pressure pump too high (e.g. a faulty metering valve)	+	0	0	+	+	−	✓
$F_{\mathrm{inj},5}$	opening of the pressure control valve is too large	−	0	0	0	0	0	✓
$F_{\mathrm{inj},6}$	opening of the pressure control valve is too small	+	0	0	0	0	0	✓
$F_{\mathrm{inj},7}$	pressure sensor signal is too high	+	0	0	+	−	−	✓
$F_{\mathrm{inj},8}$	pressure sensor signal is too low	−	0	0	+	+	+	✓

are not necessary for the diagnosis. However, since the symptom $S_{\mathrm{inj},4}$ is calculated in a wider operation range, it may be advantageous to use it in the diagnosis procedure in order to be able to diagnose the faults $F_{\mathrm{inj},5}$ and $F_{\mathrm{inj},6}$ earlier.

Table 5.3.2 shows that all faults can be distinguished from each other if the fault detection results are combined in a single fault-detection system.

The fault-detection method for the common-rail system was described for a four-cylinder engine with a three-piston high pressure pump. However, the method can also be adapted to engines with three and five to eight cylinders and high pressure pumps with one or two pistons, Clever (2011).

As the highest frequency respectively the shortest period is $180°\mathrm{CS}$, see (5.3.15), the sampling period should be $\leq 90°\mathrm{CS}$ according to Shannon's sampling theorem. Experiments with a sampling period of $60°\mathrm{CS}$ have shown the same results as in Fig. 5.3.10, Clever (2011). The described fault-diagnosis method was filed to patent, Clever and Isermann (2010).

5.4 Turbochargers with wastegate and variable geometry

As all modern diesel engines and increasingly also gasoline engines are equipped with charging units which are mechanically driven superchargers or exhaust-gas turbochargers, the monitoring and fault diagnosis of these highly stressed components is mandatory. Usually, the charging pressure $p_{2\mathrm{i}}$ and charging air temperature $T_{2\mathrm{i}}$ is measured. Together with the air mass flow rate $\dot{m}_{\mathrm{air}}$ this allows to determine the

power of the compressor with efficiency maps. These maps can be used for fault detection. In the case of turbochargers, the supervision of the turbine power additionally requires the measurements of the exhaust pressure p_3 and exhaust temperature T_3 upstream the turbine, in the exhaust manifold, and several efficiency maps. For modeling turbochargers the steady-state maps delivered from manufacturers can be used. However, they are frequently not precise in the required engine operation range and do not consider the pulsating exhaust gas flow. Therefore more comprehensive and dynamic models of turbochargers will be considered.

As a basis for model-based fault detection the nonlinear dynamic models of variable-geometry turbine (VGT) chargers and wastegate (WG) chargers will be considered in the next section. The models are derived for the diesel engine shown in Fig. 5.4.1a) with the variables shown in Fig. 5.4.1b).

5.4.1 Models of VGT turbochargers

Dynamic models of turbochargers can be derived based on thermodynamic changes of state or based on fluid dynamic approaches by using Euler's equation for turbomachinery. Both modeling ways are described in Isermann (2014), Zahn (2012) and Sidorow (2014). For fault detection thermodynamic models for the turbine and compressor power will now be preferred, because they need fewer parameters to be identified, Sidorow (2014). It is assumed that the engine is in normal warm state.

The dynamic behavior at one operating point $[u_{\mathrm{inj}}, n_{\mathrm{eng}}]$ results from the balance equation for the angular momentum

$$J_{\mathrm{tc}} \frac{d\omega_{\mathrm{tc}}(t)}{dt} = M_{\mathrm{t}}(t) - M_{\mathrm{c}}(t) - M_{\mathrm{f}}(t) \tag{5.4.1}$$

with torques M_{t} for the turbine, M_{c} for the compressor and M_{f} for the friction, see Fig. 5.4.2. Turning the equation to powers with $P = M\omega$ leads to

$$J_{\mathrm{tc}}\, \omega_{\mathrm{tc}}(t)\, \frac{d\omega_{\mathrm{tc}}(t)}{dt} = P_{\mathrm{t}}(t) - P_{\mathrm{c}}(t) - P_{\mathrm{f}}(t). \tag{5.4.2}$$

The friction torque is assumed to be dominated by viscous friction $M_{\mathrm{f}} = c_{\mathrm{f}}\, \omega_{\mathrm{tc}}$ and therefore

$$P_{\mathrm{f}}(t) = c_{\mathrm{f}}(T_{\mathrm{oil}})\, \omega_{\mathrm{tc}}^2. \tag{5.4.3}$$

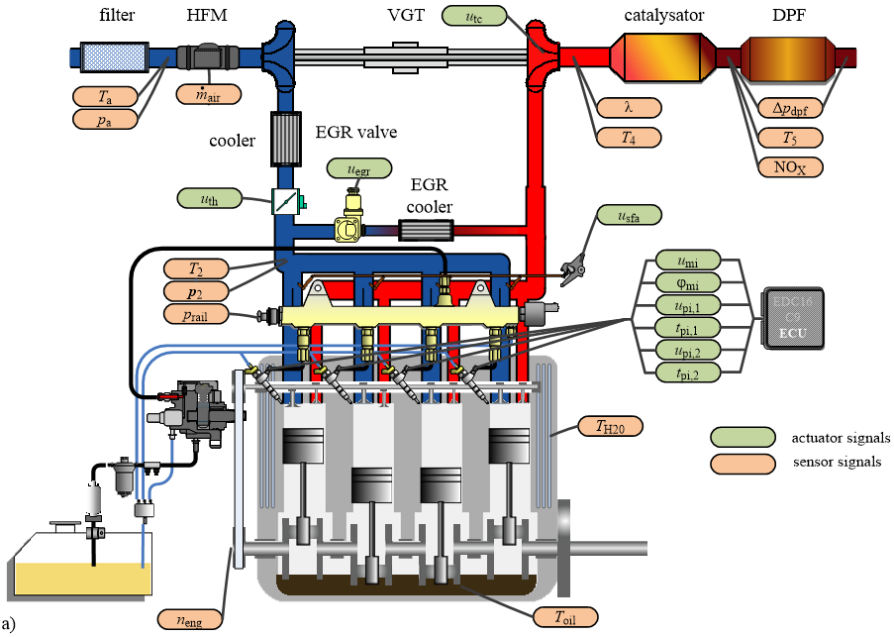
The *turbine power* follows according to

$$P_{\mathrm{t}} = \dot{m}_{\mathrm{t}}\, c_{\mathrm{pt}}\, T_3^*\, \eta_{\mathrm{t,is}} \left(1 - \left(\frac{p_4}{p_3} \right)^{\frac{\kappa-1}{\kappa}} \right) \tag{5.4.4}$$

with the mass flow determined by a throttle equation

$$\dot{m}_{\mathrm{t}} = A_{\mathrm{t,eff}}\, p_3 \sqrt{\frac{2}{R T_3^*}}\, \psi(\frac{p_4}{p_3}) \tag{5.4.5}$$

$$\psi(\frac{p_4}{p_3}) = \left(\frac{p_4}{p_3} \right)^{\frac{1}{\kappa}} \sqrt{\frac{c_{\mathrm{pt}}}{R} \left[1 - \left(\frac{p_4}{p_3} \right)^{\frac{\kappa-1}{\kappa}} \right]}. \tag{5.4.6}$$



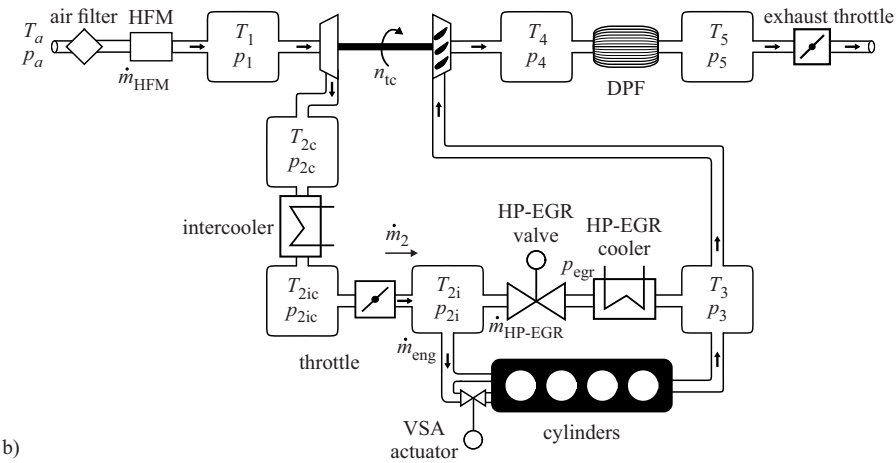


Fig. 5.4.1. a Scheme of the investigated diesel engine with common-rail direct injection and VGT turbocharger, Opel/GM, 1.9*l*, 110 kW, 315 Nm. Actuator and sensor signals as used on the IAT test bench. **b** Used variables for the diagnosis of turbochargers.

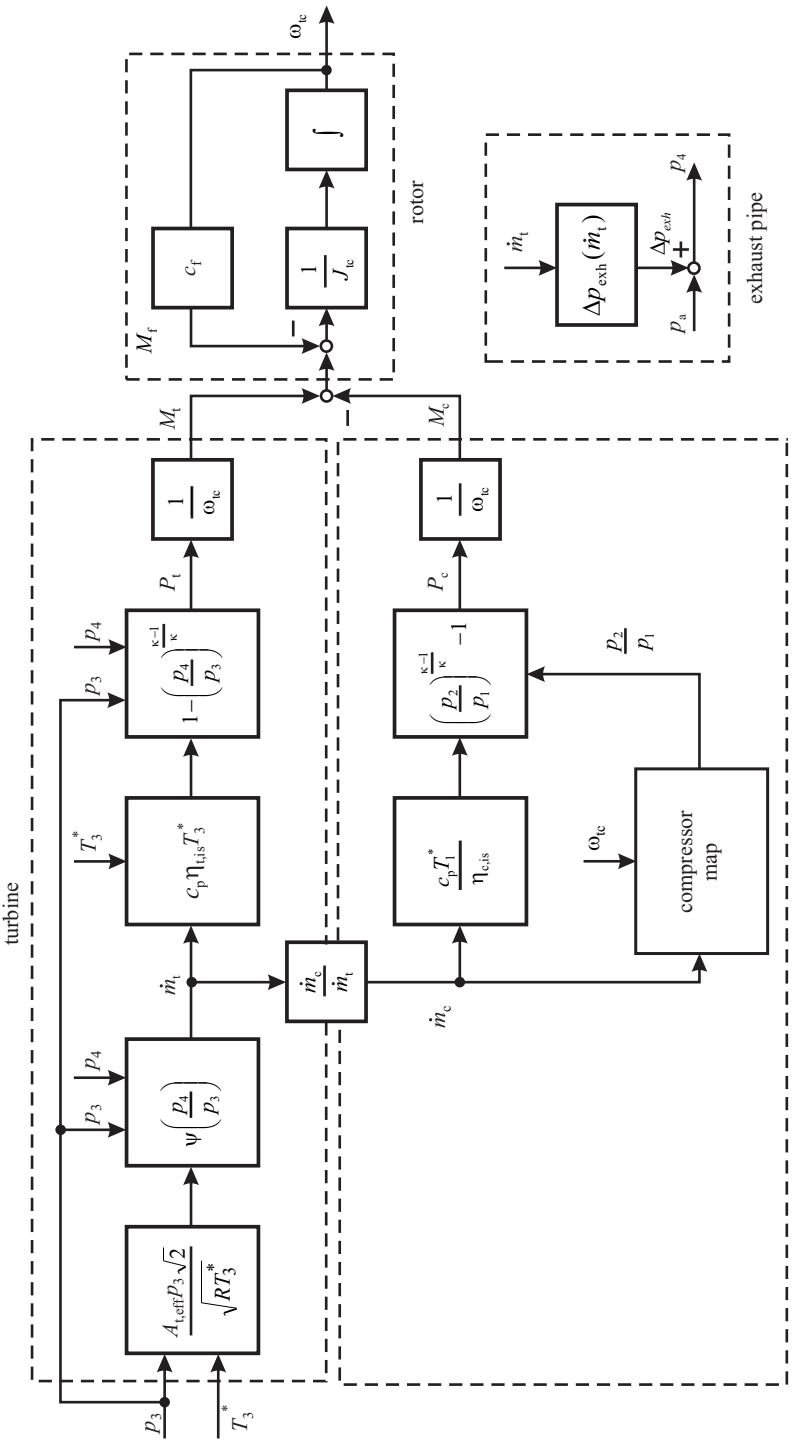


Fig. 5.4.2. Signal-flow chart for the thermodynamic model of a turbocharger (without internal heat transfer).

The effective cross section area $A_{\mathrm{t,eff}}(s_\mathrm{t}, p_4)$ is a function of the actuator position s_t and the pressure p_4 after the turbine. T_3^* is the gas temperature before the turbine reduced by the heat transfer loss to the compressor and the environment.

The isentropic efficiency of the turbine is defined as

$$\eta_{\mathrm{t,is}} = \frac{\frac{T_4}{T_3^*} - 1}{\left(\frac{p_4}{p_3}\right)^{\frac{\kappa-1}{\kappa}} - 1}. \tag{5.4.7}$$

It is difficult to determine this efficiency for low turbocharger mass flows and speeds, because the heat transfer to the compressor becomes effective in these cases. Therefore it is determined experimentally by a map $\eta_{\mathrm{t,is}}(s_\mathrm{t}, u_\mathrm{ref})$, see Fig. 5.4.3, where u_ref is the turbine blade speed ratio, Guzzella and Onder (2010),

$$u_\mathrm{ref} = \frac{u}{c_{\mathrm{u,max}}} = \frac{d_{\mathrm{t}3}\pi}{\sqrt{2c_\mathrm{p}T_3\left(1 - \left(\frac{p_4}{p_3}\right)^{\frac{\kappa-1}{\kappa}}\right)}}\omega_\mathrm{tc},$$

which is proportional to the turbocharger angular speed ω_tc, see also Isermann (2014).

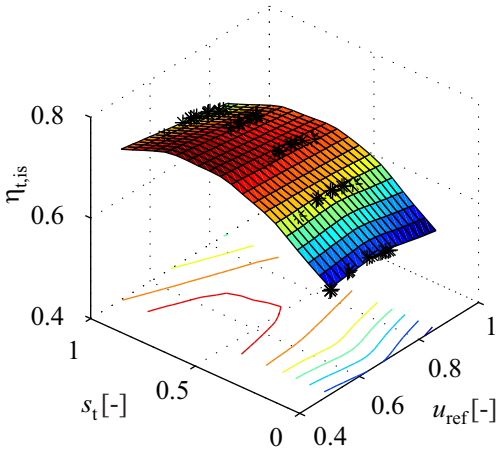


Fig. 5.4.3. Turbine isentropic efficiency map, approximated with a LOLIMOT net model, Sidorow et al (2011).

Hence, following variables have to be measured to calculate the turbine power $P_\mathrm{t}(u_\mathrm{inj}, n_\mathrm{eng})$

$$\mathbf{z}_\mathrm{t}^T = \left[p_3,\, p_4,\, T_3,\, s_\mathrm{t}\right]. \tag{5.4.8}$$

p_4 can be determined by the measured pressure drop of the particulate filter

$$p_4 = p_1 + \Delta p_\mathrm{pf}(\dot{m}_\mathrm{t}). \tag{5.4.9}$$

For larger mass flows, $T_3^* = T_3$ can be assumed. The *compressor power* is

$$P_\mathrm{c} = \dot{m}_\mathrm{c}\, c_\mathrm{pc}\, \frac{T_1^*}{\eta_\mathrm{c,is}} \left(\left(\frac{p_2}{p_1} \right)^{\frac{\kappa-1}{\kappa}} - 1 \right) \tag{5.4.10}$$

where $\dot{m}_\mathrm{c}$ is taken from a compressor mass flow map $\dot{m}_\mathrm{c}(n_\mathrm{tc}, p_2/p_1)$, see Fig. 5.4.4, or is directly measured with $\dot{m}_\mathrm{c} = \dot{m}_\mathrm{air}$.

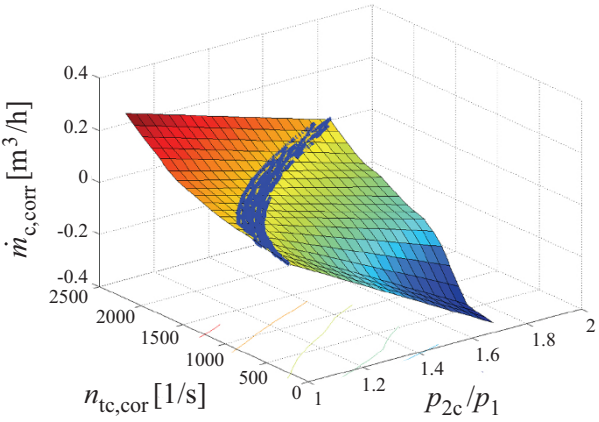


Fig. 5.4.4. Compressor mass flow map.

T_1^* is the inlet temperature, taking into account the heat transfer from the turbine

$$T_1^* = T_1 + \frac{k_{13} A_\mathrm{c}}{c_\mathrm{pc}\, \dot{m}_\mathrm{c}} (T_3 - T_1) \tag{5.4.11}$$

with the heat transfer coefficient k_{13} and the effective surface area A_c.

The isentropic efficiency of the compressor is defined as

$$\eta_\mathrm{c,is} = \frac{\left(\frac{p_2}{p_1} \right)^{\frac{\kappa-1}{\kappa}} - 1}{\frac{T_2}{T_1^*} - 1} \tag{5.4.12}$$

and it is experimentally determined in form of a map $\eta_\mathrm{c,is}(\dot{m}_\mathrm{c}, n_\mathrm{tc})$, see Fig. 5.4.5.

The required measurements needed to calculate the compressor power $P_\mathrm{c}(u_\mathrm{inj}, n_\mathrm{eng})$ are

$$\mathbf{z}_\mathrm{c}^\mathrm{T} = \left[\dot{m}_\mathrm{air},\, T_1,\, p_1,\, p_2 \right]. \tag{5.4.13}$$

To make the turbocharger characteristics independent of the changing environmental conditions, the variables are usually stated as referenced or corrected quantities using fluid mechanical laws. The reference conditions for the compressor inlet are $T_\mathrm{1ref} = 193\mathrm{K}$, $p_\mathrm{1ref} = 0.981\mathrm{bar}$ and for the turbine inlet they are $T_\mathrm{3ref} = 873\mathrm{K}$. This leads to the corrected rotational speed

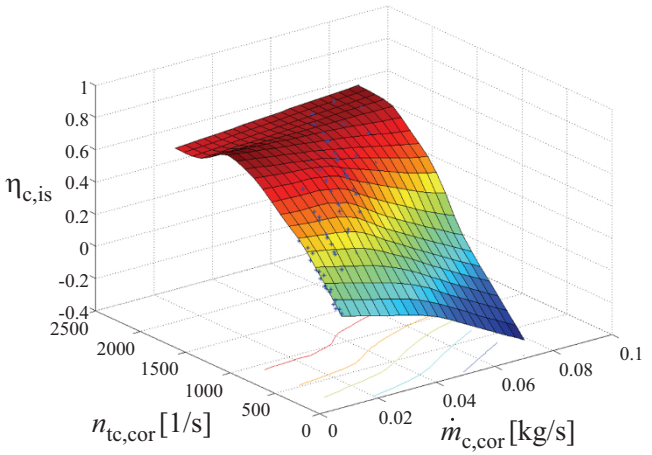


Fig. 5.4.5. Compressor isentropic efficiency map.

$$\omega_{\mathrm{tc,cor}} = \omega_{\mathrm{tc}} \sqrt{\frac{T_{\mathrm{ref}}}{T}} \tag{5.4.14}$$

and the corrected mass flow

$$\dot{m}_{\mathrm{cor}} = A \, \frac{p_{\mathrm{in,ref}}}{\sqrt{T_{\mathrm{in,ref}}}} \sqrt{\frac{2}{R}} \psi. \tag{5.4.15}$$

The heat transfer from the turbine to the compressor is relatively complicated. It is described in Sidorow (2014), Isermann (2014). However, as it can be neglected for large mass flows, Zahn (2012), it can be assumed in the above equations that $T_3^* = T_3$ and $T_1^* = T_1$.

Introducing (5.4.3), (5.4.4), (5.4.10) and $M = P/\omega_{\mathrm{tc}}$ in (5.4.1) allows to calculate the turbocharger speed. The turbocharger speed is then a variable which is fed back to the efficiency maps $\eta_{\mathrm{t,is}}$ and $\eta_{\mathrm{c,is}}$, see Fig. 5.4.2. With the assumption of a steady-state and covering the rotor friction by a mechanical efficiency

$$P_{\mathrm{t,m}} = \eta_{\mathrm{m,tc}} \, P_{\mathrm{t}} \tag{5.4.16}$$

it holds $P_{\mathrm{t,m}} = P_{\mathrm{c}}$ and from (5.4.10) one obtains for the calculation of the charging pressure

$$\frac{p_2}{p_1} = 1 + \frac{\dot{m}_{\mathrm{t}}}{\dot{m}_{\mathrm{c}}} \, \frac{c_{\mathrm{pt}}}{c_{\mathrm{pc}}} \, \frac{T_3^*}{T_1^*} \, \eta_{\mathrm{tc}} \left(1 - \left(\frac{p_4}{p_3} \right)^{\frac{\kappa-1}{\kappa}} \right)^{\frac{\kappa}{\kappa-1}} \tag{5.4.17}$$

with the turbocharger overall efficiency

$$\eta_{\mathrm{tc}} = \eta_{\mathrm{t,is}} \, \eta_{\mathrm{c,is}} \, \eta_{\mathrm{m,tc}}. \tag{5.4.18}$$

It can be assumed that $\dot{m}_{\mathrm{t}}/\dot{m}_{\mathrm{c}} = 1.05$ because of blow-by losses.

The identification of the turbocharger models includes the determination of the efficiency maps and the estimation of some parameters. As shown in Fig. 5.4.6, the obtained dynamic turbocharger models agree well with measurements, Sidorow (2014), obtained for a four-cylinder diesel engine, Opel/GM Z19DTH 1.9*l*, 110 kW, 4000 rpm.

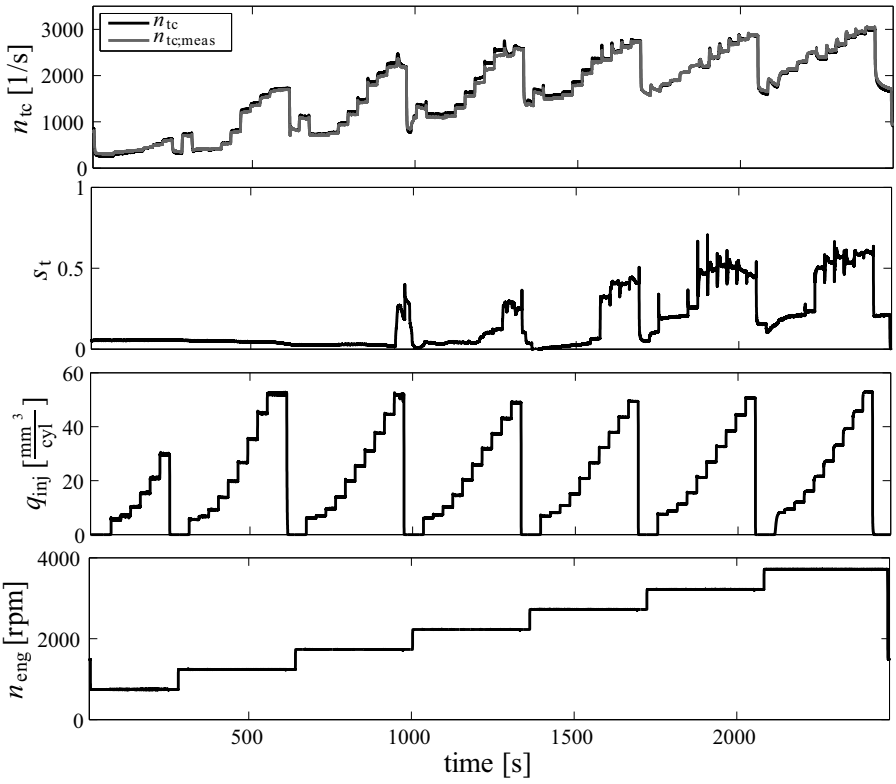


Fig. 5.4.6. Comparison of modeled and measured dynamic behavior of a turbocharger for combinations of changing injection masses and changing engine speeds.

5.4.2 Model-based symptom generation

The different nonlinear models of turbochargers allow to generate residuals by parity equations. Based on the equations of the powers and the speed following output residuals are specified as a first group

$$r_{\mathrm{P,c}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}) = P_{\mathrm{c}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}, \mathbf{z}_{\mathrm{c}}) - P_{\mathrm{c,ref}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}, \mathbf{z}_{\mathrm{c,ref}}) \tag{5.4.19}$$

$$r_{\mathrm{P,t}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}) = P_{\mathrm{t}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}, \mathbf{z}_{\mathrm{t}}) - P_{\mathrm{t,ref}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}, \mathbf{z}_{\mathrm{t,ref}}) \tag{5.4.20}$$

$$r_{\mathrm{P,f}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}) = P_{\mathrm{f}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}, \omega_{\mathrm{tc}}) - P_{\mathrm{f,ref}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}, \omega_{\mathrm{tc,ref}}) \tag{5.4.21}$$

$$r_{\mathrm{n,tc}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}) = n_{\mathrm{tc}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}, \mathbf{z}_{\mathrm{t}}, \mathbf{z}_{\mathrm{c}}) - n_{\mathrm{tc,ref}}(u_{\mathrm{inj}}, n_{\mathrm{eng}}, \mathbf{z}_{\mathrm{t,ref}}, \mathbf{z}_{\mathrm{c,ref}}). \tag{5.4.22}$$

For example, $P_{\mathrm{c}}(\mathbf{z}_{\mathrm{c}})$ is the calculated power according to (5.4.10) with the actual measured variables $\mathbf{z}_{\mathrm{c}}(t)$ under the influence of possible faults and $P_{\mathrm{c,ref}}(\mathbf{z}_{\mathrm{c}})$ results from the same model with the reference variables $\mathbf{z}_{\mathrm{c,ref}}$ under normal conditions, without faults, both at the same operation point $[u_{\mathrm{inj}}, n_{\mathrm{eng}}]$ and warm engine. n_{tc} follows from (5.4.2). Figure 5.4.2 shows the corresponding signal flow.

A next group of residuals is based on the pressures p_2, p_3 and p_4, see Sidorow et al (2011), Sidorow (2014). For example, a residual for the charging pressure is

$$r_{\mathrm{p2}}\left(u_{\mathrm{inj}}, n_{\mathrm{eng}}\right) = p_2\left(u_{\mathrm{inj}}, n_{\mathrm{eng}}, \mathbf{z}_{\mathrm{t}}, \mathbf{z}_{\mathrm{c}}\right) - p_{2,\mathrm{ref}}, \tag{5.4.23}$$

where p_2 is determined with (5.4.17). The resulting fault-symptom table is shown in Table. 5.4.1. It illustrates that with the turbocharger powers and speed and the charging pressure the faults can be isolated or a group of faults can be determined. However, the obtained symptoms depend also on the operating area $(M_{\mathrm{eng}}, n_{\mathrm{eng}})$. In Sidorow (2014) it was shown that different fault-symptom tables were obtained for six operating areas.

Table 5.4.1. Fault-symptom table for the VGT turbocharger, exhaust path and intake.

Faults		Symptoms						
		S_{Pt}	S_{Pc}	S_{Pf}	S_{n}	S_{p2}	S_{p3}	S_{p4}
Air path	restriction air filter	0	0	0	0	–	0	–
	blowby tube removed	0	0	–	–	–	0	–
	leakage intake 5 mm after intercooler	0	0	+	+	–	–	0
	leakage intake 7 mm after intercooler	+	+	+	+	–	–	0
	no cooling airflow of the intercooler	0	0	0	0	–	0	+
	restriction behind intercooler	0	0	–	–	–	0	–
Exhaust path	leakage exhaust	0	0	0	0	–	0	–
	HP-EGR valve blocked closed	–	–	–	–	–	+	0
Turbocharger	compressor blades damaged	0	0	0	0	–	0	0
	VGT blocked middle position	+	+	+	+	0	0	–

5.4.3 Wastegate turbocharger

If a turbocharger with wastegate actuation for charging pressure control is considered for the operation state of a *closed wastegate*, see Schwarte (2007), then the full exhaust gas flow streams through the turbine and simpler model equations can be used for fault detection. It is assumed that in the diesel engine as depicted in Fig. 5.4.1

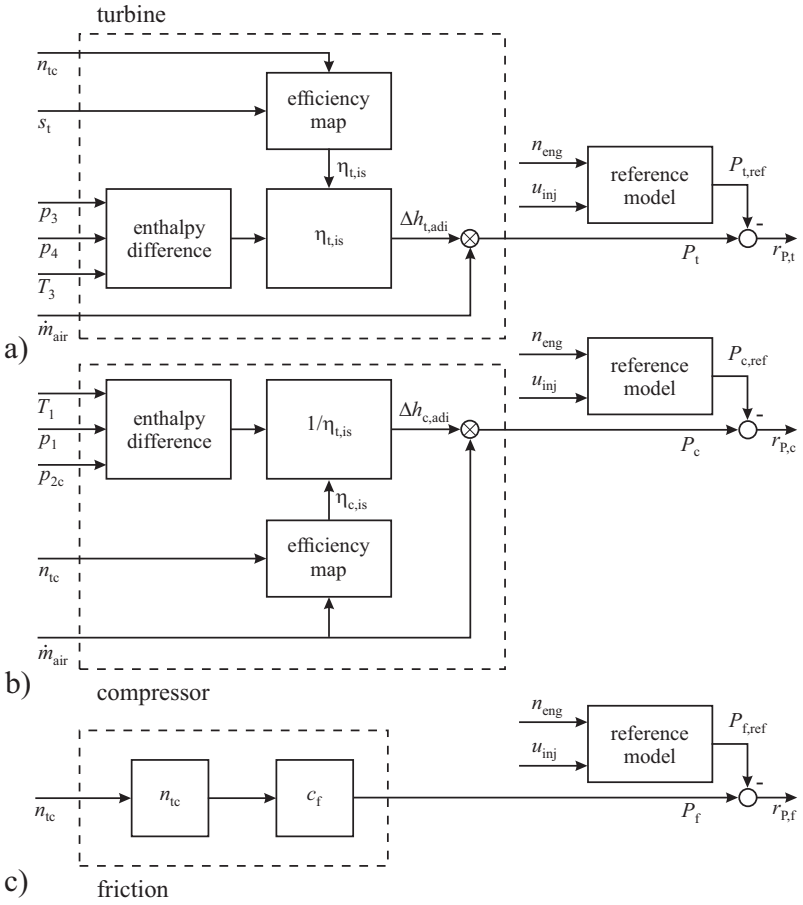


Fig. 5.4.7. Signal flow diagrams for the generation of some turbocharger output residuals for **a** turbine power, **b** compressor power, **c** friction power.

following measurements are available: air mass flow $\dot{m}_{\mathrm{c}} = \dot{m}_{\mathrm{air}}$, charging pressure $p_{2\mathrm{i}}$ and temperature $T_{2\mathrm{i}}$, pressure p_{egr} and temperature T_{egr} after the EGR cooler.

The dynamic behavior of the turbocharger speed follows from (5.4.2) by using (5.4.4) and (5.4.10) for the turbine power P_{t} and compressor power P_{c}. For closed wastegate $\dot{m}_{\mathrm{c}} = \dot{m}_{\mathrm{t}}$ can be assumed and (5.4.17) leads to

$$\frac{p_2}{p_1} = \left(1 + \frac{c_{\mathrm{pt}}}{c_{\mathrm{pc}}} \frac{T_3}{T_1} \eta_{\mathrm{c,is}} \eta_{\mathrm{t,is}} \eta_{\mathrm{m,tc}} \left(1 - \left(\frac{p_1}{p_3}\right)^{\frac{k-1}{k}}\right)\right)^{\frac{k}{k-1}} - \Delta p_{\mathrm{ic}} \tag{5.4.24}$$

where Δp_{ic} is the pressure loss in the charge air cooler

$$\Delta p_{\mathrm{ic}} = c_{\mathrm{r,ic}} \dot{m}_{\mathrm{c}}^2. \tag{5.4.25}$$

Further simplifying assumptions made are $p_1 = p_4 = p_a$, $T_1^* = T_1$, $T_3^* = T_3$. The exhaust temperature is calculated by

$$T_3 \approx T_{2\mathrm{i}} + k_1 + k_2 \lambda_\mathrm{m} \tag{5.4.26}$$

$$\lambda_\mathrm{m} = \frac{\dot{m}_\mathrm{f}}{\dot{m}_\mathrm{c} + \dot{m}_\mathrm{egr}}. \tag{5.4.27}$$

Hence, the mainly influencing variables for the charging pressure are p_3, T_3 and the efficiencies.

The simplified models are now used to define the inputs and outputs of LOLIMOT net models

$$p_{2\mathrm{i}} = f_{\mathrm{p2i}}\left(p_3, T_3, \lambda_\mathrm{m}\right) \tag{5.4.28}$$

$$p_3 = f_{\mathrm{p3}}\left(\dot{m}_\mathrm{t}, T_{2\mathrm{i}}, \lambda_\mathrm{m}\right) \tag{5.4.29}$$

where

$$T_{2\mathrm{i}} = \frac{T_{2\mathrm{ic}}\,\dot{m}_\mathrm{c} + T_\mathrm{egr}\,\dot{m}_\mathrm{egr}}{\dot{m}_\mathrm{c} + \dot{m}_\mathrm{egr}} \tag{5.4.30}$$

$$\dot{m}_\mathrm{c} = \dot{m}_1 - \frac{V_2}{R\,T_{2\mathrm{i}}}\,\frac{dp_{2\mathrm{i}}}{dt} \tag{5.4.31}$$

$$\dot{m}_\mathrm{egr} = \dot{V}_\mathrm{egr}\,\rho_\mathrm{egr} = \dot{V}_\mathrm{egr}\,\frac{p_\mathrm{egr}}{R\,T_\mathrm{egr}} \tag{5.4.32}$$

$$\dot{V}_\mathrm{egr} = \dot{V}_{2\mathrm{i}} - \dot{V}_\mathrm{c} = 0.5\,n_\mathrm{eng}\,V_\mathrm{D}\,\eta_\mathrm{v} - \frac{\dot{m}_\mathrm{c}}{\rho_{2\mathrm{i}}} \tag{5.4.33}$$

$$\dot{m}_\mathrm{t} = \dot{m}_\mathrm{c} + \dot{m}_\mathrm{f} \tag{5.4.34}$$

with η_v the volumetric efficiency coefficient, V_D the displacement of all cylinders and $\rho_{2\mathrm{i}} = p_{2\mathrm{i}}/R\,T_{2\mathrm{i}}$.

The pressure p_egr before the EGR valve differs from the exhaust pressure p_3 by the pressure losses in the EGR cooler. This is modeled by a further net model

$$p_\mathrm{egr} = f_\mathrm{pegr}(p_3, n_\mathrm{eng}). \tag{5.4.35}$$

Based on these models and the measured variables $p_{2\mathrm{i}}$, $T_{2\mathrm{i}}$ after the charge-air cooler and p_egr and T_egr after the EGR cooler, two residuals can be calculated for the case of a closed wastegate valve, the exhaust pressure residual

$$r_\mathrm{pegr} = p_\mathrm{egr} - f_\mathrm{pegr} = S_\mathrm{pegr} \tag{5.4.36}$$

and the charging pressure residual

$$r_{\mathrm{p2i}} = p_{2\mathrm{i}} - f_{\mathrm{p2i}} = S_{\mathrm{p2i}}. \tag{5.4.37}$$

The model for the normal behavior f_pegr follows from (5.4.35) and $f_{2\mathrm{i}}$ from (5.4.28). These symptoms show different deviations e.g. for decreased turbocharger efficiency or air mass sensor fault or leakage before the compressor as summarized later in Table 5.5.1. Because the intake system and the turbocharger-exhaust system are strongly coupled, the exhaust system will be considered further in the next section to yield some more symptoms.

5.5 Exhaust system

5.5.1 Analytical redundancies for air mass flow

It is assumed that two variables are measurable, the position u_{egr} of the EGR valve and the continuous excess air factor λ (broad-band sensor) respectively oxygen concentration ν_{O2}. These sensors in the exhaust system are standard sensors for modern diesel engines and allow to generate analytical redundancy to the air mass flow sensor in the intake system.

a) Dynamically corrected HFM air mass flow

Taking into account the intake volume V_{int} the real mass flow after the intercooler before mixing it with recirculated exhaust gas, Fig. 5.4.1, is

$$\dot{m}_{\mathrm{air1,eng}} = \dot{m}_2 - \frac{V_{\mathrm{int}}}{R\,T_{2\mathrm{ic}}}\,\frac{dp_{2\mathrm{ic}}}{dt} \tag{5.5.1}$$

where $p_{2\mathrm{ic}}$ and $T_{2\mathrm{ic}}$ are measured.

b) Charged air mass flow

The engine is charged by fresh air flow and exhaust gas flow. Therefore it holds for the volume flows

$$\begin{aligned}\dot{V}_{\mathrm{air,eng}} &= \dot{V}_{\mathrm{eng}} - \dot{V}_{\mathrm{egr,eng}} \\ &= 0.5\,\lambda_v\,n_{\mathrm{eng}}\,V_{\mathrm{D}} - \dot{m}_{\mathrm{egr}}\,\frac{R\,T_{\mathrm{egr}}}{p_{2\mathrm{ic}}}\end{aligned} \tag{5.5.2}$$

and the air mass flow becomes

$$\begin{aligned}\dot{m}_{\mathrm{air,2,eng}} &= \dot{V}_{\mathrm{air,eng}}\,\frac{p_{2\mathrm{ic}}}{R\,T_{2\mathrm{ic}}} \\ &= 0.5\,\lambda_v\,n_{\mathrm{eng}}\,V_{\mathrm{D}}\,\frac{p_{2\mathrm{ic}}}{R\,T_{2\mathrm{ic}}} - \dot{m}_{\mathrm{egr}}\,\frac{T_{\mathrm{egr}}}{T_{2\mathrm{ic}}}\end{aligned} \tag{5.5.3}$$

Based on an analysis of the throttling effects in the EGR valve leaning on Laval-nozzle-models a look-up table function $f_{\mathrm{egr}}\left(\frac{p_{2\mathrm{ic}}}{p_{\mathrm{egr}}}\right)$ was experimentally identified leading to the mass flow

$$\dot{m}_{\mathrm{egr}} = A_{\mathrm{egr}}(u_{\mathrm{egr}})\,f_{\mathrm{egr}}\left(\frac{p_{2\mathrm{ic}}}{p_{\mathrm{egr}}}\right). \tag{5.5.4}$$

where A_{egr} is the effective opening of the exhaust valve.

c) Air mass flow based on exhaust oxygen content

If the fuel flow $\dot{m}_f$ and its H/C-ratio is known, the air mass flow is for stationary behavior related by

$$\dot{m}_{\mathrm{air,3,eng}} = \frac{14.5 + 4.6\,\nu_{\mathrm{O2}}}{1 - 4.8\,\nu_{\mathrm{O2}}}\,\dot{m}_f \tag{5.5.5}$$

see Sect. 5.2.

Now three residuals can be determined

$$\left.\begin{aligned} r_{\mathrm{a1,2}} &= \dot{m}_{\mathrm{air1}} - \dot{m}_{\mathrm{air2}} = S_{\mathrm{a1,2}} \\ r_{\mathrm{a2,3}} &= \dot{m}_{\mathrm{air2}} - \dot{m}_{\mathrm{air3}} = S_{\mathrm{a2,3}} \\ r_{\mathrm{a1,3}} &= \dot{m}_{\mathrm{air1}} - \dot{m}_{\mathrm{air3}} = S_{\mathrm{a1,3}} \end{aligned}\right\} \tag{5.5.6}$$

(An alternative is to normalize these residuals by dividing them through the two averaged air flows)

Figure 5.5.1 shows the reaction of 3 differently determined air mass flows. Without faults they agree well. However, for a 5 mm diameter leakage in the intake after the intercooler the calculated flows $\dot{m}_{\mathrm{air2}}$ and $\dot{m}_{\mathrm{air3}}$ deviate from $\dot{m}_{\mathrm{air1}}$ and indicate the smaller air flow mass into the cylinders.

This means also that because of the three independent determinations for the air mass flow a 2-out-of-3 voter can be established to form a fault-tolerant air mass flow sensor system, see Sect. 8.5.4.

5.5.2 Combined fault detection for wastegate turbocharger and air mass flow

If symptoms of the wastegate turbocharger module from Sect. 5.4.3 and a part of the exhaust module for determining the air mass flow are joined together, the fault-symptom table, Table 5.5.1 results. Note however, that the two turbocharger symptoms can only be used if the wastegate is closed, i.e. at loads with medium to high torque and low to medium speed. All considered faults show different patterns except air mass flow sensor flow too low (sensor fault) and leakage between air mass flow sensor and compressor. All other faults result at least in one differing sign (weakly isolating). Hence, by combining the symptoms of the different fault-detection modules, fault diagnosis and locating faults is considerably improved.

