Lino Guzzella
Christopher H. Onder

Introduction to Modeling and Control of Internal Combustion Engine Systems

Second Edition

Springer

Lino Guzzella • Christopher H. Onder

Introduction to Modeling and Control of Internal Combustion Engine Systems

Lino Guzzella • Christopher H. Onder

Introduction to Modeling and Control of Internal Combustion Engine Systems

With 204 Figures

Springer

Professor Dr. Lino Guzzella
guzzella@imrt.mavt.ethz.ch
Dr. Christopher H. Onder
onder@imrt.mavt.ethz.ch
ETH Zürich
Institut für Mess- und Regeltechnik
Sonneggstr. 3
8092 Zürich
Switzerland

ISBN 978-3-642-06085-4 ISBN 978-3-662-08003-0 (eBook)
DOI 10.1007/978-3-662-08003-0

Library of Congress Control Number: 2004107593

springeronline.com

Originally published by Springer-Verlag Berlin Heidelberg New York in 2004
Softcover reprint of the hardcover 1st edition 2004

Typesetting: data delivered by authors
Cover design: medio Technologies AG, Berlin
Printed on acid free paper 62/3020/M - 5 4 3 2 1 0

Preface

Who should read this text?

This text is intended for students interested in the design of classical and novel IC engine control systems. Its focus lies on the control-oriented mathematical description of the physical processes involved and on the model-based control system design and optimization.

This text has evolved from a lecture series held during the last several years in the mechanical engineering (ME) department at ETH Zurich. The presumed audience is graduate ME students with a thorough understanding of basic thermodynamic and fluid dynamics processes in internal combustion engines (ICE). Other prerequisites are knowledge of general ME topics (calculus, mechanics, etc.) and a first course in control systems. Students with little preparation in basic ICE modeling and design are referred to [52], [81], [163], and [175].

Why has this text been written?

Internal combustion engines represent one of the most important technological success stories in the last 100 years. These systems have become the most frequently used sources of propulsion energy in passenger cars. One of the main reasons that this has occurred is the very high energy density of liquid hydrocarbon fuels. As long as fossil fuel resources are used to fuel cars, there are no foreseeable alternatives that offer the same benefits in terms of cost, safety, pollutant emission and fuel economy (always in a total cycle, or "well-to-wheel" sense, see e.g., [5] and [55]).

Internal combustion engines still have a substantial potential for improvements; Diesel (compression ignition) engines can be made much cleaner and Otto (spark ignition) engines still can be made much more fuel efficient. Each goal can be achieved only with the help of control systems. Moreover, with the systems becoming increasingly complex, systematic and efficient system

design procedures have become technological and commercial necessities. This text addresses these issues by offering an introduction to model-based control system design for ICE.

What can be learned from this text?

The primary emphasis is put on the ICE (torque production, pollutant formation, etc.) and its auxiliary devices (air-charge control, mixture formation, pollutant abatement systems, etc.). Mathematical models for some of these processes will be developed below. Using these models, selected feedforward and feedback control problems will then be discussed.

A model-based approach is chosen because, even though more cumbersome in the beginning, it after proves to be the most cost-effective in the long run. Especially the control system development and calibration processes benefit greatly from mathematical models at early project stages.

The appendix contains a brief summary of the most important controller analysis and design methods, and a case study that analyzes a simplified idle-speed control problem. This includes some aspects of experimental parameter identification and model validation.

What cannot be learned from this text?

This text treats ICE systems, i.e., the load torque acting on the engine is assumed to be known and no drive-train or chassis problems will be discussed.

Moreover, this text does not attempt to describe *all* control loops present in engine systems. The focus is on those problem areas in which the authors have had the opportunity to work during earlier projects.

Acknowledgments

Many people have implicitly helped us to prepare this text. Specifically our teachers, colleagues and students have helped to bring us to the point where we felt ready to write this text. Several people have helped us more explicitly in preparing this manuscript: Alois Amstutz, with whom we work especially in the area of Diesel engines, several of our doctoral students whose dissertations have been used as the nucleus of several sections (we reference their work at the appropriate places), Simon Frei, Marzio Locatelli and David Germann who worked on the idle-speed case study and helped streamlining the manuscript, and, finally, Brigitte Rohrbach and Darla Peelle, who translated our manuscripts from "Germlish" to English.