5.5.3 Particulate filter and catalyst

The supervision of *particulate filters* (PF) is integrated in the regeneration control. The loading state is usually observed by the pressure drop over the PF

$$\Delta p_{\mathrm{PF}} = f(\dot{V}_{\mathrm{eg}}) \tag{5.5.7}$$

where $\dot{V}_{\mathrm{eg}}$ is the exhaust gas volume flow. The pressure drop

$$\Delta\Delta p_{\mathrm{PF}} = \Delta_{\mathrm{PF}}(\dot{V}_{\mathrm{eg}}) - \Delta p_{\mathrm{Pf,empty}}(\dot{V}_{\mathrm{eg}}) \tag{5.5.8}$$

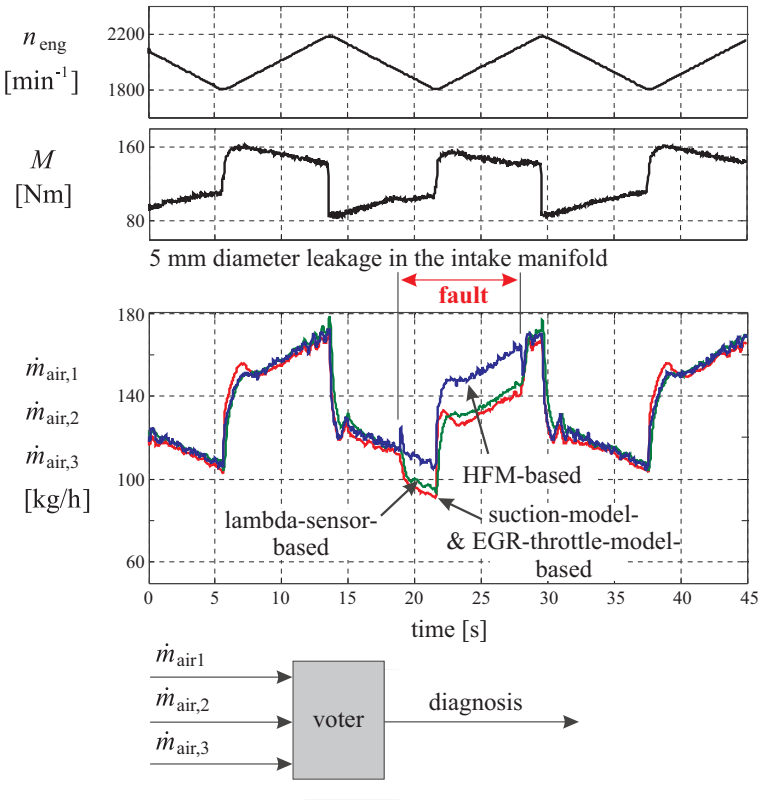


Fig. 5.5.1. Signals of three independent methods for air mass flow reconstruction during dynamic engine operation. A leakage of 5 mm diameter in the intake manifold after the intercooler happens for 18 s < t < 29 s.

Table 5.5.1. Fault-symptom table for combined turbocharger and exhaust/airmass detection module. Wastegate closed. Symptoms: + positive; − negative; 0 no reaction.

faults	turbocharger symptoms WG closed		exhaust/airmass symptoms full operation range		
	S_{p2i}	S_{pegr}	$S_{a1,2}$	$S_{a2,3}$	$S_{a1,3}$
decreased turbo-charger efficiency	−	0	0	0	0
leaky wastegate, leaky exhaust pipe	−	−	0	0	0
air mass flow too high	0	−	+	0	−
air mass flow too low	0	+	−	0	−
leakage between compressor and engine	−	−	+	0	+
leakage between air mass flow sensor and compressor	0	+	−	0	−
restriction after turbine (cat, filter)	−/ 0	−	0	0	0
EGR too high (faulty valve)	0	0	−	+	0

compared with that of an empty PF serves as an indicator for the loading state. It increases with the operating time and if it exceeds a certain threshold the PF is assumed to be fully loaded. The start of regeneration is usually based on a combined evaluation of the pressure drop, the calculated accumulated particulate mass based on a loading model and the driven distance. For more details see Isermann (2014), Sect. 8.9 and the cited references.

Supervision and fault detection of an NO_x-*storage catalyst* (lean NO_x trap, LNT) to reduce NO_x is also based on a loading and regeneration phase. The control of this accumulator type catalyst distinguishes two phases. In normal lean burn operation $\lambda > 1$ NO is oxidized to NO_2 and then adheres in form of nitrate (NO_3) to a metal oxide in the loading phase. Then a regeneration phase follows with rich exhaust gas $\lambda < 1$ by retarded injection and air throttling. The loading phase lasts 30 to 300 s and the regeneration phase 2 to 10 s. End of the loading phase can be detected by a NO_x sensor or λ-sensor after the catalyst and by a loading model.

The fault detection of NO_x storage catalysts depends on the available measurements. Usually only output signals are available as an NO_x or λ-sensor. The amplitudes and time periods of the loading and regeneration phases can then be used as features.

5.6 Overall diesel engine fault diagnosis

The described model-based fault-detection and diagnosis methods followed the component-oriented detection modules depicted in Fig. 5.0.1. Table 5.6.1 shows the fault-symptom table for four engine parts. It is an extract of the Tables 5.1.1, 5.2.1, 5.3.2 and 5.4.1. The generated symptoms are mainly based on the available measurements for the considered corresponding components. Most of the symptoms are related to the respective process parts. Sometimes also measurements from other components were used. The patterns of the symptoms are in many cases different, except in some cases with same effects to the symptoms. For some components the generated symptoms are also influenced by faults in other components. For example, the symptoms of the intake system A1 influence also the symptoms of the VGT turbocharger A4 for the case of intake system faults. The overall fault-symptom table can be used to extend the fault-symptom relations and may improve the fault-diagnosis coverage and, thus, improve an overall engine diagnosis.

References

Antory D, Kruger U, Irwin G, McCullough G (2005) Fault diagnosis in internal combustion engines using non-linear multivariate statistics. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 219(4):243–258

Barta J, Hampson G (2016) Real-time combustion diagnostics and control for improved engine management. MTZ industrial 6(1):26–31

Table 5.6.1. Overall model-based fault-symptom table of a diesel engine with distributor pump or common-rail injection (extract)(+ positive; − negative; 0 no change).

Engine		A1: intake system			A2/A3: fuel & combustion						A3: common rail & injection				A4: VGT turbocharger				
process	symptoms	$S_{\eta v}$	$S_{A\dot{m}}$	S_{Ap}	S_{MME}	S_{ER}	S_{ERCi}	S_{MF}	$S_{\Delta MF}$	S_{MFCi}	$S_{inj,1}$	$S_{inj,2}$	$S_{inj,3}$	$S_{inj,4}$	S_{Pt}	S_{Pc}	S_{Pf}	S_{n}	S_{p2}
part	residuals	$r_{\eta v}$	$r_{A\dot{m}}$	r_{Ap}	r_{MME}	r_{ER}	r_{ERCi}	r_{MF}	$r_{\Delta MF}$	r_{MFCi}	$r_{inj,1}$	$r_{inj,2}$	$r_{inj,3}$	$r_{inj,4}$	r_{Pt}	r_{Pc}	r_{Pf}	r_{n}	r_{p2}
	faults																		
intake system	removed crank case tube	+	0	0											0	0	-	-	-
	leakage between intercooler & block	-	0	0											0	0	+	+	-
	restriction after intercooler	0	-	-											0	0	-	-	-
	swirl flaps closed	0	-	-															
	EGR open	++	+	+															
	EGR valve leaky	+	0	0															
distributor pump injection & combustion	pump pressure too low				-	0	0	+	0	0									
	clogging of nozzles				0	+	-	0	+	-									
	nozzle needle stuck open				0	+	-	0	+	+									
	nozzle needle stuck closed				0	++	–	0	++	–									
	cylinder mechanics				0	+	-	0	0	0									
common-rail injection	low delivery at pump piston										0	+	0	+					
	reduced injection, one injector										0	0	+	0					
	pressure sensor too high										+	0	0	+					
VTG-turbo charger & intake & exhaust	air filter restriction														0	0	0	0	-
	no cooling intercooler														0	0	0	0	-
	VGT stuck middle position														+	+	+	+	0
	leakage exhaust														0	0	0	0	-
	compressor blades damaged														0	0	0	0	-

Clever S (2010) Modellgestützte Fehlererkennung und Diagnose für Common-Rail-Einspritzsysteme. In: Isermann R (ed) Elektronisches Management motorischer Fahrzeugantriebe, Vieweg-Teubner, Wiesbaden

Clever S (2011) Model-based fault detection and diagnosis for a common rail engine. Diss. Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, 1202. VDI Verlag, Düsseldorf

Clever S, Isermann R (2008) Signal- and process-model-based fault detection and diagnosis of a common rail injection system. In: FISITA 2008 World Automotive Congress, München

Clever S, Isermann R (2010) Model-based fault detection and diagnosis for common-rail injection systems. MTZ worldwide 71(2):32–39

Eck C, Sidorow A, Konigorski U, Isermann R (2011) Fault detection for common rail diesel engines with low and high pressure exhaust gas recirculation. In: 10th International Conference on Engines & Vehicles (ICE 2011), Naples, Italy

Fischer G (2000) Expertenmodell zur Berechnung der Reibungsverluste von Ottomotoren. Dissertation Technische Universität Darmstadt. dissertation.de, Berlin

Führer J, Sinsel S, Isermann R (1993) Erkennung von Zündaussetzern aus Drehzahlsignalen mit Hilfe eines Frequenzbereichsverfahrens. In: Proceedings 13. Tagung Elektronik im Kraftfahrzeug, Haus der Technik Essen, Germany

Guzzella L, Onder C (2010) Introduction to modeling and control of internal combustion engine systems, 2nd edn. Springer, Berlin

Heywood JB (1988) Internal combustion engine fundamentals. McGraw-Hill, New York

Isermann R (2014) Engine Modeling and Control. Springer, Berlin

Isermann R, Münchhof M (2011) Identification of dynamic systems. Springer, Berlin, Heidelberg

Kimmich F (2004) Modellbasierte Fehlererkennung und Diagnose der Einspritzung und Verbrennung von Dieselmotoren. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 12, 549. VDI Verlag, Düsseldorf

Kimmich F, Isermann R (2002) Model-based fault detection for the injection, combustion and engine-transmission. In: 15th IFAC World Congress, Barcelona, Spanien

Kimmich F, Schwarte A, Isermann R (2001) Model-based fault detection for diesel engines. In: Aachen Colloquium, Automobile and Engine Technology, Aachen, Germany

Kimmich F, Schwarte A, Isermann R (2005) Fault detection for modern diesel engines using signal- and process model-based methods. Control Engineering Practice – CEP 13(2):189–203

Leonhardt S (1996) Modellgestützte Fehlererkennung mit neuronalen Netzen - Überwachung von Radaufhängungen und Diesel-Einspritzanlagen. Dissertation Technische Hochschule Darmstadt. Fortschr.-Ber. VDI Reihe 12, 295. VDI Verlag, Düsseldorf

Pischinger S, Heywood JB (1989) A study of flame development and engine performance with breakdown ignition systems in a visualization engine. SAE transactions 97:858–876

Robert Bosch GmbH (ed) (2005) Diesel-Engine Management, 4th edn. Bentley publishers, Cambridge

Schmidt M, Kimmich F, Straky H, Isermann R (2000) Combustion supervision by evaluating the crankshaft speed and acceleration. In: SAE 2000 World Congress, Detroit, Michigan, USA

Schwarte A (2007) Modellbasierte Fehlererkennung und Diagnose des Ansaug-und Abgassystems von Dieselmotoren. Dissertation Technische Universität Darmstadt 2006. Fortschr.-Ber. VDI Reihe 8, 634. VDI Verlag, Düsseldorf

Schwarte A, Kimmich F, Isermann R (2002) Model-based fault detection and diagnosis for diesel engines. MTZ worldwide 63(7–8)

Schwarte A, Hack L, Isermann R, Nitzke HG, Jeschke J, Piewek J (2004) Automatisierte Applikation von Motorsteuergeräten mit kontinuierlicher Motorvermessung. In: AUTOREG 2004, VDI, Wiesloch, Germany, vol VDI-Berichte 1828

Sidorow A (2014) Model-based fault diagnosis of the air path and turbocharger of a diesel engine. Diss. Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, 1234. VDI Verlag, Düsseldorf

Sidorow A, Isermann R, Cianflone F, Landsmann G (2011) Model-based fault detection of the air and exhaust path of diesel engines including turbocharger models. sae 2011-01-0700. In: SAE World Congress, Detroit, MI, USA

Töpfer S, Wolfram A, Isermann R (2002) Semi-physical modelling of nonlinear processes by means of local model approaches. In: Proc. 15th IFAC World Congress, Barcelona, Spain

Zahn S (2012) Arbeitsspielaufgelöste Modellbildung und Hardware-in-the-Loop-Simulation von Pkw-Dieselmotoren mit Abgasturbolader. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 12, 760. VDI Verlag, Düsseldorf

Part III

Diagnosis of Electric Drives, Motors and Actuators

6

Diagnosis of electric motors

The increasing electrification of the powertrain, the chassis and the body means a strong enlargement of electric drives and electric actuators, as discussed in Chap. 1. Important electric motors at the *internal combustion engine* are, for example, the *auxiliary drives*, like the electrical generator, starter and cooling fan and several *electric actuators*, like for the electrical throttle, the exhaust-gas recirculation valve and the variable geometry turbocharger. For hybrid and electrical vehicles the *electric drives* are part of the powertrain.

Examples for electric motors in the *chassis* are the *electric power steering* or *electro-mechanical brakes*. Many smaller electric motors are parts of the body, like windshield wipers, window drives, and electrical seat adjustment devices. Thus, the overall functions of a vehicle depends highly on the quality and reliability of electric motors. Therefore, the automatic supervision and fault diagnosis of their present and also future state is an important issue.

The most important types of electric motors can generally be divided into:

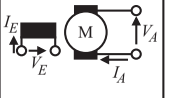
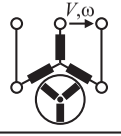
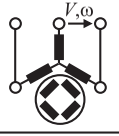
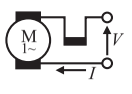
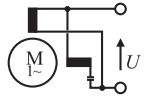
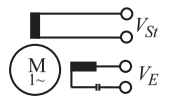
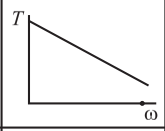
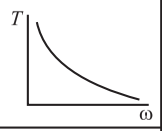
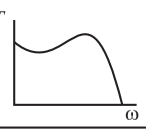
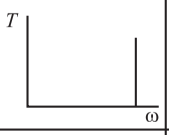
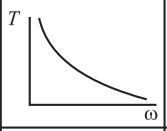
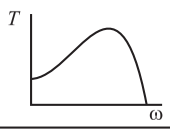
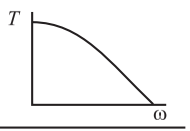
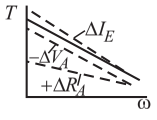
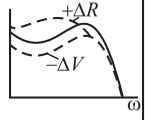
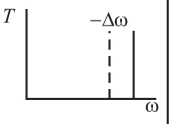
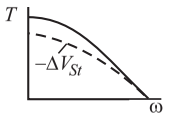
1. DC motors
   - series-wound motors
   - shunt-wound motors
   - permanent-field motors
2. Three-phase AC motors
   - induction motors (asynchronous motors)
   - synchronous motors
3. Single-phase AC motors
   - commutator motors (universal motors)
   - squirrel-cage motors.

For vehicles mostly DC motors with and without brushes are applied for low power and three-phase motors in form of synchronous motors or asynchronous motors for higher power applications. Table 6.0.1 gives an overview of some basic types, illustrating torque characteristics and corresponding control inputs.

As static and dynamic models of the various electric motors are required for model-based fault detection, the reader is referred to well-known basic books on

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7_6

Table 6.0.1. General survey of electric motors with small power, Isermann (2005)

motor		DC shunt-wound motor	DC series-wound motor	three-phase asynchronous motor (AC)	three-phase permanent magnet synchronous motor (PMSM)	single-phase (universal motor)	single-phase asynchronous motor with condensator	single-phase asynchronous motor (Ferraris motor)
circuit diagram								
torque–speed characteristics								
torque characteristics for manipulation	different manipul. variables:							
	normal:							
	manipulated variables	ΔV_A armature voltage ΔI_E excitation current ΔR_A armature resistance	ΔV voltage	ΔV voltage $\Delta\omega$ frequency ΔR rotor resistance	$\Delta\omega$ frequency	ΔV voltage ΔR armature resistance		ΔV_{St} manipulated voltage

electric drives such as Fraser and Milne (1994), Leonhard (1996), Sarma (1996), Schröder (1995), Schröder (2009), Stölting (2004).

In the following model-based fault detection and diagnosis is described for direct-current motors (DC), alternating-current asynchronous motors (AC), and alternating-current synchronous motors (SM). First, the basic equations as required for model-based control are compiled and then the fault-detection methods are developed and demonstrated by experiments. These sections are shortened versions of chapters in Isermann (2011).

6.1 Direct-current motor (DC)

6.1.1 Models of a DC motor with brushes

A permanently excited DC motor with a rated power of $P = 550\,\mathrm{W}$ at rated speed $n = 2500\,\mathrm{rpm}$ is considered, Höfling (1996). This DC motor has a two-pair brush commutation, two pole pairs, and an analog tachometer for speed measurement; it operates against a hysteresis brake as load, see Fig. 6.1.1. The measured signals are the armature voltage U_A, the armature current I_A and the speed ω. A servo amplifier with pulse-width-modulated armature voltage as output and speed and armature current as feedback allows a cascaded speed control system. The three measured signals first pass analog anti-aliasing filters and are processed by a digital signal processor (TXP 32 CP, 32-bit fpt, 50 MHz) and an Intel Pentium host PC. Also the hysteresis brake is controlled by a pulse-width servo amplifier. Usually, such DC motors can be described by linear dynamic models.

However, experiments have shown that this model with constant parameters does not match the process in the whole operational range. Therefore, two nonlinearities are included so that the model fits the process better. The resulting first-order differential equations are:

$$L_\mathrm{A}\dot{I}_\mathrm{A}(t) = -R_\mathrm{A}I_\mathrm{A}(t) - \Psi\omega(t) - K_\mathrm{B}|\omega(t)|I_\mathrm{A}(t) + U_\mathrm{A}^*(t) \tag{6.1.1}$$

$$J\dot{\omega} = \Psi I_\mathrm{A}(t) - M_{\mathrm{F}1}\omega(t) - M_{\mathrm{F}0}\ \mathrm{sign}\ (\omega(t)) - M_\mathrm{L}(t) \tag{6.1.2}$$

Figure 6.1.2 depicts the resulting signal flow diagram. The term $K_\mathrm{B}|\omega(t)|I_\mathrm{A}(t)$ compensates for the voltage drop at the brushes in combination with a pulse-width-modulated power supply. The friction is included by a viscous- and a dry-friction term $M_{\mathrm{F}1}\omega$ and $M_{\mathrm{F}0}\mathrm{sign}(\omega)$, see also Isermann (2005). The parameters are identified by least-squares estimation in the continuous-time domain, Höfling (1996). Table 6.1.1 gives the nominal values. Most of them ($R_\mathrm{A}, \Psi, K_\mathrm{B}, M_{\mathrm{F}1}, M_{\mathrm{F}0}$) influence the process gain, and the other two (L_A, J) the time constants. The signals U_A^*, I_A and ω are measured with a sampling frequency of 5 kHz, and state-variable filtered by a fourth-order low-pass filter with Butterworth characteristic and a cut-off frequency of 250 Hz.

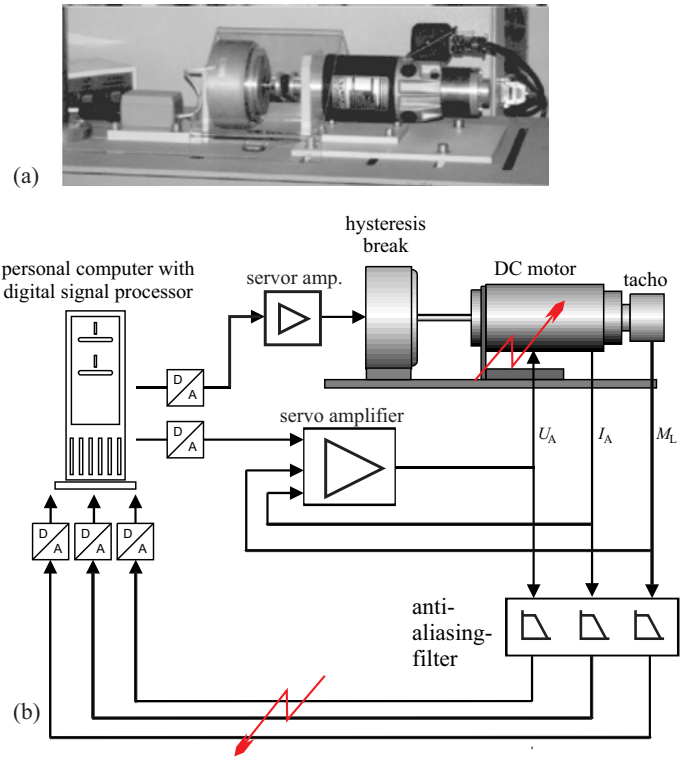


Fig. 6.1.1. DC motor test bench with hysteresis brake. **a** test bench. **b** scheme of equipment.

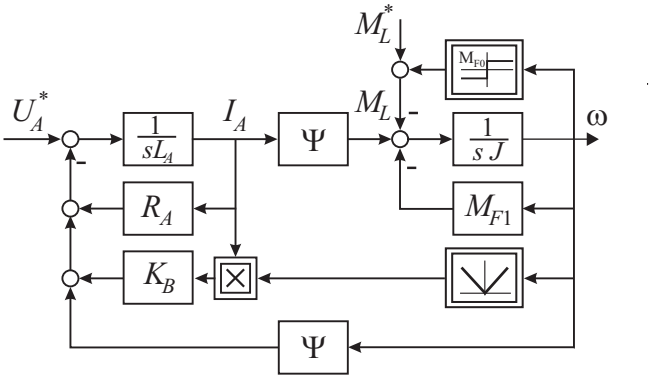


Fig. 6.1.2. Signal flow diagram of the considered DC motor.

Table 6.1.1. Data for the DC motor

armature resistance	$R_\mathrm{A} = 1.52\ \Omega$
armature inductance	$L_\mathrm{A} = 6.82 \cdot 10^{-3}\ \Omega$ s
magnetic flux	$\Psi = 0.33$ V s
voltage drop factor	$K_\mathrm{B} = 2.21 \cdot 10^{-3}$ V s / A
inertia constant	$J = 1.92 \cdot 10^{-3}$ kg m^2
viscous friction	$M_{\mathrm{F}1} = 0.36 \cdot 10^{-3}$ Nm s
dry friction	$M_{\mathrm{F}0} = 0.11$ Nm

6.1.2 Fault detection with parity equations

For the detection and isolation of sensor (output) and actuator (input) faults a set of structured parity equations with state-space models is applied.

As the differential equations (6.1.1) and (6.1.2) are nonlinear, the design procedure for a linear parity space cannot be applied directly. But defining $U_\mathrm{A}^* - K_\mathrm{B}|\omega(t)|I_\mathrm{A}$ as voltage input U_A and as load input $M_\mathrm{L} = M_{\mathrm{F}0}\,\mathrm{sign}\,\omega$ leads to a linear description. The linear state-space representation then becomes

$$\begin{aligned}\dot{\mathbf{x}} &= \begin{bmatrix} \dot{I}_\mathrm{A} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} -\frac{R_\mathrm{A}}{L_\mathrm{A}} & -\frac{\Psi}{L_\mathrm{A}} \\ \frac{\Psi}{J} & -\frac{M_\mathrm{F}}{J} \end{bmatrix} \begin{bmatrix} I_\mathrm{A} \\ \omega \end{bmatrix} + \begin{bmatrix} \frac{1}{L_\mathrm{A}} & 0 \\ 0 & -\frac{1}{J} \end{bmatrix} \begin{bmatrix} U_\mathrm{A} \\ M_\mathrm{L} \end{bmatrix} \\ \mathbf{y} &= \begin{bmatrix} I_\mathrm{A} \\ \omega \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x} \end{aligned} \tag{6.1.3}$$

A corresponding signal flow diagram is depicted in Fig. 6.1.2.

An observability test reveals that both outputs (I_A and ω) can also observe each other. This is a precondition for a parity space of full order (here: 2). Then, $\mathbf{W}$, the design matrix in Equation (10.52) in Isermann (2006) is chosen such that a set of *structured residuals* is obtained, where residual $r_1(t)$ is independent of $M_\mathrm{L}(t)$, $r_2(t)$ of $U_\mathrm{A}(t)$, $r_3(t)$ of $\omega(t)$ and $r_4(t)$ of $I_\mathrm{A}(t)$, see also Höfling (1996), Pfeufer (1994), Füssel (2002):

$$\mathbf{W} = \begin{bmatrix} R_\mathrm{A} & \Psi & L_\mathrm{A} & 0 & 0 & 0 \\ -\Psi & M_{\mathrm{F}1} & 0 & J & 0 & \\ \alpha & 0 & \beta & 0 & JL_A & 0 \\ 0 & \alpha & 0 & \beta & o & JL_A \end{bmatrix} \tag{6.1.4}$$

with $\alpha = \Psi^2 + R_\mathrm{A}\ M_{\mathrm{F}1}$;
$\beta = L_\mathrm{A}\ M_{\mathrm{F}1} + J\ R_\mathrm{A}$.
The residuals, using three measured signals, then follow as:

$$\begin{aligned} r_1(t) &= L_\mathrm{A}\ \dot{I}_\mathrm{A}(t) + R_\mathrm{A}\ I_\mathrm{A}(t) + \Psi\ \omega(t) - U_\mathrm{A}(t) \\ r_2(t) &= J\ \dot{\omega}(t) - \Psi\ I_\mathrm{A}(t) + M_{\mathrm{F}1}\ \omega(t) + M_\mathrm{L}(t) \\ r_3(t) &= J\ L_\mathrm{A}\ \ddot{I}_\mathrm{A}(t) + (L_\mathrm{A}\ M_{\mathrm{F}1} + J\ R_\mathrm{A})\ \dot{I}_\mathrm{A}(t) \\ &\quad + (\Psi^2 + R_\mathrm{A} M_{\mathrm{F}1}) I_\mathrm{A}(t) - J\ \dot{U}_\mathrm{A}(t) - M_{\mathrm{F}1}\ U_\mathrm{A}(t) - \Psi\ M_\mathrm{L}(t) \\ r_4(t) &= J\ L_\mathrm{A}\ \ddot{\omega}(t) + (L_\mathrm{A}\ M_{\mathrm{F}1} + J\ R_\mathrm{A})\ \dot{\omega}(t) + (\Psi^2 + R_\mathrm{A}\ M_{\mathrm{F}1})\ \omega(t) \\ &\quad - \Psi\ U_\mathrm{A}(t) + L_\mathrm{A}\ \dot{M}_\mathrm{L}(t) + R_\mathrm{A}\ M_\mathrm{L}(t). \end{aligned} \tag{6.1.5}$$

The same residual equations can be also obtained via transfer functions as described in Example 10.3 in Isermann (2006). If an additive fault of the measured signals and of M_L occurs, all residuals except the decoupled one are deflected. The scheme of the structured residuals is not touched by the compensation for the nonlinear voltage drop of the brushes, as its magnitude is small enough. Two parameters R_A and M_F1, however, depend on the present motor temperature. The behavior of R_A and its effect on residual r_1 is depicted in Figure 6.1.3. Therefore, the use of adaptive parity equations improves the residual performance, see Höfling (1996) and Isermann (2006).

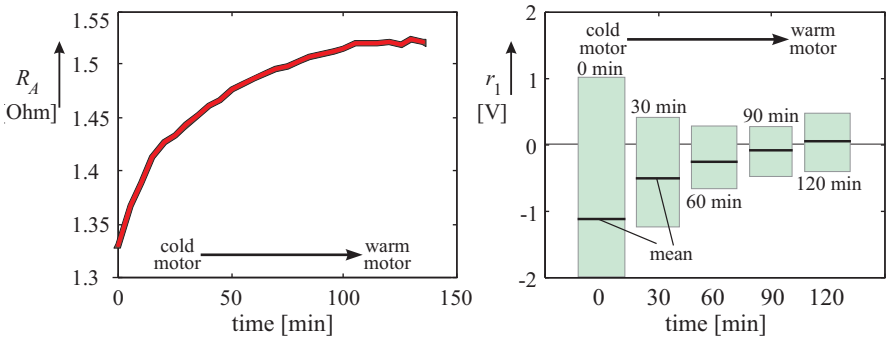


Fig. 6.1.3. Influence of the motor temperature on resistance R_A and residual r_1.

The residuals are now examined with regard to their sensitivity to additive and parametric faults. As r_1 and r_2 comprise all parameters and all signals, it is sufficient to consider only these two, although r_3 or r_4 can also be taken. From (6.1.5) it yields

$$\begin{aligned} r_1(t) &= \Delta\, L_\mathrm{A}\, \dot{I}_\mathrm{A}(t) + \Delta\, R_\mathrm{A}\, I_\mathrm{A}(t) + \Delta\, \Psi\, \omega(t) \\ &\quad + L_\mathrm{A}\, \Delta\, \dot{I}_\mathrm{A}(t) + R_\mathrm{A}\, \Delta\, I_\mathrm{A}(t) + \Psi\, \Delta\, \omega(t) - \Delta\, U_\mathrm{A}(t) \\ r_2(t) &= +\Delta\, J\, \dot{\omega}(t) - \Delta\, \Psi\, I_\mathrm{A}(t) + \Delta\, M_\mathrm{F1}\, \omega(t) \\ &\quad + J\, \Delta\, \dot{\omega}(t) - \Psi\, \Delta\, I_\mathrm{A}(t) + M_\mathrm{F1}\, \Delta\, \omega(t) + \Delta\, M_\mathrm{L}(t). \end{aligned} \tag{6.1.6}$$

In the presence of residual noise, e.g. of r_1 with a magnitude of about 1 V and an armature current of 3 A, a resistance change must be at least 0.3 Ω in order to deflect the residual significantly. Therefore, the two linear parameters R_A and M_F1 are selected to be tracked according to a single parameter estimation together with parity equations, as described in Isermann (2006), Sect. 10.5. The forgetting factor is chosen as $\lambda = 0.99$.

6.1.3 Fault detection with parameter estimation

The parameter estimation is based on the two differential equations (6.1.1) and (6.1.2) in the simplified form

$$\dot{I}_{\mathrm{A}}(t) = -\hat{\theta}_1\, I_{\mathrm{A}}(t) - \hat{\theta}_2\, \omega(t) + \hat{\theta}_3\, U_{\mathrm{A}}(t) \tag{6.1.7}$$

$$\dot{\omega}(t) = \hat{\theta}_4\, I_{\mathrm{A}}(t) - \hat{\theta}_5\, \omega(t) - \hat{\theta}_6\, M_{\mathrm{L}}(t) \tag{6.1.8}$$

with the process coefficients

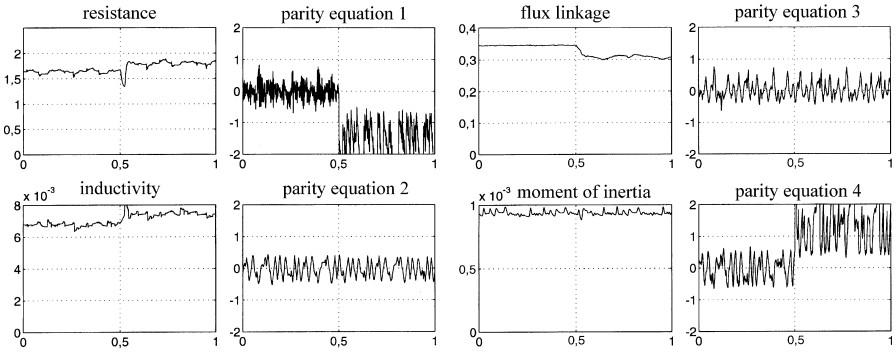
$$R_{\mathrm{A}} = \frac{\hat{\theta}_1}{\hat{\theta}_3};\ L_A = \frac{1}{\hat{\theta}_3};\ \Psi = \frac{\hat{\theta}_2}{\hat{\theta}_3} \text{ and } \Psi = \frac{\hat{\theta}_4}{\hat{\theta}_6};\ J = \frac{1}{\hat{\theta}_6};\ M_{\mathrm{F1}} = \frac{\hat{\theta}_5}{\hat{\theta}_6}. \tag{6.1.9}$$

Applying the recursive parameter estimation method DSFI (discrete square-root filtering in information form), Isermann and Münchhof (2011), with forgetting factor $\lambda = 0.99$ yields the parameters $\hat{\theta}_{\mathrm{i}}$ by using three measured signals. Then, all process coefficients can be calculated with (6.1.9). Experimental results with idle running ($M_{\mathrm{L}} = 0$) resulted in standard deviations of the process coefficients in the range of $2\% < \sigma_\theta < 6.5\%$, Höfling (1996).

6.1.4 Experimental results for fault detection

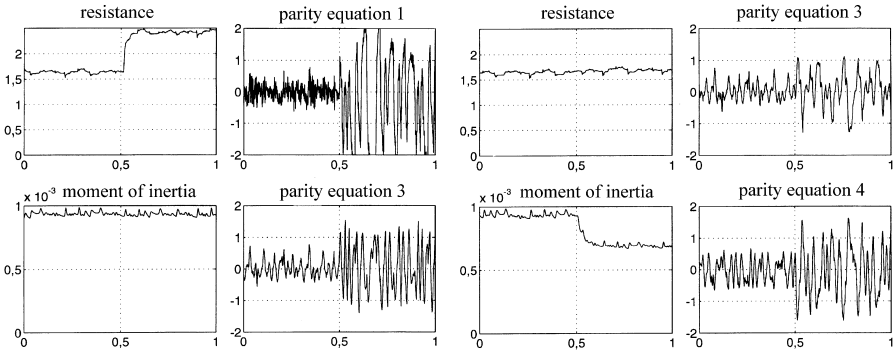
Based on many test runs, five different faults are now selected to show the detection of additive and multiplicative faults with parity equations and recursive parameter estimation, Höfling (1994). The time histories depict the arising faults at $t = 0.5\,\mathrm{s}$. The faults are step changes and were artificially produced. Figure 6.1.4 f) shows the parameter estimates and the residuals of parity equations. The residuals are normalized by division through their thresholds. Therefore, exceeding of 1 or -1 indicates the detection of a fault. In the cases a) to d) and f) the DC motor is excited by a pseudo random binary signal (PRBS) of the armature voltage U_{A} which is a requirement for dynamic parameter estimation, as shown in Fig. 6.1.4 f). In case e) the input is constant. The results can be summarized as:

a) A sensor-gain fault of the voltage sensor U_{A} leads as expected to a change of residual 1 (and 3, 4) but not of residual 2, which is independent of U_{A}. The parameter estimates show (incorrect) changes for R_{A}, L_{A} and Ψ, because the gain of the voltage sensor is not modeled.
b) An offset fault in the speed sensor ω leads to a change of the residuals r_4, r_1 and r_2, but r_3 remains uneffected, because it is independent of ω. The parameter estimate of Ψ shows an (incorrect) change.
c) A multiplicative change of the armature resistance R_{A} yields a corresponding change of the parameter estimate $\hat{R}_{\mathrm{A}}$. However, the residuals increase their variance drastically and exceed their thresholds.
d) A change of the ratio of inertia is correctly given by the parameter estimate $\hat{J}$. But all residuals, except r_1, exceed their thresholds by increasing their variance.
e) The same fault in R_{A} as in c) is introduced, but the input U_{A} is kept constant. The parameter estimate $\hat{R}_{\mathrm{A}}$ does not converge to a constant value and the parity equation residuals r_1 and r_4 change their mean, however, with large variance.
f) A brush fault leads to an increase of R_{A} and L_{A} but not of Ψ. The residuals show an increase of the variance.



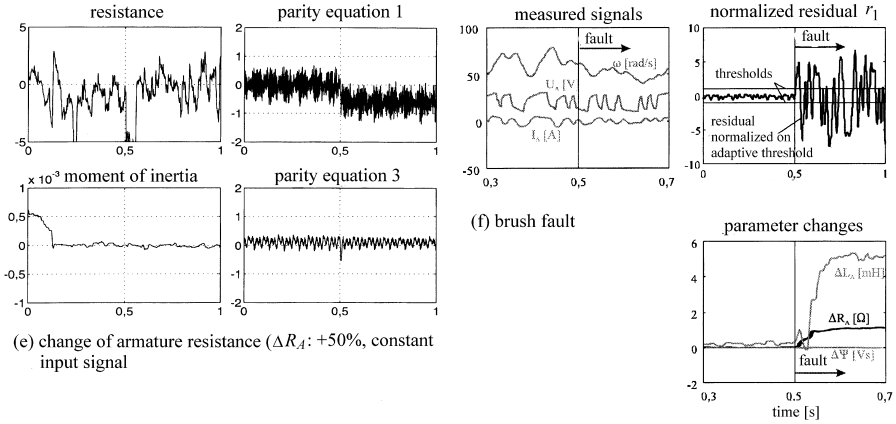
(a) sensor-gain fault (ΔU_A: +10%)

(b) DC value sensor fault ($\Delta\omega$: 10%)



(c) change of armature resistance (ΔR_A: +50%)

(d) change of moment of interia (ΔJ: -30%)



(e) change of armature resistance (ΔR_A: +50%, constant input signal

(f) brush fault

Fig. 6.1.4. Time histories of signals, residuals of parity equations and parameter estimation at fault occurrence

- parameter estimates: R_A resistance, L_A inductivity, Ψ flux linkage, J_A moment of inertia
- parity equations: r_1, r_2, r_3, r_4.

Table 6.1.2. Fault-symptom table for the fault detection of a DC motor with dynamic input excitation $U_\mathrm{A}(t)$ in the form of a PRBS. $+$ positive deflection; $++$ strong positive deflection; 0 no deflection; $-$ negative deflection; $--$ strong negative deflection; $\pm$ increased variance

Faults			Symptoms								
			parameter estimation					parity equations			
			R_A	L_A	Ψ	J	M_F1	r_1	r_2	r_3	r_4
parametric faults	armature resistance	ΔR_A	$++$	0	0	0	0	$\pm$	0	$\pm$	$\pm$
	brush fault		$++$	$+$	0	0	0	$\pm$	0	$\pm$	$\pm$
	change of inertia	ΔJ	0	0	0	$++$	0	0	$\pm$	$\pm$	$\pm$
	change of friction	ΔM_F1	0	0	0	0	$++$	0	$\pm$	$\pm$	$\pm$
	voltage sensor gain fault	ΔU_A	$\pm$	$\pm$	$\pm$	0	0	$-$	0	$-$	$-$
additive faults	speed sensor offset fault	$\Delta\omega$	0	0	$-$	0	0	$+$	$+$	0	$+$
	current sensor offset fault	ΔI	$\pm$	$\pm$	$\pm$	0	0	$+$	$-$	$+$	0

Table 6.1.2 summarizes the effects of some investigated faults on the parameter estimates and parity residuals.

These investigations have shown:

1) *Additive faults* like the offsets of sensors are well detected by the parity equations. They react fast and do not need an input excitation for a part of the faults. However, they have a relatively large variance, especially if the model parameters do not fit well to the process.
2) *Multiplicative faults* are well detected by parameter estimation, also for small faults. Because of the inherent regression method the reactions are slower but smoothed. But they require an input excitation for dynamic process models.

Therefore, it is recommended to combine both methods, as shown in Isermann (2006), Sect. 14.3. The parity equations are used to detect changes somewhere in the process and if the fault detection result is unclear a parameter estimation is started, eventually by a dynamic test signal for some seconds. If the motor operates dynamically anyhow (as for servo systems and actuators), then the parameter estimation can be applied continuously, but with a supervision scheme, see Isermann et al (1992).

Höfling (1996) has shown that a considerable improvement can be obtained by continuously estimating the armature resistance with a single parameter estimation using parity equations in order to reach the temperature dependent resistance parameters, Höfling and Isermann (1996). Furthermore, adaptive thresholds are recommended, to compensate for model uncertainties.

The model-based fault-detection system with parity equations and parameter estimation is now the basis for the fault-diagnosis procedure. As described in Chap. 2 the methods for fault diagnosis can be divided in classification and inferencing. A

first simple classification is the use of fault-symptom tables and pattern recognition as in Table 6.1.2. Also decision trees belong to the class of classification methods. However, a combination with a neuro-fuzzy structure gives them a learning behavior of fuzzy if-then rules with AND operators, forming an adaptive inference method, called SELECT, Füssel (2002). This was applied to the DC motor test bench, see Isermann (2011).

6.1.5 Conclusions

The detailed theoretical and experimental investigations with the permanently excited DC motor in idle running or with load have demonstrated that it is possible to detect 14 different faults by measurement of only three signals and combination of the parity equation and parameter estimation approach. Additive faults, like offsets of sensors, are easily detectable by parity equations in normal operation without extra input excitation signals. Multiplicative faults, like parameter deviations of the motor are better detected by parameter estimation, but require appropriate input excitation signals, at least for short times. Applying parameter estimation in addition to parity equations allows to track the strongly temperature dependent armature resistance. The described methods can be transformed to other types of DC motors, depending on their construction, and also to single-phase AC motors. Further, by applying the self-learning neuro-fuzzy system SELECT all faults could be diagnosed with a 98% correct classification rate. A selection of faults, especially in the mechanical parts can also be detected by applying only signal models for structure-borne vibrations, Filbert (2003).

6.2 Alternating-current motor (AC)

Alternating-current motors in the form of induction or asynchronous motors consist usually of three windings placed in stator slots that are interconnected with the individual phases of a three-phase voltage supply system either in delta- or Y-connection, see Fig. 6.2.1a). A rotating magnetic field is generated where the angular velocity depends on the power supply frequency f and on the number of pole pairs p within the stator. Depending on different rotor constructions, *induction motors* and *synchronous motors* can be distinguished. In the following, induction motors with a squirrel-cage rotor are considered. This type has a simple construction, is very robust, and is also cheap and needs less maintenance. The speed can be controlled by a field-oriented approach with variable frequency and amplitude generated by a voltage-source DC-link converter. Statistics on failure rates of AC motors show that about 50% are due to bearings, short cuts in stator windings count with 16% and broken rotor bars about 5%, Thorsen and Dalva (1995), see also Filbert (2003). A model-based procedure for fault diagnosis of AC drives will be treated in the following, developed by Wolfram (2002). The developed methods can be directly used for synchronous motors, see Sect. 6.3.

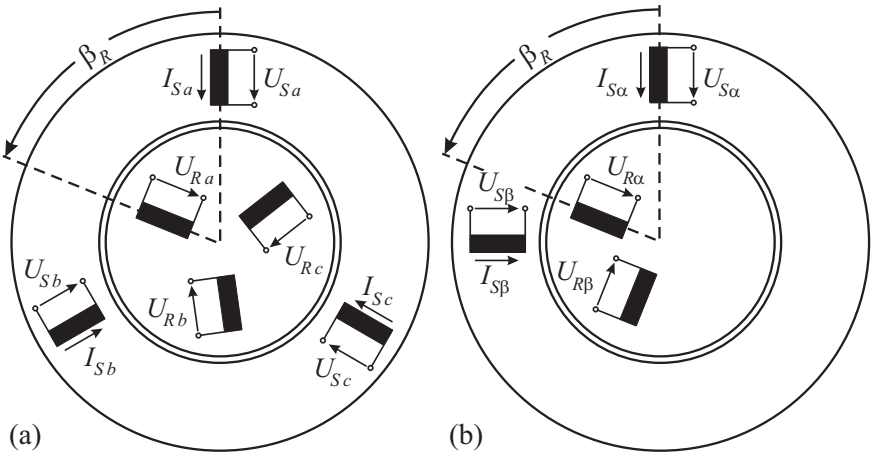


Fig. 6.2.1. Schematic configuration of the stator and rotor for AC motors. **a** three-phase representation with one pole pair ($p = 1$) per phase. **b** two-phase equivalent circuit. (Clarke transformation)

6.2.1 Models of induction motors (asynchronous motors)

a) Electrical subsystem

Detailed models of induction motors are derived e.g., in Isermann (2005), Leonhard (1996), Lyshevski (2000). For each rotor and stator winding the voltage and current equations are established, resulting in six coupled differential equations for a three-phase induction motor. However, by transforming the three-phase system $(U_{\mathrm{Sa}}, U_{\mathrm{Sb}}, U_{\mathrm{Sc}})$ into a two-phase system $(U_{S\alpha}, U_{S\beta})$ via the *Clarke–transform* a considerable simplification can be reached, Fig. 6.2.1b). If the rotor flux is taken as the reference coordinate system by using the *Park–transform*, the two-phase system is represented by $(U_{\mathrm{Sq}}, U_{\mathrm{Sd}})$, see the basic books on electric drives or e.g. Isermann (2005), Wolfram (2002). Then, two equations result for the rotor flux Ψ_{Rd} and electric motor torque M_{el}:

$$T_{\mathrm{R}} \frac{d\Psi_{\mathrm{Rd}}}{dt}(t) + \Psi_{\mathrm{Rd}}(t) = M I_{\mathrm{Sd}}(t) \quad \text{with} \quad T_{\mathrm{R}} = \frac{L_{\mathrm{R}}}{R_{\mathrm{R}}} \tag{6.2.1}$$

$$M_{\mathrm{el}}(t) = \frac{3}{2} p \frac{M}{L_{\mathrm{R}}} \Psi_{\mathrm{Rd}}(t) I_{\mathrm{Sq}}(t) \tag{6.2.2}$$

L_{R}	rotor self-inductance
R_{R}	rotor resistance
R_{R}	stator resistance
L_{R}	stator self-inductance
M	mutual inductance between stator and rotor
$I = I_{\mathrm{Sd}} + iI_{\mathrm{Sq}}$	stator current vector

p	number of pole pairs
Ψ_{Rd}	rotor flux.

The flux linkage Ψ_{Rd} depends on I_{Sd} and the torque M_{el} on I_{Sq}, i.e. from each component of the stator current vector, see the signal flow chart in Fig. 6.2.2. This two-phase system is the basis for the field-oriented control shown in Fig. 6.2.3. It consists of two cascaded control loops for the flux and the speed, with the two current components as sub (minor) control variables.

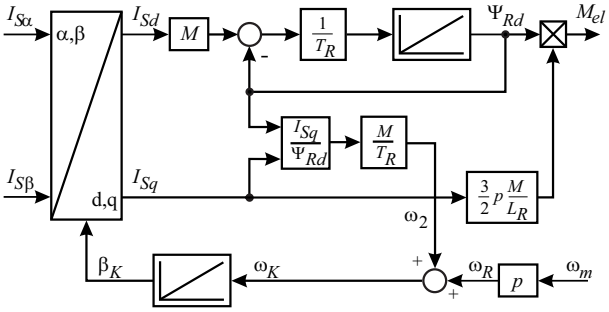


Fig. 6.2.2. Calculation of the electrical torque M_{el} based on inputs I_{Sd}, I_{Sq} and ω_{m}.

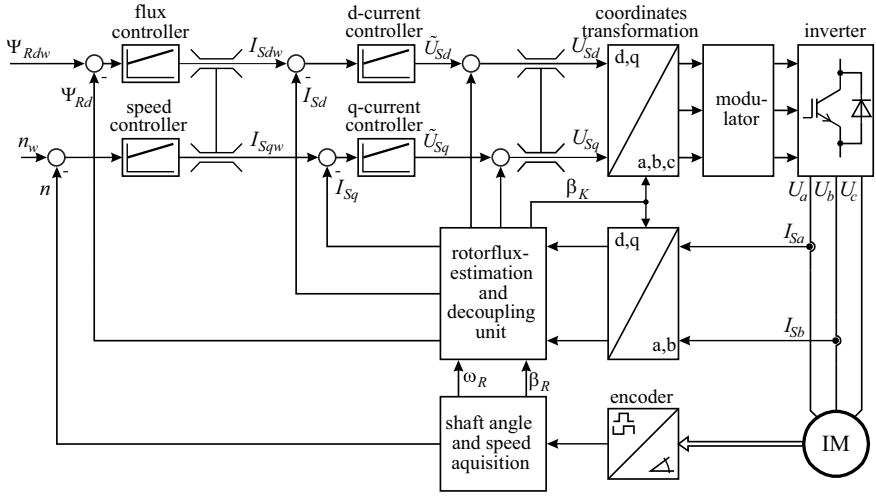


Fig. 6.2.3. Field-oriented control structure for induction motors.

The dynamic behaviors of the electrical induction motor subsystems are

$$U_{\mathrm{Sd}} = \left(R_{\mathrm{S}} + R_{\mathrm{R}} \frac{M^2}{L_{\mathrm{R}}^2} \right) I_{\mathrm{Sd}} + \sigma L_{\mathrm{S}} \frac{dI_{\mathrm{Sd}}}{dt} - \sigma L_{\mathrm{S}} \omega_{\mathrm{K}} I_{\mathrm{Sq}} - \frac{R_{\mathrm{R}} M}{L_{\mathrm{R}}^2} \Psi_{\mathrm{Rd}} \tag{6.2.3}$$

$$U_{\mathrm{Sq}} = \left(R_{\mathrm{S}} + R_{\mathrm{R}} \frac{M^2}{L_{\mathrm{R}}^2} \right) I_{\mathrm{Sq}} + \sigma L_{\mathrm{S}} \frac{dI_{\mathrm{Sq}}}{dt} + \sigma L_{\mathrm{S}} \omega_{\mathrm{K}} I_{\mathrm{Sd}} + \frac{M}{L_{\mathrm{R}}} \omega_{\mathrm{R}} \Psi_{\mathrm{Rd}}, \tag{6.2.4}$$

$$\sigma = 1 - \frac{M^2}{L_{\mathrm{S}} L_{\mathrm{R}}}. \tag{6.2.5}$$

Herewith, the electrical rotor speed is $\omega_{\mathrm{R}} = p\omega_{\mathrm{m}}$, where ω_{m} is the mechanical rotor speed and the speed of the flux is ω_{K} with regard to the stator reference coordinate system.

b) Mechanical subsystem

The dynamic behavior of the mechanical part is obtained by establishing the angular momentum balance:

$$J \frac{d\omega_{\mathrm{m}}(t)}{dt} = M_{\mathrm{el}}(t) - M_{\mathrm{f}}(t) - M_{\mathrm{L}}(t) \tag{6.2.6}$$

J ratio of inertia of motor and load
M_{f} friction torque
M_{L} load torque
ω_{m} mechanical rotor speed.

The friction torque usually consists of a Coulomb term and a viscous term:

$$M_{\mathrm{f}} = M_{\mathrm{f0}} \operatorname{sign} \omega_{\mathrm{m}}(t) + M_{\mathrm{f1}} \omega_{\mathrm{m}}. \tag{6.2.7}$$

The load torque depends on the connecting power consuming machine, like a vehicle and can frequently be approximated by a polynomial:

$$M_{\mathrm{L}} = M_{\mathrm{L0}} + M_{\mathrm{L1}} \omega_{\mathrm{m}} + M_{\mathrm{L2}} \omega_{\mathrm{m}}^2. \tag{6.2.8}$$

c) Thermal subsystem

Within the stator and rotor several power losses P_{LS} and P_{LR} arise which lead to a heating of the induction motor parts. The main heat sources are ohmic losses and iron losses, which can further be split up into hysteresis and eddy current losses. With the stator and rotor heat capacity

$$C_{\mathrm{S}} = m_{\mathrm{S}} c_{\mathrm{Sp}}$$
$$C_{\mathrm{R}} = m_{\mathrm{R}} c_{\mathrm{Rp}}$$

where m is the respective mass and c_{P} the specific heat value, two first-order differential equations result for the stator temperature $\vartheta_{\mathrm{S}}(t)$ and rotor temperature $\vartheta_{\mathrm{R}}(t)$. With further simplifications about the heat transfer through the air gap and air cooling a second-order model results for the stator temperature:

$$\Delta\vartheta_{\mathrm{S}}(s) = \frac{(b_{\mathrm{S1}} s + b_{\mathrm{S0}})\, P_{\mathrm{LS}}(s) + b_{\mathrm{R0}} P_{\mathrm{LR}}(s)}{a_2 s^2 + a_1 s + a_0}, \tag{6.2.9}$$

see Wolfram (2002), Wolfram and Isermann (2001).

6.2.2 Signal-based fault detection of the power electronics

The power supply of cars is usually based on a high voltage or low voltage direct current network. Variable speed AC motors then need a motor-side three-phase DC-AC converter (inverter) that generates the three-phase system with variable frequency and amplitude, as depicted in Fig. 6.2.4.

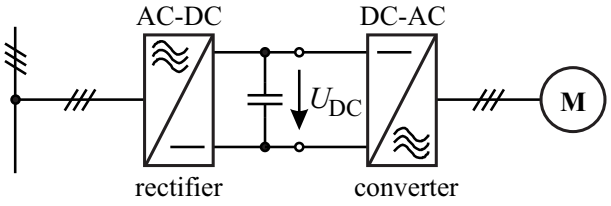


Fig. 6.2.4. Voltage source DC-link converter scheme for feeding AC motors.

In the following it is shown how faults in these power electronics can be detected with *signal-based methods*. Measured variables are the DC-link (battery) voltage U_{DC}, the phase currents I_{S1}, I_{S2}, I_{S3}, which are identical to I_{Sa}, I_{Sb}, I_{Sc} of Fig. 6.2.1, see Fig. 6.2.5. For the phase voltages U_{S1}, U_{S2}, U_{S3} only the setpoints of the PWM converter are available.

Figure 6.2.5 shows a scheme of a considered PWM DC-AC inverter. Inverter faults and also stator winding faults generate characteristic harmonics of the stator current vector. The stator current

$$\mathbf{I}_{\mathrm{S}}(t) = \frac{2}{3}\left(I_{\mathrm{S1}}(t) + I_{\mathrm{S2}}(t)e^{-i\frac{2\pi}{3}} + I_{\mathrm{S3}}(t)e^{-i\frac{4\pi}{3}}\right) \tag{6.2.10}$$

is transformed into an orthonormal α- and β-component coordinate system

$$\mathbf{I}_{\mathrm{S}}(t) = I_{\mathrm{S}\alpha}(t) + iI_{\mathrm{S}\beta}(t) = I_{\mathrm{S0}}(t)e^{i\varphi(t)}. \tag{6.2.11}$$

In the fault-free case the trajectory of the current vector forms a circle, which deforms to an ellipse in the case of a stator winding fault, see Fig. 6.2.6a) and to other trajectories for inverter faults and current sensor faults, Fig. 6.2.6b)-d). In the case of these faults the spectrum of the current vector contains a positive and negative frequency, Wolfram (2002), Wolfram and Isermann (2000).

$$\mathbf{I}_{\mathrm{S}}(t) = \bar{I}_{\mathrm{S1}}e^{i(\omega_{\mathrm{S}}t+\varphi_1)} + \bar{I}_{\mathrm{S}-1}e^{i(-\omega_{\mathrm{S}}t+\varphi_{-1})} \tag{6.2.12}$$

where ω_{S} is the stator angular frequency.

The vector trajectory $\mathbf{I}_{\mathrm{S}-1}$ is also circular, but with smaller radius and opposite direction in the case of an ellipse. By monitoring I_{S1} and $I_{\mathrm{S}-1}$ obtained through a Fourier series analysis it is possible to detect the mentioned faults, Wolfram and Isermann (2000). This can be applied to AC motors with constant grid frequency f_{S}. However, in the case of a field-oriented control with variable frequency f_{S} the current

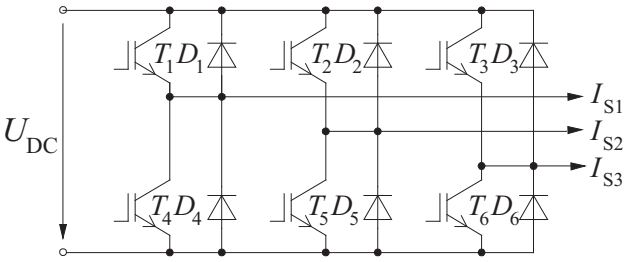


Fig. 6.2.5. PWM inverter for DC–AC converter.

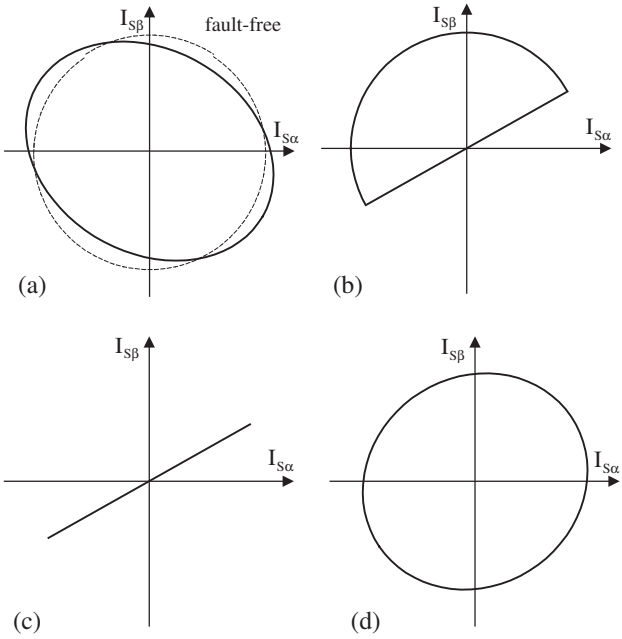


Fig. 6.2.6. Stator current vector trajectories. **a** stator winding fault. **b** inverter IGBT (Insulated Gate Bipolar Transistor) fault. **c** phase 2 disconnection. **d** current sensor fault.

vector is influenced by control dynamics. Because also the stator voltage $\mathbf{U}_{\mathrm{S}}(t)$ as output of the current controllers, see Fig. 6.2.3, shows corresponding frequencies, at least for high speeds of the motor the Fourier analysis can be performed with $\mathbf{U}_{\mathrm{S}}(t)$. Therefore

$$r_{\mathrm{U}-1}(t) = |\bar{U}_{\mathrm{S}-1}(t)| \tag{6.2.13}$$

is taken as the fault feature for a stationary speed. However, a speed-dependent threshold is required, which has to be determined experimentally. An additional feature is the DC value

$$r_{\mathrm{U0Hz}} = |\bar{U}_{\mathrm{SOHz}}| \tag{6.2.14}$$

for detecting offsets in the voltage vector $\mathbf{U}_{\mathrm{S}}$.

As in the case of a disconnected phase or a defective IGBT valve the three currents become different, the effective values of the effective values (rms values)

$$I_{\mathrm{Si,eff}} = \frac{1}{T_\mathrm{p}} \int_0^{T_\mathrm{p}} I_\mathrm{Si}^2(t)dt$$

of the three currents $I_\mathrm{Si}(t)$ are calculated by taking the squares $I_\mathrm{Si}^2(t)$ with subsequent low-pass filtering. Then mutual residuals are formed:

$$\begin{aligned} r_{12}(t) &= I_{\mathrm{S1,eff}}^2(t) - I_{\mathrm{S2,eff}}^2(t) \\ r_{23}(t) &= I_{\mathrm{S2,eff}}^2(t) - I_{\mathrm{S3,eff}}^2(t) \\ r_{31}(t) &= I_{\mathrm{S3,eff}}^2(t) - I_{\mathrm{S1,eff}}^2(t). \end{aligned} \tag{6.2.15}$$

In order to include all measured currents the current sum is used as a further residual:

$$r_\mathrm{S0} = |I_\mathrm{S0}| = |I_\mathrm{S1}(t) + I_\mathrm{S2}(t) + I_\mathrm{S3}(t)| \tag{6.2.16}$$

which is usually zero in a fault-free situation.

Finally, Table 6.2.1 shows the fault-symptom relations for different faults. All considered faults can be isolated. In general, these results indicate a strong isolability. Only within the groups of open phases, defective valve and ground cuts is there a weak isolability.

Table 6.2.1. Fault-symptom table for the PWM DC-AC converter and stator windings

	Symptoms					
Faults	$\lvert r_{12}\rvert$	$\lvert r_{23}\rvert$	$\lvert r_{31}\rvert$	$\lvert r_{\mathrm{U}-1}\rvert$	$\lvert r_{\mathrm{UOHz}}\rvert$	$\lvert I_\mathrm{S0}\rvert$
open phase 1	++	0	++	++	0	0
open phase 2	++	++	0	++	0	0
open phase 3	0	++	++	++	0	0
defective valve 1	++	+	++	+	++	0
defective valve 2	++	++	+	+	++	0
defective valve 3	+	++	++	+	++	0
stator winding shortcut	≈ 0	≈ 0	≈ 0	+	0	0
offset fault sensor 1 or 2	≈ 0	≈ 0	≈ 0	0	+	+/++
gain fault sensor 1 or 2	≈ 0	≈ 0	≈ 0	+	0	+/++
faulty current sensor	0	0	0	0	0	+/++
ground cut phase 1	++	+	++	++	0	+/++
ground cut phase 2	++	++	+	++	0	+/++
ground cut phase 3	+	++	++	++	0	+/++

6.2.3 Model-based fault detection of the AC motor

It is assumed that the following measurements and calculated variables are available:

- U_{Sq}, U_{Sd} voltages of the q- and d-systems
- I_{Sq}, I_{Sd} currents of the q- and d-systems
- $\omega_{\mathrm{K}} = \omega_{\mathrm{S}}$ stator angular frequency
- $\omega_{\mathrm{R}} = p\omega_{\mathrm{m}} = \omega_{\mathrm{K}} - \omega_2$ rotor angular frequency.

ω is a slip frequency, calculated by

$$\omega_2 = \frac{M}{T_{\mathrm{R}}} \frac{I_{\mathrm{Sq}}}{\Psi_{Rd}}. \tag{6.2.17}$$

In order to apply parity equations for the fault detection of the AC motor, nonlinear dynamic models are required, which are obtained by nonlinear process identification methods, Wolfram and Isermann (2001).

The AC motor is the type VEM K21R90S (Normmotor) with four poles and rated values 400 V, 2.62 A, 1.1 kW, 1420 rpm (50 Hz), see Wolfram (2002).

a) Electrical part of the AC motor

As a basis for obtaining dynamic models (6.2.3) and (6.2.4) are used for the d- and q-subsystems. It has to be taken into account that phase voltages are not exactly known. For the practical experiments the d-current control stays closed, whereas the q-current control is opened in order to introduce an excitation signal U_{Sq}, see Fig. 6.2.3. Therefore the rotor flux reference value Ψ_{Rdref} stays constant. Discretizing (6.2.4) with the discrete time $k = t/T_0$, where T_0 is sampling time, leads to

$$I_{\mathrm{Sq}}(k) = \Theta_1 U_{\mathrm{Sq}}(k) + \Theta_2 \omega_{\mathrm{K}}(k) I_{\mathrm{Sd}}(k) + \Theta_3 \omega_{\mathrm{R}}(k) \Psi_{\mathrm{Rd}} + \Theta_4 I_{\mathrm{Sq}}(k-1) \tag{6.2.18}$$

where Θ_{i} are parameters, which depend on physical parameters. The product $\omega_{\mathrm{K}}(k)$ $I_{\mathrm{Sd}}(k)$ can be neglected and $\Psi_{\mathrm{Rd}} = \text{const.}$ can be assumed. The parameters Θ_{i} further depend on the operating point through ω_{K} and I_{Sq}. Therefore a local linear model is defined:

$$I_{\mathrm{Sq}}(k) = w_1(\mathbf{z}) U_{\mathrm{Sq}}(k) + w_2(\mathbf{z}) \omega_{\mathrm{R}}(k) + w_3(\mathbf{z}) I_{\mathrm{Sq}}(k-1). \tag{6.2.19}$$

The operating-point dependence is expressed by the weighting vector

$$\mathbf{z}^T = \left[\omega_{\mathrm{K}}(k) I_{\mathrm{Sq}}(k-1)\right]. \tag{6.2.20}$$

Hence, this is a semi-physical model because the structure stems from physical-based modeling. The weighting parameters $w_{\mathrm{i}}(\mathbf{z})$ are estimated with the LOLIMOT identification method, Isermann and Münchhof (2011).

However, the parameters depend also on the temperature of the AC motor. Therefore, the stator temperature ϑ_{S} is measured and two correction factors $k_1(\vartheta_{\mathrm{S}})$ and $k_2(\vartheta_{\mathrm{S}})$ are introduced in (6.2.19):

$$I_{\mathrm{Sq}}(k) = w_1(\mathbf{z}) k_1(\vartheta_{\mathrm{S}}) U_{\mathrm{Sq}}(k) + w_2(\mathbf{z}) k_2(\vartheta_{\mathrm{S}}) \omega_{\mathrm{R}}(k) + w_3(\mathbf{z}) I_{\mathrm{Sq}}(k-1). \tag{6.2.21}$$

These correction factors are estimated and $k_1(\vartheta_{\mathrm{S}})$ is approximated by a second-order polynomial and $k_2(\vartheta_{\mathrm{S}})$ with a linear dependence.

The dynamic behavior of the AC motor was identified by excitation of U_{Sq} with an APRBS, an amplitude-modulated PRBS, and sampling time $T_0 = 1.5\,\mathrm{ms}$ (667 Hz). The obtained generalization results in Fig. 6.2.7 show a very good agreement with six local models and two correction characteristics for stator temperatures $\vartheta_{\mathrm{S}} \in [25°\mathrm{C}, 60°\mathrm{C}]$.

In a similar way the d-system can be identified. (6.2.3) and experimental trials lead to

$$U_{\mathrm{Sd}}(k) = w_0(\mathbf{z}) + w_1(\mathbf{z})\omega_{\mathrm{K}}(k) + w_2(\mathbf{z})I_{\mathrm{Sq}}(k). \tag{6.2.22}$$

As I_{Sd} is constant, its derivative is zero. Therefore, the d-model is static. The temperature dependence is again considered by correction factors. Figure 6.2.8 shows relatively good results with a model having six local linear models and 18 correction characteristics.

These nonlinear precise models can now be used to apply parity equations for fault detection. The output residuals

$$r_{\mathrm{q}} = I_{\mathrm{Sq}} - \hat{I}_{\mathrm{Sq}} \tag{6.2.23}$$

$$r_{\mathrm{d}} = U_{\mathrm{Sd}} - \hat{U}_{\mathrm{Sd}}. \tag{6.2.24}$$

are low-pass filtered and their variances are formed, compare Fig. 6.2.9. Furthermore, r_{12}, r_{23}, r_{31} and $|I_{\mathrm{S0}}|$ are determined from the phase currents, (6.2.15) and (6.2.16). As the models are more precise in the case of stationary behavior, an adaptive threshold is used for dynamic states in dependence on the current I_{Sq}, which is proportional to the torque-generating dynamics. Table 6.2.2 presents the fault-symptom relation for different faults. The AC motor faults as stator winding defect, broken rotor bar, and rotor eccentricity are strongly isolated and can therefore be diagnosed. However, broken rotor bar and broken end ring are only weakly isolated and cannot be clearly separated. For the other faults the same results are obtained as in Table 6.2.1 for the PWM converter.

b) Mechanical subsystem of the AC motor

The dynamic behavior of the rotor speed $\omega_{\mathrm{R}}(t)$ follows from (6.2.6). Faults in the mechanical part express themselves especially in the friction parameters M_{f0} and M_{f1} and eventually in the ratio of inertia J. However, these parameters also depend on the connected load like general drive-trains with gears. Therefore this dynamics equation depends on the load and corresponding available measurements, see e.g. Isermann (2005). The electrical torque M_{el} can be determined with (6.2.1) and (6.2.2) using the current I_{Sd} and I_{Sq} of the d- and q-systems, which are known within a field-oriented controller, and the mechanical rotor speed ω_{m}. Figure 6.2.2 summarizes the resulting signal flow and computations. Some electrical parameters or groups of parameters can then be estimated. As the mechanical subsystem is slower than the electrical subsystem, a larger sampling time can be chosen, e.g. $T_0 = 10\,\mathrm{ms}$ (100 Hz).

An application with a circular pump as load is reported in Isermann and Münchhof (2011).

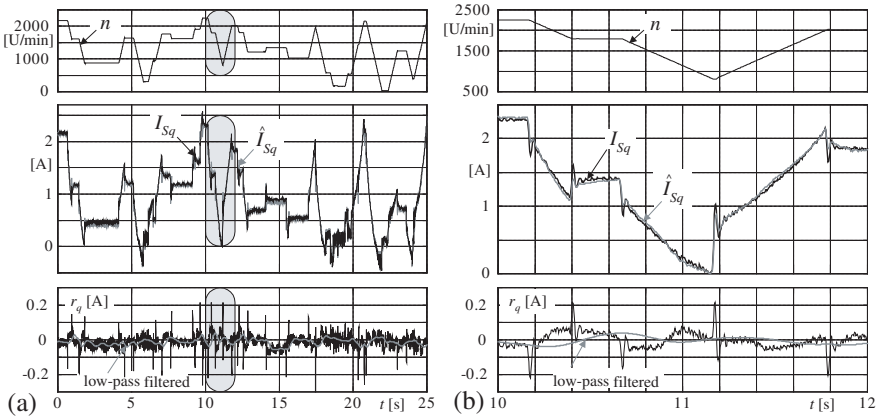


Fig. 6.2.7. Generalization data for LOLIMOT identification of the q-system. **a** Input: U_{Sq} (APRBS). Output: I_{Sq}, n and output error $r_{\mathrm{q}} = I_{\mathrm{Sq}} - \hat{I}_{\mathrm{Sq}}$ for $\delta_{\mathrm{S}} = 25°\mathrm{C}$. **b** zoomed signals.

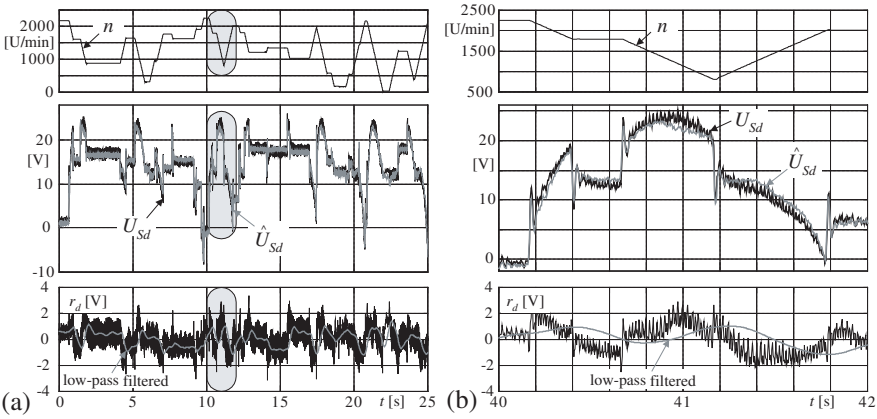


Fig. 6.2.8. Generalization data for LOLIMOT identification of the d-system. **a** Input: U_{Sq} (APRBS). Output: U_{Sd}, n and output error $r_{\mathrm{d}} = U_{\mathrm{Sd}} - \hat{U}_{\mathrm{Sd}}$ for $\delta_{\mathrm{S}} = 25°\mathrm{C}$; **b** zoomed signals.

c) Thermal subsystem

The thermal state of the AC motor is indicated by the temperature of the rotor and the stator. Overheating arises because of defective cooling, high friction and overload. The generated stator heat power due to ohmic losses is

$$P_{\mathrm{LS}} = \frac{3}{2} R_{\mathrm{S}} \left(I_{\mathrm{Sd}}^2 + I_{\mathrm{Sq}}^2 \right) \tag{6.2.25}$$

and the rotor losses. If the rotor power losses are neglected, only one part of the transfer function of (6.2.9) has to be considered. The corresponding z-transfer function

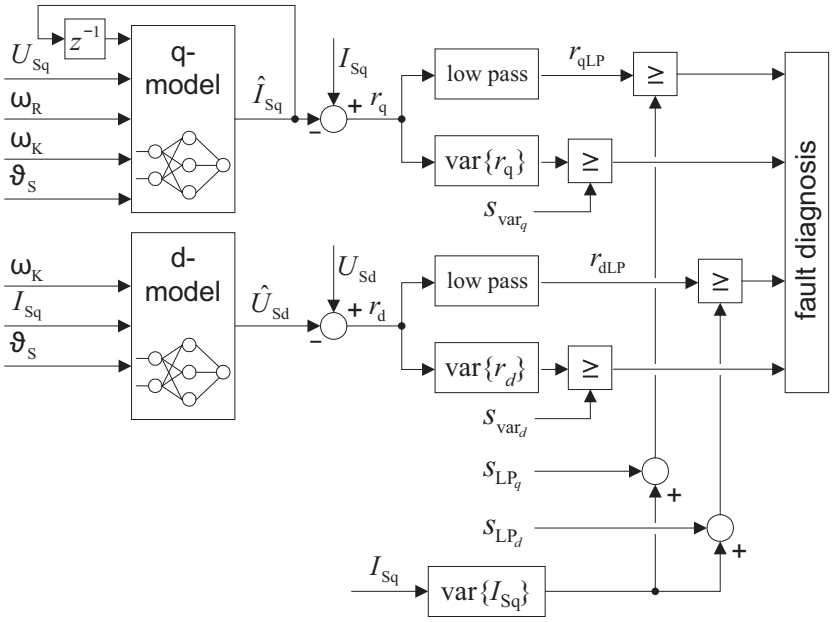


Fig. 6.2.9. Residual generation for the AC motor with nonlinear parity equations, S: thresholds.

Table 6.2.2. Fault-symptom table for the AC motor

	Symptoms							
Faults	$\|r_{qLP}\|$	var $\{r_q\}$	$\|r_{dLP}\|$	var $\{r_d\}$	$\|r_{12}\|$	$\|r_{23}\|$	$\|r_{31}\|$	$\|I_{S0}\|$
open phase 1	−−	++	++	++	++	0	++	0
open phase 2	−−	++	++	++	++	++	0	0
open phase 3	−−	++	++	++	0	++	++	0
defective valve 1	−	+	++	++	++	+	++	0
defective valve 2	−	+	++	++	++	++	0	0
defective valve 3	−	+	++	++	+	++	++	0
stator winding shortcut	−−	0	+	+	≈ 0	≈ 0	≈ 0	0
broken rotor bar	++	+	+	++	0	0	0	0
broken end ring	+	+	+	+	0	0	0	0
rotor eccentricity	+	0	0	0	0	0	0	0
gain fault sensor 1 or 2	−/+	0	0	0	0	0	0	+/++
offset sensor fault 1 or 2	+	0	0	0	0	0	0	+/++
fault current sensor	0	0	0	0	0	0	0	+/++
ground cut phase 1	−−	+	+	+	++	+	++	+/++
ground cut phase 2	−−	+	+	+	++	++	+	+/++
ground cut phase 3	−−	+	+	+	+	++	++	+/++

$$G(z) = \frac{\Delta\hat{\vartheta}_{\mathrm{S}}(z)}{P_{\mathrm{LS}}(z)} = \frac{\beta_1 z^{-1} + \beta_2 z^{-2}}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}} \tag{6.2.26}$$

possesses two poles which belong to a large time constant $T_1 \approx 17\,\mathrm{min}$ and a small one $T_2 \approx 2\,\mathrm{min}$, which were determined by parameter estimation (large sampling time $T_0 = 30\,\mathrm{s}$). Then, an output residual as the difference of the measured and estimated stator temperature

$$r_\vartheta = \vartheta_{\mathrm{S}} - \hat{\vartheta}_{\mathrm{S}} \tag{6.2.27}$$

can be formed. If this residual exceeds a threshold of for example $+5°\,\mathrm{C}$, a cooling defect can be detected.

In the case of dynamic operations parameter estimation can be applied. Especially the time constant $\hat{T}_1$ then indicates faults of the thermal system.

Figure 6.2.10 shows the increase of the stator temperature for a defective fan wheel and Table 6.2.3 the estimates of the time constants. For cooling faults or overload the stator temperature increases considerably, indicated by a larger gain of (6.2.26) or larger residual r_ϑ. Also, the large time constant T_1 shows a strong increase. The small constant T_2 remains approximately constant. This can be used for more details of a cooling fault diagnosis, as shown in Table 6.2.3.

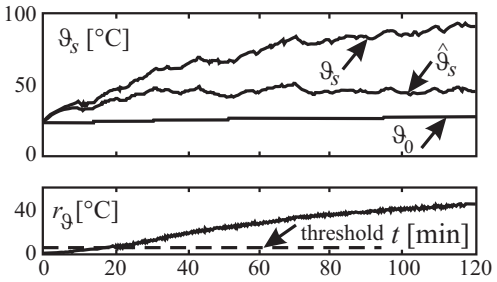


Fig. 6.2.10. Stator temperature ϑ_{S} and temperature residual r_ϑ in the case of a defective fan wheel. $\hat{\vartheta}_{\mathrm{s}}(P_{\mathrm{LS}}, t)$ is the model output for normal behavior.

Table 6.2.3. Parameter estimates of the thermal subsystem for different cooling defects

time constants	fault-free	defective fan wheel	covered cooling slots	covered motor
$\hat{T}_1$ [min]	16.6	69.3	31.2	36.1
$\hat{T}_2$ [min]	1.99	1.86	1.82	1.72

d) AC motor at standstill

Some faults in induction motors do not effect immediately a complete failure of the drive. Thus, e.g. in the case of broken rotor bars further operation is possible in

principle. In consideration of the higher currents in the adjacent bars and the resulting mechanical stresses due to thermal overload and unbalance, further rotor or end-ring breakage may occur, Thorsen and Dalva (1995).

Hence, the monitoring of these faults does not have to be performed permanently. Basically, it is sufficient to supervise the induction motor after certain periods of time, which is described in Wolfram (2002) and Isermann (2011).

6.2.4 Conclusions

Figure 6.2.11 gives an overall scheme of the signal- and process-model-based fault detection and diagnosis of a speed-controlled AC drive. The investigations have shown that with the support of physical modeling it is possible to detect several different faults in the PWM inverter, the AC-motor stator and rotor, and the motor mechanics. The use of nonlinear output *parity equations* is especially attractive for the electrical subsystems. However, it required relatively precise process models. These models can be obtained with a multi-model approach by applying local linear models and the LOLIMOT *nonlinear parameter estimation*. For the fault detection of the motor mechanics *parameter estimation* with linearity in the parameters is preferable and includes also process parameters from the drive train. The model-based fault-diagnosis approach requires only four sensors and some variables which are available within the field-oriented control, determines up to 14 symptoms and can diagnose about 10 different faults. The described fault-detection methods can be directly transferred to *synchronous motors*, see next section, Sect. 6.3.

6.3 Alternating-current synchronous motors (SM)

6.3.1 Types of three-phase synchronous motors

In view of the electrical and magnetical properties the stator of synchronous motors is similarly constructed to that of induction motors, whereas the rotor is designed in a manner, that a constant magnetic field along one axis is produced ($p = 1$). In case of more than one pole pairs ($p > 1$) the magnetic field of the rotor is spread over a corresponding number of axis.

Synchronous motors equipped with *salient-pole rotors* comprise of distinct poles and have a single excitation coil wound on a core. The rotor field winding is fed with direct current through slip rings.

In view of the excitation field produced by the rotor windings, no slip between the stator field and the electrical speed $\omega_{\mathrm{R}} = p\omega_{\mathrm{m}}$ emerges. Hence, the synchronous motor always rotates with the speed that results from the supply frequency divided by the number of pole pairs. The speed can therefore only be adjusted by changing the frequency of the feeding symmetrical voltage system. As in the rotor no alternating magnetical fields appear, no eddy currents are induced and the core need not to be designed laminated.

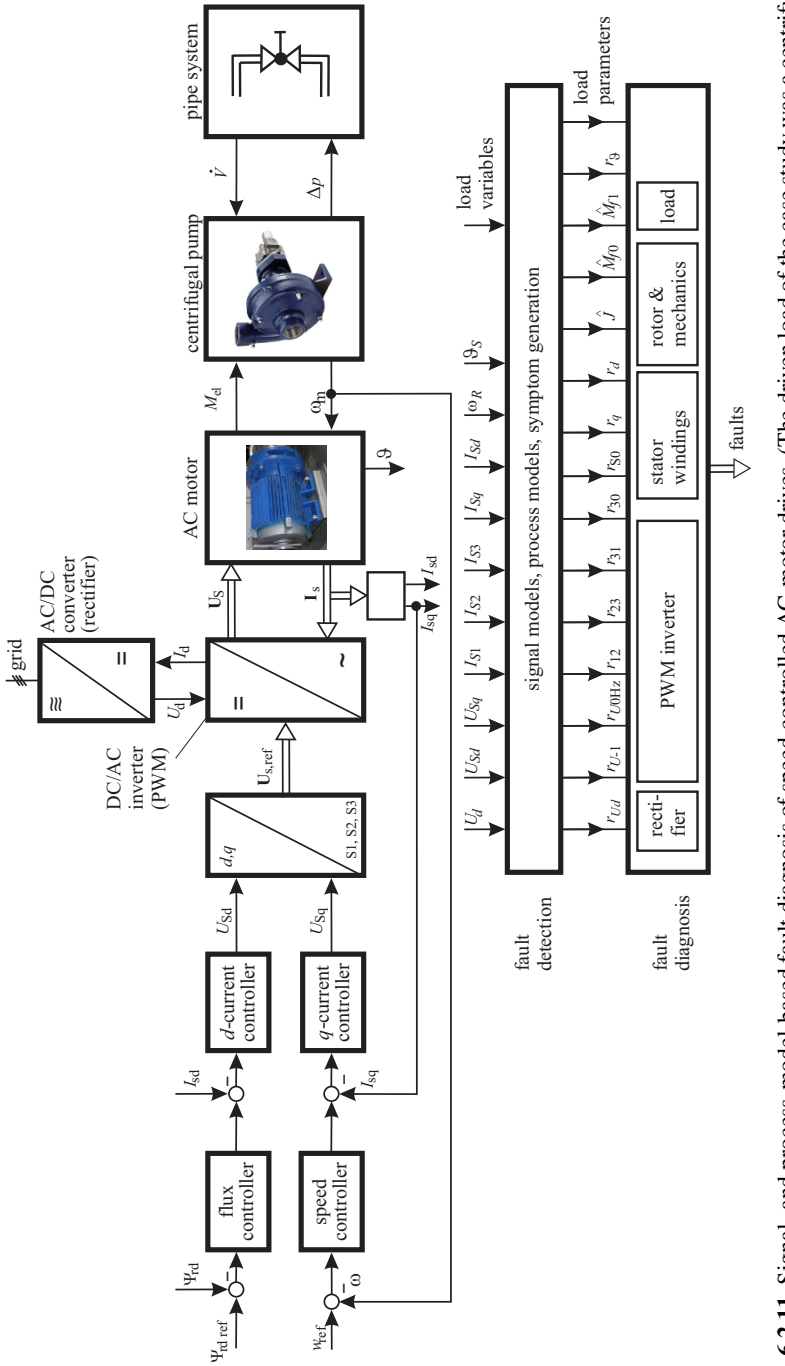


Fig. 6.2.11. Signal- and process-model-based fault diagnosis of speed-controlled AC-motor drives. (The driven load of the case study was a centrifugal pump, but can also be a drive-train.)