Zurich,
May 2004

Lino Guzzella
Christopher H. Onder

Contents

1

Introduction

This chapter contains some general remarks on electronic engine control systems and introduces the most common control problems encountered in spark ignition (Otto or gasoline) and compression ignition (Diesel) engine systems. The intention is to show the general motivation for using control systems and to give the reader an idea of the problems that can be tackled by feedforward and feedback systems for both SI and CI engines.

The emphasis in this chapter is on qualitative arguments. The mathematically precise formulation is deferred to subsequent chapters. Those readers not familiar with modern electronic sensors, actuators and control hardware for automotive applications may want to consult either [7], [90], or [104].

1.1 Control Systems for IC Engines

1.1.1 Relevance of Engine Control Systems

Future cars are expected to incorporate approximately one third of their parts value in electric and electronic components. These devices help to reduce the emission of pollutant species and fuel consumption, to increase safety, and to improve the drivability and comfort of passenger cars. As the electronic control systems become more complex and powerful, an ever increasing number of mechanical functions are being replaced by electric and electronic devices. An example of such an advanced vehicle is shown in Fig. 1.1.

In such a system, the engine is only one part within a larger structure. Its main input and output signals are the commands issued by the electronic control unit (ECU) or directly by the driver, and the load torque transmitted through the clutch onto the engine's flywheel. Figure 1.2 shows a possible substructure of the vehicle control system. In this text, only the "ICE" (i.e., the engine and the corresponding hardware and software needed to control the engine) will be discussed.

L. Guzzella et al., *Introduction to Modeling and Control of Internal Combustion Engine Systems*

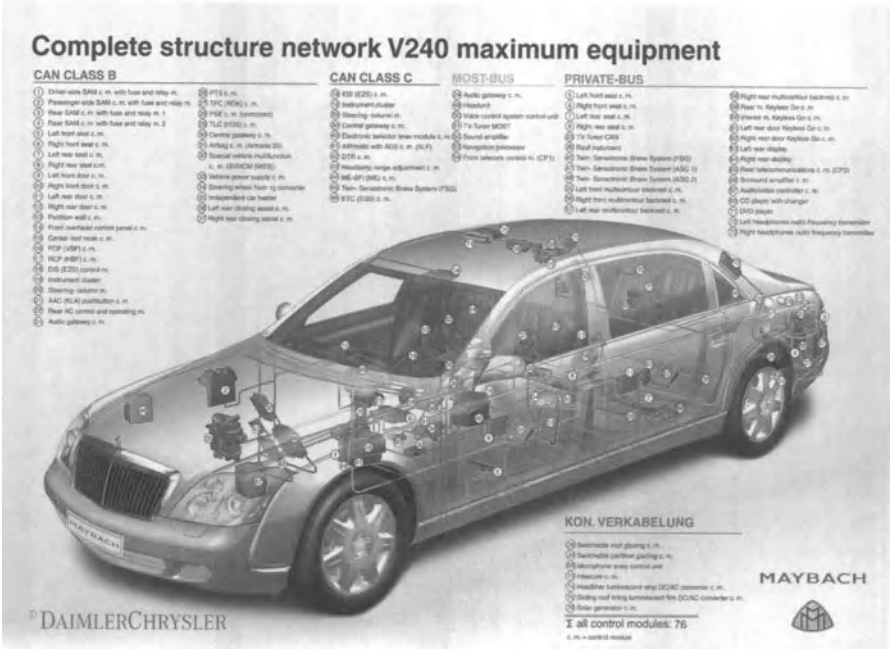


Fig. 1.1. Wiring harness of a modern vehicle (Maybach), reprinted with the permission of DaimlerChrysler AG.

Control systems were introduced in ICE on a larger scale with the advent of three-way catalytic converters for the pollutant reduction of SI engines. Good experiences with these systems and substantial progress in microelectronic components (performance and cost) have opened up a path for the application of electronic control systems in many other ICE problem areas. It is clear that the realization of the future, more complex, engine systems, e.g., hybrid power trains or homogeneous charge compression ignition engines, will not be possible without sophisticated control systems.

1.1.2 Electronic Engine Control Hardware and Software

Typically, an electronic engine control unit (ECU) includes standard microcontroller hardware (process interfaces, RAM/ROM, CPU, etc.) and at least one additional piece of hardware, which is often designated as a *time processing unit* (TPU), see Fig. 1.3. This TPU synchronizes the engine control commands with the reciprocating action of the engine.[1] Notice also that clock rates of ECU microprocessors are typically much lower than those of desktop computers due to electromagnetic compatibility considerations.