Synchronous motors with salient pole rotor are mainly employed, in order to achieve high numbers of pole pairs and therefore low speeds. Regarding the large centrifugal force which is proportional to the square of the speed, salient pole rotors are not applicable for higher speeds. They are usually applied as high power generators in water power stations.

For low power applications *permanently magnetic rotors* are employed. The advantage is that the slip-ring construction for salient pole rotors can be omitted. Furthermore, the power loss produced in the rotor windings does not appear so that high efficiency rates can be achieved. For the permanent magnets often rare earth magnetic materials such as Sm-Co are employed, since they provide high flux densities (1T). Figure 6.3.1 illustrates a scheme.

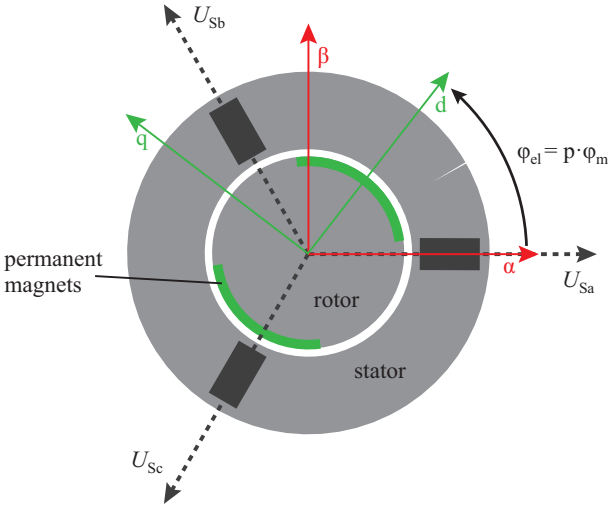


Fig. 6.3.1. Scheme of a synchronous motor with permanent magnets (one pole pair, $p = 1$), Keßler (2015). (α, β are the coordinates of the stator-fixed two-phase system using the Park-Clarke transformation for the stator-fixed three-phase system)

The electrical properties of permanent magnet synchronous motors (PMSM) depend on the design of the rotor, Teigelkötter (2012). In the case of "surface-mounted permanent magnet synchronous motor (SPMSM)" the magnetic material is glued on the rotor. It has a relatively small magnetic resistance (reluctance) and therefore the inductance of the stator windings is approximately independent on the rotor position which results in simpler models.

If the permanent magnets are embedded in the rotor, this is called "interior permanent magnet synchronous motor (IPMSM)". The stator inductances then depend on the rotor position.

6.3.2 Models and control of permanent magnet synchronous motors (PMSM)

The sine commutation for the PMSM is generated by an inverter with a pulse-width modulation (PWM) and the DC link voltage U_{DC} as input voltage, see Fig. 6.2.4. The input voltage of each motor phase can be modeled by a mean value model extended by two voltage drops

$$U_{\mathrm{S}j0} = d_j\, U_{\mathrm{DC}} - U_{\mathrm{TG}} - U_{\mathrm{VL}} \text{ with } j = a, b, c. \tag{6.3.1}$$

The first voltage drop U_{TG} appears because of non-ideal switching times. The second voltage drop U_{VL} depends on the voltage losses of the diodes and IGBT's. For the detailed model see Keßler (2017). The phase to ground voltages $U_{\mathrm{S}j0}$ can now be used as input of the motor model. The relation between the phase to ground voltages and the phase voltages can be described by

$$U_{\mathrm{S}j0} = U_{\mathrm{S}j} + U_{\mathrm{SP}} \text{ with } j = a, b, c \tag{6.3.2}$$

with the starpoint voltage U_{SP}. For the stator voltage in a stator-fixed coordinate system holds

$$U_{Sj} = R_S\, I_{Sj} + \frac{d\Psi_{Sj}}{dt} \text{ with } j = a, b, c. \tag{6.3.3}$$

The equations for the stator voltage in the rotor-fixed coordinate system become with neglection of reluctance effects and therefore the assumption $L_q = L_d = L_S$

$$U_{\mathrm{Sd}} = R_{\mathrm{S}} I_{\mathrm{Sd}} + L_{\mathrm{S}} \frac{dI_{\mathrm{Sd}}}{dt} - L_{\mathrm{S}} \omega_{\mathrm{el}} I_{\mathrm{Sq}}, \tag{6.3.4}$$

$$U_{\mathrm{Sq}} = R_{\mathrm{S}} I_{\mathrm{Sq}} + L_{\mathrm{S}} \frac{dI_{\mathrm{Sq}}}{dt} + L_{\mathrm{S}} \omega_{\mathrm{el}} I_{\mathrm{Sd}} + \omega_{\mathrm{el}} \Psi_{\mathrm{Rd}}, \tag{6.3.5}$$

compare (6.2.3) and (6.2.4). Hence, differential equations of first order with mutual couplings result. ω_{el} is the electrical angular speed in the stator windings and relates to the mechanical angular speed of the rotor ω_{mech} by

$$\omega_{\mathrm{el}} = p \cdot \omega_{\mathrm{mech}} = 2\pi p n_{\mathrm{mech}}. \tag{6.3.6}$$

The electrical inner torque can then be determined through the electrical power with $P_{\mathrm{el}} = \omega_{\mathrm{m}} M_{\mathrm{el}}/p$ to

$$M_{\mathrm{el}} = \frac{3}{2} p \Psi_{\mathrm{Rd}} I_{\mathrm{Sq}} \tag{6.3.7}$$

see e.g. Teigelkötter (2012).

The control of inverter-fed permanent magnet synchronous motors can be performed similarly to the rotor-flux-oriented control of induction motors. The equations comprising the behavior between the stator currents and the stator voltages within the rotor-oriented reference frame (neglecting the reluctance effects) are described by (6.3.4) and (6.3.5).

The relationship between the stator current components and the parallel stator voltage components can be specified apart from coupling terms and the EMF term by

first-order-lags. Since both equations are not ideally decoupled, a decoupling unit has to eliminate all the terms leading to deviations from the first order lag characteristic. Note the similarity to DC motors, Sect. 6.1.

In case of feeding the synchronous machine with a variably adjustable three-phase voltage system the manipulated variables are the stator voltages U_{Sd} and U_{Sq}. Controlled variables are the speed n as well as both stator currents I_{Sd} and I_{Sq}. Thus, the control can be performed by means of a cascade structure with secondary current control loops, as depicted in Fig. 6.3.2

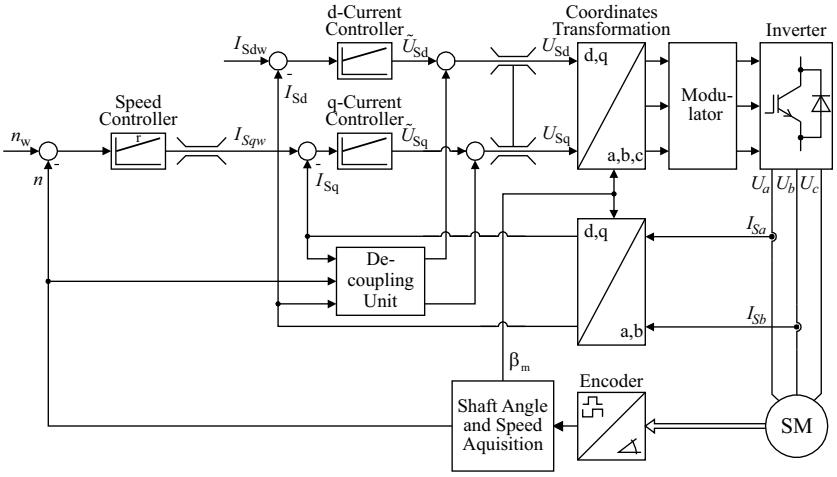


Fig. 6.3.2. Field-oriented scheme of permanent magnet synchronous motors.

As the flux producing current component I_{Sd} does not affect the torque, the reference value $I_{\mathrm{sdw}} = 0$ is chosen. In this case the motor is driven as efficient as possible with the smallest possible stator current.

Then, (6.3.5) simplifies to

$$U_{\mathrm{Sq}} = R_{\mathrm{S}} I_{\mathrm{Sq}} + L_{\mathrm{S}} \frac{dI_{\mathrm{Sq}}}{dt} + \omega_{\mathrm{el}} \Psi_{\mathrm{Rd}}. \tag{6.3.8}$$

After considering small changes around the operation point the last EMF term vanishes because it can be assumed as constant. The dynamic current behavior then can be described by the transfer function

$$G_{\mathrm{IU}}(s) = \frac{\Delta I_{\mathrm{Sq}}(s)}{\Delta U_{\mathrm{Sq}}(s)} = \frac{1}{R_{\mathrm{S}}} \left(1 + \frac{L_{\mathrm{S}}}{R_{\mathrm{S}}} s\right)^{-1}. \tag{6.3.9}$$

If a PI-current controller is used

$$G_{\mathrm{C}}(s) = \frac{\Delta U_{\mathrm{Sq}}(s)}{\Delta I_{\mathrm{Sq}}(s)} = \frac{K_{\mathrm{c}}}{T_{\mathrm{I}} s} \left(1 + T_{\mathrm{I}} s\right) \tag{6.3.10}$$

and pole-zero cancelation with $T_{\mathrm{I}} = L_{\mathrm{S}}/R_{\mathrm{S}}$ is applied, the closed-loop behavior is

$$G_{\mathrm{II}}(s) = \frac{\Delta I_{\mathrm{Sq}}(s)}{\Delta I_{\mathrm{Sqw}}(s)} = \frac{1}{1 + T_{\mathrm{q}} s} \tag{6.3.11}$$

with the closed-loop time constant

$$T_{\mathrm{q}} = \frac{L_{\mathrm{S}}}{K_{\mathrm{c}}}. \tag{6.3.12}$$

The closed-loop current behavior can therefore be described by a first order lag. The torque follows from (6.3.7). If a reference value for I_{Sd} different from zero is chosen ($I_{\mathrm{Sd}} \neq 0$), an additional positive magnetic field for $I_{\mathrm{Sd}} > 0$ or in the opposite direction an additional negative magnetic field for $I_{\mathrm{Sd}} < 0$ relative to the field generated by the permanent magnet is produced. According to the magnetization curve of the permanent magnet, the rotorflux can be adapted by specifying a certain value I_{Sdw}. In this manner field weakening is even possible. This has to be done carefully in order to avoid the degaussing of the permanent magnets, Vas (1990), Leonhard (1996), Novotny and Lipo (1996).

An analysis of the steady-state behavior of synchronous motors shows that the rotor displacement angle δ between the supply stator voltage and the pole position depends on the torque. With increasing torque the displacement angle increases until $\delta > 90°$. For $\delta = 90°$ the torque decreases until zero for $\delta = 180°$, see Leonhard (1996), Sarma (1996).

6.3.3 Model-based fault detection of a PMSM motor

The fault detection of the power electronics for PMSM motors follows the signal-based methods described in Sect. 6.2.2 for AC motors.

The development of fault-detection methods for the permanently excited motor depends on the available signals. Because of the field-oriented control following calculated or measured variables are usually available

- $U_{\mathrm{Sq}}, U_{\mathrm{Sd}}$ voltages of the q- and d-system
- $I_{\mathrm{Sq}}, I_{\mathrm{Sd}}$ currents of the q- and d-system
- $\omega_{\mathrm{el}} = \omega_{\mathrm{S}}$ supply angular frequency of the PWM inverter
- ω_{m} mechanical rotor angular speed, or
- φ_{m} mechanical rotor angle.

The currents I_{Sq} and I_{Sd} are determined by measurements of the stator currents I_{Sa} and I_{Sb} and use of the Clarke transformation and Park transformation, Leonhard (1996), Teigelkötter (2012).

Because of the similarity of the basic equations to generate parity equations to those of the AC motor described in Sect. 6.2.2 and 6.2.3, the same residuals $r_{\mathrm{q}}, r_{\mathrm{d}}, r_{12}, r_{23}, r_{31}$ and $|I_{\mathrm{so}}|$ can be calculated and similar symptoms are obtained as in Table 6.2.2. Hence, these symptoms are based on the q-current and d-voltage

of the field-controlled PMSM and on the stator currents: $I_{\mathrm{S1}} = I_{\mathrm{Sa}}$, $I_{\mathrm{S2}} = I_{\mathrm{Sb}}$, $I_{\mathrm{S3}} = I_{\mathrm{Sc}}$.

Another possibility is to use only the stator currents and voltages according to Keßler (2017). The stator current residuals can be determined by comparing the measured phase currents and modeled phase currents, which can be calculated by (6.3.3)

$$r_{I\mathrm{a}} = \hat{I}_{\mathrm{Sa}} - I_{\mathrm{Sa}} \tag{6.3.13}$$

$$r_{I\mathrm{b}} = \hat{I}_{\mathrm{Sb}} - I_{\mathrm{Sb}} \tag{6.3.14}$$

$$r_{I\mathrm{c}} = \hat{I}_{\mathrm{Sc}} - I_{\mathrm{Sc}}. \tag{6.3.15}$$

These residuals are in output error form, which means that they are less influenced by measurement noise. In addition three residuals in equation error form can be determined by the equations (6.3.1) to (6.3.3)

$$r_{U_{\mathrm{ab}}} = \Delta\hat{U}_{\mathrm{Sab0}} - \Delta\hat{U}_{\mathrm{ab}} = (\hat{U}_{\mathrm{Sa0}} - \hat{U}_{\mathrm{Sb0}}) - (\hat{U}_{\mathrm{Sa}} - \hat{U}_{\mathrm{Sb}}) \tag{6.3.16}$$

$$r_{U_{\mathrm{ac}}} = \Delta\hat{U}_{\mathrm{Sab0}} - \Delta\hat{U}_{\mathrm{ac}} = (\hat{U}_{\mathrm{Sa0}} - \hat{U}_{\mathrm{Sc0}}) - (\hat{U}_{\mathrm{Sa}} - \hat{U}_{\mathrm{Sc}}) \tag{6.3.17}$$

$$r_{U_{\mathrm{bc}}} = \Delta\hat{U}_{\mathrm{Sbc0}} - \Delta\hat{U}_{\mathrm{bc}} = (\hat{U}_{\mathrm{Sb0}} - \hat{U}_{\mathrm{Sc0}}) - (\hat{U}_{\mathrm{Sb}} - \hat{U}_{\mathrm{Sc}}). \tag{6.3.18}$$

Residuals in equation error form are more influenced by measurement noise. However, in each of the three residuals the influence of one phase is eliminated. Thus, they allow a fault isolation.

Because of the additional model of the inverter by (6.3.1) the motor model is more precise and the generated residuals (6.3.16 - 6.3.18) can detect even small faults relatively fast.

A further residual follows by comparison of the torques of the load $M_{\mathrm{load}}(n_{\mathrm{mech}})$ and the torque of the motor $M_{\mathrm{el}}(n_{\mathrm{mech}})$

$$r_{\Delta\mathrm{M}} = M_{\mathrm{el}}(n_{\mathrm{mech}}) - M_{\mathrm{load}}(n_{\mathrm{mech}}). \tag{6.3.19}$$

The torque of the motor results from (6.3.7) by measurement of I_{Sq} and the load torque from measured variables of the load if possible, compare Fig. 6.2.11.

The additional residuals (6.3.13) to (6.3.18) allow to detect also a demagnetization fault and a better isolation of the winding shorts.

Finally, Table 6.3.1 depicts the resulting fault-symptom relations.

Another case study for the fault diagnosis of an *electrical cooling fan drive* for combustion engines is treated in Pagel (2016). The considered synchronous motor consists of a three-phase internal stator and an external rotor with permanent ferrite magnets. The stator voltage is electronically commutated via a B6-bridge with three half bridges supplied by an automotive DC-battery-board net.

Based on dynamic models of the electrical motor, the driven fan and the thermal behavior of stator windings and permanent magnets and use of available measurements different faults could be experimentally detected, like winding and capacitor faults, demagnetization through overheating, and defective bearings, fan blades and blockage.

Table 6.3.1. Fault-symptom table with additional symptoms for the PMSM. + positive deflection; − negative deflection; +/− sign of deflection depends on sign of fault; +− increased variance; 0 no deflection.

Faults	Symptoms										
	$\lvert r_{I_a}\rvert$	$\lvert r_{I_b}\rvert$	$\lvert r_{I_c}\rvert$	$s_{I_{ab}} = r_{I_a} - r_{I_b}$	$s_{I_{ac}} = r_{I_a} - r_{I_r}$	$s_{I_{bc}} = r_{I_b} - r_{I_c}$	$\lvert s_{I_{ab}}\rvert + \lvert s_{I_{ac}}\rvert + \lvert s_{I_{bc}}\rvert$	$\lvert r_{U_{ab}}\rvert$	$\lvert r_{U_{ac}}\rvert$	$\lvert r_{U_{bc}}\rvert$	$\lvert r_{\Delta M}\rvert$
switching fault a	++	+	+	−/+	−/+	0	0	+	+	0	+
switching fault b	+	++	+	−/+	0	−/+	0	+	0	+	+
switching fault c	+	+	++	0	−/+	−/+	0	0	+	+	+
IGBT/diode fault a	++	+	+	−/+	−/+	0	0	+	+	0	+
IGBT/diode fault b	+	++	+	−/+	0	−/+	0	+	0	+	+
IGBT/diode fault c	+	+	++	0	−/+	−/+	0	0	+	+	+
winding short a	++	+	+	+	+	0	+	+	+	0	+
winding short b	+	++	+	+	0	+	+	+	0	+	+
winding short c	+	+	++	0	+	+	+	0	+	+	+
demagnetization	++	++	++	+−	+−	+−	0	+	+	+	+
change of friction	0	0	0	0	0	0	0	0	0	0	++

References

Filbert D (2003) Technical diagnosis for the quality control of electrical low power motors (in German). Technisches Messen 70(9):417–427

Fraser C, Milne J (1994) Electro-mechanical engineering – an integrated approach. IEEE Press, Piscataway, NJ

Füssel D (2002) Fault diagnosis with tree-structured neuro-fuzzy systems. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, 957. VDI Verlag, Düsseldorf

Höfling T (1994) Zustandsgrößenschätzung zur Fehlererkennung. In: Isermann R (ed) Überwachung und Fehlerdiagnose, VDI, Düsseldorf, pp 89–108

Höfling T (1996) Methoden zur Fehlererkennung mit Parameterschätzung und Paritätsgleichungen. Dissertation Technische Hochschule Darmstadt. Fortschr.-Ber. VDI Reihe 8, 546. VDI Verlag, Düsseldorf

Höfling T, Isermann R (1996) Fault detection based on adaptive parity equations and single-parameter tracking. Control Engineering Practice – CEP 4(10):1361–1369

Isermann R (2005) Mechatronic systems – fundamentals, 2nd edn. Springer, London

Isermann R (2006) Fault-diagnosis systems – An introduction from fault detection to fault tolerance. Springer, Heidelberg

Isermann R (2011) Fault-diagnosis applications. Springer, Heidelberg

Isermann R, Münchhof M (2011) Identification of dynamic systems. Springer, Berlin, Heidelberg

Isermann R, Lachmann KH, Matko D (1992) Adaptive control systems. Prentice Hall International UK, London

Keßler P (2015) Einstellbare Verkopplungsregelung für ein Duplex-Aktorsystem einer achsparallelen elektrischen Servolenkung (EPS). In: Bäker B, Unger A (eds) Diagnose in mechatronischen Fahrzeugsystemen IX, TUDpress, Dresden, pp 217–231

Keßler P (2017) Position fault tolerance for an electrical power steering system. In: Submitted to the IFAC World Congress, Toulouse, France

Leonhard W (1996) Control of electrical drives, 2nd edn. Springer, Berlin

Lyshevski S (2000) Electromechanical systems, electric machines, and applied mechatronics. CRC Press, Boca Raton, FL

Novotny D, Lipo T (1996) Vector control and dynamics of AC drives. Clarendon Press, Oxford

Pagel M (2016) Model-based diagnosis of electrical cooling fan drive systems. Internal report. Institute of Automatic Control and Mechatronics. Technische Hochschule Darmstadt

Pfeufer T (1994) Improvement of flexibility and reliability of automobiles actuators by model-based algorithms. In: IFAC SICICA, Budapest, Hungary

Sarma M (1996) Electric machines. Steady-state theory and dynamic performance. PWS Press, New York

Schröder D (1995) Elektrische Antriebe, vol 1: Grundlagen. Springer, Berlin

Schröder D (2009) Elektrische Antriebe - Regelung von Antriebssystemen, 4th edn. Springer, Berlin

Stölting H (2004) Electromagnetic actuators. In: Janocha H (ed) Actuators, Springer, Berlin

Teigelkötter J (2012) Energieeffiziente elektrische Antriebe: Grundlagen, Leistungselektronik, Betriebsverhalten und Regelung von Drehstrommotoren. Springer Vieweg

Thorsen O, Dalva M (1995) A survey of the reliability with an analysis of faults on variable frequency drives in industry. In: Proc. European Conference on Power Electronics and Applications EPE '95, pp 1033–1038

Vas P (ed) (1990) Vector control of AC machines. Clarendon Press, Oxford

Wolfram A (2002) Komponentenbasierte Fehlerdiagnose industrieller Anlagen am Beispiel frequenzumrichtergespeister Asynchronmaschinen und Kreiselpumpen. Fortschr.-Ber. VDI Reihe 8, 967. VDI Verlag, Düsseldorf

Wolfram A, Isermann R (2000) On-line fault detection of inverter-fed induction motors using advanced signal processing techniques. In: IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS´2000), Budapest, Hungary

Wolfram A, Isermann R (2001) Fault detection of inverter-fed induction motors using a multimodel approach based on neuro-fuzzy models. In: Proc. European Control Conference, Porto, Portugal

7

Diagnosis of actuators

Actuators usually transform low-powered manipulated variables (e.g. analog voltages $0...12\,\mathrm{V}$, applied DC currents $0...20\,\mathrm{mA}$ or $4...20\,\mathrm{mA}$, pneumatic pressures $0.2...1$ or $1...8\,\mathrm{bar}$, or hydraulic pressures $0...150\,\mathrm{bar}$) into process (engine) input variables of a much higher power level. Frequently the process input variable is a flow of energy or matter, or a force or torque. The power needed for actuating is provided by an auxiliary energy supply, which feeds the power amplifier for the actuator. The auxiliary energy can be electric, pneumatic or hydraulic.

Table 7.0.1 gives an overview of actuators and some drives for combustion engines. The application of electrical actuators has increased in the last two decades because of their precise and fast positioning. However, pneumatic actuators are still applied especially for VGT and wastegate turbochargers and for swirl flaps because of the hot environment and for reasons of cost. They are usually driven with low vacuum pressure $< 1\,\mathrm{bar}$. In the case of truck engines pneumatic actuators with up to $8\,\mathrm{bar}$ are used for turbochargers because of the large actuation forces. Hydraulic actuators are mainly applied for camshaft phasing.

In many cases the actuators are composed of a signal transformer, an actuator drive, an actuator transformer (gear, spindle) and an actuating device or valve, compare Fig. 7.0.1 and Janocha (2004). Actuators can operate in open loop or closed loop (e.g. position or flow-control). A survey of basic structures of actuators, different types, characteristics and mathematical models is given in Isermann (2005).

Actuators play an important role in modern power trains and vehicles. In the following sections the fault detection of a selection of electrical and fluidic actuators is considered.

7.1 Electric actuators

7.1.1 Electromagnetic actuator

Electromagnets are frequently used as actuators in combustion engines, transmissions, brakes, suspensions and in the body. One distinguishes switching magnets

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7_7

Table 7.0.1. Actuators and drives for combustion engines SI: gasoline engine, CI: diesel engine.

actuator type	components	air system	injection system	combustion system	exhaust system	cooling, lubrication	auxiliaries
A	electrical actuators	• el. throttle • flaps (charge, manifold)	• low pressure valve • high pressure valve (rail)	• valve lift (SI)	• VGT actuator • EGR valve	• coolant flow valve • heating flow valve	
	pneumatic actuators	• flaps (charge, motion)			• wastegate • VGT actuator • EGR valve		
	hydraulic actuators			• camshaft positioner			
B	injection valves		• electro-magnetic • piezoelectric		• urea injection (SCR), (CI) • diesel fuel (PF regeneration)		
	magnetic switching valves		• shutoff valve (CI)		• EGR-bypass valve • SCR-dosing unit (CI)		
C	electrical drives					• el. fan	• el. generator • el. starter
	electrical pumps	• air vacuum pump	• fuel pump		• secondary air pump (SI) • urea pump (CI)	• coolant pump	• air condition compressor
	mechanical pumps	• compressor	• injection pump			• oil pump	• air condition compressor • hydraulic steering pump

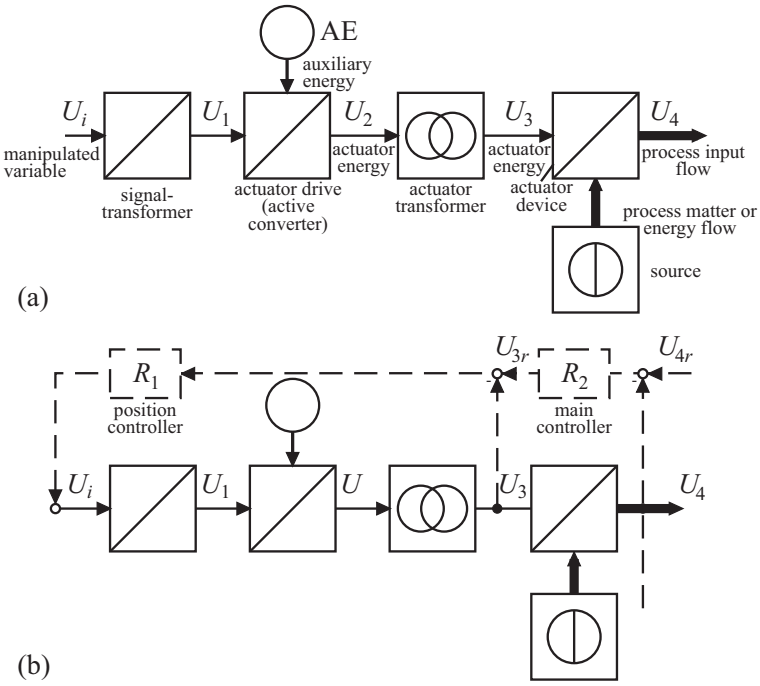


Fig. 7.0.1. Basic structures of actuators: a) open-loop controlled actuator; b) closed-loop controlled actuator.

and proportional magnets. The *switching magnets* have the task to switch, e.g. a valve from one end position to the other end position. Therefore, only switching and reaching of the end positions are of interest. *Proportional magnets* have to hold any position between the two end positions. They act usually against a spring, compare Fig. 7.1.1. Because of the required precise positioning they should have a linear characteristic between the input voltage U and the magnetic force F and small electrical and mechanical (frictional) hysteresis. Usually, the proportional actuator is operating in a closed loop for position control, where the controlled variable is the measured position of the armature or another output variable like pressure. To design the force characteristic $F(z)$ independently of the air gap the magnetic yoke is frequently designed conically around the operating point of the armature, Isermann (2005), Kallenbach et al (2008), see Fig. 7.1.1. The nonlinear force-position characteristic of a switching magnet, which has no constant characteristic $F(z)$, can also be linearized by a nonlinear compensation as shown in Fig. 7.1.2.

a) Position control

As an example a simple solenoid designed as in Fig. 7.1.1 is considered with the following data, Raab (1993):

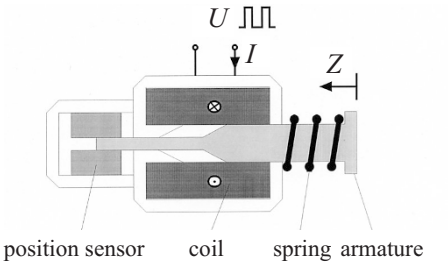


Fig. 7.1.1. Scheme of low-cost DC solenoid drive.

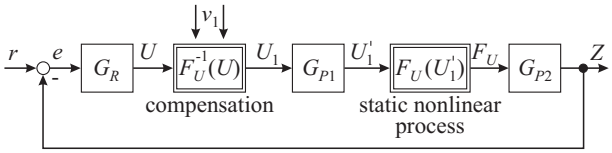


Fig. 7.1.2. General structure of a series correction (compensation) for nonlinear process static behavior within a position feedback control.

armature length	125 mm
armature diameter	25 mm
coil length	60 mm
coil resistance	$R = 22.4\,\Omega$
inductivity	$L(Z = 0)$ mm = 0.87 H
	$L(Z = 25)$ mm = 1.18 H
voltage	$U = 24$ V (DC)
spring constant	$c_{\mathrm{F}} = 1620$ N/m
position sensor	inductive, accuracy 0.5%
	measurement range: 40 mm
	time constant: 2.5 ms

An appropriate function for describing the nonlinear force–current characteristic, Kallenbach et al (2008), see Fig. 7.1.3, is the polynominal approximation

$$F(I, Z) = I \sum_{i=0}^{2} \frac{K_{\mathrm{i}}}{(Z_0 - Z)^i} \quad \text{with} \quad Z_0 = 26\,\mathrm{mm}. \tag{7.1.1}$$

The resulting statics of the linearized actuator are shown in Fig. 7.1.4 where a typical hysteresis characteristic becomes obvious. Its gradient represents the local gain K_{P} of the actuator, which can after linearization be assumed constant. The position-dependent width of the hysteresis characteristic is a measure for frictional forces and magnetic hysteresis.

In addition to the compensation of the nonlinear force characteristic an adaptive friction compensation can be applied, based on the parameter estimation of the Coulomb-friction coefficient, as shown in Isermann (2005), Isermann and Keller (1993) and Raab (1993).

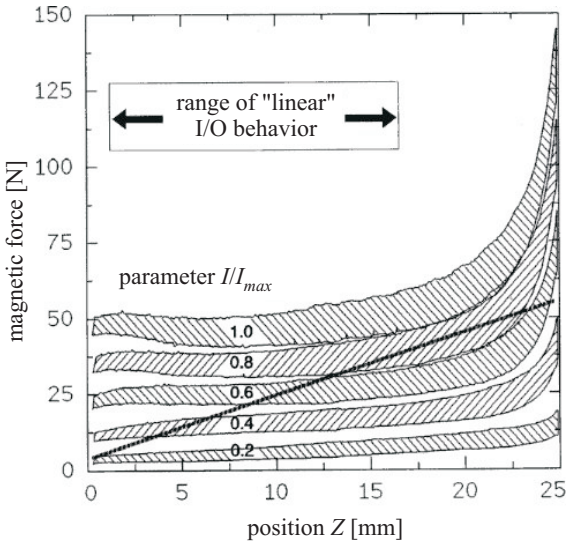


Fig. 7.1.3. Position-dependent nonlinear force–position characteristic of the (switching) solenoid drive. The straight line represents the linear spring characteristic.

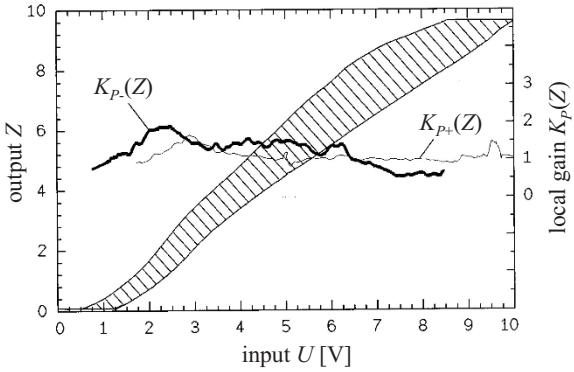


Fig. 7.1.4. Hysteresis characteristic and position-dependent local gain of the *linearized* (compensated) solenoid drive (1 V $\triangleq$ 2.5 mm).

The linearized system including the compensation of the nonlinear force-position characteristic can now be described by two equations. For the inner current circuit holds approximately

$$T_{\mathrm{I}}\dot{I}(t) + I(t) = K_{\mathrm{I}}U(t) \tag{7.1.2}$$

and for the mechanical subsystem with $\Delta Z = z$

$$m\ddot{z}(t) + d\dot{z}(t) + cz(t) = K_{\mathrm{Mag}}I(t) - F_{\mathrm{C}}\,\mathrm{sign}(\dot{z}) + F_{\mathrm{L}}(t). \tag{7.1.3}$$

The I/O behavior of the actuator can be modeled as a third-order system. The unknown parameters are obtained during a pre-identification phase, exciting the actuator with a special input signal with sampling time $T_0 = 2.5\,\text{ms}$. Taking the effect of Coulomb friction into account, the output error parameter estimation with (7.1.3) leads to the following direction-dependent (corresponding) transfer functions:

$$G_+(s) = \frac{z(s)}{U(s)} = \frac{382.4}{(s+116.4)\,(s^2+40.4\,s+3329.4)} e^{-0.0025\,s} \tag{7.1.4}$$

$$G_-(s) = \frac{z(s)}{U(s)} = \frac{220.0}{(s+47.9)\,(s^2+47.9\,s+3444.5)} e^{-0.0025\,s}. \tag{7.1.5}$$

The index $+/-$ denotes the direction of the armature motion and the additional dead-time describes the effect of asynchronous PWM (pulse width modulation) generation.

Figure 7.1.5 shows the obtained control performance using a numerical optimized discrete-time position controller

$$G_{\text{q}}\left(q^{-1}\right) = \frac{\Delta U(k)}{r(k)} = \frac{2.231 - 4.204q^{-1} + 2.000\,q^{-2}}{(1-q^{-1})\,(1-0.616q^{-1})} \tag{7.1.6}$$

of proportional-integral derivative (PID) type with first-order lag, $T_0 = 2.5\,\text{ms}$ where q^{-1} is a shift operator for one sampling time $u(k)q^{-1} = u(k-1)$. Although there is a change in the actuator's dynamic behavior, the controller designed for the slower negative motion (worst case) is robust enough for positive motions. The dynamic features are suitable and stability is obtained even in the positioning range $17\,\text{mm} < Z < 25\,\text{mm}$, where the behavior is unstable with linear control and uncompensated actuator characteristic.

b) Fault detection with parameter estimation

The electromagnetic actuator is now considered for the linearized operation range $0 - 25\,\text{mm}$. From the equations for the current circuit (7.1.2) and the mechanical subsystem (7.1.3) a third-order differential equation follows:

$$z^{(3)}(t) + a_2^*\ddot{z}(t) + a_1^*\dot{z}(t) + a_0^* z(t) = b_0^*\Delta U(t) + c_0^*(t). \tag{7.1.7}$$

The parameters of the continuous-time representation

$$\boldsymbol{\Theta}^T(t) = [a_2^* a_1^* a_0^* b_0^* c_0^*] \tag{7.1.8}$$

depend thereby on the physical process coefficients

$$\mathbf{p}^T = [T_1, D, \omega_0, K_{\text{P}}, c_0^*] \tag{7.1.9}$$

with, for example, Isermann (2005),

$$D = \frac{d}{2\sqrt{mc}}, \omega_0 = \sqrt{\frac{c}{m}}. \tag{7.1.10}$$

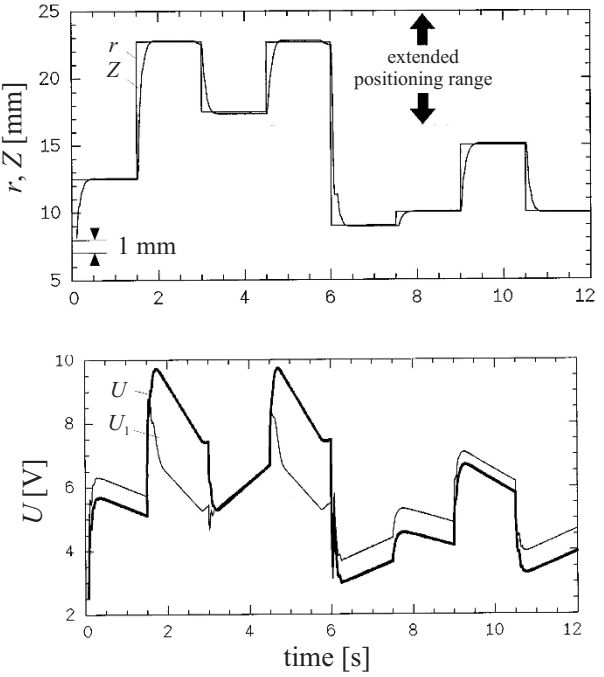


Fig. 7.1.5. Closed-loop position control of the solenoid with correction of the nonlinear actuator characteristic, but without friction compensation, $T_0 = 2.5$ ms. U_1; after compensation of the nonlinearity, Fig. 7.1.3.

These process coefficients can be expressed in terms of the parameter estimates $\boldsymbol{\Theta}$. Hence, after estimation of the model parameters $\boldsymbol{\Theta}$ by measuring the voltage U and the position Z all process coefficients $\mathbf{p}$ can be calculated, Raab (1993).

In the following, some experimental results are shown for artificially generated actuator faults:

F1: too large spring pretension
F2: decrease of the spring constant (by break or aging, change from c = 1650 to 1200 N m^{-1})
F3: increase of friction (increase of surface roughness and jamming)
F4: fault in the current circuit (weak controller gain).

The parameters were estimated by output error minimization using specific excitation signals. Sampling time was $T_0 = 0.2$ ms. Figure 7.1.6 and Table 7.1.1 show the results for different faults. Based on the deviations (symptoms) all faults can be identified. This can be performed by a pattern recognition or a systematic treatment of fault-symptom trees. In all cases different patterns of coefficient changes result. This enables a unique diagnosis of the four faults based on parameter estimation.

This means that an electromagnetic proportional acting electromagnet can be diagnosed based on measurements of the PWM voltage U and the position Z. How-

ever, the sampling time has to be in range of ms and, therefore, relatively small. Instead of the position another manipulated quantity like a pressure can be used. If the parameter estimation is reduced to steady-state conditions fault detection can be based on the static coefficients only, see Table 7.1.1, and the sampling time can then be larger.

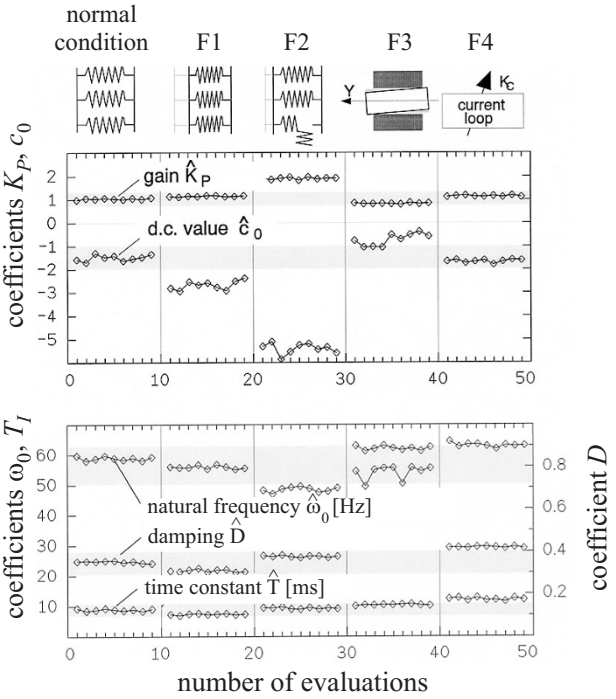


Fig. 7.1.6. Parameter estimates for an electromagnetic actuator with different faults (positive motion direction.)

Table 7.1.1. Changes of process coefficients for an electromagnet depending on different faults

	static coefficients		dynamic coefficients		
Fault type	K_{P+}	c_{0+}	ω_{0+}	D_+	T_1
F1	0	−−	0	0	0
F2	++	−−	−	+	0
F3	−	+	0	++	0
F4	0	0	0	+	++

c) Reconstruction of the armature position of a proportional magnet with voltage and current measurement

A simplified model of an electro-magnetic solenoid actuator is

$$U(t) = RI(t) + \frac{d\Psi(I,z)}{dt} \tag{7.1.11}$$

with dynamic changes of the flux linkage according to

$$\frac{d\Psi(I,z)}{dt} = \underbrace{\frac{\partial\Psi(I,z)}{\partial I}\frac{dI}{dt}}_{U_{\mathrm{I}}(t)} + \underbrace{\frac{\partial\Psi(I,z)}{\partial z}\frac{dz}{dt}}_{U_{\mathrm{z}}(t)}. \tag{7.1.12}$$

$U_{\mathrm{I}}(t)$ is the inducted voltage by the changing current and $U_{\mathrm{z}}(t)$ the induced voltage through the armature motion.