[1] The reciprocating or *event-based* behavior of all ICE also has important consequences for the controller design process. These problems will be addressed in Chapters 3 and 4.

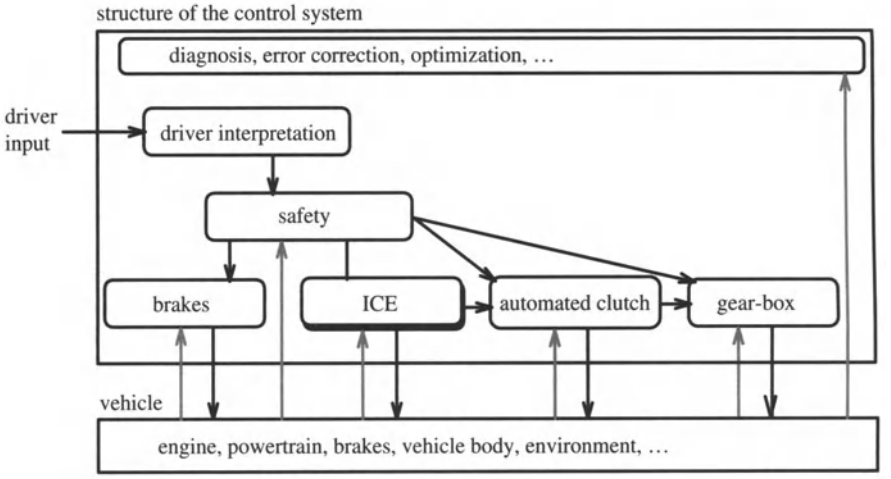


Fig. 1.2. Substructure of a complete vehicle control system.

ECU software has typically been written in assembler code, with proprietary real-time kernels. In the last few years there has been a strong tendency towards standardized high-level programming interfaces. Interestingly, the software is structured to reflect the primary physical connections of the plant to be controlled [57].

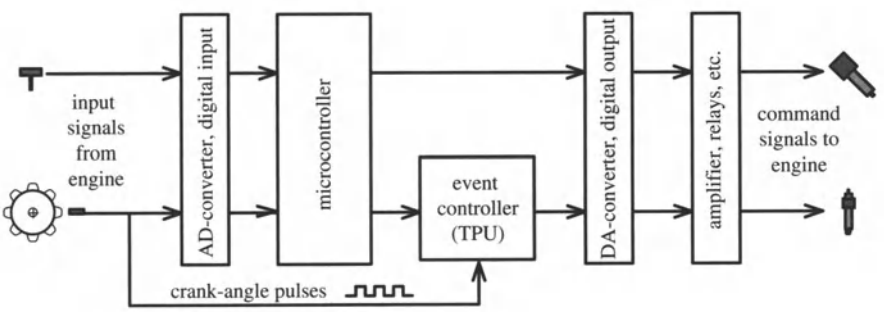


Fig. 1.3. Internal structure of an electronic engine controller unit.

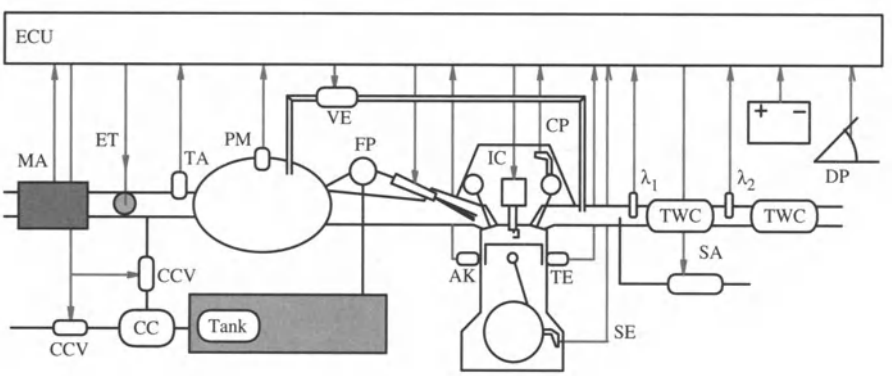
1.2 Overview of SI Engine Control Problems

1.2.1 General Remarks

The majority of modern passenger cars are still equipped with port (indirect) injection spark-ignited gasoline engines. The premixed and stoichiomet-

ric combustion of the Otto process permits an extremely efficient exhaust gas purification with three-way catalytic converters, and produces very little particulate matter (PM). A standard configuration of such an engine is shown in Fig. 1.4.