If the armature is not moving, it holds

$$\frac{d\Psi(I)}{dt} = \frac{d\Psi(I)}{dI}\frac{dI}{dt} = L_d\frac{dI}{dt} \tag{7.1.13}$$

where L_d is the differential inductivity.

Now (7.1.12) is represented by

$$\frac{d\Psi(I,z)}{dt} = L_{\mathrm{I}}(I)\frac{dI}{dt} + L_{\mathrm{z}}\frac{dz}{dt}. \tag{7.1.14}$$

$L_{\mathrm{I}}(t)$ is estimated experimentally for the non-moving armature, and approximated by a polynomial. L_{z} is modeled as a constant parameter. Then one obtains with (7.1.11)

$$U(t) = RI(t) + L_{\mathrm{I}}(I)\frac{dI}{dt} + L_{\mathrm{z}}\frac{dz}{dt} \tag{7.1.15}$$

which allows to calculate the armature position after a dynamic voltage change

$$z_{\mathrm{rec}}(t) = \frac{1}{L_{\mathrm{z}}}\int_0^t \left(U(t) - RI(t) - L_{\mathrm{I}}\frac{dI(t)}{dt}\right)dt + z_0. \tag{7.1.16}$$

For more details see Straky et al (1999), Moseler et al (2000). The initial position z_0 is determined from the initial values of the current and the flux linkage change. In Straky et al (1999) it is shown that a good agreement between a simulated model output and the reconstructed position is obtained with sampling time $T_0 \approx 0.5\,\mathrm{ms}$. Therefore, a residual

$$r_{\mathrm{z}}(t) = z_{\mathrm{sim}}(t) - z_{\mathrm{rec}}(t) \tag{7.1.17}$$

can be used to detect faults like increased friction, blockage or winding faults. The method requires a dynamic voltage change like a step function and uses only measurement of the voltage and the current, however, with small sampling time.

7.1.2 Electrical automotive throttle valve actuator

Since about 1990, electrically driven throttle valves became a standard component for gasoline engines. They control the air mass flow through the intake manifold to the cylinders. The electrical throttles are manipulated by the accelerator-pedal sensors via an electronic control unit and additional control inputs from idle-speed control, traction control and cruise control. In many vehicles it is the first drive-by-wire component, replacing the former mechanical linkage, Streib and Bischof (1996). Because the electrical throttles are safety-related components, reliability and safety are of high importance. In the following, a fault-detection and diagnosis method is described, developed by Pfeufer (1999).

a) Structure and models of the actuator

Figure 7.1.7 shows a cross-sectional view of the actuator. A permanently excited DC motor with brush commutation drives the throttle through a two-stage gear in the opening or closing direction. It operates against the main helic spring. A second spring works in the closing region in the opposite direction, in order to open the throttle in the case of voltage loss into a limp-home position (a mechanical redundancy). The motor is manipulated by a pulse-width-modulated (PWM) armature voltage $U_{\mathrm{A}}(-12 \text{ to } +12\,\mathrm{V})$. The measured variables are the armature voltage U_{A}, the armature current I_{A} and the angular throttle position $\varphi_K(0 \text{ to } 90°)$. This throttle position is measured by two redundant wiper-potentiometers operating in two different directions. Some technical data are given in Table 7.1.2. The position controller was a model-based sliding-mode controller or PID controller with time lag and sampling time $T_0 = 1.5\,\mathrm{ms}$, Pfeufer (1999), Pfeufer et al (1996).

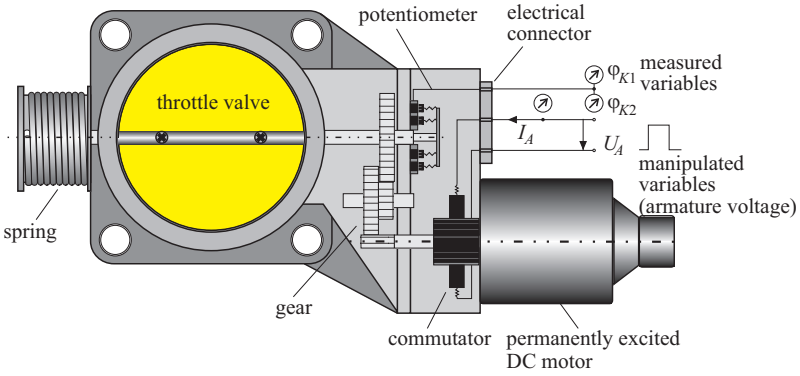


Fig. 7.1.7. Scheme of the electrical throttle.

Theoretical modeling of the throttle valve leads to the following basic equations, compare (7.1.1), (7.1.2):

Electrical part,

$$U_{\mathrm{A}}(t) = R_{\mathrm{A}} I_{\mathrm{A}}(t) + \Psi \omega_{\mathrm{A}}(t) + c_{0\mathrm{e}} \tag{7.1.18}$$

$$M_{\mathrm{el}}(t) = \Psi I_{\mathrm{A}}(t). \tag{7.1.19}$$

Mechanical part (related to motor axle),

$$\nu J \dot{\omega}_{\mathrm{k}} = M_{\mathrm{el}}(t) - M_{\mathrm{mech}}(t) \tag{7.1.20}$$

$$M_{\mathrm{mech}}(t) = \frac{1}{\nu}\left(c_{\mathrm{S1}} \varphi_{\mathrm{k}}(t) + M_{\mathrm{S0}} + M_{\mathrm{F}}\right) \quad (\varphi_{\mathrm{k}} > \varphi_{\mathrm{k0}}) \tag{7.1.21}$$

$$M_{\mathrm{F}}(t) = M_{\mathrm{F0}} \operatorname{sign} \omega_{\mathrm{k}}(t) + M_{\mathrm{F1}} \omega_{\mathrm{k}}(t). \tag{7.1.22}$$

The signals used are:

R_{A}	armature resistance
Ψ	magnetic flux linkage
ν	gear ratio ($\nu = 16.42$)
J	moment of inertia of the motor
M_{F0}	Coulomb-friction torque
M_{F1}	viscous-friction torque
c_{S1}	spring constant
M_{S0}	spring pretension
$\omega_{\mathrm{k}} = \dot{\varphi}_{\mathrm{k}}$	throttle angular speed
ω_{A}	motor angular speed. $\omega_{\mathrm{A}} = \nu \omega_{\mathrm{k}}$.

Table 7.1.2. Technical data of the electromagnetic throttle actuator (Bosch, DV-E4). Permanently excited DC motor with brushes. 1 pole pair, 12 commutation segments, 2 ball bearings; 12 V, 1.4 A, reversion current: 7.5 A; throttle diameter: 70 mm, needle bearings.

Parameter			Reference value
armature resistance	R_{A}	$[\Omega]$	1.2
inductance	L_{A}	[μ H]	600
magnetic flux linkage	Ψ	[Nm/A]	0.029
moment of inertia of the motor	J	[kg m^2/rad]	0.0000092
spring constant	c_{S1}	[Nm/ $^\circ$]	$0.002 - 0.0021$
spring pretension	M_{S0}	[Nm]	0.29 ± 0.03
Coulomb-friction torque	M_{F0}	[Nm]	ca. 0.18
gear ratio	ν	[-]	16.42

Compare the general equations for a DC motor in Section 6.1 and Fig. 7.1.8. The armature inductance can be neglected, because the electrical time constant $T_{\mathrm{el}} = L_{\mathrm{A}}/R_{\mathrm{A}} \approx 1$ ms is much smaller than the mechanical dynamics. The constant $c_{0\mathrm{e}}$ takes additive faults into account.

Dependent on the input excitation either the Coulomb friction or the viscous friction turned out to be dominant. Figure 7.1.9 depicts the principle of the applied fault detection and diagnosis.

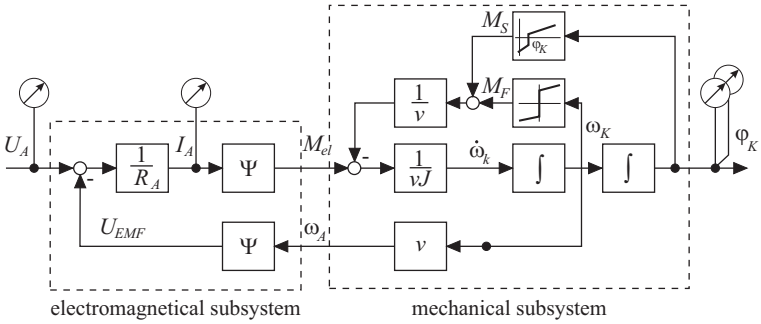


Fig. 7.1.8. Signal flow diagram of the electromagnetic throttle actuator.

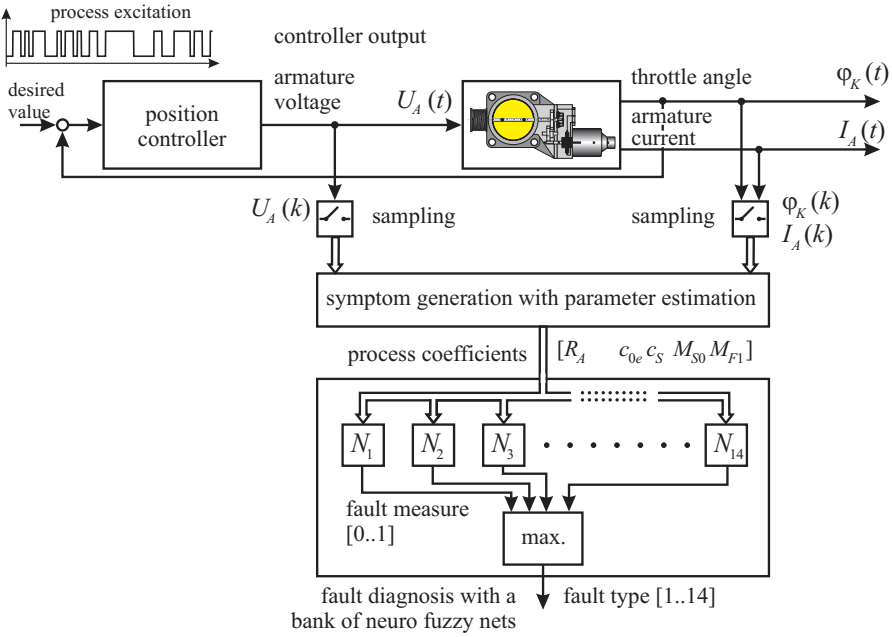


Fig. 7.1.9. Schematic of the fault-diagnosis system of the electromagnetic throttle actuator. N: normalization.

b) Input test cycle

In order to achieve optimal diagnostic results, the test cycle is composed of different phases, where each phase is designed to get a deep insight into a specific subsystem or technical component of the whole throttle valve actuator, see Fig. 7.1.9.

At the beginning of the test (phases 1…4) the throttle valve actuator is controlled in an open loop by directly manipulating the armature voltage of the DC motor. In this first step, an open circuit, a short circuit or a sneak path between the electrical

lines of the actuator are detected. Furthermore, offset faults in the measured signals armature voltage U_{A}, armature current I_{A} and throttle position φ_{K} are detected.

After this first stage the diagnosis proceeds in the closed-loop mode (phases 5 and 6). Now the manipulated variable of the diagnostic algorithm is given by the throttle position setpoint. This stage consists of two phases. In the first one, the whole range of operation is covered by a triangular setpoint. Herewith the redundantly measured position signals are checked for plausibility, and an insight into the mechanics of the actuator is gained by the estimation of some parameters of the mechanical subsystem. In phase 6, the test object is excited with a high dynamic signal in order to achieve a large variation of the speed of the DC motor. Then the parameters of the electrical subsystem of the motor are determined by continuous-time parameter-estimation algorithms. Furthermore, dynamic deviations from the nominal process behavior are detected with a parity equation using the model obtained from parameter estimation.

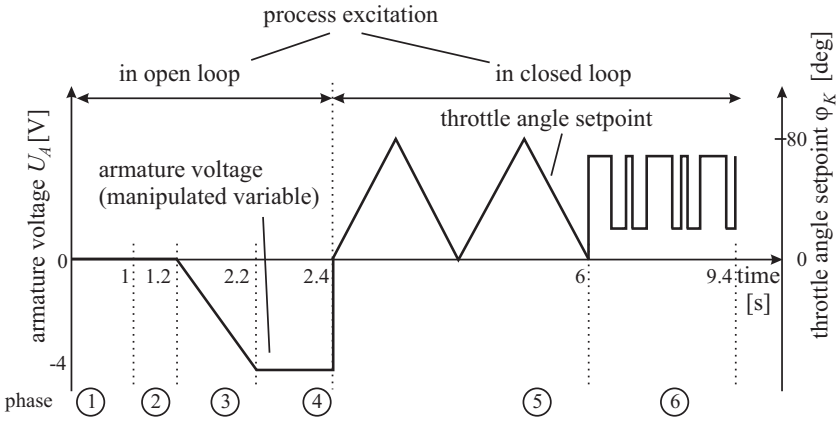


Fig. 7.1.10. Test cycle for the automatic fault diagnosis of the throttle valve (e.g. at service stations or end of production line).

Fault detection with parameter estimation

c) Parameter estimation for the dynamic behavior

The parameter estimation is carried out with recursive least-squares estimation in the form of discrete square-root filtering (DSFI), see e.g. Isermann and Münchhof (2011). The basic model equation is

$$y(t) = \boldsymbol{\psi}^T(t)\hat{\boldsymbol{\theta}} + e(t) \tag{7.1.23}$$

and the data vector and the parameter estimation for the electrical part are

$$y(t) = U_A(t) \tag{7.1.24}$$

$$\boldsymbol{\psi}^T(t) = [I_\mathrm{A}(t)\ \nu\omega_\mathrm{k}(t)1] \tag{7.1.25}$$

$$\hat{\theta}^T = [\hat{\Theta}_1\ \hat{\Theta}_2\ \hat{\Theta}_3] \tag{7.1.26}$$

and for the mechanical part

$$y(t) = \dot{\omega}_\mathrm{k}(t) \tag{7.1.27}$$

$$\boldsymbol{\psi}^T(t) = [I_\mathrm{A}(t)\ \varphi_\mathrm{k}(t)\ \omega_\mathrm{k}(t)1] \tag{7.1.28}$$

$$\hat{\theta}^T = [\hat{\Theta}_4\ \hat{\Theta}_5\ \hat{\Theta}_6\ \hat{\Theta}_7]. \tag{7.1.29}$$

Because of a fast input excitation the Coulomb-friction term is neglected and only the viscous-friction parameter M_F1 is estimated under the condition that $|\omega_\mathrm{k}| > 1.5\,\mathrm{rad/s}$.

The relation between the physical process coefficients and the parameter estimates are

$$\begin{aligned} &\hat{\Theta}_1 = R_\mathrm{A};\ \ \hat{\Theta}_2 = \Psi;\ \ \hat{\Theta}_3 = c_{0\mathrm{e}} \\ &\hat{\Theta}_4 = \frac{\Psi}{\nu J};\ \ \hat{\Theta}_5 = -\frac{c_\mathrm{S1}}{\nu^2 J};\ \ \hat{\Theta}_6 = -\frac{M_\mathrm{F1}}{\nu^2 J};\ \ \hat{\Theta}_7 = -\frac{M_\mathrm{S0}}{\nu^2 J}. \end{aligned} \tag{7.1.30}$$

As the gear ratio ν is known, the ratio of inertia follows from

$$J = \frac{\hat{\Theta}_2}{\nu\hat{\Theta}_4}. \tag{7.1.31}$$

All other process coefficients can be directly determined from the parameter estimates $\hat{\Theta}_\mathrm{i}$.

For the parameter estimation the actuator operates in closed loop and the setpoint is changed with a PRBS between 10 and 70 deg. The derivatives $\omega_\mathrm{k} = \dot{\varphi}_\mathrm{k}$ and $\dot{\omega}_\mathrm{k} = \ddot{\varphi}_\mathrm{k}$ are determined by a state-variable filter with sampling time $T_{0\mathrm{SVF}} = 2\,\mathrm{ms}$. The sampling time for the parameter estimation is $T_0 = 6\,\mathrm{ms}$. The resulting parameter estimates converge fast and the largest equation error is $\leq 5\%$ or $\leq 3.5\,\mathrm{deg}$ for the electrical part and $\leq 7...12\%$ for the mechanical part, Pfeufer (1999). Table 7.1.3 shows the deviations of the seven parameter estimates after introducing 14 different faults in several throttle actuators of the same type based on measurements of $U_\mathrm{A}, I_\mathrm{A}, \omega_\mathrm{k}$ and φ_k. All faults lead to different patterns, except F11 and F1, F2, which are not isolable.

Parameter estimation for the static behavior

In order to obtain more precise information on the mechanical part and especially the friction phenomena only the static behavior is considered for slow continuous input change according to a triangular upward and downward motion (phase 5 in Fig. 7.1.10). Setting $\dot{\omega}_\mathrm{k} = 0$ and neglecting the viscous friction (7.1.18) to (7.1.22) leads with $t = kT_0$ to

Table 7.1.3. Process parameter deviations for different actuator faults: 0: no significant change; +: increase; −: decrease.

		Features						
		parameter estimates						
	Faults	R_A	Ψ	$c_{0\mathrm{e}}$	J	$c_{\mathrm{S}1}$	$M_{\mathrm{F}1}$	$M_{\mathrm{S}0}$
F1	increase spring pretension	0	0	0	0	0	0	+
F2	decrease spring pretension	0	0	0	0	0	0	−
F3	commutator shortcut	−	−	0	+	+	+	0
F4	armature winding shortcut	0	−	0	+	+	+	0
F5	armature winding break	+	−	0	0	+	+	+
F6	additional serial resistance	+	0	0	0	0	0	0
F7	additional parallel resistance	−	−	0	0	+	+	0
F8	increased gear friction	0	0	0	+	+	+	0
F9	offset fault U_A	0	0	+/−	0	0	0	0
F10	offset fault I_A	0	0	−/+	0	0	0	+/−
F11	offset fault φ_k	0	0	0	0	0	0	−/+
F12	scale fault U_A	+/−	+/−	+/−	+/−	+/−	+/−	+/−
F13	scale fault I_A	−/+	0	0	+/−	+/−	+/−	+/−
F14	scale fault φ_k	0	−/+	0	−/+	−/+	−/+	−/+

$$\begin{aligned} I_\mathrm{A}(t) &= \frac{1}{\nu\Psi}\left(c_{\mathrm{S}1}\varphi_\mathrm{k}(k) + M_{\mathrm{S}0} + M_{\mathrm{F}0}\operatorname{sign}\omega_\mathrm{k}(k)\right) \\ &= \boldsymbol{\Psi}^T(k)\boldsymbol{\Theta}. \end{aligned} \tag{7.1.32}$$

Because of the direction-dependent Coulomb friction for the opening and closing two estimations are made:

$$\boldsymbol{\Psi}_1^T(k) = [\varphi_\mathrm{k}^+(k)\ \ 1] \qquad \boldsymbol{\Psi}_2^T(k) = [\varphi_\mathrm{k}^-(k)\ \ 1]$$

$$\hat{\boldsymbol{\Theta}}^+(k) = [\hat{\Theta}_1\ \ \hat{\Theta}_2] \qquad \hat{\boldsymbol{\Theta}}^-(k) = [\hat{\Theta}_3\ \ \hat{\Theta}_4]$$

with

$$\hat{\Theta}_1 = \frac{c_{\mathrm{S}1}}{\nu\Psi} \qquad \hat{\Theta}_2 = \frac{M_{\mathrm{S}0} + M_{\mathrm{F}0}}{\nu\Psi}\hat{\Theta}_3 = \frac{c_{\mathrm{S}1}}{\nu\Psi} \qquad \hat{\Theta}_4 = \frac{M_{\mathrm{S}0} - M_{\mathrm{F}0}}{\nu\Psi}. \tag{7.1.33}$$

The magnetic flux linkage Ψ is known from (7.1.30). The physical process parameters then result as

$$\begin{aligned} c_{\mathrm{S}1} &= \nu\Psi\frac{\hat{\Theta}_1 + \hat{\Theta}_3}{2} \\ M_{\mathrm{S}0} &= \nu\Psi\frac{\hat{\Theta}_2 + \hat{\Theta}_4}{2} \\ M_{\mathrm{F}0} &= \nu\Psi\frac{\hat{\Theta}_2 - \hat{\Theta}_4}{2}. \end{aligned} \tag{7.1.34}$$

The parameter estimation is performed with recursive DSFI and $T_0 = 6\,\mathrm{ms}$ for each motion based on measurements of I_A and φ_k. Figure 7.1.11 shows the results for

a fault-free case. The spring pretension M_{S0} leads to a positive offset of the linear spring characteristic and the dry friction shifts the friction characteristic by M_{F0}^{+} and M_{F0}^{-} such that a hysteresis characteristic results. A comparison with the electrical torque $M_{\mathrm{el}}' = \nu\Psi I_{\mathrm{A}}$ related to the throttle axle indicates a good agreement with the estimated hysteresis characteristic. Changes of the spring constant cause a change of the slope c_{S1} and changes of the friction or the pretension lead to shifts of the characteristics. (The oscillations of the calculated electrical torque are due to the closed-loop behavior in connection with adhesive friction or stick–slip effects, which are not modeled. The range around the point of returns, where adhesion works, is omitted in the parameter estimation for simplifying reasons).

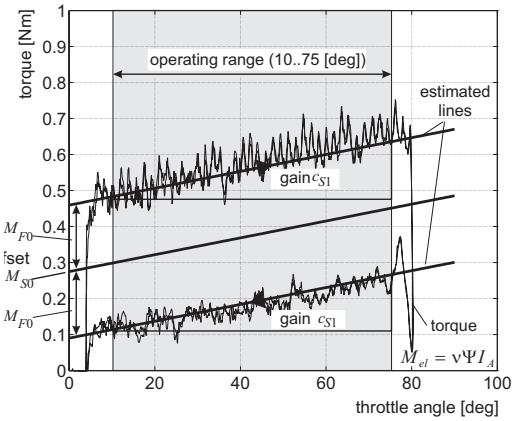


Fig. 7.1.11. Estimated static behavior of the throttle actuator (fault-free case).

c) Fault detection with parity equations

The application of several parity equations with structured residuals according to Sect. 6.1.2 was not successful because the residuals have shown too much variance due to the difficulties in modeling the dynamic behavior of the mechanical system. This was especially due to the various kinds of dry and viscous friction and stick–slip effects of the motor brushes, the two-stage gear and pretensioned spring. Therefore only one parity equation is used for the electrical part. (7.1.18) leads to the voltage residual

$$r(t) = U_{\mathrm{A}}(t) - R_{\mathrm{A}} I_{\mathrm{A}}(t) - \Psi\nu\omega_{\mathrm{k}}(t). \tag{7.1.35}$$

The residual is calculated in continuous time and $\omega_{\mathrm{k}}(t) = \dot{\varphi}_{\mathrm{k}}(t)$ is determined by a state-variable filter (SVF). Because the SVF has to be applied to all measurements, all signals are low-pass filtered with an SVF with equal parameters. Simulations indicate that the residual shows a short burst for a sudden position sensor offset

and an increase of the variance after changing the resistance R_A or the gain of the position sensor, after about 200 ms, Pfeufer (1999).

However, the parameters R_A and Ψ change with temperature. The ranges are for -40 to $+120°\mathrm{C}$, $R_\mathrm{A} = 1.0 \ldots 1.81\,\Omega$ (rated value 1.3 Ω) and $\Psi = 0.0314 \ldots 0.0224$ Vs (rated value 0.028 Vs). This means that these parameters should be taken from parameter estimation. An alternative is to apply adaptive parity equations, e.g. described in Höfling (1996), Isermann (2006) and Pfeufer (1999). The armature resistance changes follow from

$$\Delta\hat{R}_\mathrm{A} = \frac{\sum_{\mathrm{i}=0}^{N} \lambda^j I_\mathrm{A}(k-i) r(k-i)}{\sum_{\mathrm{i}=0}^{N} \lambda^j I_\mathrm{A}(k-i) I_\mathrm{A}(k-i)} \tag{7.1.36}$$

where $\lambda \leq 1$ is a forgetting factor. $\hat{R}_\mathrm{A}$ is updated after a threshold ΔR_Ath is exceeded. As there exists a linear relationship between $\Psi(\vartheta)$ and $R_\mathrm{A}(\vartheta)$ under the influence of the temperature $\vartheta, \Psi(R_\mathrm{A})$ can be stored in a look-up table.

The theoretical and experimental investigations have shown that a detailed fault detection is especially possible with parameter-estimation methods. The parameter estimation of the static behavior gives best results for the mechanical part. Parameter estimation for the dynamic behavior allows one to detect faults in the mechanical part as well as the electrical part. Parity equations were only successful for the electrical part in combination with parameter estimation. The application of the parity equation for online real-time detection of position sensor faults and reconfiguration of the position control is described in Pfeufer (1999). Table 7.1.4 summarizes the conditions and applicability for the application of the fault-detection methods.

Table 7.1.4. Conditions and applicability for the application of fault-detection methods of the throttle actuator.

Data evaluation	Parameter estimation dynamic	Parameter estimation static	Parity equation	Application
online real-time	dynamic excitation	not suitable	with parameter estimation	onboard
offline	dynamic excitation	rampwise test signals	with parameter estimation	quality control

d) Fault diagnosis

The experiments with the test cycle allows one to generate 30 different symptoms mainly by parameter estimation. In order to perform a fault diagnosis the observed symptoms can either be compared with a fault-symptom table, Table 7.1.3, or systematically evaluated with a fuzzy-logic diagnosis system, as described in Isermann (2006). The basis consists of fuzzy IF-THEN rules like

$$\mathrm{IF}\ \{s_1 \text{ is } A_{11}\ \mathrm{AND}\ s_2 \text{ is } A_{21}\}\ \mathrm{THEN}\ \{f \text{ is } F_1\}\,.$$

Some examples are given in Fig. 7.1.12. The membership functions of the attributes (symptoms) have simple shapes like triangles, trapezoids or ramps. Fuzzy-logic operators for the premise are min/max compositions. As it is sufficient to have one rule for each fault and to represent the faults as singletons $f_j \in 0 \ldots 1$, the degree of the fulfillment of the premise gives directly the degree of the conclusion, such that more complicated evaluations of aggregation and defuzzification are avoided. The diagnosis system was tested with different actuators and allowed to diagnose 38 different implemented faults (by measurement of only three signals).

In addition a learning neuro-fuzzy system SARAH (system for adaptive rule acquisition with Hebbian learning), Ayoubi (1996), was applied with a classification rate close to 100% for 22 rules.

A test equipment based on a PC and DSP signal processing cards is described in Pfeufer (1999) and Isermann (2011).

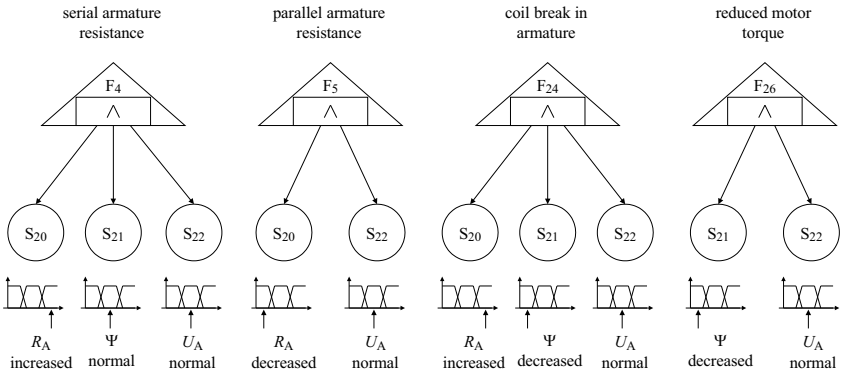


Fig. 7.1.12. Extract of fault-symptom relations with membership functions for the symptoms.

e) Conclusions

The developed fault-detection and diagnosis methods for the throttle actuator can be applied as follows:

- quality control during manufacturing (e.g. end-of-line test) with special test cycles or for problematic returned actuators (offline data processing)
- function test and troubleshooting in service stations with built-in actuator (offline data processing)
- onboard fault detection during normal operation (online)
- onboard fault detection and control reconfiguration with redundant position sensors (fault tolerance).

For a comprehensive fault detection and diagnosis of the electrical and mechanical part mainly symptom generation with parameter estimation is suitable. Parity equations could only be applied for the electrical part.

The described methods for the example of an electrical throttle can be also used for other DC motor driven actuators which act on a mechanical part.

7.1.3 Brushless DC motor

a) Structure and models

Brushless DC motors with permanent magnetic excitation are increasingly applied for actuators in automobiles. The following example shows the results of a case study for an actuator system in the fuselage of passenger aircraft, where the air pressure control is realized with DC motor driven outflow valves. The outflow valve is made fault tolerant by two brushless DC motors which operate over the gear to a lever mechanism moving the flap, Figure 7.1.13.

The two DC motors form a duplex system with dynamic redundancy and cold standby. Therefore, a fault detection for both DC motors is required to switch from the possibly faulty one to the standby motor.

In the following, it is shown how the fault detection of a brushless DC motor can be realized by combining parameter estimation and parity equations with implementation on a low-cost microcontroller, Moseler (2001), Moseler et al (1999), Moseler and Isermann (2000), Moseler and Müller (2000). The brushless DC motor in combination with a gear and a moved mechanical part is another frequently used actuator principle for automobiles, e.g. for EGR or VGT actuators.

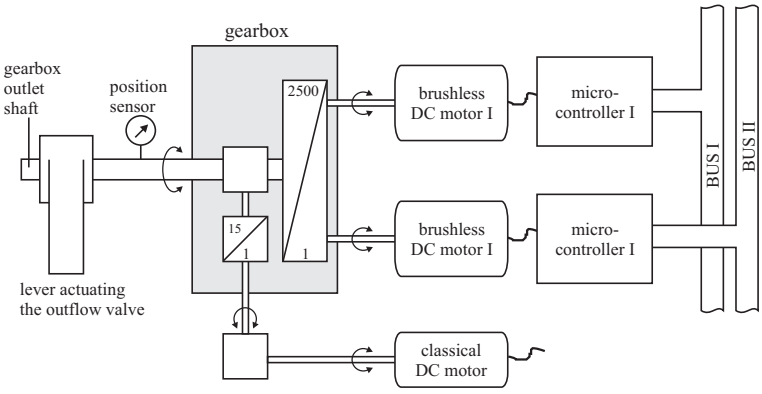


Fig. 7.1.13. Redundant DC motor drive system for the outflow valve.

Figure 7.1.14 depicts the structure of the electronic commuted DC motor. The stator possesses three coils which are Y-connected and driven by a PWM (pulse-width modulation) inverter. The rotor has four permanent magnets. The position of the rotor magnets is measured by three Hall sensors mounted on the stator. These determine the switching sequence of six MOSFET transistors of the PWM inverter. This switching scheme is implemented in a separated programmable logic array. The

PWM inverter obtains a fixed voltage U_B and the resulting current I_B from the DC power supply and generates square wave voltages through the commutation logic via the six transistors to the three coils (phases).

The advantage of the electronic commutation is that no brushes exist, which are subject to wear and are a source of electromagnetic disturbances. Therefore, the reliability is relatively high. Possible faults in the brushless DC motor may originate from hall sensors, commutation circuits and transistors (overheating), stator coil windings, mechanical defects of bearings and magnets (eccentricity, striping), and electromagnetic disturbances. Usually, only measured signals for the supply voltage U_B, the input current I_B of the six-phase full bridge circuit and the angular rotor speed ω are available.

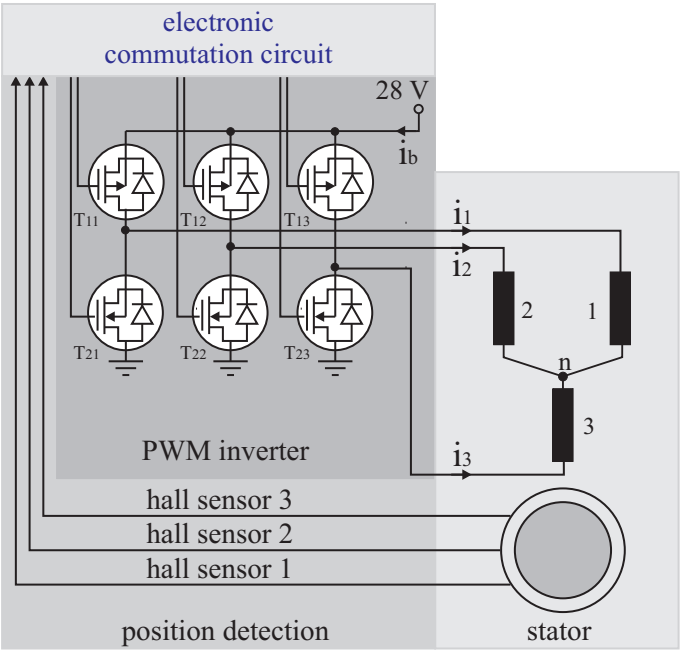


Fig. 7.1.14. Scheme of the brushless electronic commutated DC motor with PWM inverter (B6-bridge circuit).

A detailed model of the brushless DC motor for all three phases is given in Isermann (2005) and Moseler et al (1999). It could be shown that for the case of fault detection averaged values (by low-pass filter) of the voltage $U_\mathrm{B}(t) = U(t)$ and the current $I_\mathrm{B}(t) = I(t)$ to the stator coils can be assumed. This leads to the voltage equation of the electrical subsystem, compare Fig. 7.1.8

$$U(t) - k_\mathrm{E}\omega_\mathrm{r}(t) = RI(t) \tag{7.1.37}$$

with R the overall resistance and k_{E} the magnetic flux linkage. The generated rotor torque is proportional to the effective magnetic flux linkage $k_{\mathrm{T}} < k_{\mathrm{E}}$.

$$M_{\mathrm{r}}(t) = k_{\mathrm{T}} I(t). \tag{7.1.38}$$

(In ideal cases $k_{\mathrm{E}} = k_{\mathrm{T}}$.) The mechanical part is then described by

$$J_{\mathrm{r}} \dot{\omega}_{\mathrm{r}}(t) = k_{\mathrm{T}} I(t) - M_{\mathrm{f}}(t) - M_{\mathrm{L}}(t) \tag{7.1.39}$$

with the ratio of inertia J_{r}, the Coulomb friction torque

$$M_{\mathrm{f}}(t) = c_{\mathrm{f}} \operatorname{sign} \omega_{\mathrm{r}}(t) \tag{7.1.40}$$

and the load torque $M_{\mathrm{L}}(t)$. The gear ratio ν of the gear between the DC motor and the lever of the outflow valve, see Figure 7.1.13, relates the motor shaft position φ_{r} to the flap position φ_{g}

$$\varphi_{\mathrm{g}} = \varphi_{\mathrm{r}} / \nu \tag{7.1.41}$$

with $\nu = 2500$. The load torque of the flap is a function of the position φ_{g}

$$M_{\mathrm{L}} = c_{\mathrm{s}} f(\varphi_{\mathrm{g}}) \tag{7.1.42}$$

and is approximately known around the steady-state operation point. (For the experiments the flap was replaced by a lever with a spring.) The resulting signal flow is similar to a DC motor, as shown in Fig. 7.1.8.

b) Fault detection with parameter estimation

For fault detection the following measurements are available: $U(t)$, $I(t)$, $\omega_{\mathrm{r}}(t)$, $\varphi_{\mathrm{g}}(t)$. Using the notation

$$y(t) = \boldsymbol{\psi}^T(t) \boldsymbol{\theta} \tag{7.1.43}$$

two equations were used for *parameter estimation*:

- electrical subsystem,

$$y(t) = U(t), \boldsymbol{\psi}^T(t) = [I(t)\, \omega_{\mathrm{r}}(t)]; \quad \boldsymbol{\theta}^T = [R\, k_{\mathrm{E}}] \tag{7.1.44}$$

- mechanical subsystem,

$$\begin{aligned} y(t) &= k_{\mathrm{T}} I(t) - c_{\mathrm{s}} f(\varphi_{\mathrm{g}}(t) - J_{\mathrm{r}} \dot{\omega}_{\mathrm{r}}(t)) \\ \boldsymbol{\psi}^T(t) &= [\operatorname{sign} \omega_{\mathrm{r}}(t)];\ \boldsymbol{\theta}^T = [c_{\mathrm{f}}] \ \ (J_{\mathrm{r}} \text{ known}). \end{aligned} \tag{7.1.45}$$

Hence, three parameters $\hat{R}$, $\hat{k}_{\mathrm{E}}$ and $\hat{c}_{\mathrm{f}}$ are estimated. Various parameter estimation methods were applied like: RLS (recursive least squares), DSFI (discrete square-root filtering), FSDFI (fast DSFI), NLMS (normalized least mean squares) and compared with regard to computational effort in floating-point and integer word realization and estimation performance. The floating-point implementation is standard for, e.g. 16-bit signal processors and in this case RLS, DSFI or FDFSI can be used. However, integer word implementation is (still) required if reliable and certified low-cost microcontrollers like the 16-bit Siemens C 167 have to be used. Then, only NLMS is feasible, Moseler (2001).

c) Fault detection with parity equations

The *parity equations* are obtained from the basic two equations (7.1.37) and (7.1.39) by assuming known parameters (obtained from parameter estimation)

$$r_1(t) = U(t) - RI(t) - k_\mathrm{E}\omega_r(t) \tag{7.1.46}$$

$$r_2(t) = k_\mathrm{T}I(t) - J_\mathrm{r}\dot{\omega}_\mathrm{r}(t) - c_\mathrm{f}\,\mathrm{sign}\,\omega_\mathrm{r}(t) - c_\mathrm{s}f(\varphi_\mathrm{g}) \tag{7.1.47}$$

$$r_3(t) = U(t) - \frac{R}{k_\mathrm{T}}(J_\mathrm{r}\dot{\omega}_\mathrm{r}(t) + c_\mathrm{s}f(\varphi_\mathrm{g}) + c_\mathrm{f}\,\mathrm{sign}\,\omega_\mathrm{r}(t) + k_\mathrm{E}\omega r(t)) \tag{7.1.48}$$

$$r_4(t) = \varphi_\mathrm{g}(t) - \varphi_\mathrm{r}(t)/\nu. \tag{7.1.49}$$

Each of the residuals is decoupled from one measured signal. r_1 is independent from φ_g, r_2 from U, r_3 from I, r_4 from all but φ_r. (φ_r is assumed to be correct. It can directly be supervised by a logic evaluation within the motor electronics.) Figure 7.1.15 shows measured signals, parameter estimates and residuals for five different implemented faults.

The actuator was operating in closed loop with slow triangle changes of the reference variable (setpoint). The fault-detection methods, including state-variable filters (SVF) were implemented on a digital signal processor TI TMS 320 C40 with signal sampling period $T_0 = 1$ ms. The results for fault detection are summarized in Table 7.1.5. The sign and size of changes for the parameter estimates with FDSFI clearly allow one to identify the parametric faults and for the parity residuals the respective additive (offset) sensor faults. But there are also cross-couplings: for parametric faults some residuals show changes and for sensor-additive faults some parameter estimates change (except for φ_g), which can all be interpreted by the equations used. According to Gertler (1998) the symptom pattern is weakly isolating as a parametric fault of R and an additive fault in U differ only in one symptom. However, all faults can be isolated. Including the standard deviation of the symptoms isolability can be improved, Moseler (2001). By processing eight symptoms with a rule-based fuzzy-logic diagnosis system, finally 10 different faults could be diagnosed, Moseler (2001), Moseler and Vogt (2000).

Table 7.1.5. Parameter deviations and parity equation residuals for different actuator faults (0 no significant change; + increase; ++ large increase; − decrease; −− large decrease).

	Parameter estimates			Residual parity equations			
Faults	$\hat{R}$	$\hat{k}_\mathrm{E}$	$\hat{c}_\mathrm{f}$	r_1	r_2	r_3	r_4
increasing R	+	0	0	+	0	+	0
increasing c_f	0	0	++	0	−−	++	0
offset U	+	+	0	++	0	++	0
offset φ_g	0	0	0	0	0	0	0
offset I_b	++	−	−−	++	++	0	0

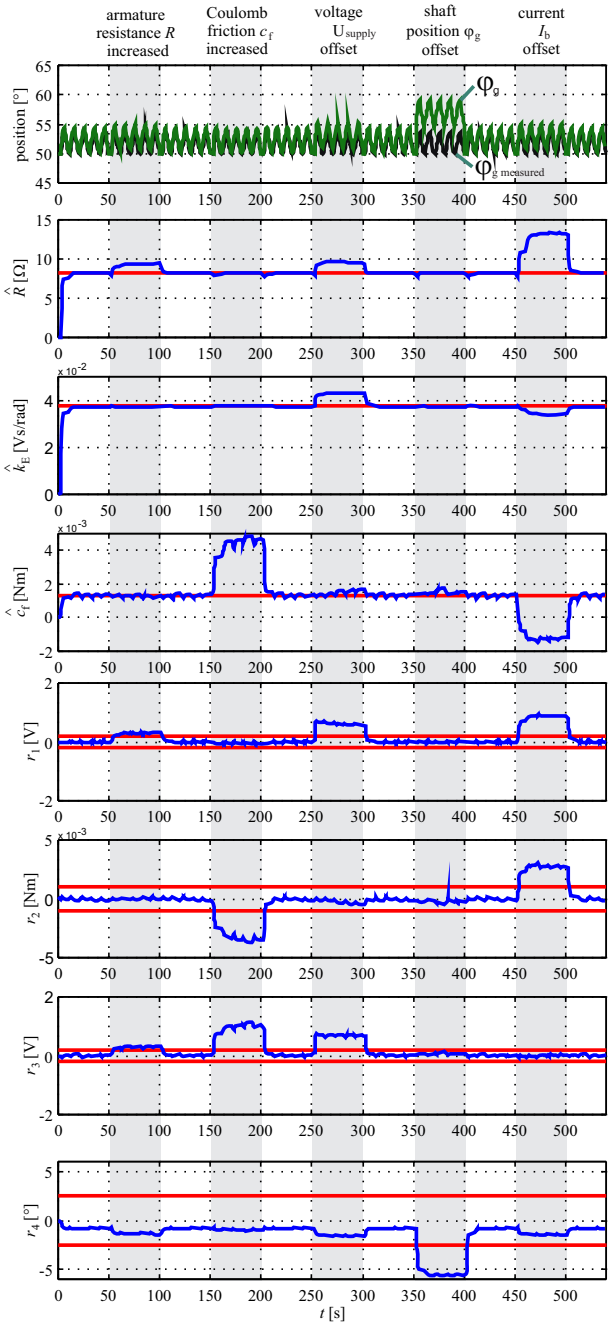


Fig. 7.1.15. Resulting symptoms from parameter estimation and parity equations by measuring $U(t)$, $I(t)$, $\omega(t)$, $\varphi_g(t)$ and $\varphi_r(t)$.

Because the position sensors of the rotor φ_r and the shaft φ_g yield redundant information, sensor fault detection for φ_g was used to reconfigure the closed loop after failure of φ_g by using φ_r as the control variable, Moseler (2001). The described combined fault-detection methodology needs about 8 ms calculation time on a 16-bit microcontroller. Therefore, online implementation in a smart actuator is possible by only measuring four easy accessible variables U, I and ω_r and φ_g.

d) Conclusions

The fault-diagnosis approach of this actuator has shown that the brushless DC motor can be modeled as for DC motors with brushes. Parameter estimation is primarily suitable to detect parametric faults, and parity equations to detect additive faults. If the input signal U stays approximately constant, only parity equations should be applied, which then may indicate faults. Then for isolating or diagnosing the faults a test signal on U can be applied for short time to gain deeper information. Hence, by applying both parameter estimation and parity equations a good fault coverage can be obtained.

7.2 Pneumatic actuators

Pneumatic actuators attached to combustion engines are mainly applied for charge motion flaps, EGR valves, wastegate and VGT turbochargers. Their advantage is the robustness, especially for high temperatures and relatively low weight and cost. In passenger cars they usually operate with vacuum pressure < 1 bar, either taken from the intake manifold for gasoline engines or from rotary pumps for diesel engines. This vacuum supply pressure is anyhow available if pneumatic brake boosters are installed. Pneumatic actuators for truck engines allow to operate with higher air pressure of e.g. 8 bar which is available from the pneumatic brake system. It enables considerable larger actuation forces.

7.2.1 Design of pneumatic actuators

The general set-up and mathematical model of pneumatic actuators are described e.g. in Goedecke (1989), Raab (1993), Tang and Walker (1995), Backé and Klein (2004), Janocha (2004), Isermann (2005), Schwarte (2007) and with application to fault detection in Isermann (2011). Figure 7.2.1 depicts a schematic of typical pneumatic actuators, either with a double-rod cylinder or a diaphragm drive with spring return.

The supply pressure p_0 can either be a vacuum $p_\mathrm{s} \approx 0.2...0.8$ bar from the intake manifold or generated by a vacuum pump (passenger cars) or $p_\mathrm{s} \approx 2...8$ bar delivered by an air compressor (trucks).

The electro-pneumatic servo valves are either proportional-acting valves or switching valves. *Proportional-acting servo valves* allow a continuous manipulation of the

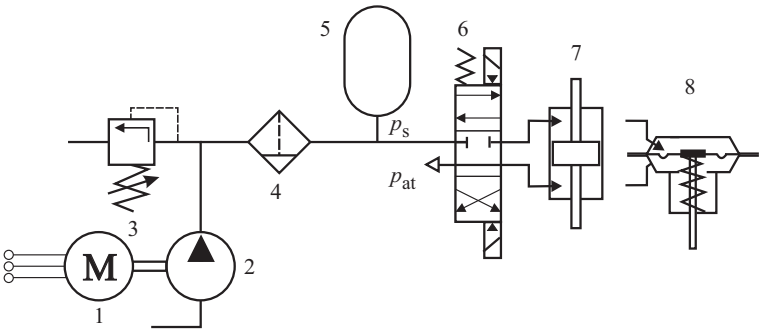


Fig. 7.2.1. Schematic of a pneumatic actuator for linear motion with power supply 1: Electrical motor 2: air pump or compressor 3: pressure relief valve (if required) 4: air filter with water trap (if required) 5: air storage (accumulator) 6: $4/3$ proportional valve with electrical solenoid actuation and spring return 7: double-rod cylinder 8: diaphragm drive with spring return.

airflow. They are manipulated electrically, e.g. by one or two solenoids as shown in Fig. 7.2.1. *Switching servo valves* are electromagnetic devices which operate with pulse-width modulation (PWM) at high frequencies in order to generate a certain position through their low-pass behavior and therefore also a quasi-continuous manipulation. Hence, these servo valves function as electro-pneumatic transmitters. The valve stem of the pneumatic actuator acts either directly or over a lever design on an air flap, vane manipulation mechanism of a VGT turbocharger, by-pass valve of a wastegate turbocharger or EGR valve.

Pneumatic valves have in principle a design as shown in Fig. 7.2.2. The pneumatic subsystem comprises a chamber sealed by a diaphragm that is acting on the valve stem. For the case of flow valves a body (plug) is mounted at the tip of the valve stem which in conjunction with its counterpart, the valve seat, controls the flow. Depending on the precision accuracy and the kind of fluid, different geometries are used. The stem passes through a gland (e.g., stuffing box) in order to seal the system.

Figure 7.2.2 depicts a signal flow for a cascaded control system with a minor and a main controller of a pneumatic actuator. If the position z_a of the pneumatic actuator is measured an usually proportional acting position controller G_{C1} is used, which provides a voltage U_1 for the servovalve. This servovalve then generates a chamber pressure p_{a1} for the actuator. Depending on the use of a vacuum pump or an air compressor the supply pressure p_s is then $p_\mathrm{s} < p_\mathrm{at}$ or $p_\mathrm{s} > p_\mathrm{at}$ where p_at is the atmospheric pressure. The other chamber of the diaphragm (or piston) can then be open to the atmosphere. The (minor) position controller obtains its reference value from the main controller G_{C2} which controls the intended main variable, e.g. an (EGR) flow $\dot{V}$ or a (charging) pressure p.

The measurement and control of the position z of the pneumatic actuator is advantageous for the control performance, because the position controller compensates hysteresis effects through friction and changing external forces on the valve stem.

In this case, according to Fig. 7.2.2, following variables can be used for control and diagnosis: W_2, $\dot{V}$ or p, and z. If z_a cannot be measured electronically only W_2, $\dot{V}$ or p can be used.

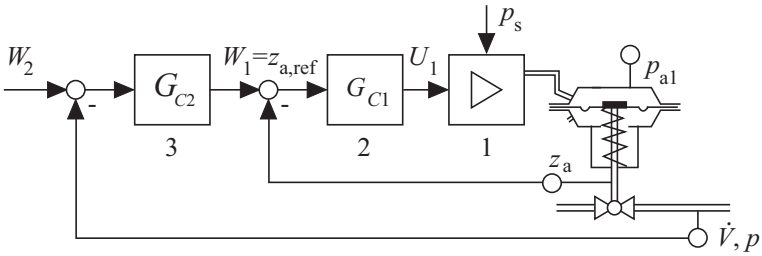


Fig. 7.2.2. Pneumatic flow valve with position and flow or pressure controller. p_1: diaphragm chamber pressure, p_s: supply air pressure; z_a stem position, p fluid pressure, $\dot{V}$ flow rate, 1: servo valve, 2: position controller, 3: flow rate or pressure controller.

7.2.2 Models of pneumatic actuators

A $3/3$ way air flow servo valve and the diaphragm chamber of a pneumatic actuator are presented in a simplified scheme in Fig. 7.2.3. If the servo valve opens to a vacuum pressure, supply air flows out of the chamber with mass flow $\dot{m}_1$ (bleeding) and the chamber pressure p_{a1} lowers such that the difference pressure $(p_{a2} - p_{a1})$ with $p_{a2} > p_{a1}$ moves the diaphragm and the actuator stem upward against a spring. p_{a2} is e.g. the atmospheric pressure $p_{a2} = p_a$ and $p_{a1} = p_1$ is the pressure after valve. The spring has such a pretension that the diaphragm only moves for $p_{a1} < p_{a10}$ (e.g. $p_{a10} = 0.7$ bar) such that the actuator does not open in case of external forces acting on the actuator stem. If the servo valve opens a channel to the atmosphere, p_1 increases and the spring force moves the stem downwards. The voltage of the electromagnet is pulse-width-modulated with high frequency, such that the resulting pressure p_1 is a function of the average effective voltage U_1. Figure 7.2.4 depicts a resulting characteristic $z_a(p_{a1})$ where the full range of the position is manipulated for vacuum pressures $0.4 < p_{a1} < 0.7$ bar.

The basic mathematical models of pneumatic actuator components are derived in the following according to, e.g. in Raab (1993), Isermann (2005), Schwarte (2007). The mass flow through the *servo valve* can be described by the flow through an orifice

$$\dot{m}_1 = A_v \Psi p_1 \sqrt{\frac{2}{RT_1}} \tag{7.2.1}$$

with the outflow function

$$\Psi = \sqrt{\frac{\kappa}{\kappa - 1}\left[\left(\frac{p_2}{p_1}\right)^{\frac{2}{\kappa}} - \left(\frac{p_2}{p_1}\right)^{\frac{\kappa+1}{\kappa}}\right]} \tag{7.2.2}$$

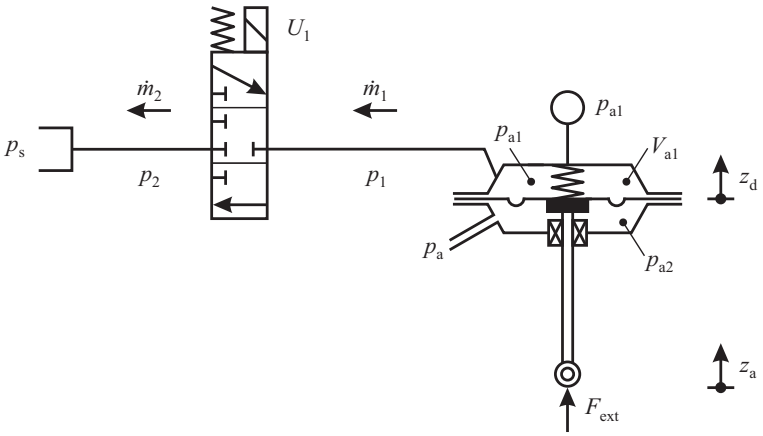


Fig. 7.2.3. Schematic of a pneumatic actuator with a diaphragm operating with vacuum supply pressure.

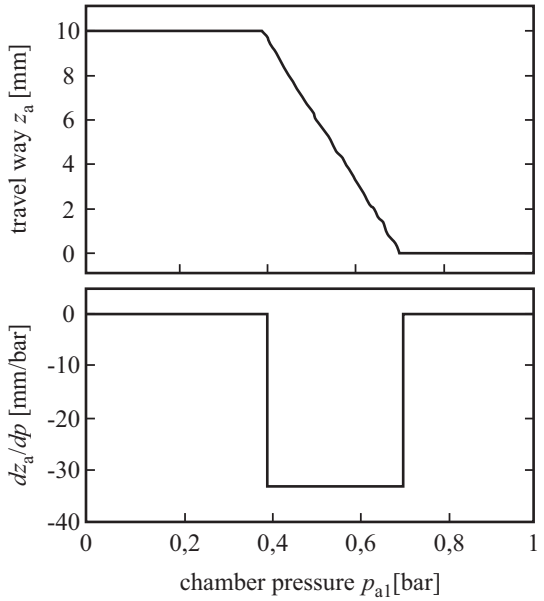


Fig. 7.2.4. Position characteristic (idealized) of a pneumatic actuator with vacuum pressure supply, Schwarte (2007).

($\kappa = c_\mathrm{p}/c_\upsilon = 1.4$ for air)

Here, it is assumed that $p_1 > p_2$, i.e. the air flows out of the actuator chamber. The outflow function has a maximum for $(p_2/p_1)_\mathrm{crit} = 0.53$ with $\Psi_\mathrm{max} = 0.484$. For constant p_1 and decreasing p_2, Ψ increases until Ψ_max and holds this value also for smaller p_2 because of reaching the sound velocity, Isermann (2005). Therefore for overcritical $p_2/p_1 < 0.53$ the mass flow $\dot{m}_1$ only depends on p_1 and T_1 with $\Psi = 0.484 = \text{const}$. An approximation of the mass flow is obtained with

$$\dot{m}_1 = \begin{cases} A_\upsilon c p_1 & \frac{p_2}{p_1} < b \\ A_\upsilon c p_1 \sqrt{1 - \left(\frac{p_2/p_1 - b}{1-b}\right)^2} & b < \frac{p_2}{p_1} \leq 1. \end{cases} \tag{7.2.3}$$

With this equation, pneumatic resistances can be described, where the parameters c and b are determined experimentally, Minxue et al (1986).

Figure 7.2.5 shows the corresponding flow characteristics of a pneumatic valve. The cross sectional area $A_\upsilon(U_\upsilon)$ depends on the opening characteristic of the servo valve.

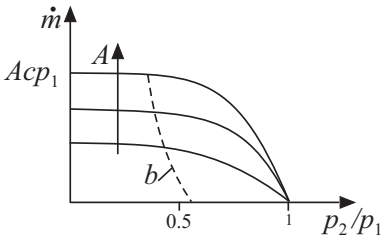


Fig. 7.2.5. Flow characteristic for a pneumatic valve.

The *actuator chamber* is assumed as a gas storage without flow resistances, such that the chamber pressure is $p_\mathrm{a1}(t) = p_1(t)$. The surface of the diaphragm (respectively piston in a cylinder) be A_d and its travel way z_d. Then a mass balance of this pneumatic part gives

$$\begin{aligned} \dot{m}_1(t) &= \frac{d}{dt}\left(V_\mathrm{a1}(t)\rho_\mathrm{a1}(t)\right) = \dot{V}_\mathrm{a1}(t)\rho_\mathrm{a1}(t) + V_\mathrm{a1}(t)\dot{\rho}_\mathrm{a1}(t) \\ &= A_\mathrm{d}\rho_\mathrm{a1}(t)\dot{z}_\mathrm{d}(t) + \left(V_\mathrm{a10} - A_\mathrm{d} z_\mathrm{d}(t)\right)\dot{\rho}_\mathrm{a1}(t) \end{aligned} \tag{7.2.4}$$

where V_a10 is the filled chamber volume ($p_\mathrm{a1} = p_\mathrm{at}$). Inserting the gas equation $\rho_\mathrm{a1} = p_\mathrm{a1}/RT_\mathrm{a1}$ yields with isothermal change of state i.e. $T_\mathrm{a1} = \text{const}$.

$$\left(V_\mathrm{a10} - A_\mathrm{d} z_\mathrm{d}(t)\right)\dot{p}_\mathrm{a1}(t) + A_\mathrm{d}\dot{z}_\mathrm{d}(t) p_\mathrm{a1}(t) = R\,T_\mathrm{a1}\,\dot{m}_1(t). \tag{7.2.5}$$

Hence, the parameters of this first order differential equation are time variant and depend on the position of the diaphragm. (7.2.1) and (7.2.2) lead to a first part of the signal flow graph in Fig. 7.2.6 and (7.2.5) to a second part.

The mechanical part of the actuator, consisting on the diaphragm, the actuator stem and further mechanical linkage is governed by the balance of forces:

$$m_{\mathrm{a}}\ddot{z}_{\mathrm{a}}(t) + d_{\mathrm{a}}\dot{z}_{\mathrm{a}}(t) + c_{\mathrm{s}}z_{\mathrm{a}}(t) + f_{\mathrm{c}}\,\mathrm{sign}\,\dot{z}_{\mathrm{a}}(t) = A_{\mathrm{d}}p_{\mathrm{a1}}(t) - F_{\mathrm{ext}}(z) \qquad (7.2.6)$$

where m_{a} is the mass of the stem and connected parts, c_{s} the spring constant, d_{a} and f_{c} the coefficients of the viscous and dry friction of the stuffing box and guidance, F_{ext} the external force through e.g. a flowing fluid and $z_{\mathrm{a}} = z_{\mathrm{d}}$.

The equation of the motion is strongly nonlinear because of the usually considerable influence of the dry friction.

The stationary behavior with $d/dt = 0$ yields the position characteristic of the actuators

$$z_{\mathrm{a}} = \frac{1}{c_{\mathrm{s}}}\left(A_{\mathrm{d}}p_{\mathrm{a1}} - f_{\mathrm{c}}\,\mathrm{sign}\,\dot{z}_{\mathrm{a}} - F_{\mathrm{ext}}(z_{\mathrm{a}})\right). \qquad (7.2.7)$$

Neglecting external forces the position z_{a} is linearly dependent on the chamber pressure p_{a1}, but with a direction-dependent dry-friction term, which results in a hysteresis curve. It can be approximated by

$$\begin{aligned} z &= z_0 - C_{1\mathrm{p}}p_{\mathrm{a1}} + f_{\mathrm{c}-} = z_{00-} - C_{1\mathrm{p}}p_{\mathrm{a1}} \quad \text{for} \quad \dot{z}_{\mathrm{a}} < 0 \\ z &= z_0 - C_{1\mathrm{p}}p_{\mathrm{a1}} - f_{\mathrm{c}+} = z_{00+} - C_{1\mathrm{p}}p_{\mathrm{a1}} \quad \text{for} \quad \dot{z}_{\mathrm{a}} > 0, \end{aligned} \qquad (7.2.8)$$

see Deibert (1997).

In the case of relatively slow motions and small masses $m_{\mathrm{a}} = 0$ can be assumed. (7.2.6) adds the last part to the signal flow graph in Fig. 7.2.6. The signal flow shows in general proportional behavior and several nonlinearities and feedback paths. The main dynamic delay results from the storage volume of the actuator. As well the stationary as the dynamic behavior vary considerably with the supply pressure p_{s} and the position z_{a}.

Figure 7.2.7 shows measured step responses of a diaphragm actuator. They show in the case of upward motion a time delay of $T_{\mathrm{tsv}} = 4...8$ ms which results from the servo valve, and approximately first order delay with time constants of about $T_{\mathrm{a1}} \approx 0.3...0.5$ s for small difference pressures of $p_{\mathrm{s}} - p_{\mathrm{at}} = 150$ and 200 mbar. For larger difference pressure than 250 mbar the mechanical restrictions are reached with approximately integral behavior. The proportional behavior arises because for small pressure differences the actuator chamber pressure p_{a1} reaches relatively early the steady- state air flow $\bar{\dot{m}}_1 = 0$ as $\Psi(p_{\mathrm{s}}/p_{\mathrm{a1}}) \to 0$. The downward motion is strongly determined by the spring force and therefore less dependent on the pressure difference.

The strong dependency on the vacuum supply pressure p_{s}, which varies continuously if the intake manifold pressure for gasoline engines is used, is disadvantageous for the actuator control. The situation improves considerably if the supply pressure is kept constant as it is possible with controlled vacuum pumps used for diesel engines.

The physically derived process model summarized in the signal flow graph of Fig. 7.2.6 indicates that the model is rather complicated and depends on parameters which are not exactly known. Therefore, Fig. 7.2.8 shows a much simpler model which represents the main effects.

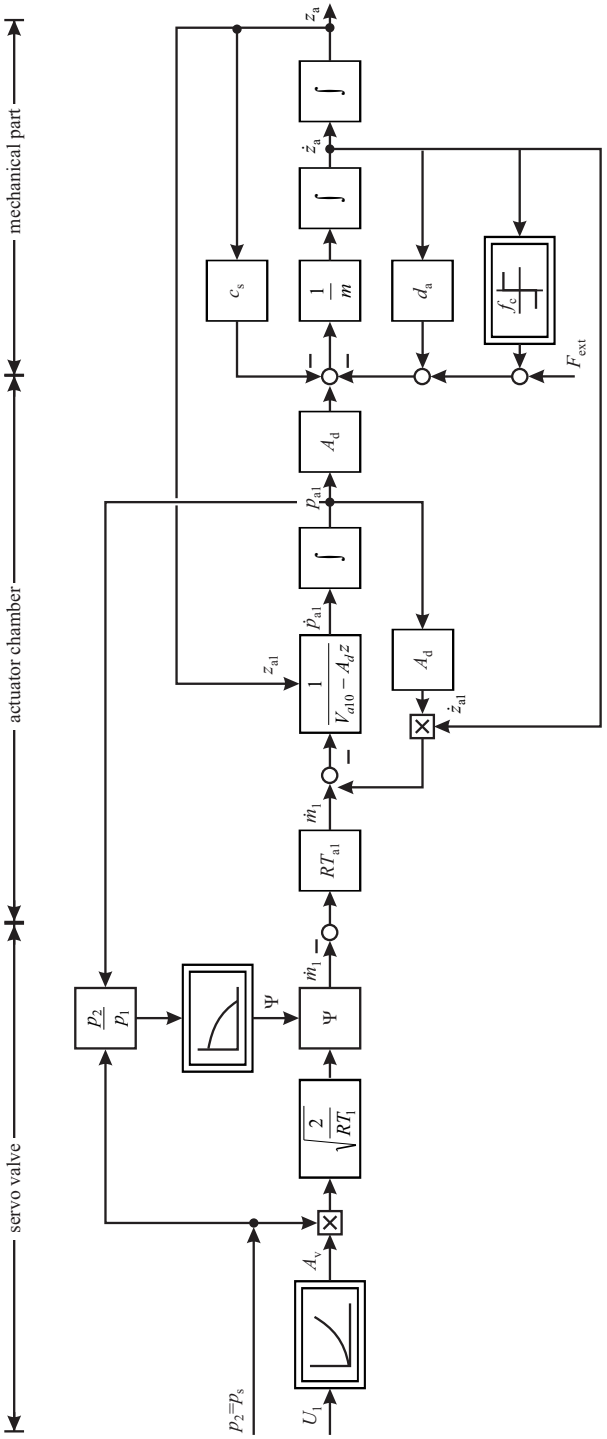


Fig. 7.2.6. Signal flow graph of a pneumatic actuator according to Fig. 7.2.3.

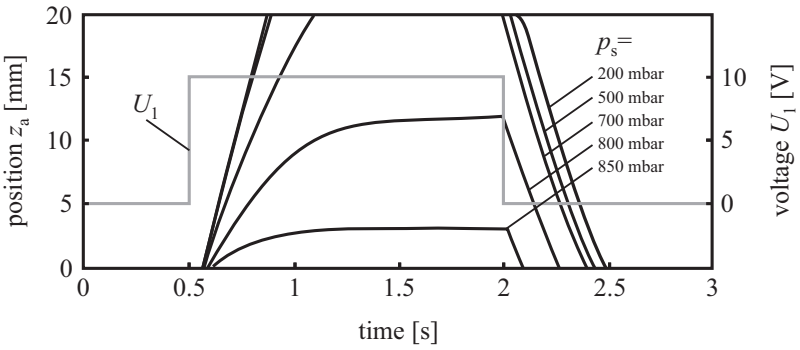


Fig. 7.2.7. Measured step responses of a pneumatic actuator for different supply vacuum pressures and motion directions, Raab (1993). Maximum travel way: 20 mm, $A_d = 27\,cm^2$, $V_{a1} = 83\,cm^3$, $c_s = 2000$ N/m, pretension spring force: $F_{S0} = 30$ N.

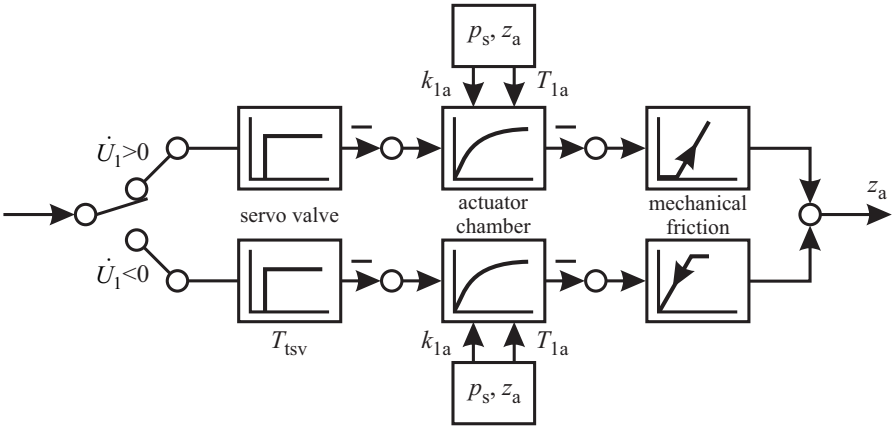


Fig. 7.2.8. Simplified signal flow for a pneumatic actuator with vacuum supply pressure and resulting direction dependent behavior.

The servo valve is modeled with a deadtime, the actuator chamber with a proportional first order delay and the mechanics by a hysteresis function. Two paths take into account the different behavior for the two motion directions. A corresponding differential equation for each path becomes

$$T_{1a}\dot{z}_a(t) + z_a(t) = K_{1a}U_1(t - T_{tsv}) - f_{1a}\,\mathrm{sign}\,\dot{z}_a(t) \tag{7.2.9}$$

with $f_{1a} = f_c/c_s$ and operation point dependent gain and time constant

$$K_{1a} = f_K(p_s, z_a);\;\; T_{1a} = f_T(p_s, z_a).$$

A corresponding transfer function without the friction term is

$$G_a(s) = \frac{z_a(s)}{U_1(s)} = \frac{K_{1a}}{(1 + T_{1a}s)}e^{-T_{tsv}s}. \tag{7.2.10}$$

Models and model-based control of a pneumatic actuator with an overpressure supply of $p_\mathrm{s} = 8$ bar for a VGT turbocharger of truck engines is treated in Schaffnit (2002). It is shown that an adaptive compensation of the dry friction improves the control performance considerably. Measured variables are the chamber pressure $p_{\mathrm{a}1}$, charging pressure $p_{2\mathrm{VGT}}$ and servo-valve voltage U_1.

7.2.3 Fault detection of pneumatic actuators

Faults of pneumatic actuators may arise in the air supply, air supply line, diaphragm or piston, valve stem sealing and plug-seat combinations. Table 7.2.1 lists some possible faults. The possibilities for the fault detection of pneumatic actuators depends highly on the available measurements. If no position controller is attached, the only electrical signals may come from a superimposed electronic controller, like a flow (or pressure) controller in the form of the flow rate $\dot{V}$ and its reference value W_2, compare Fig. 7.2.2. If an additional electrical signal is added, like the chamber pressure $p_{\mathrm{a}1}$, the fault coverage improves. The situation improves further if the position controller is electronic because then the manipulated variable U_1 and the valve position z are available, see Table 7.2.1.

Table 7.2.1. Fault detection coverage for pneumatic actuators in closed loop, dependent on instrumentation. x: yes, especially for large faults. −: no.

valve part	faults		available measurements $W_1, U_1, z_\mathrm{a}, W_2, \dot{V}$ symptoms $K_{1\mathrm{a}}$	$T_{1\mathrm{a}}$	$f_{1\mathrm{a}}$	r_z	$\lvert e_1\rvert$	$\lvert e_2\rvert$	$U_1, W_2, \dot{V}$ symptoms $\lvert e_2\rvert$	$U_{1\mathrm{lim}}$
servo valve	F_1	coil windings	x	−	−	−	x	x	x	−
	F_2	friction	x		x	x	x	x	x	x
pneumatic part	F_3	supply pressure	x	−	−	x	x	x	x	x
	F_4	line leakage	x	x	−	x	x	x	x	x
mechanical part	F_5	stem friction	−	−	x	x	x	x	x	−
	F_6	valve cross section area	−	−	−	−	−	x	x	x
	F_7	position sensor	x	−	−	x	−	−	−	x
process flow	F_8	flow or pressure sensor	−	−	−	−	−	−	−	x

Based on measurements of U_1 and z_a the parameters $K_{1\mathrm{a}}$, $T_{1\mathrm{a}}$ and $f_{1\mathrm{a}}$ can be estimated during transients. Also an output residual can be determined with the parity

equation

$$r_{\mathrm{z}}(t) = T_{1\mathrm{a}}\dot{z}_{\mathrm{a}}(t) + z_{\mathrm{a}}(t) - K_{1\mathrm{a}}U_1(t - T_{\mathrm{tsv}}) + f_{1\mathrm{a}}\,\mathrm{sign}\,\dot{z}_{\mathrm{a}}(t). \tag{7.2.11}$$

Deviations of their values from normal allow to detect the faults given in Table 7.2.1. Closed loop position control allows to monitor the control deviation

$$e_1(t) = W_1(t) - z_{\mathrm{a}}(t) = z_{\mathrm{a,ref}}(t) - z_{\mathrm{a}}(t) \tag{7.2.12}$$

and closed loop flow (or pressure) control

$$e_2(t) = W_2(t) - \dot{V}(t). \tag{7.2.13}$$

Remaining control deviations give hints for faults in the closed loop like too high friction or backlash or detuned controllers. A further improvement for the detectability of actuator faults is reached if the actuator pressure $p_{1\mathrm{a}}$ is available. Then e.g. the static characteristic $z_{\mathrm{a}}(p_{1\mathrm{a}})$ can be determined.

If no position control is provided and also the position z_{a} is not measured but the main control variable $\dot{V}$, its reference value W_2 and the voltage U_1 of the servo valve the methods for closed loop supervision, treated in Isermann (2006) can be applied. Except the control deviation $e_2(t)$, a permanently sticking to the limit $U_1 = U_{1\mathrm{lim}}$ (saturation) of the manipulated variable, detection of oscillations and measures of control performance can be observed. However the diagnosis of specific component faults is then hardly possible. Closed loops cover usually small faults, like leakages, friction and sensor offsets. Therefore mostly large faults are indicated by measures for the closed loop behavior, see Table 7.2.1.

A fault-detection method for pneumatic actuators based on changes of the dynamic behavior without detailed modeling is proposed by Schwarte (2007). Features in form of areas under the curve of the output variables after rectangular test signal excitations allows to detect small leakages. Additional methods with parity equations and parameter estimation for industrial pneumatic valves are described in Deibert (1997), Füssel (2002), Isermann (2011), Rahman et al (2016).

7.3 Hydraulic actuators

Hydraulic actuators are applied if high manipulation forces or torques combined with a large power density are required. For passenger cars this holds at the chassis mainly for hydraulic power steering and active suspensions and at the combustion engine for camshaft positioning. In the case of trucks especially the various working machines and tools operate with hydraulics.

7.3.1 Camshaft phasing

The main usage for hydraulic actuators at the combustion engine is the camshaft positioning relatively to the crankshaft with a positioning range of 40 to 60° CS

(corresponds to 20 to 30° CAM of the camshaft). An electro-hydraulic servovalve manipulates the oil flow to and from the camshaft actuator. This actuator has e.g. a rotary design with a sliding vane wheel or an axial piston / cylinder design with a helical teethgear, see e.g. Heisler (1995), Hannibal et al (2002), van Basshuysen and Schäfer (2007), Kopp (2006), Meinig and Bohner (2013). Thus, the hydraulic camshaft actuation can be sketched as depicted in Fig. 7.3.1. The housing of the camshaft actuator with its chain or belt driven pulley, the stator, is connected with two oil channels to the servo control valve. The rotor is attached to the camshaft and turns via sliding vanes relatively to the pulley depending on the pressure difference in two oil chambers. The servo-control valve is connected with the engine oil circuit and actuated with PWM voltage from the ECU. Therefore it operates like a proportional valve. The spool of the servo-valve operates against a spring which holds it without electrical actuation in a basis position. Then the full oil pressure moves the vanes of the rotor to one restriction, such that in the case of the inlet camshaft a late valve timing is obtained. Without oil pressure the actuator is mechanically locked, leading to a mechanical stiff connection. The camshaft phasing control loop consists of the angle sensor, the servo-valve and the camshaft actuator, as shown in Fig. 7.3.1. The camshaft phase $\Delta\varphi_{\mathrm{cam}}$ is usually determined from the difference of angular position sensor signals of the camshaft and crankshaft, Hannibal et al (2002).

7.3.2 Models of a hydraulic camshaft phasing system

As an example the camshaft phasing system with a sliding vane actuator of Fig. 7.3.1 is considered. The derived models lean on Münchhof (2006) and Isermann (2005). The magnetic force F_{m} of an electromagnet depending on the applied current I_{a} is generally a nonlinear function $F_{\mathrm{m}}(z_{\mathrm{v}})$ depending on the design of the armature and its counter part, where z_{v} is the spool valve position. However, it can be assumed that its characteristic can be approximated at least piecewise by a linear relation, compare (7.1.3) and Fig. 7.1.3,

$$F_{\mathrm{m}} = c_{\mathrm{z}}(z_v) I_{\mathrm{a}} \tag{7.3.1}$$

The spool operates against a spring with force

$$F_{\mathrm{s}} = c_{\mathrm{s}} z_v. \tag{7.3.2}$$

Neglecting dynamics and the magnetic hysteresis, it holds $F_{\mathrm{m}} = F_{\mathrm{s}}$ and the position of the spool follows to

$$z_v = \frac{c_{\mathrm{z}}}{c_{\mathrm{s}}} I_{\mathrm{a}} = k_{\mathrm{I}} I_{\mathrm{a}} \tag{7.3.3}$$

for an applied current and with $U_{\mathrm{a}} = R I_{\mathrm{a}}$

$$z_v = \frac{c_{\mathrm{z}}}{c_{\mathrm{s}} R} U_{\mathrm{a}} = k_{\mathrm{U}} U_{\mathrm{a}} \tag{7.3.4}$$

for an applied voltage. Usually this voltage is manipulated by pulse width modulation (PWM) and changing the duty cycle.

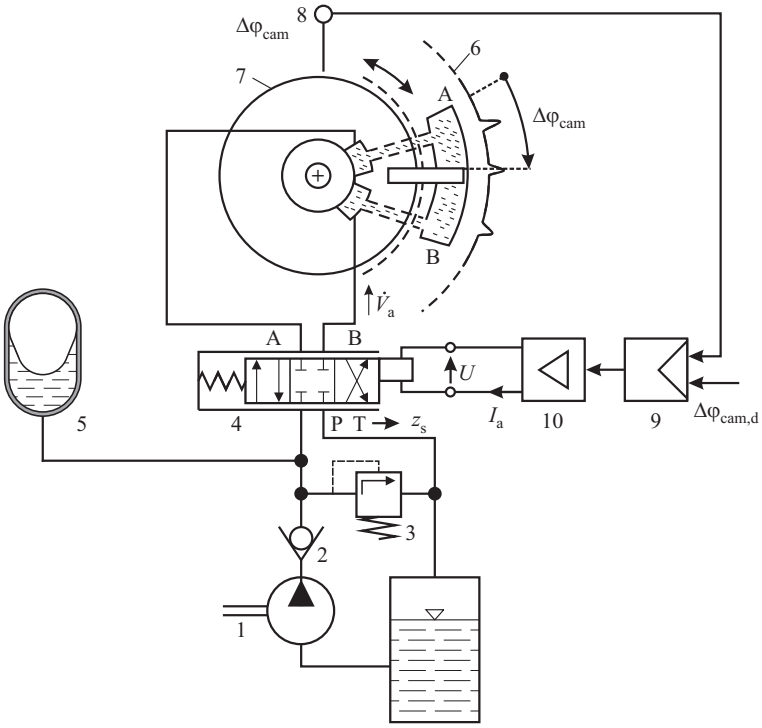


Fig. 7.3.1. Scheme of a hydraulic camshaft phasing system. 1 oil pump, 2 nonreturn valve, 3 pressure-relief valve, 4 4/3-servo control valve with connections to oil supply P, oil sump T, chamber A and B, 5 accumulator (optional), 6 pulley (stator), chain or belt driven, 7 rotor attached to camshaft, 8 phase sensor, 9 camshaft phase controller (ECU), 10 amplifier, driver.

The spool has overlapped valves. Therefore, the two ways to the actuator chambers A and B are closed in the center position where

$$\Delta\varphi_{\mathrm{cam}} = \varphi_{\mathrm{cam}} - \varphi_{\mathrm{cam},0}, \tag{7.3.5}$$

such that the position of the phasing actuator is freezed and the enclosed fluid has the same pressure $p_{\mathrm{A}} = p_{\mathrm{B}}$ on both sides. After the spool passes the valve lapping with $z_\upsilon > z_{\upsilon 0}$ e.g. the orifice opens to the supply pressure with $p_{\mathrm{s}} = 1 \ldots 5$ bar and the other orifice B opens to the oil sump with $p_{\mathrm{a}} \approx 1$ bar. The volume flows through the valve orifices can with $z'_\upsilon = z_\upsilon - z_{\upsilon 0}$ be described by

$$\dot{V}_{\mathrm{j}}\left(z'_\upsilon\right) \underset{j=A,B}{=} \alpha_{\mathrm{D}} A_\upsilon\left(z'_\upsilon\right) \sqrt{\frac{2}{\rho}} \sqrt{|\Delta p_\upsilon|}\,\mathrm{sign}\left(\Delta p_\upsilon\right) \tag{7.3.6}$$

with α_{D} a contraction number, A_υ the effective opening area and the pressure difference

$$\Delta p_{\upsilon \mathrm{A}} = p_{\mathrm{s}} - p_{\mathrm{A}} \text{ and } \Delta p_{\upsilon \mathrm{B}} = p_{\mathrm{B}} - p_{\mathrm{a}} \tag{7.3.7}$$

where p_a is the ambient pressure. For the volume flows holds

$$\dot{V}_\mathrm{A} = -\dot{V}_\mathrm{B} \tag{7.3.8}$$

if leakages over the vanes are neglected.

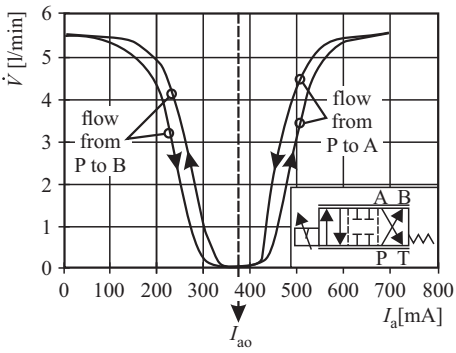


Fig. 7.3.2. Volume flow characteristic of a 4/3 proportional servo-valve for oil pressure p_s until 5 bar, Hannibal et al (2002).

Figure 7.3.2 depicts the volume flow characteristic $\dot{V}(I_\mathrm{a})$ for a spoolvalve. In the center position around $I_\mathrm{a} = I_\mathrm{a0}$ the volume flows $\dot{V}_\mathrm{A}$ and $\dot{V}_\mathrm{B}$ are zero. The oil is then trapped in the chambers and the rotor is fixed in a certain position. The characteristics also show a hysteresis behavior which originates from the electromagnet and dry friction of the spoolvalve.

Changes of the armature current $\Delta I_\mathrm{a}(t) = I_\mathrm{a}(t) - I_\mathrm{a0}$ lead to a change $\Delta z'_\upsilon(t) = z_\upsilon(t) - z_{\upsilon 0}$ of the spool valve and corresponding flows $\Delta\dot{V}_\mathrm{A}(t)$ and $\Delta\dot{V}_\mathrm{B}(t)$. The resulting pressure changes $\Delta p_\mathrm{A}(t)$ and $\Delta p_\mathrm{B}(t)$ and consequently the pressure difference

$$\Delta p_\mathrm{rot}(t) = \Delta p_\mathrm{A}(t) - \Delta p_\mathrm{B}(t) \tag{7.3.9}$$

generate a rotor torque ΔM_cam which turns the rotor and the camshaft with a change of the angular speed $\Delta\dot{\varphi}_\mathrm{cam}(t)$.

The flow through the spool valve can after linearisation be approximated by

$$\dot{V}_\mathrm{A} = k_{\dot{\mathrm{v}}} z'_\upsilon - k_\mathrm{c} p_\mathrm{A} \tag{7.3.10}$$

$$\dot{V}_\mathrm{B} = -\left(k_{\dot{\mathrm{v}}} z'_\upsilon - k_\mathrm{c} p_\mathrm{B}\right) \tag{7.3.11}$$

with the coefficients following from valve characteristics

$$k_{\dot{\mathrm{v}}} = \frac{\partial \dot{V}_\mathrm{j}}{\partial z_\upsilon} \; ; \; k_\mathrm{c} = \frac{\partial \dot{V}_\mathrm{j}}{\partial p_\mathrm{A}} \tag{7.3.12}$$

which depend on the supply pressure p_s.