The torque of a stoichiometric SI engine is controlled by the quantity of air/fuel mixture in the cylinder during each stroke (the quality, i.e., the air/fuel ratio, remains constant). Typically, this quantity is varied by changing the intake pressure and, hence, the density of the air/fuel mixture. Thus, a throttle plate is used upstream in the intake system. This solution is relatively simple and reliable, but produces substantial "pumping losses" that negatively affect the part-load efficiency of the engine. Novel approaches, such as electronic throttle control, variable valve timing, etc., which offer improved fuel economy and pollutant emission, will be discussed below.



AK	knock sensor	PM	manifold pressure sensor	VE	EGR valve
CP	camshaft sensor	ET	electronic throttle	SA	secondary air valve
IC	ignition command	TA	intake air temperature sensor	TWC	3-way catalyst
MA	air mass-flow sensor	TE	cooling water temperature sensor	ECU	controller
SE	engine speed sensor	CC	active carbon canister	CCV	CC control valves
FP	fuel pressure control	$\lambda_{1,2}$	air/fuel ratio sensors	DP	driver pedal

Fig. 1.4. Overview over a typical SI engine system structure.

A simplified control-oriented substructure of an SI engine is shown in Fig. 1.5. The main blocks are the fuel path P_φ and the air path P_α, which define the mixture entering the cylinder, and the combustion block P_χ that determines the amount of torque produced by the engine.

Other engine outputs are the knock signal y_ζ (as measured by a knock sensor P_ζ) and the engine-out air/fuel ratio y_λ (as measured by a λ sensor P_λ mounted as close as possible to the exhaust valves). The engine speed ω_e is the output of the block P_Θ, taking into account the rotational inertia of the engine, whose inputs are the engine torque T_e and the load torque T_l.

The four most important control loops are indicated in Fig. 1.5 as well:

- the fuel-injection feedforward loop;
- the air/fuel ratio feedback loop;
- the ignition angle feedforward[2] loop; and
- the knock feedback loop.

In addition, the following feedforward or feedback loops are present in many engine systems:[3]

- idle and cruise speed control;
- exhaust gas recirculation (for reducing emission during cold-start or for lean-burn engines);
- secondary air injection (for faster catalyst light-off); and
- canister purge management (to avoid hydrocarbon evaporation).

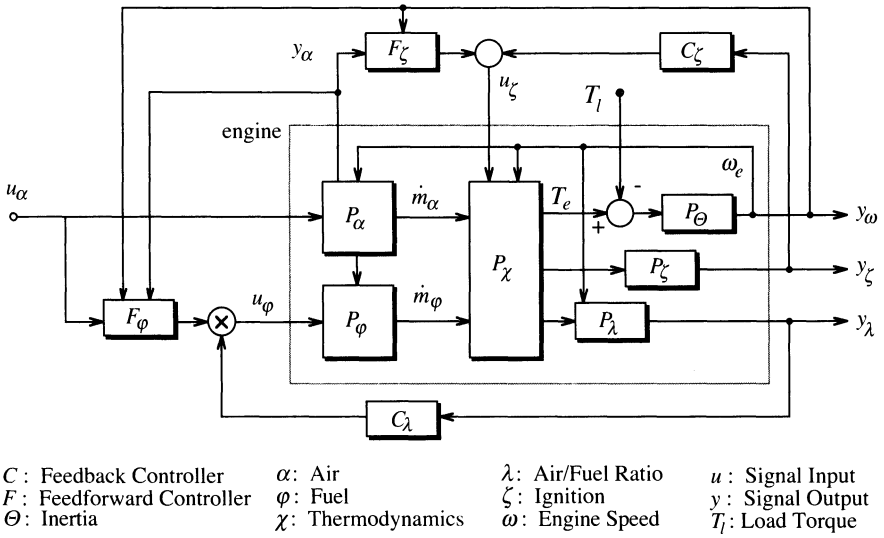


Fig. 1.5. Basic SI engine control substructure.

1.2.2 Main Control Loops in SI Engines

Air/Fuel Ratio Control

The air/fuel ratio control problem has been instrumental in paving the road for the introduction of several sophisticated automotive control systems. For this reason, it is described in some detail.

[2] Closed-loop control has been proposed in [50] using the spark plug as an ion current sensor.

[3] Modern SI engines can include several other control loops.