The dynamic behavior of the oil chamber follows from a *mass balance* of the oil

$$\frac{dm_{\mathrm{A}}(t)}{dt} = \dot{m}_{\mathrm{A,in}}(t) - \dot{m}_{\mathrm{A,out}}(t) \tag{7.3.13}$$

where the mass inflow is with (7.3.11)

$$\dot{m}_{\mathrm{A,in}} = \dot{V}_{\mathrm{A}}\rho_{\mathrm{A}} = \left(k_{\dot{\mathrm{V}}} z'_{\upsilon} - k_{\mathrm{c}} p_{\mathrm{A}}\right)\rho_{\mathrm{A}} \tag{7.3.14}$$

$$\dot{m}_{\mathrm{A,out}} = \dot{V}_{\mathrm{A}}\rho_{\mathrm{A}} = A_{\mathrm{a}} r_{\mathrm{a}} \Delta\dot{\varphi}_{\mathrm{cam}}\rho_{\mathrm{A}}. \tag{7.3.15}$$

Herewith A_{a} is the effective area of the vanes, r_{a} is a representative radius and the change of the oil volume in chamber A is

$$V_{\mathrm{A}} = A_{\mathrm{a}} r_{\mathrm{a}} \Delta\varphi_{\mathrm{cam}} \tag{7.3.16}$$

where $\Delta\varphi_{\mathrm{cam}}$ is the rotor angle relative to the stator, see Fig. 7.3.1. For the stored mass holds, see Isermann (2005), (10.4.43),

$$\frac{dm_{\mathrm{A}}(t)}{dt} = \frac{d}{dt}\left(V_{\mathrm{A}}\rho_{\mathrm{A}}(t)\right) = V_{\mathrm{A}}\frac{d\rho_{\mathrm{A}}(t)}{dt} = V_{\mathrm{A}}\frac{\rho_{\mathrm{A}}}{\beta}\frac{dp_{\mathrm{A}}}{dt} \tag{7.3.17}$$

where β is the compressibility module (bulk modulus)

$$\beta = -V\left(\frac{dp}{dV}\right)_{\mathrm{r=const}}. \tag{7.3.18}$$

Inserting these equations in (7.3.13) yields

$$\frac{A_{\mathrm{a}} r_{\mathrm{a}} \Delta\varphi_{\mathrm{cam}}}{k_{\mathrm{c}}\beta}\frac{dp_{\mathrm{A}}(t)}{dt} + p_{\mathrm{A}}(t) = \frac{k_{\dot{\mathrm{V}}}}{k_{\mathrm{c}}} z'_{\upsilon}(t) - \frac{A_{\mathrm{a}} r_{\mathrm{a}}}{k_{\mathrm{c}}}\Delta\dot{\varphi}_{\mathrm{cam}}(t) \tag{7.3.19}$$

Hence, the pressure in chamber A behaves as a first order delay with *chamber time constant*

$$T_{\mathrm{p}} = \frac{A_{\mathrm{a}} r_{\mathrm{a}} \Delta\varphi_{\mathrm{cam}}}{k_{\mathrm{c}}\beta} \tag{7.3.20}$$

which is small for small rotor angle positions $\Delta\varphi_{\mathrm{cam}}$ and large β, i.e. large oil stiffness.

Correspondingly it holds for chamber B

$$\frac{A_{\mathrm{a}} r_{\mathrm{a}} \Delta\varphi_{\mathrm{cam}}}{k_{\mathrm{c}}\beta}\frac{dp_{\mathrm{B}}(t)}{dt} + p_{\mathrm{B}}(t) = \frac{-k_{\dot{\mathrm{V}}}}{k_{\mathrm{c}}} z'_{\upsilon}(t) + \frac{A_{\mathrm{a}} r_{\mathrm{a}}}{k_{\mathrm{c}}}\Delta\dot{\varphi}_{\mathrm{cam}}(t). \tag{7.3.21}$$

The *angular momentum balance* of the rotor with the attached camshaft for changes $\Delta\varphi_{\mathrm{cam}}$ becomes

$$J_{\mathrm{cam}}\Delta\ddot{\varphi}_{\mathrm{cam}}(t) = A_{\mathrm{a}} r_{\mathrm{a}} \Delta p_{\mathrm{rot}}(t) - M_{\mathrm{cam,f}}\left(\Delta\dot{\varphi}(t)\right) \tag{7.3.22}$$

where $M_{\mathrm{cam.f}}$ is the friction torque of the camshaft. The driving torque of the camshaft with the valve train oscillates strongly. For a four-cylinder engine the peak values are $+20$ Nm and -10 Nm for $n_{\mathrm{eng}} = 1000$ rpm and $+30$ Nm and -22 Nm for $n_{\mathrm{eng}} = 6000$ rpm, see Hannibal et al (2002). According to Heywood (1988) the

load on the camshaft is at low speeds primarily due to spring forces and at higher speeds dominated by inertia masses of the component masses. It decreases slightly with the speed. An average value of the friction torque is about $+5\,\mathrm{Nm}$. Hence it is assumed that for the relative turning of the rotor with camshaft and relative to the stator the torque can be approximated by a Coulomb and speed dependent viscous term.

$$M_{\mathrm{cam,f}}\left(\Delta\varphi\right) = M_{\mathrm{cam,f0}} + M_{\mathrm{cam,f1}}\Delta\dot{\varphi}_{\mathrm{cam}}. \tag{7.3.23}$$

Insertion of (7.3.22) in (7.3.23) yields

$$J_{\mathrm{cam}}\Delta\ddot{\varphi}_{\mathrm{cam}}(t) + M_{\mathrm{cam,f1}}\Delta\dot{\varphi}_{\mathrm{cam}} + M_{\mathrm{cam,f0}} = A_{\mathrm{a}}r_{\mathrm{a}}\Delta p_{\mathrm{rot}}(t). \tag{7.3.24}$$

If the viscous friction is small compared to the Coulomb friction, it holds

$$J_{\mathrm{cam}}\Delta\ddot{\varphi}_{\mathrm{cam}}(t) = A_{\mathrm{a}}r_{\mathrm{a}}\Delta p_{\mathrm{rot}}(t) - M_{\mathrm{cam,f0}} \tag{7.3.25}$$

with Δp_{rot} according to (7.3.9), see Fig. 7.3.3. Because of $p_{\mathrm{B}} \approx p_{\mathrm{a}}$ it is $\Delta p_{\mathrm{rot}} \approx \Delta p_{\mathrm{A}}(t)$.

For small deviations of $\Delta\dot{\varphi}_{\mathrm{cam}}$ it holds

$$\frac{J_{\mathrm{cam}}}{M_{\mathrm{cam,f0}}}\Delta\Delta\ddot{\varphi}_{\mathrm{cam}}(t) = \Delta\Delta p_{\mathrm{rot}}(t) \approx \Delta\Delta p_{\mathrm{A}}(t). \tag{7.3.26}$$

Hence, an integrating element results with integration time

$$T_I = \frac{J_{\mathrm{cam}}}{M_{\mathrm{cam,f0}}}. \tag{7.3.27}$$

Combining these equations leads to a signal flow chart in Fig.7.3.3. It shows an integral behavior $\Delta\varphi_{\mathrm{cam}}\left(U_{\mathrm{a}}(t)\right)$ with two nonlinear first order delays of the actuator chambers and rotor with camshaft. Therefore it is a dynamic system of 3rd order. Usually it can be assumed that the pressure in the chamber opening to the oilsump reaches the ambient pressure fast, such that $p_{\mathrm{B}} = p_{\mathrm{a}}$. If for the case of motion the torque generating chamber is considered and the relations are linearized one obtains a signal flow as in Fig. 7.3.4.

Hence, the dynamic model for the camshaft positioning can be approximately described by the transfer function

$$G_{\mathrm{U}\varphi}(s) = \frac{\Delta\varphi_{\mathrm{cam}}(s)}{\Delta U_{\mathrm{a}}(s)} = \frac{K_{\mathrm{U}\varphi}}{\left(1 + T_{\mathrm{p}}s\right)T_{\mathrm{I}}s^2} \tag{7.3.28}$$

and if it can be assumed that T_{p} is negligible, it holds

$$G'_{\mathrm{U}\varphi}(s) = \frac{K'_{\mathrm{U}\varphi}}{T_{\mathrm{I}}s^2}. \tag{7.3.29}$$

Therefore, the dynamic behavior can be approximately described as a double integrating element. Because of the double integral behavior the phase position controller can be, e.g. a PD-controller with parameters depending on oil temperature and engine speed.

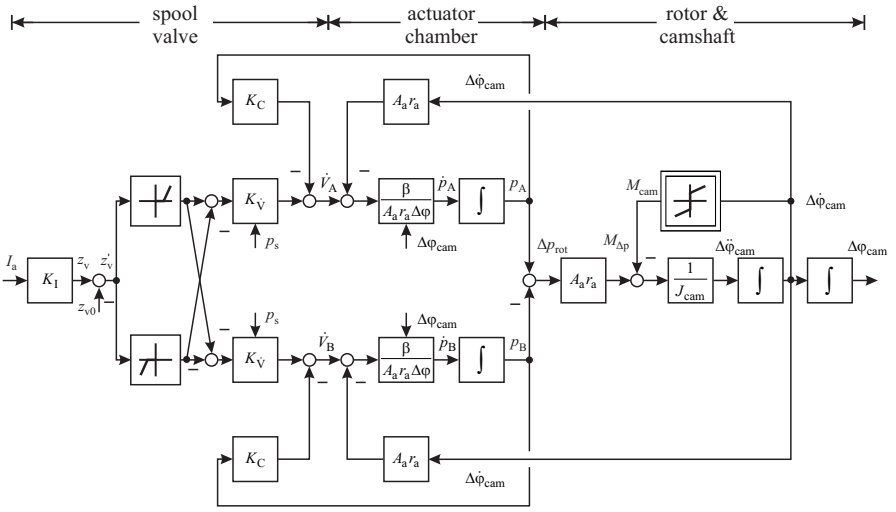


Fig. 7.3.3. Signal flow chart of a hydraulic camshaft phasing system with sliding vane actuator. $\Delta\varphi_{\mathrm{cam}}$ is the change of the camshaft rotor relative to the stator.

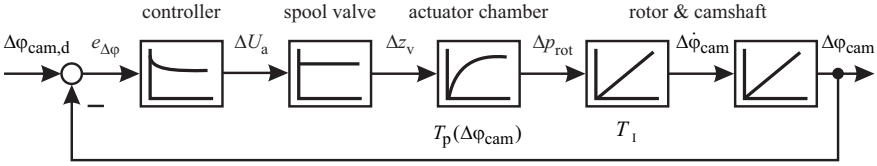


Fig. 7.3.4. Camshaft phase control loop with simplified, linear signal flow for the hydraulic camshaft system.

7.3.3 Fault detection

Model-based fault detection and diagnosis for electro hydraulic spool valves and a cylindric hydraulic actuator is with many details treated in Isermann and Münchhof (2011). The results can be in principle be transferred to the hydraulic camshaft phasing system. Possible faults are congestions and leaks in the oil supply and return line, eroded valve spool edges, valve spool grooving, internal leakage a the vane seal, gas enclosures, sensor faults, controller faults and faults of the oil pump. If enough measurements are available, as for well instrumented hydraulic servo axis systems, many of these faults can be detected and diagnosed by model-based fault-detection methods. However in the case of variable camshaft phasing systems usually the only measurements are the relative position $\Delta\varphi_{\mathrm{cam}}$, the valve spool armature voltage U_{a} in form the PWM signal or the spool current I_{a} and the oil supply pressure p_{s}. In addition it has to be considered that the used signals stem from a closed loop structure. It is shown in Isermann (2006) that closed loops compensate several faults as long as the actuator does not reach its end position and the control deviation $e(t)$ stays close to zero. It is further concluded that parity equation for linear processes can be

applied in the same way as for open loop. Therefore an output residual can be taken between the measured position and its calculated value with the positioning model (7.3.28)

$$r_{\Delta\varphi}(t) = \Delta\varphi_{\mathrm{cam}}(t) - \Delta\hat{\varphi}_{\mathrm{cam}}(U_{\mathrm{a}}, t) \tag{7.3.30}$$

which becomes in the Laplace domain

$$r_{\Delta\varphi}(s) = \Delta\varphi_{\mathrm{cam}}(s) - \hat{G}_{\mathrm{U}\varphi}(s)\Delta U_{\mathrm{a}}(s). \tag{7.3.31}$$

A further simple performance measure of a closed loop is the control deviation

$$e_{\varphi}(t) = \Delta\varphi_{\mathrm{cam}}(t) - \Delta\varphi_{\mathrm{cam,r}}(t) \tag{7.3.32}$$

where $\Delta\varphi_{\mathrm{cam,r}}$ is the reference value of the position control.

Both deviations can be monitored with regard to exceeding a threshold. In addition the parameters of the controlled process $G_{\mathrm{U}\varphi}(s)$ or $G'_{\mathrm{U}\varphi}(s)$ can be estimated based on measurements of $\Delta\varphi_{\mathrm{cam}}(t)$ and $U_{\mathrm{a}}(t)$ for the case of dynamic changes of the reference variable $\Delta\varphi_{\mathrm{cam,r}}$ and compared with normal values. As shown in Isermann (2006), Table 12.4, it is then possible to detect following faults: sluggish control behavior, oscillations, instability, friction, backlash, sensor offset. However, its not easy to locate and diagnose all faults with only a few measured signals.

References

Ayoubi M (1996) Nonlinear system identification based on neural networks with locally distributed dynamics and application to technical processes. Dissertation Technische Hochschule Darmstadt. Fortschr.-Ber. VDI Reihe 8, 591. VDI Verlag, Düsseldorf

Backé W, Klein A (2004) Fluid power actuators. In: Aktoren – Grundlagen und Anwendungen, Springer, Berlin

van Basshuysen R, Schäfer F (2007) Handbuch Verbrennungsmotor, 4th edn. Vieweg, Wiesbaden

Deibert R (1997) Methoden zur Fehlererkennung an Komponenten im geschlossenen Regelkreis. Dissertation Technische Hochschule Darmstadt. Fortschr.-Ber. VDI Reihe 8, 650. VDI Verlag, Düsseldorf

Füssel D (2002) Fault diagnosis with tree-structured neuro-fuzzy systems. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, 957. VDI Verlag, Düsseldorf

Gertler J (1998) Fault detection and diagnosis in engineering systems. Marcel Dekker, New York

Goedecke W (1989) Servopneumatische Antriebssysteme. O+P Ölhydraulik und Pneumatik 33(8)

Hannibal W, Knecht A, Stephan W (2002) Nockenwellenverstellungen für Ottomotoren. Die Bibliothek der Technik 247, Verlag Moderne Industrie, Landsberg/Lech

Heisler H (1995) Advanced engine technology. Society of Automotive Engineers, Warrendale

Heywood JB (1988) Internal combustion engine fundamentals. McGraw-Hill, New York

Höfling T (1996) Methoden zur Fehlererkennung mit Parameterschätzung und Paritätsgleichungen. Dissertation Technische Hochschule Darmstadt. Fortschr.-Ber. VDI Reihe 8, 546. VDI Verlag, Düsseldorf

Isermann R (2005) Mechatronic systems – fundamentals, 2nd edn. Springer, London

Isermann R (2006) Fault-diagnosis systems – An introduction from fault detection to fault tolerance. Springer, Heidelberg

Isermann R (2011) Fault-diagnosis applications. Springer, Heidelberg

Isermann R, Keller H (1993) Intelligente Aktoren. atp – Automatisierungstechnische Praxis 35:593–602

Isermann R, Münchhof M (2011) Identification of dynamic systems. Springer, Berlin, Heidelberg

Janocha H (ed) (2004) Actuators. Basics and Applications. Springer, Berlin

Kallenbach E, Eick R, Quandt P, Ströhla T, Feindt K, Kallenbach M (2008) Elektromagnete: Grundlagen, Berechnung, Entwurf und Anwendung, 3rd edn. Teubner, Stuttgart

Kopp C (2006) Variable Ventilsteuerung für Pkw-Dieselmotoren mit Direkteinspritzung. PhD thesis, Otto-von-Guericke-Universität Magdeburg

Meinig U, Bohner J (2013) New camshaft phaser module for automobile engines. MTZ worldwide 74(7-8):48–52

Minxue C, Kolvenbach H, Ohligschläger O (1986) Charakterisierung kompressibel durchströmter Widerstandsnetzwerke. O + P, Ölhydraulik und Pneumatik 30(12)

Moseler O (2001) Mikrocontrollerbasierte Fehlererkennung für mechatronische Komponenten am Beispiel eines elektromechanischen Stellantriebs. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, 980. VDI Verlag, Düsseldorf

Moseler O, Isermann R (2000) Application of model-based fault detection to a brushless DC motor. IEEE Trans on Industrial Electronics 47(5):1015–1020

Moseler O, Müller M (2000) A smart actuator with model-based FDI implementation on a microcontroller. In: 1st IFAC Conference on Mechatronic Systems, Darmstadt, Germany

Moseler O, Vogt M (2000) FIT- filtering and identification. In: Proc. 12th IFAC Symposium on System Identification (SYSID), Santa Barbara, CA, USA

Moseler O, Heller T, Isermann R (1999) Model-based fault detection for an actuator driven by a brushless DC motor. In: 14th IFAC World Congress, Beijing, China, vol P, pp 193–198

Moseler O, Straky H, Isermann R (2000) Verfahren zur Rekonstruktion der Ankerbewegung eines elektromagnetischen Aktors. Patent DE 10034830 C2 18.7.2000

Münchhof M (2006) Model-based fault detection for a hydraulic servo axis. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, 1105. VDI Verlag, Düsseldorf

Pfeufer T (1999) Modellgestützte Fehlererkennung und Diagnose am Beispiel eines Fahrzeugaktors. Fortschr.-Ber. VDI Reihe 8, 749. VDI Verlag, Düsseldorf

Pfeufer T, Isermann R, Rehm L (1996) Quality assurance of mechanical-electronical automobile actuator using an integrated model-based diagnosis control (in German). In: Proc. VDI-Conference Elektronik im Kraftfahrzeug, vol VDI-Bericht Nr. 1287, pp 145–159

Raab U (1993) Modellgestützte digitale Regelung und Überwachung von Kraftfahrzeugaktoren. Fortschr.-Ber. VDI Reihe 8, 313. VDI Verlag, Düsseldorf

Rahman B, Busch G, Ahmed Q, Rizzoni G (2016) Structural analysis-based fault diagnosis of pneumatic system. In: 2016 ßASME Dynamic Systems and Control Conference, Minneapolis, MN, USA

Schaffnit J (2002) Simulation und Control Prototyping zur Entwicklung von Steuergerätefunktionen für aufgeladene Nutzfahrzeug-Dieselmotoren. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 12, 492. VDI Verlag, Düsseldorf

Schwarte A (2007) Modellbasierte Fehlererkennung und Diagnose des Ansaug-und Abgassystems von Dieselmotoren. Dissertation Technische Universität Darmstadt 2006. Fortschr.-Ber. VDI Reihe 8, 634. VDI Verlag, Düsseldorf

Straky H, Moseler O, Isermann R (1999) Modellgestützte fehlererkennung an elektromagnetischen aktoren. In: Nordmann R, Isermann R (eds) Kolloquium Aktoren in Mechatronischen Systemen - 11. März 1999. Fortschr.-Ber. VDI Reihe 8, Nr. 743, VDI Verlag, Düsseldorf, pp 58–80

Streib HM, Bischof H (1996) Electronic throttle control (ETC): A cost effective system for improved emissions, fuel economy, and driveability. In: SAE International Congress and Exposition, Warrendale, PA, 960338

Tang J, Walker G (1995) Variable structure control of a pneumatic actuator. ASME J Dyn Sys, Meas, Control 117(1):88–92

Part IV

Fault-Tolerant Systems

8

Fault-tolerant components

8.1 Safety-related systems

Partially and highly automated driving automobiles are characterized by automatic controlled longitudinal and lateral movement and at least a reduced, permanent acting driver. This requires an increased supervision of all active systems and a fault-tolerant design of the safety-relevant components of the chassis and the powertrain.

As for all *safety-related systems*, all aspects of reliability, availability, maintainability and safety (RAMS) have to be considered. To meet safety requirements, special procedures were developed in different technical disciplines like railway, aircraft, space, military, nuclear and, later, automotive systems, IEC 26262 (2011). These procedures are covered by the terms *system integrity* or *system dependability*. Safety and reliability are generally achieved by a combination of: fault avoidance, fault removal, fault tolerance, fault detection, fault diagnosis, automatic supervision and protection.

To investigate the effect of faults on the reliability and safety during the design and type certification, a range of analysis methods were developed, mainly reliability analysis, event tree analysis (ETA), fault tree analysis (FTA), failure mode and effect analysis (FMEA), hazard analysis (HA) and risk classification, see e.g. IEC 61508 (1997), Storey (1996), Isermann (2006).

The *reliability* can be improved by oversizing, maintenance, protection for mechanical, hydraulic, electrical and some electronic components and by redundancy especially for electrical components, electronic hardware and software.

Because not all faults and failures can be avoided totally after careful design, manufacturing and testing, high-integrity systems should have the ability of *fault tolerance*. This means that faults are compensated in such a way that they do not lead to system failures. Within the powertrain this relates to important sensors and actuators of the combustion engine and to transmissions.

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7_8

8.2 Basic fault-tolerant structures

Fault-tolerance methods generally use *redundancy*. This means that in addition to the considered module one or more modules are connected, usually in parallel. These redundant modules are either *identical* or *diverse*. Such redundant schemes can be designed for hardware, software, information processing, mechanical and electrical components, like sensors, actuators, microcomputers, buses, power supplies, etc.

There exist mainly two basic approaches for fault tolerance, static redundancy and dynamic redundancy. The corresponding configurations are first considered for electronic hardware and then for other components, Leveson (1995), Isermann (2006).

Figure 8.2.1 shows a scheme for *static redundancy*. It uses three or more parallel modules which have the same input signal and are all active. Their outputs are connected to a voter who compares these signals and decides by majority which signal value is the correct one. If a triple modular redundant system is applied and the fault in one of the modules generates a wrong output, this faulty module is masked (i.e. not taken into account) by the 2-out-of-3 voting. Hence, a single faulty module is tolerated without any effort for specific fault detection. n redundant modules can tolerate $(n-1)$ / 2 faults (n odd).

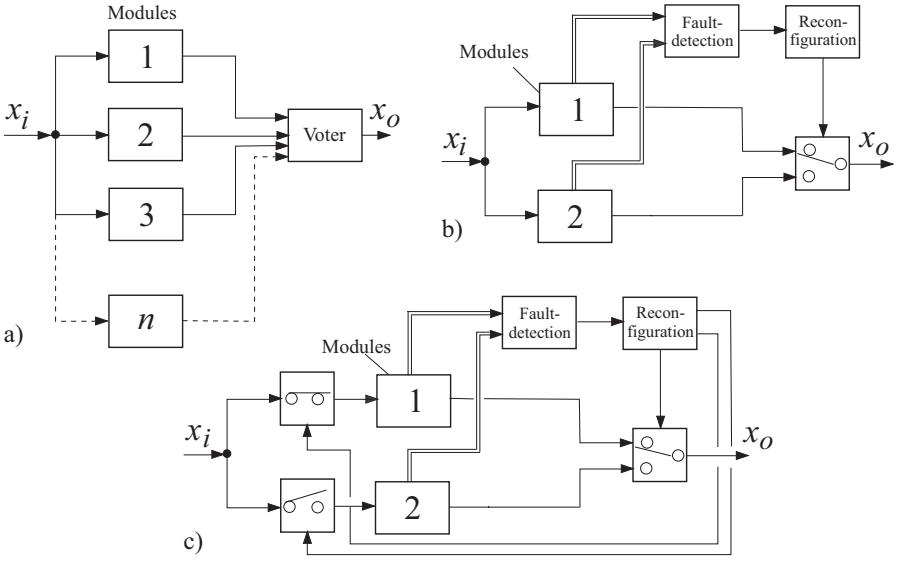


Fig. 8.2.1. Fault-tolerant schemes for electronic hardware. **a** static redundancy: multiple redundant modules with majority voting and fault masking, m out of n system (all modules are active); **b** dynamic redundancy: standby module which is continuously active, "hot standby", **c** dynamic redundancy: standby module that is inactive, "cold standby".

Dynamic redundancy needs less modules on cost of more information processing. A minimal configuration consists of two modules, Fig. 8.2.1b) and c). One module is usually in operation and if it fails the standby or backup unit takes over. This requires a fault detection to observe if the operation modules become faulty. Simple fault-detection methods use the output signal only for, e.g. consistency checking (range of the signal), comparison with redundant modules or use of information redundancy in computers like parity checking or watchdog timers. After fault detection it is the task of the reconfiguration to switch to the standby module and to remove the faulty one.

In the arrangement of Fig. 8.2.1b) the standby module is continuously operating, called "*hot standby*". Then, the transfer time is small on cost of operational aging (wear out) of the standby module. Dynamic redundancy, where the standby system is out of function and does not wear, is shown in Fig 8.2.1c), called "*cold standby*". This arrangement needs two more switches at the input and more transfer time due to a start-up procedure. For both schemes the performance of the *fault detection* is essential.

Similar redundant schemes as for electronic hardware exist for *software fault tolerance*. Here, tolerance against mistakes in coding or errors of calculations is meant. The simplest form of a *static redundancy* is repeated running $(n \geq 3)$ of the same software and majority voting for the result. However, this only helps for some transient faults. As software faults in general are systematic and not random, a duplication of the same software does not help. Therefore, the redundancy must include *diversity of software*, like other programming teams, other languages, or other compilers. With $(n \geq 3)$ diverse programs a multiple redundant system can be established followed by majority voting as Fig 8.2.1a). However, if only one processor is used, calculation time is increased, and using n processors may be too costly.

Dynamic redundancy by using standby software with diverse programs can be realized by using recovering blocks. This means that in addition to the main software module other diverse software modules exist, Leveson (1995), Storey (1996).

Fault tolerance can also be designed for purely *mechanical* and *electrical systems*. Static redundancy is very often used in all kind of homogeneous and inhomogeneous materials (e.g. metals and fibers) and special mechanical constructions like lattice-structures, spoke-wheels, dual tyres or in electrical components with multiple wiring, multiple coil windings, multiple brushes for DC motors, multiple contacts for potentiometers. This quite natural built-in fault tolerance is generally characterized by a parallel configuration. However, the inputs and outputs are not signals but, e.g. forces, electrical currents or energy flows and a voter does not exist. All elements operate in parallel and if one element fails (e.g. by breakage) the others take over a higher force or current, following the physical laws of compatibility or continuity. Hence, this is a kind of "stressful degradation". Mechanical and electrical systems with dynamic redundancy as depicted in Fig. 8.2.1b), c) can also be built. Both cold or hot standby may be applied.

Fault tolerance with dynamic redundancy and cold standby is especially attractive for *mechatronic systems* where more measured signals and embedded computers are already available and therefore fault detection can be improved considerably by

applying process-model-based approaches. Table 8.2.1 summarizes the appropriate fault-tolerance methods for the case of electronic hardware.

Table 8.2.1. Fail behavior of electronic hardware for different redundant structures. FO: fail-operational; F: fail; FS: fail-safe not considered.

		Static redundancy		Dynamic redundancy		
Structure	Number of elements	Tolerated faults	Fail behavior	Tolerated failures	Fail behavior	Discrepancy detection
Duplex	2	0	F	0 1	F FO-F	2 comparators fault-detection
Triplex	3	1	FO-F	2	FO-FO-F	fault detection
Quadruplex	4	1	FO-F	3	FO-FO-FO-F	fault detection
duo-duplex	4	1	FO-F	–	–	–

Mainly because of costs, space and weight, a suitable compromise between the degree of fault tolerance and the number of redundant components has to be found for *automotive systems*. In contrast to fly-by-wire systems, only one single failure usually must be tolerated (presently) for hazardous cases, Jonner et al (1996), mainly because a safe state can be reached easier and faster. This means that not for all components of automobiles very stringent fault-tolerance requirements are needed. Following steps of degradation are distinguished:

- Fail-operational (FO): One failure is tolerated, i.e. the component stays operational after one failure. This is required if no safe state exists immediately after the component fails,
- Fail-silent (FSIL): After one (or several) failure(s) the component behaves quiet externally, i.e. stays passive by switching off and therefore does not influence other components in a wrong way,
- Fail-safe (FS): After one (or several) failure(s) the component possesses directly a safe state (passive fail-safe, without external power) or is brought to a safe state by a special action (active fail-safe, with external power).

For vehicles it is proposed to subdivide FO in “long time” and “short time”. Considering these degradation steps for various components one has first to check if a safe state exists. For automobiles (usually) a safe state is stand still (or low speed) at a non-hazardous place. For components of automobiles a fail-safe status is (usually) a mechanical backup (i.e. a mechanical or hydraulic linkage) for direct manipulation by the driver. Passive fail-safe is then reached, e.g. after failure of electronics if independent of the electronics the vehicle comes to a stop, e.g. by a closing spring in the

throttle or by actions of the driver via mechanical backup. However, if no mechanical backup exists as for drive-by-wire steering or braking after failure of electronics, only an action by other electronics (switch to a still operating module) can bring the vehicle (in motion) to a safe-state. This requires redundancy and the availability of electric power.

Generally, a graceful degradation is envisaged, where less critical functions are dropped to maintain the more critical functions available, using priorities, IEC 61508 (1997). Table 8.2.1 shows degradation steps to fail-operational for different redundant structures of electronic hardware. As the fail-safe status depends on the controlled system and the kind of components, it is not considered here.

For *flight-control* computers usually at least a triplex structure with dynamic redundancy (hot standby) is used, which leads to FO-FO-FS, such that two failures are tolerated and a third one allows the pilot to operate manually, Rauch (1995), Reichel (2004). If the fault tolerance has to cover only one fault to stay fail-operational (FO-F), a triplex system with static redundancy or a duplex system with dynamic redundancy is appropriate. If fail-safe can be reached after one failure (FS), a duplex system with two comparators is sufficient. However, if one fault has to be tolerated to continue fail-operational and after a next fault it is possible to switch to a fail-safe status (FO-FS), either a triplex system with static redundancy or a duo-duplex system may be used. The duo-duplex system has the advantages of simpler failure detection and modularity.

8.3 Fault tolerance for control systems

For automatically controlled systems the appearance of faults and failures in the actuators, the process and the sensors will usually effect the operating behavior. With *feedforward control* generally all small or large faults influence the output variables and therefore more or less the operation. If the system operates with *feedback control*, small additive or multiplicative faults in the actuator or process are in general covered by the control actions because of usual robustness properties. This property is therefore a *passive control-loop fault-tolerance*. However, permanent additive and gain sensor faults will immediately lead to deviations from the reference values. For large changes of the behavior in actuators, process and sensors the dynamic control behavior becomes either too sluggish or too less damped or even unstable. Then, either a very robust control system or an *active fault-tolerant control system* is required to save the operation. In the last case it consists of fault-detection methods and a reconfiguration mechanisms which modifies the controller. Dependent on the kind of faults, the reconfiguration may change the structure and/or parameters of the controller. This can also include the switch to other manipulated variables or actuators or sensors, if available. However, fault detection in closed loops needs careful consideration, Isermann (2006), see also Sect. 2.6.

Examples are fault-tolerant flight control system with reconfiguration to other control surfaces after failure of actuators of ailerons, elevators and rudders, see e.g.

Rauch (1995), Chandler (1997), Patton (1997), Chen et al (1999). Fault-tolerant control for lateral vehicle control is treated in Suryanaryanan and Tomizuka (2000).

8.4 Fault management

In the frame of automated systems a fault tolerance can be designed for several components as actuators, sensors, communication channels, microcomputers, and control functions. Figure 8.4.1 illustrates a general automatic fault-management system with two manipulated and two controlled variables. The second actuator, for example, replaces the first actuator in the case of a fault if a similar manipulation effect can be reached. Hence, the components can be identical or diverse or redundant in itself. The fault-detection module operates in closed loop without or with test signals. The fault management then imitates a reconfiguration to the redundant component. This can occur by hard or soft switch-over, change of operation status or switch to feedforward control. For more details see Isermann (2011).

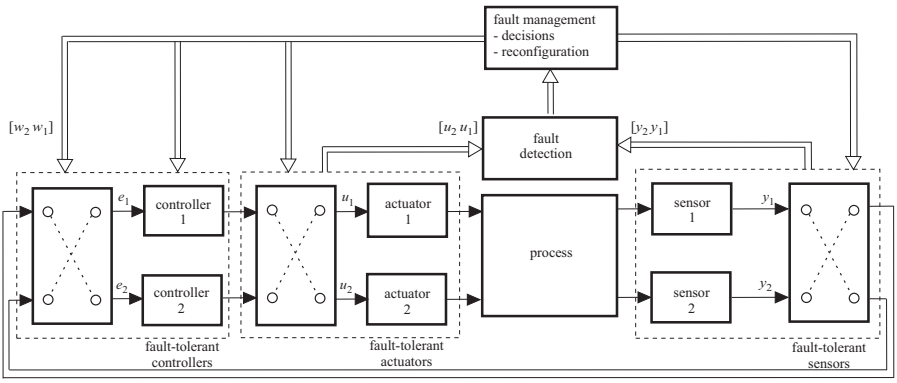


Fig. 8.4.1. Fault-tolerant control system with automatic fault management, shown for two actuators, two sensors and two controllers.

8.5 Fault-tolerant sensors

The discussion on high-integrity systems and automatic controlled systems shows that a comprehensive overall fault tolerance can be obtained by fault-tolerant components and fault-tolerant control. In the following, first some examples for fault-tolerant sensors are described.

A fault-tolerant sensor configuration should be at least fail-operational (FO) for one sensor fault. This can be obtained by applying *hardware redundancy* with the same type of sensors or by *analytical redundancy* with different other sensors and process models.

8.5.1 Hardware sensor redundancy

Sensor systems with static redundancy are realized for example with a triplex system and a voter, Fig. 8.2.1a). A configuration with dynamic redundancy needs at least two sensors and a fault detection for each sensor, Fig. 8.2.1b). Usually, only hot standby is feasible. Another less powerful possibility is a plausibility check for two sensors, also by using signal models (e.g. variance), to select the more plausible one.

The fault detection can be performed by self-tests, e.g. by applying a known measurement value to the sensor. Another way are self-validating sensors, Henry and Clarke (1993) where the sensor, transducer, and a microprocessor form an integrated, decentralized unit with self-diagnostic capability. The self-diagnosis takes place within the sensor or transducer and uses several internal measurements.

8.5.2 Analytical sensor redundancy

As a simple example a process with one input u and two measurable outputs y_1 and y_2 and the two processes G_1 and G_2 are considered, see Fig. 8.5.1a). If the process models G_{M1} and G_{M2} are known, two redundant signal values $\hat{y}_1$ and $\hat{y}_{1u}$ can be calculated, if G_{M2} can be inverted. Based on the three signals y_1, $\hat{y}_1$ and $\hat{y}_{1u}$ a fault-tolerant signal $\hat{y}_{1\mathrm{FT}}$ can be determined by a voter with 2-out-of-3 selection and a voter. The same principle can be applied to obtain fault-tolerant signals $\hat{y}_{2\mathrm{FT}}$ and $\hat{u}_{\mathrm{FT}}$. An application of this sensor-fault-tolerance principle is shown in Pfeufer (1999) for the position sensor of an electrical combustion engine throttle.

A similar scheme is depicted in Fig. 8.5.1b). Three redundant signals are generated based on three measurements and three residuals r_1, r_2 and r_3 are determined by comparison with the original signals. A fault-residual table in Fig. 8.5.1c) then shows unique patterns for the faults of all three sensors. Therefore, the detection of a respective sensor fault is possible and two fault-tolerant signals $y_{1\mathrm{FT}}$ and $y_{2\mathrm{FT}}$ can be determined.

One example for this combined analytical redundancy is the *yaw rate sensor* for the ESC, where additionally the steering wheel angle as input is used to reconstruct the yaw rate through a vehicle model as in Fig. 8.5.1b). Based on the lateral acceleration and the wheel speed difference of the right and left wheel (no slip) and vehicle models the yaw rate is reconstructed and given to a voter to form a fault-tolerant value as shown in Fig. 8.5.1a), van Zanten et al (1999).

Figure 8.5.2 shows the principle of a *fault-tolerant steering angle sensor*, Binasch et al (1989), Quass and Schiebel (2007). The axle of the steering wheel is equipped with a gear wheel driving two pinions. Each pinion turns a permanent magnet whose position is sensed by one of two giant magneto resistances (GMR) measuring bridges. The two sensors use the nonius/vernier principle to obtain the absolute position over $\pm 720°$. The evaluation of the sensors is integrated in two separate microcontrollers which supervise each other. In normal operation the master is connected to a bus and the slave monitors the master. In case of a fault, the master switches off and the slave is connected to the bus. Hence, the system is FO-FSIL. One fault of the sensor or microcontroller can be tolerated.

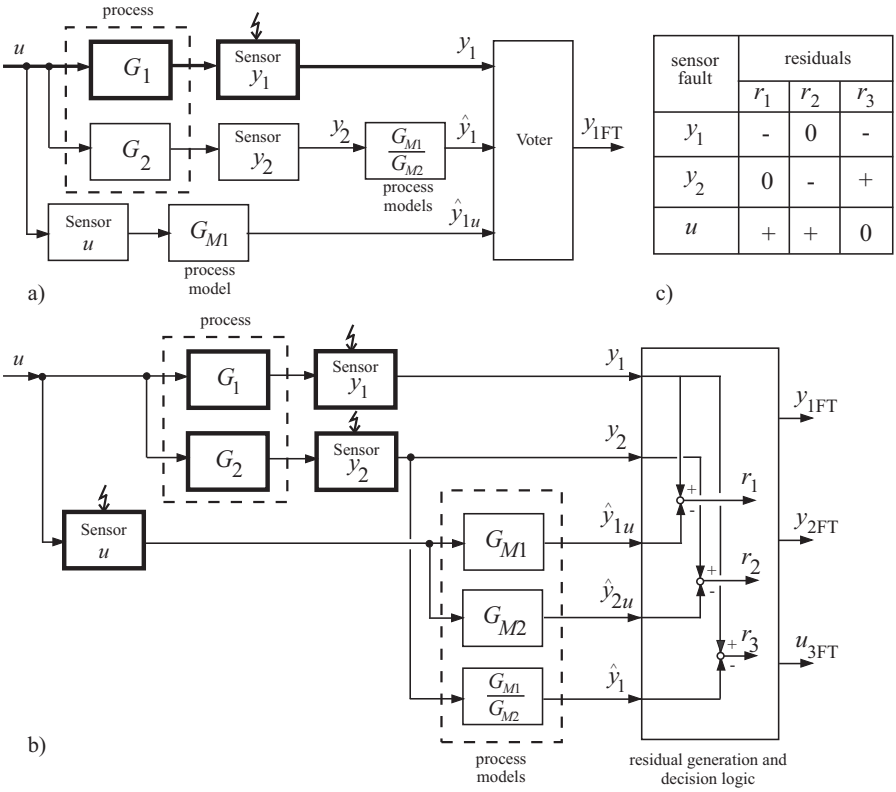


Fig. 8.5.1. Analytical sensor fault tolerance for a process with one measured input u and two outputs y_1 and y_2. **a** two redundant values for y_1 are calculated based on measured u and y_2, **b** redundant values for y_1 and y_2 based on residuals r_1, r_2 and r_3 are determined, **c** fault-symptom table.

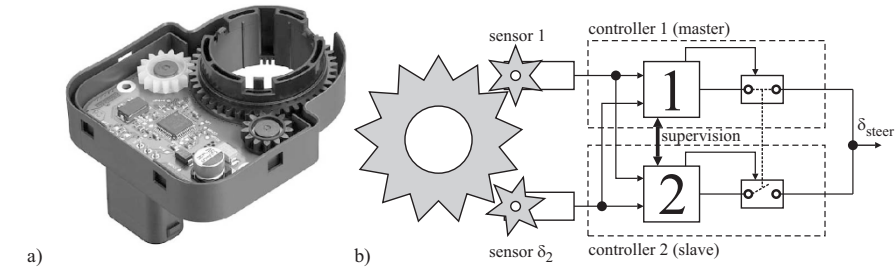


Fig. 8.5.2. Fault-tolerant steering angle sensor, Quass and Schiebel (2007). **a** system, **b** schematic and signal flow.

8.5.3 Fault-tolerant position sensor for an electrical throttle

The electrical throttle valve actuator is mainly used for gasoline engines. Its control and fault diagnosis was already treated in Sect. 7.1.2. It is another example for a fault-tolerant design, see Fig. 8.5.3.