The engine-out pollutant emission of SI engines (mainly hydrocarbon (HC), carbon monoxide (CO) and nitrogen oxide (NO_x) greatly exceed the levels mandated by most regulatory boards, and future emission legislation will require substantial additional reductions of pollutant emission levels. These requirements can only be satisfied if appropriate exhaust gas after-treatment systems are used.

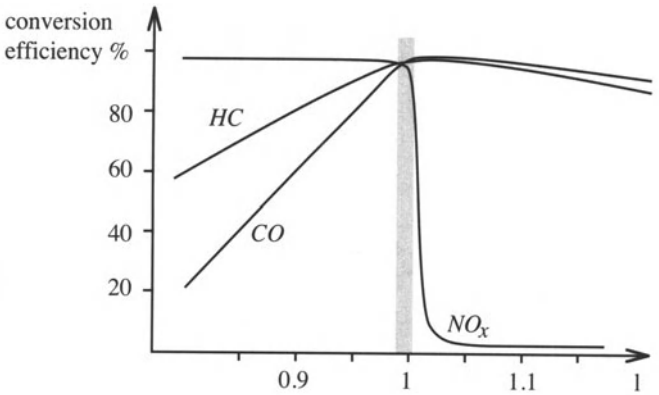


Fig. 1.6. Conversion efficiency of a TWC (after light-off, stationary behavior).

The key to clean SI engines is a three-way catalytic converter (TWC) system whose stationary conversion efficiency is depicted in Fig. 1.6. Only for a very narrow air/fuel ratio "window," whose mean value is slightly below the stoichiometric level, can all three pollutant species present in the exhaust gas be almost completely converted to the innocuous components water and carbon dioxide. In particular, when the engine runs under lean conditions, the reduction of nitrogen oxide stops almost completely, because the now abundant free oxygen in the exhaust gas is used to oxidize the unburned hydrocarbon and the carbon monoxide. Only when the engine runs under rich conditions, do the unburned hydrocarbon (HC) and the carbon monoxide (CO) act as agents reducing the nitrogen oxide on the catalyst, thereby causing the desired TWC behavior.

The mean air/fuel ratio can be kept within this narrow band only if electronic control systems and appropriate sensors and actuators are used. The air/fuel ratio sensor (λ sensor) is a very important component in this loop. A precise fuel injection system also is necessary. This is currently realized using "sequential multiport injectors." Each intake port has its own injector, which injects fuel sequentially, i.e., only when the corresponding intake valves are closed.

Finally, appropriate control algorithms have to be implemented in the ECU. The fuel-injection feedforward controller F_φ tries to quickly realize a suitable injection timing based only on the measured air-path input informa-

tion (either intake air mass flow, intake manifold pressure, or throttle plate angle and engine speed). The air/fuel ratio feedback control system C_λ compensates the unavoidable errors in the feedforward loop. While it guarantees the mean value of the air/fuel ratio to be at the stoichiometric level, it cannot avoid transient excursions in the air/fuel ratio.

Ignition Control

Another important example of a control system in SI engines is the spark angle control system. This example shows how control systems can help improve fuel economy as well.

In fact, the efficiency of SI engines is limited, among other factors, by the knock phenomenon. Knock (although still not fully understood) results from an unwanted self-ignition process that leads to locally very high pressure peaks that can destroy the rim of the piston and other parts in the cylinder. In order to avoid knocking, the compression ratio must be kept below a safe value and ignition timing must be optimized off-line and on-line.

A first optimization takes place during the calibration phase (experiments on engine or chassis dynamometers) of the engine development process. The nominal spark timing data obtained are stored in the ECU. An on-line spark timing control system is required to handle changing fuel qualities and engine characteristics. The key to this component is a knock sensor and the corresponding signal processing unit that monitors the combustion process and signals the onset of knocking.

The feedforward controller F_ζ, introduced in Fig. 1.5, computes the nominal ignition angles (realizing maximum brake torque while avoiding knock and excessive engine-out pollution levels) depending on the engine speed and load (as measured by manifold pressure or other related signals). This correlation is static and is only optimal for that engine used to obtain the ignition data during the calibration of the ECU. The feedback control system C_ζ utilizes the outcome of the knock detection system to adapt the ignition angle to a safe and fuel efficient value despite variations in environmental conditions, fuel quality, etc.