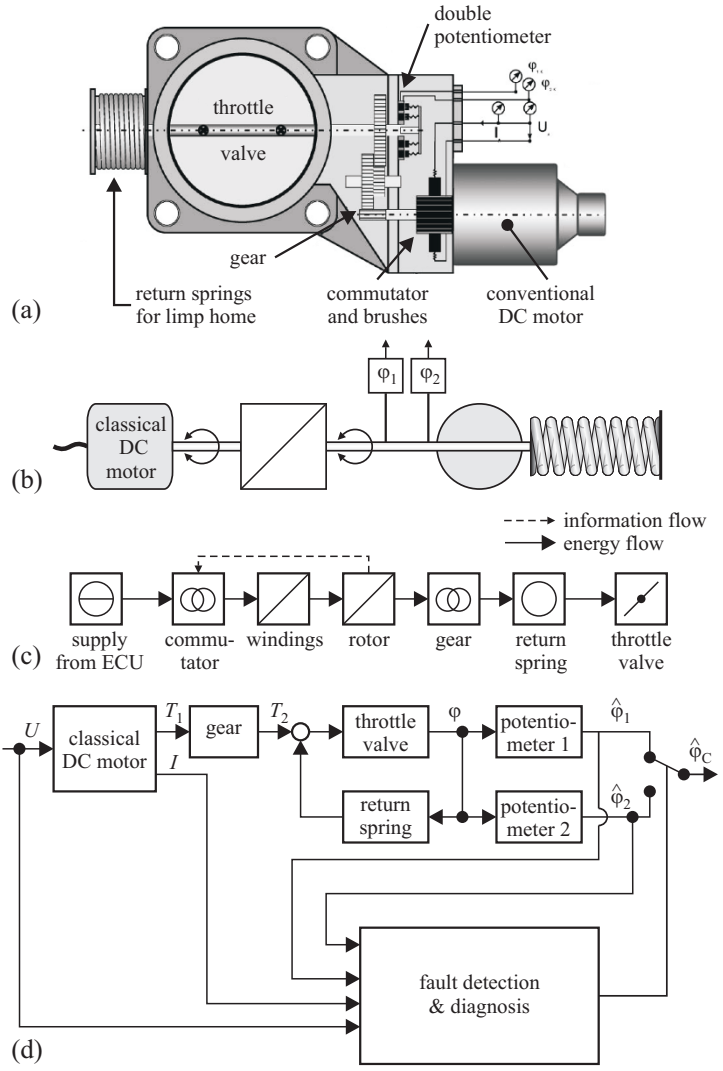


Fig. 8.5.3. Fault-tolerant electronic throttle. **a** throttle plate with gear and DC motor, **b** schematic, **c** energy flow scheme, **d** signal flow scheme Pfeufer (1999).

The electric motor, a classical DC motor with brushes, may be monitored using model-based methods. The physical model of the motor along with the mechanics also allows one to monitor the two potentiometers and enables the system to decide which of the two potentiometer readings may be faulty. Using dynamic redundancy concepts, the system is thus capable of withstanding one sensor fault despite the use of only two instead of three sensors, as would have been necessary in the case of static redundancy, see Pfeufer (1999). Figure 8.5.4 depicts the redundancy concept where sensor outputs φ_{1k} and φ_{2k} are used for the residual

$$r_1 = \varphi_{1k} - \varphi_{2k} \tag{8.5.1}$$

and the sensor output φ_{1k} and the electric motor model (7.2.1) are used to calculate $\hat{\varphi}_k = \omega_A$ for the determination of the residual r_2

$$r_2 = \frac{d}{dt}(\varphi_{1k}) - \varphi_k. \tag{8.5.2}$$

Then, one faulty sensor can be detected and the other one can be used for position control. Hence, the analytical sensor redundancy concept of Sect. 8.5.2 is used, similar to Fig. 8.5.1a).

Figure 8.5.5 shows a reconfiguration of a stuck fault of a potentiometer in closed loop, Isermann (2005), Pfeufer (1999). The system is active fail-safe (FS) with respect to motor faults. Return springs center the throttle valve in a slightly open position upon a loss of the motor torque (limp-home position). With respect to sensor faults, the system is FO-FS. It can sustain one sensor fault. After the second sensor fault, the system is switched off and the return springs once again bring the throttle into the slightly open position.

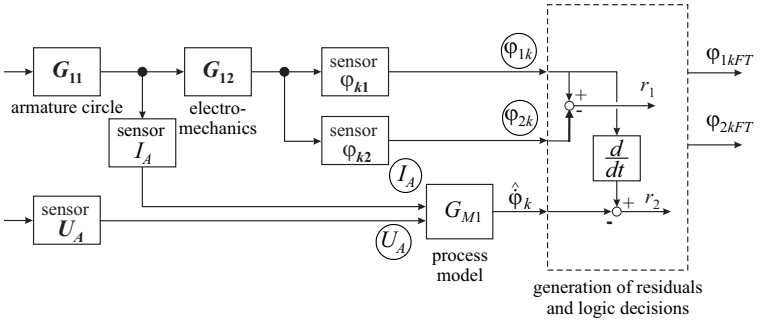


Fig. 8.5.4. Fault-tolerant position sensor system with two potentiometers and a virtual, model-based position value.

8.5.4 Fault-tolerant air intake sensor system

For *gasoline engines* the air *mass flow rate* $\dot{m}_{2,\mathrm{air}}$ into the cylinders can either be measured directly by an hot-film sensor HFM yielding $\dot{m}_1$ or it can be reconstructed

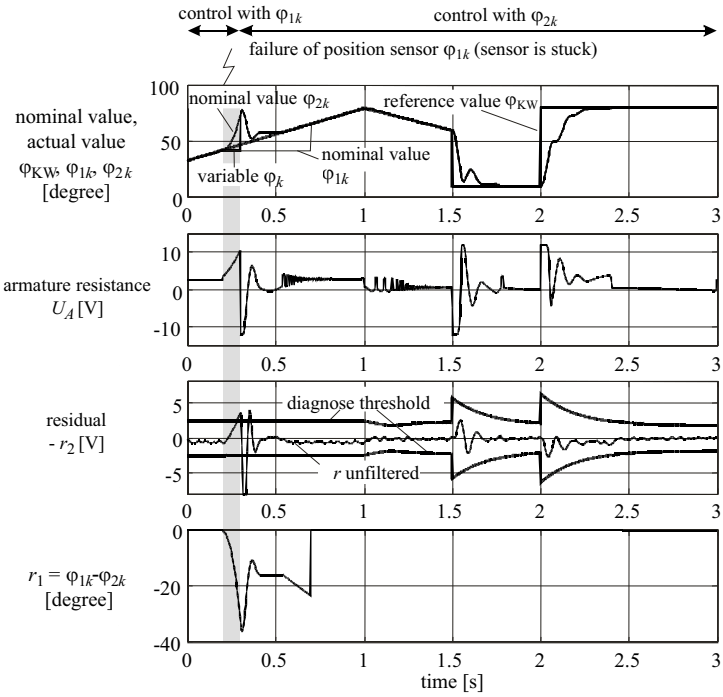


Fig. 8.5.5. Position control with stuck sensor $1k$ and reconfiguration with the second sensor $2k$ in closed loop.

by the measurement of the manifold pressure p_2 or by the measurement of the throttle position α_{th}, see Isermann (2014).

In dynamic operation it holds

$$^{1}\dot{m}_{2,\mathrm{air}} = \dot{m}_1(t) - \frac{V_{\mathrm{int}}}{RT_2}\frac{dp_2(t)}{dt} \tag{8.5.3}$$

if no additional gases $\dot{m}_{\mathrm{add}}$ enter the manifold. V_{int} is the intake volume, p_2 and T_2 the measured pressure and temperature in the manifold. Using the intake pressure p_2 as main measurement and assuming the volumetric efficiency $\eta_{\mathrm{v}}(n, \rho_2)$ is known, it is

$$^{2}\dot{m}_{2,\mathrm{air}}(t) = \frac{\eta_{\mathrm{v}}(n,\rho_2)V_{\mathrm{D}}n}{2RT_2}p_2(t) - \dot{m}_{\mathrm{add}}(t). \tag{8.5.4}$$

V_{D} is the displaced volume of all cylinders.

Based on the throttle angle α_{th} one obtains applying the flow equations for a throttle

$$^{3}\dot{m}_{2,\mathrm{air}}(t) = c_{\mathrm{th}}(\alpha_{\mathrm{th}}n)A(\alpha_{\mathrm{th}})p_a\sqrt{\frac{2}{RT_a}}\Psi\left(\frac{p_2}{p_a}\right) - \frac{V_{\mathrm{int}}}{RT_2}\frac{dp_2(t)}{dt}. \tag{8.5.5}$$

A is the cross-sectional open area of the throttle, Ψ an outflow function through a contraction and c_{th} a correction factor, which has to be calibrated experimentally.

The three ways for the determination of the airflow rate allow to build up a fault-tolerant sensor system together with a 2-out-of-3 voter, see Fig. 8.5.6, together with appropriate threshold selection of the three air flow rate variables. Thus, by majority voting it can be detected if the (main) sensor HFM is faulty and the remaining values ${}^2\dot{m}_{2,\mathrm{air}}$ and ${}^3\dot{m}_{2,\mathrm{air}}$ can be used as replacement signals for the engine ECU.

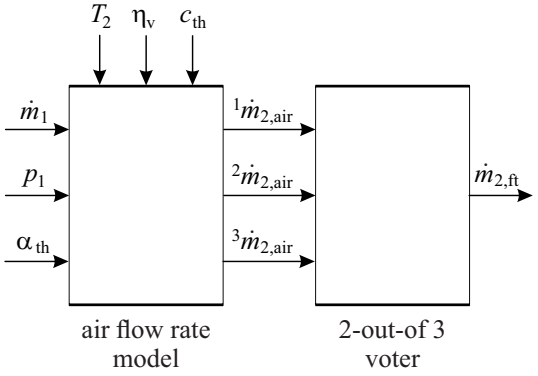


Fig. 8.5.6. Fault-tolerant system for the intake air-mass flow rate of a gasoline engine.

For *diesel engines* the air-mass flow rate can in addition to $\dot{m}_{2,\mathrm{air},1}$ based on direct measurement of $\dot{m}_1$ due to (5.5.1) also be determined by measurement of the EGR valve position u_{egr}, see (5.5.4), and the oxygen concentration ν_{O2} with continuous λ-sensor, see (5.5.5).

Figure 8.5.7 shows a scheme for the three ways of directly measured and reconstructed air mass flows. Instead of using the three values with analytical redundancy for the fault detection with three residuals (5.5.6) appropriate thresholds are applied and a 2-out-of-3 voter determines a fault-tolerant value $\dot{m}_{2,\mathrm{air,ft}}$ after a fault of the directly measured mass flow $\dot{m}_{2,\mathrm{air},1}$.

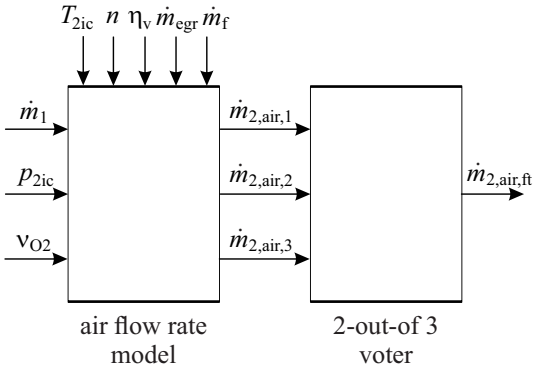


Fig. 8.5.7. Fault-tolerant system for the air-mass flow rate of a diesel engine.

8.6 Fault-tolerant actuators and drive systems

Actuators generally consist of different parts: input transformer, actuation converter, actuation transformer and actuation element (e.g. a set of dc amplifier, dc motor, gear and valve, as shown in Fig. 8.6.1a). The actuation converter converts one energy (e.g. electrical or pneumatic) into another energy (e.g. mechanical or hydraulic). Available measurements are frequently the input signal U_i, the manipulated variable U_0 and an intermediate signal U_3.

Fault-tolerant actuators can be designed by using multiple complete actuators in parallel, either with static redundancy or dynamic redundancy with cold or hot standby (Fig. 8.2.1), Münchhof et al (2009). One example for static redundancy are hydraulic actuators for fly-by-wire aircraft where at least two independent actuators operate with two independent hydraulic energy circuits. Another possibility is to limit the redundancy to parts of the actuator which have the lowest reliability. Figure 8.6.1b) shows a scheme where the actuation converter (motor) is split into separate parallel parts. Examples with static redundancy are two servo valves for hydraulic actuators, Oehler et al (1997) or three windings of an electric motor (including power electronics), Krautstrunk and Mutschler (1999).

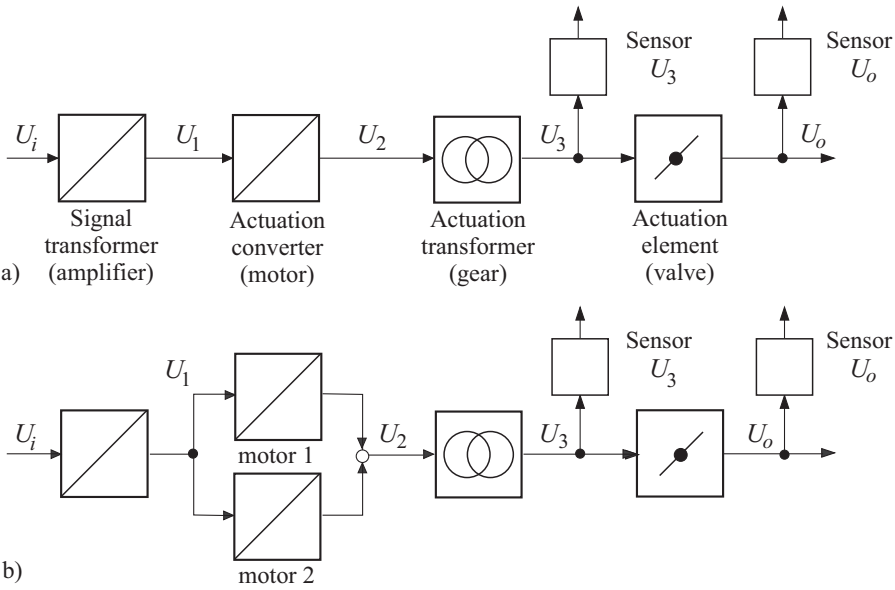


Fig. 8.6.1. Fault-tolerant actuator. **a** common actuator, **b**) actuator with duplex drive, two parallel motors and one gear. Hot or cold standby.

As cost and weight generally are higher than for sensors, actuators with fail-operational duplex configuration are to be preferred. Then either static redundant structures, where both parts operate continuously or dynamic redundant structures

with hot standby or cold standby can be chosen, see Fig. 8.2.1. For dynamic redundancy fault-detection methods of the actuator parts are required. One goal should always be that the faulty part of the actuator fails silent, i.e. has no influence on the redundant parts.

8.6.1 Fault-tolerant hydraulic systems

a) Fault-tolerant hydraulic brake

A traditional redundant system for automobiles is the *hydraulic brake* designed as a mandatory dual-circuit transmission system with, e.g. front axle/rear axle split or diagonal distribution system, Robert Bosch GmbH (2011). The braking pressure results from a combination of pedal force and (pneumatic) brake booster. The pedal force acts directly on the piston in a tandem master cylinder. Its pressure in a first chamber moves a floating intermediate piston and generates a pressure in the second chamber. Both chambers supply the respective circuits separately. Upon a fault (e.g. leak) in one circuit the other is still active, however, with reduced braking acceleration. Hence, the system is FO-F with dynamic redundancy and hot standby.

b) Fault-tolerant fuel injection pump

High pressure common rail injection pumps can be provided with two plungers acting on one common rail. The pump camshaft acts with two lobes on two plungers which operate phase-shifted on the same rail. Hence, by faults in one plunger pump element the other continues with a reduction of the maximum fuel flow. This system is a duplex system with hot standby and FO-F.

c) Fault-tolerant hydraulic actuators

Hydraulic actuators are applied in vehicles, for example, for power steering, active suspensions, oil lubrication circuits, and valve train phasing. Because of a good reliability of these hydraulic systems over the vehicle's lifetime, a fault-tolerant design was not in the focus until now. Therefore, a view is given on realizations in the field of aircraft.

Figure 8.6.2 shows the electro-hydraulic rudder actuator of the Eurofighter, see e.g. Kress (2002). The actuator consists of a proportional-acting valve, which is driven by four separate solenoid units. Each one is supplied by its own power electronics. All solenoids are acting on the same valve spool. Two control edges each supply one cylinder chamber with hydraulic fluid. The cylinder has four chambers with identical active piston areas. The system can sustain leakages and/or pressure losses in one of the two hydraulic circuits. It can also sustain internal leakages in one of the two hydraulic circuits without loosing the characteristic stiffness of hydraulic systems. As long as neither the valve spool nor the piston rod are affected, the system is FO-F.

A different prototypical realization of a redundant hydraulic actuator was developed by Münchhof (2006). Here, only the valve has been doubled since it has been found by a detailed statistical analysis of maintenance records that the valve alone makes up roughly 51 % of all faults at hydraulic servo axes. Upon the jam of one valve spool, the other valve can take over the volume flow. The entire servo axis is built from standard components, i.e. no design and construction of new, specialized components are necessary. This setup can tolerate all valve faults. As long as the piston rod is not affected and the spare valve is able to conduct any remaining parasitic flow of the faulty valve, the system is FO-F.

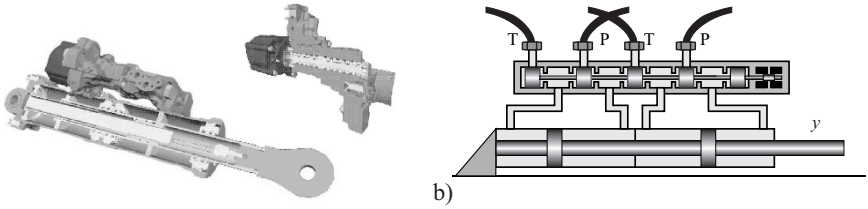


Fig. 8.6.2. Fault-tolerant electro hydraulic rudder actuator for the Eurofighter with two centralized hydraulic power supplies.

Another design with a dual-tandem ram, that is typical for aeronautical applications, Moog Aircraft Group (1996), is shown in Fig 8.6.3. This figure shows an actuation system for the F/A-18 horizontal stabilizer, which is a secondary control surface. In this example, the hydraulic cylinder is doubled and directly supplied with hydraulic fluid by two fixed-displacement pumps that are driven by two brushless DC motors. Bypass valves allow the piston to move even if the motor axle, respectively pump jams. The big advantage of this setup is that in the case of removal of the component, no hydraulic connections must be loosened. Furthermore, in the case of a leakage of the hydraulic piping/components, only the hydraulic oil of the component specific hydraulic circuit will spill, leaving all other hydraulic circuits of the plane unaffected. Provided that the valves are still functional and can disconnect the piston chambers from the pump, the system is FO-FSIL.

Many other architectures for electro-hydraulic actuators in aeronautical applications have been assessed in Sadeghi and Lyons (1992). The big lead in X-by-wire functionalities and the accompanying use of fault-tolerant mechatronic components in aeronautical applications is typical and can be explained easily: Fly-by-wire functionalities for airplanes could be introduced with little increase in risk as control surfaces by themselves are redundant. Almost all maneuvers can be realized by different combinations of control surfaces.

8.6.2 Fault-tolerant electrical actuators and drives

Until now electrical actuators in automobiles are mostly not designed as redundant units. However, the trend to partially or fully automatic driving will require more re-

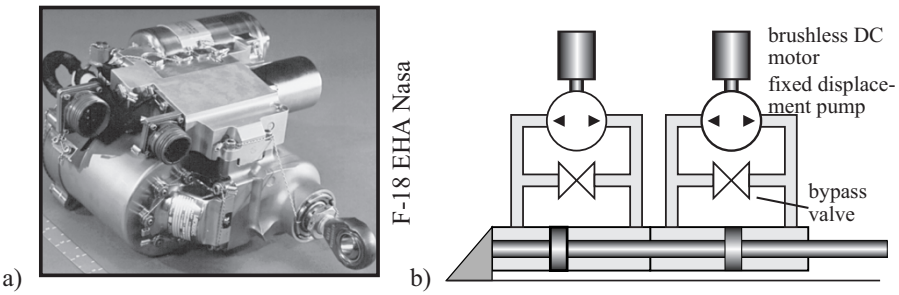


Fig. 8.6.3. Fault-tolerant electro-hydraulic actuator with integrated, decentralized power supply for secondary flight control surface.

dundant actuators if they are safety-relevant. First, a realized duplex actuator system for aircrafts will be considered and then some examples for fault-tolerant drives in general and for electrical steering are discussed.

a) Electrical duplex actuator system

Figure 8.6.4 depicts a *duplex actuator system* for the cabin pressure control in passenger aircraft. Two brushless DC motors (BLCD) act on a common gear. These DC motors and their microcontrollers operate alternatively from flight to flight. A model-based fault-detection system was developed to detect different faults of the DC motors, Moseler et al (1999), Moseler (2001), Isermann (2006). In the case of a fault in one motor the other one is switched to be active. This system is therefore FO-F with dynamic redundancy and cold-standby.

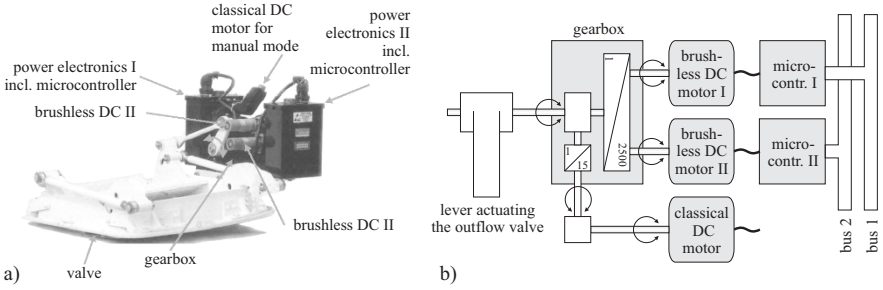


Fig. 8.6.4. Fault-tolerant electrical actuator of the aircraft cabin outflow valve. **a** system, **b** schematic.

b) Fault-tolerant electrical drives

Fault-tolerant frequency converter
The *power electronics for frequency-controlled AC asynchronous (AS) and synchronous (SM) drives* have usually one inverter leg per motor phase. Upon the loss of one leg, the three-phase motor becomes single-phased and cannot generate a rotating magnetic field any longer. To overcome this fault situation, one can control each phase separately with a full H-bridge. If each phase of the motor is connected to a full H-bridge, the three-phase motor, upon the loss of one inverter leg or phase, can be operated as a two-phase motor and can still produce a rotating magnetic field. The major disadvantage is the fact that for separate H-bridges, each winding needs two wires to establish the connection to the power converter, making the wiring more expensive. For more details see Green et al (2003). The combination of a standard three-phase PMSM (permanent-magnet synchronous motor) with an inverter with full H-bridges has been presented in Green et al (2003), Krautstrunk (2005). The system is now also FO-FSIL with respect to inverter or motor phase faults.

Multi-phase motors
An alternative to using full H-bridges for each phase is to design so called multi-phase motors, that have more than three phases. Although the first designs date back to the late 1960s, they have only recently been in the focus of research, as they are well suited for applications in the area of ship, locomotive, and electric/hybrid vehicle propulsion, see, e.g. Levi (2008), and the more electric aircraft, an initiative gathering momentum in the late 1990s with the aim to control aircraft subsystems with electrical actuators in place of mechanical, hydraulic or pneumatic means.

One general reason for the introduction of multi-phase motors in applications demanding the highest power is that semiconductors are not yet capable of switching the high currents that traditional three-phase motors would need to satisfy such high power demands. By increasing the number of phases, the power-per-switch can be limited to values that can be borne by the semiconductors. Also, multi-phase machines can easily sustain phase losses. Upon the loss of one phase, an n-phase motor becomes an $(n-1)$ phase motor $(n \geq 4)$ and hence shows a smaller loss in the power-rating and the uniformity of the circumferential torque distribution as the number of phases n increases. Furthermore, multi-phase motors show improvements in the noise characteristics and their torque production can easily be enhanced by the injection of higher-frequency harmonics.

The design of a four-phase fault-tolerant PMSM aircraft actuator is shown in Atkinson et al (2005). A five-phase permanent magnet motor has been realized by Biansch et al (2008) and has been investigated experimentally for post-fault operation. Depending on the number of windings, the topology of the windings and the allowable loss in torque, the system has a dynamic redundancy with hot standby and can sustain one or more faults. Thus, the system is at least FO-FSIL with respect to phase faults.

Duplex motors in parallel or serial connection

Instead of designing fault-tolerant components of electrical drives one may also take separate complete motors. This is especially of interest for bearing faults which may lead to a blockage of the rotor and the faulty motor can be detached by a clutch. As shown in Fig. 8.6.5 either a parallel or serial configuration can be chosen, Reuß and Isermann (2004), Münchhof et al (2009).

There are two load transfer modes possible for the parallel connection, Fig. 8.6.5a), which are chosen depending on the severity of the fault, Reuß and Isermann (2004). In the case of a *minor fault*, the second motor first powers up. Once the desired rotational velocity is reached, the first clutch disengages and thus separates the defective motor from the load and the second clutch engages and establishes the power flow between the spare motor and the load. In case of a *severe* fault, the first clutch disengages instantaneously. The load is thus free-running while the second motor powers up and reaches the desired final velocity. Once the drive is up to speed, the second clutch closes and the load is driven by the second motor. As each motor can be decoupled from the load, it is also possible to conduct special motor tests from time to time without affecting the motion of the load.

In the case of a serial connection, two motors are mounted one behind the other as illustrated in Fig. 8.6.5b). Here, at least the rear motor can be detached from the load, while the front motor is always connected to the load. Thus, if the rotor of the front motor blocks, the entire fault-tolerant drive fails. These duplex motors may be considered for electrically driven vehicles as part of the power train. Embedded motors in the wheels are already redundant drives.

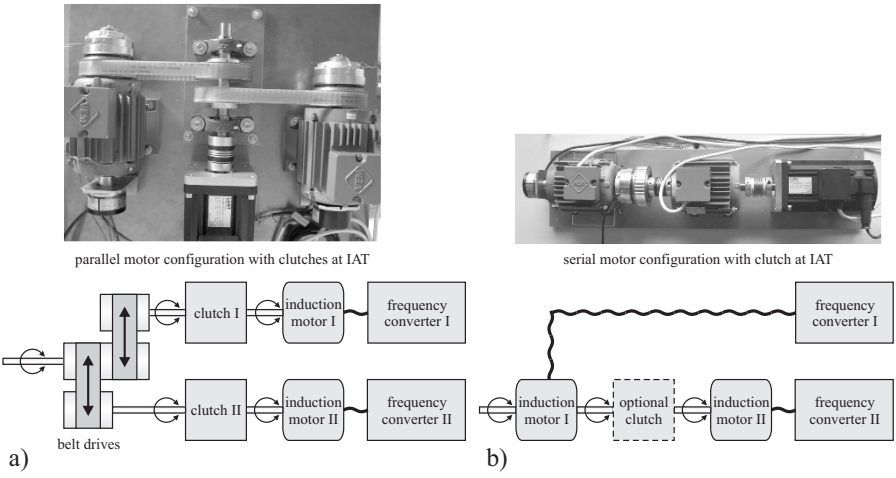


Fig. 8.6.5. Fault-tolerant electro drives. **a** parallel motor configuration, **b** serial motor configuration.

Fault-tolerant electrical power steering systems
For small and medium class passenger cars the hydraulic power steering system (HPS) was replaced by an electrical power steering column EPS (6-8 kN, 0.2-0.5 kW), pinion EPS (7-12 kN, 0.5-0.7 kW) and rack EPS (9-16 kN, 0.5-1 kW). Advantages are less installations effort, fuel saving, improved adaptation and the use of electrical input commands for driver assistance systems (automatic parking, lane keeping) and automatic driving. In the case of failures of the electrical power assisting system, the driver can take over the required steering torque for small and medium cars. However, for larger cars and light commercial vehicles the EPS should be fail-operational with regard to failures.

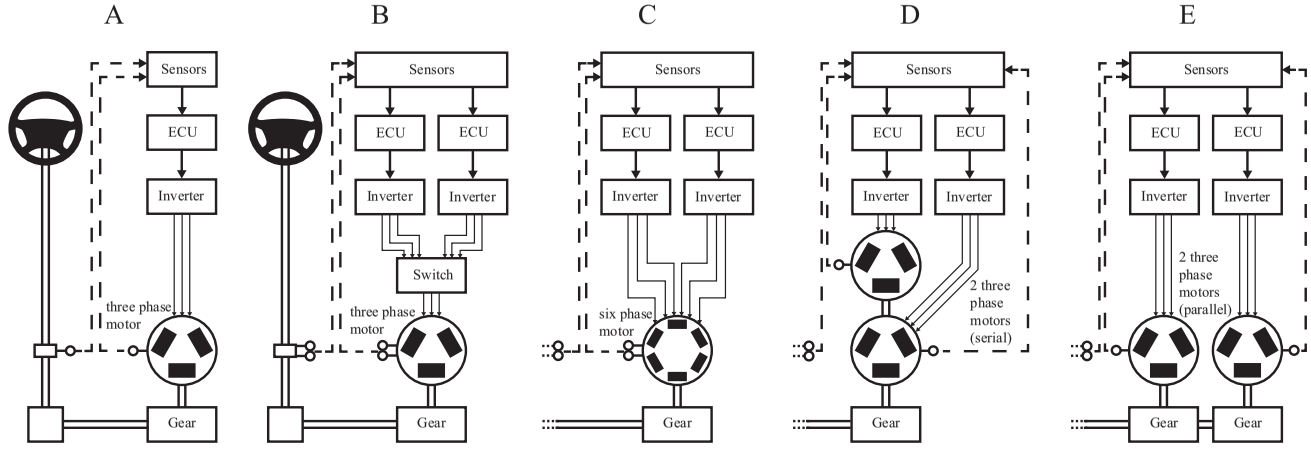
In the case of *automatic steering* the driver is out of the loop. Automatic closed-loop systems cover usually smaller additive and parametric faults of the controlled process. However, if the faults become larger, either sluggish, less damped or unstable behavior may result. The not engaged driver may then not perform the right steering command in time. Therefore, it will be required that EPS systems in the case of automatic controlled steering must have *fail-operational functions* in the case of certain faults, i.e. have to be fault-tolerant.

For most components of the steering actuator system hardware redundancy is required in form of dynamic redundancy with hot or cold standby according to Fig. 8.2.1b) and c). This means that a selection of components has to be doubled. Figure 8.6.6 depicts redundancy structures with different degrees of redundancy. At first the torque sensor can be duplicated or an analytical redundancy concept can be programmed, Schöttler (2007). In a next step the torque sensor, the ECU and inverter are duplicated, see Fig. 8.6.6B. The arrangement can be cold or hot standby. In the case of a fault in one channel the other channel stays active and the faulty channel is switched off. Case Fig. 8.6.6C has a further redundance in the windings by a multi-phase configuration, Hayashi (2013), Yoneki et al (2013). A serial connection of two motors is shown in Fig. 8.6.6D. Figure 8.6.6E depicts a duplication of the complete EPS actuator. In this parallel arrangement also the gear is duplicated, Beck and Isermann (2010), Keßler (2015).

The degree of redundancy increases from case A to E, however on cost of hardware extent, installation space, cost and weight. The selection of the redundancies also depends on fault-statistics for the different components. Case C seems to be a reasonable compromise, however requires a special motor design. In the case of a winding fault the power is reduced. In the cases D and E the power can be distributed differently to both motors in the normal operating range.

References

Atkinson GJ, Mecrow BC, Jack AG, Atkinson DJ, Sangha P, Benarous M (2005) The design of fault tolerant machines for aerospace applications. In: Proc. IEEE International Conference on Electric Machines and Drives, pp 1863–1869



faults	redundancy	redundancy	redundancy	redundancy	redundancy
Sensors	-	✓	✓	✓	✓
ECU	-	✓	✓	✓	✓
Inverter	-	✓	✓	✓	✓
Motor	-	-	✓	✓	✓
Gear	-	-	-	-	(✓)

Fig. 8.6.6. Redundancy structures for EPS systems.

Beck M, Isermann R (2010) Fehlertoleranzstrategien für mechatronische Systeme - Abschlussbericht. Arbeitsgemeinschaft industrieller Forschungsvereinigungen "Otto von Guericke" e.V. (BMWi), Frankfurt

Biansch N, Bolognani S, Pre MD (2008) Impact of stator winding of a five-phase permanent-magnet motor on postfault operations. IEEE Transactions on Industrial Electronics 55(5):1978–1987

Binasch G, Grünberg P, Saurenbach F, Zinn W (1989) Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. Physical review B 39(7):4828

Chandler P (1997) Reconfigurable flight control at wright laboratory. AGARD Advisory Report (Report 360, Aerospace 2020)

Chen J, Patton R, Chen Z (1999) Active fault-tolerant flight control systems design using the linear matrix inequality method. Trans Inst of Meas and Control 21(2 and 3):77–84

Green S, Atkinson DJ, Mecrow BC, Jack AG, Green B (2003) Fault tolerant, variable frequency, unity power factor converters for safety critical PM drives. IEE Proceedings – Electric Power Applications 150(6):663–672

Hayashi J (2013) Road map of the motor for an electric power steering system. In: 4. ATZ-Konferenz chassis.tech plus, München, Germany

Henry M, Clarke D (1993) The self-validating sensor: rationale, definitions, and examples. Control Engineering Practice – CEP 1(2):585–610

IEC 26262 (2011) Functional safety standard for automotive electrical/electronic systems. Beuth Verlag, Berlin

IEC 61508 (1997) Functional safety of electrical/electronic/programmable electronic systems. International Electrotechnical Commission, Switzerland

Isermann R (2005) Fehlertolerante mechatronische Systeme. In: VDI Tagung Mechatronik 2005, Wiesloch, Germany

Isermann R (2006) Fault-diagnosis systems – An introduction from fault detection to fault tolerance. Springer, Heidelberg

Isermann R (2011) Fault-diagnosis applications. Springer, Heidelberg

Isermann R (2014) Engine Modeling and Control. Springer, Berlin

Jonner W, Winner H, Dreilich L, Schunck E (1996) Electrohydraulic brake system - the first approach. In: SAE Technical paper Series, Warrendale, PA, 960991

Keßler P (2015) Einstellbare Verkopplungsregelung für ein Duplex-Aktorsystem einer achsparallelen elektrischen Servolenkung (EPS). In: Bäker B, Unger A (eds) Diagnose in mechatronischen Fahrzeugsystemen IX, TUDpress, Dresden, pp 217–231

Krautstrunk A (2005) Fehlertolerantes Aktorkonzept für sicherheitsrelevante Anwendungen. Shaker Verlag, Aachen, Germany

Krautstrunk A, Mutschler P (1999) Remedial strategy for a permanent magnet synchronous motor drive. In: 8th European Conference on Power Electronics and Applications, EPE'99, Lausanne, Switzerland

Kress R (2002) Robuste Fehlerdiagnoseverfahren zur Wartung und Serienabhnahme elektrohydraulische Aktuatoren. Doctoral thesis. Technische Universität Darmstadt, Fachbereich Maschinenbau, Darmstadt

Leveson N (1995) Safeware. System safety and computer. Wesley Publishing Company, Reading, MA

Levi E (2008) Multiphase electric machines for variable-speed applications. IEEE Transactions on Industrial Electronics 55(5):1893–1909

Moog Aircraft Group (1996) Redundant Electrohydrostatic Actuation System - Application: F/A-18 C/D Horizontal Stabilizer. Moog Aircraft Group

Moseler O (2001) Mikrocontrollerbasierte Fehlererkennung für mechatronische Komponenten am Beispiel eines elektromechanischen Stellantriebs. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, 980. VDI Verlag, Düsseldorf

Moseler O, Heller T, Isermann R (1999) Model-based fault detection for an actuator driven by a brushless DC motor. In: 14th IFAC World Congress, Beijing, China, vol P, pp 193–198

Münchhof M (2006) Model-based fault detection for a hydraulic servo axis. Dissertation Technische Universität Darmstadt. Fortschr.-Ber. VDI Reihe 8, 1105. VDI Verlag, Düsseldorf

Münchhof M, Beck M, Isermann R (2009) Fault-tolerant actuators and drives – structures, fault-detection principles and applications. Annual Reviews in Control 33:136–148

Oehler R, Schoenhoff A, Schreiber M (1997) Online model-based fault detection and diagnosis for a smart aircraft actuator. In: Prepr. IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), Pergamon Press, Hull, UK, vol 2, pp 591–596

Patton R (1997) Fault-tolerant control: the 1997 situation. In: Prepr. IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), Pergamon Press, Hull, UK, vol 2, pp 1033–1055

Pfeufer T (1999) Modellgestützte Fehlererkennung und Diagnose am Beispiel eines Fahrzeugaktors. Fortschr.-Ber. VDI Reihe 8, 749. VDI Verlag, Düsseldorf

Quass S, Schiebel P (2007) Aspects of future steering markets and their relevance to steering sensors. In: Proceedings of IQPC – Advanced Steering Systems

Rauch H (1995) Autonomous control reconfiguration. IEEE Control Systems Magazine 15(6):37–48

Reichel R (2004) Steuersysteme im flugzeug–fly-by-wire (flight control systems–fly-by-wire). at–Automatisierungstechnik 52(12):588–595

Reuß J, Isermann R (2004) Umschaltstrategien eines redundaten Asynchronmotoren-Antriebssystems. In: SPS/IPC/DRIVES 2004: Elektrische Automatisierung, Systeme und Komponenten: Fachmesse and Kongress, Nürnberg, Germany, pp 469–477

Robert Bosch GmbH (ed) (2011) Automotive Handbook, 8th edn. Bentley publishers, Cambridge

Sadeghi T, Lyons A (1992) Fault tolerant EHA architectures. IEEE Aerospace and Electronic Systems Magazine 7(3):32–42

Schöttler F (2007) Functional safety in electrical power steering systems. In: Proceedings of IQPC Conference on Steering Systems, Frankfurt, Germany

Storey N (1996) Safety-critical computer systems. Addison Wesley Longman Ltd., Essex

Suryanaryanan S, Tomizuka M (2000) Fault-tolerant lateral control of automated vehicles based on simultaneous stabilization. In: Proc. IFAC Conference om Mechatronic Systems, Darmstadt, Germany

Yoneki S, Hitozumi E, Collerais B (2013) Fail-operational eps by distributed architecture. In: 4. ATZ-Konferenz chassis.tech plus, München, Germany, pp 421–442

van Zanten A, Erhardt R, Schramm H, Pfaff G (1999) Die Fahrdynamik-Regelung ESP vom Pkw zum Nkw. In: Proc. 3. Stuttgarter Symposium Kraftfahrwesen und Verbrennungsmotoren, expert Verlag, Renningen-Malmsheim, pp 801–814

Part V

Appendix

9

Terminology in fault detection and diagnosis

The following definitions are the result of a coordinated action within the IFAC Technical Committee SAFEPROCESS, published in Isermann and Ballé (1997). Some basic definitions can also be found in IFIP (1983), Omdahl (1988) and in German standards like DIN and VDI/VDE-Richtlinien, see references at the end of this section and Isermann (2006).

1) **States and signals**

Fault:	*Unpermitted deviation* of at least one characteristic property of the system
Failure:	*Permanent* interruption of a systems ability to perform a required function under specified operating conditions
Malfunction:	*Intermittent irregularity* in fulfilment of a systems desired function
Error:	*Deviation* between a computed value (of an output variable) and the true, specified or theoretically correct value
Disturbance:	An *unknown* (and uncontrolled) *input* acting on a system
Perturbation:	An input acting on a system which results in a *temporary departure* from a steady state
Residual:	*Fault indicator*, based on deviations between measurements and model-equation-based calculations
Symptom:	*Change* of an observable quantity from *normal behavior*.

2) **Functions**

Fault detection:	Determination of faults present in a system and time of detection
Fault isolation:	Determination of kind, location and time of detection of a fault by evaluating symptoms. Follows fault detection
Fault identification:	Determination of the size and time-variant behavior of a fault. Follows fault isolation

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7_9

Fault diagnosis:	Determination of kind, size, location and time of detection of a fault by evaluating symptoms. Follows fault detection. Includes fault detection, isolation and identification
Monitoring:	A continuous real-time task of determining the possible conditions of a physical system, recognizing and indicating anomalies of the behavior
Supervision:	Monitoring a physical system and taking appropriate actions to maintain the operation in the case of faults
Protection:	Means by which a potentially dangerous behavior of the system is suppressed if possible, or means by which the consequences of a dangerous behavior are avoided.

3) **Models**

Quantitative model:	Use of static and dynamic relations among system variables and parameters in order to describe a system's behavior in quantitative mathematical terms
Qualitative model:	Use of static and dynamic relations among system variables and parameters in order to describe system's behavior in qualitative terms such as causalities or if-then rules
Diagnostic model:	A set of static or dynamic relations which link specific input variables – the symptoms – to specific output variables – the faults
Analytical redundancy:	Use of two, not necessarily identical ways to determine a quantity where one way uses a mathematical process model in analytical form.

4) **System properties**

Reliability	Ability of a system to perform a required function under stated conditions, within a given scope, during a given period of time. Measure: MTTF = mean time to failure. MTTF = $1/\lambda$. λ is rate of failure (e.g. failures per hour)
Safety:	Ability of a system not to cause danger to persons or equipment or the environment
Availability:	Probability that a system or equipment will operate satisfactorily and effectively at any point in time. Measure:

$$A = \frac{\text{MTTF}}{\text{MTTF}+\text{MTTR}}$$

MTTR mean time to repair
MTTR $= 1/\mu$; μ : rate of repair

References

IFIP (1983) Proc. of the IFIP 9th World Computer Congress, Paris, France, September 19–23. Elsevier

Isermann R (2006) Fault-diagnosis systems – An introduction from fault detection to fault tolerance. Springer, Heidelberg

Isermann R, Ballé P (1997) Trends in the application of model-based fault detection and diagnosis in technical processes. Control Engineering Practice – CEP 5(5):638–652

Omdahl T (ed) (1988) Reliability, availability and maintainability (RAM) dictionary. ASQC Quality Press, Milwaukee, WI, USA

Index

R. Isermann, *Combustion Engine Diagnosis*, ATZ/MTZ-Fachbuch, DOI 10.1007/978-3-662-49467-7