1.2.3 Future Developments

Pollutant emission of stoichiometric SI engines is or soon will be a "problem solved" such that the focus of research and development efforts will be redirected towards the improvement of the fuel economy. The most severe drawbacks of current SI engines are evident in part-load operating conditions. As Fig. 1.7 shows, the average efficiency even of modern SI engines remains substantially below their best bsfc[4] values. This is a problem because most passenger cars on the average (and also on the governmental test cycles) utilize

[4] Brake-specific fuel consumption (usually in g fuel/kWh mechanical work).

less than 10% of the maximum engine power.[5] Not surprisingly, cycle-averaged "tank-to-wheel" efficiency data of actual passenger cars are between 12% and 18% only. The next step in the development of SI engines therefore will be a substantial improvement of their part-load efficiency.

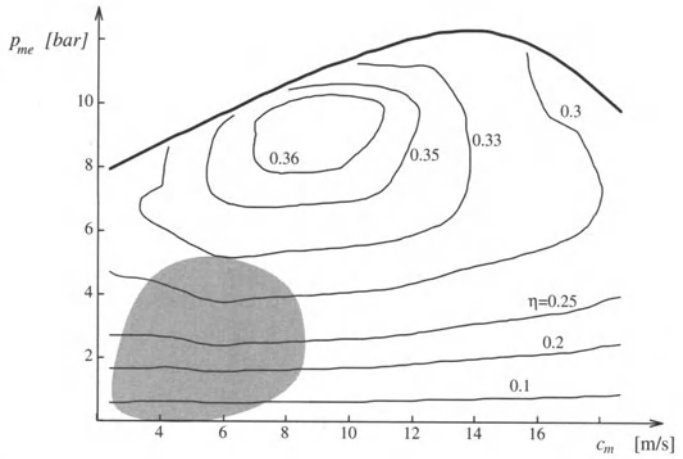


Fig. 1.7. Engine map (mean effective pressure versus mean piston speed) of a modern SI engine, gray area: part-load zone, η =const: iso-efficiency curves. For the definition of p_{me} and c_m see Sect. 3.2.1.

Several ideas have been proposed to improve the fuel efficiency of SI engines, all of which include some control actions, e.g.,

- variable valve timing systems (electromagnetic or electrohydraulic);
- downsizing and supercharging systems;
- homogeneous and stratified lean combustion SI engines;
- variable compression ratio engines; and
- engines with improved thermal management.

These systems reduce the pumping work required in the gas exchange part of the Otto cycle, reduce mechanical friction, or improve the thermodynamic efficiency in part-load conditions.

Another approach to improving part-load efficiency is to include novel power train components, such as starter-generator[6] devices, CVTs[7], etc. As

[5] Maximum engine power is mainly determined by the customer's expectation of acceleration performance and is, therefore, very much dependent on vehicle mass.

[6] These advanced electric motors and generators typically have around 5 kW mechanical power and permit several improvements like idle-load shut-off strategies or even "mild hybrid" concepts.

[7] Continuously variable transmissions allow for the operation of the engine at the lowest possible speed and highest possible load, thus partially avoiding the low efficiency points in the engine map.

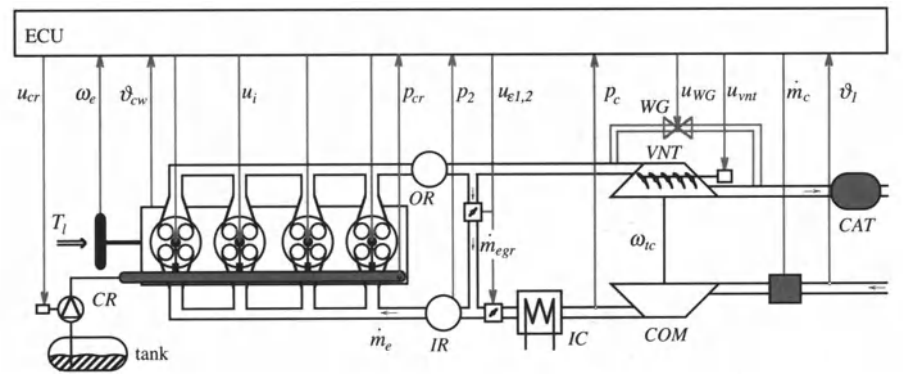
mentioned in the Introduction, these approaches will not be analyzed in this text.

1.3 Overview of CI Engine Control Problems

1.3.1 General Remarks

Diesel engines are inherently more fuel-efficient than gasoline engines, but they cannot use the pollutant abatement systems that have proved to be so successful in gasoline engines. In fact, the torque output of Diesel engines is controlled by changing the air/fuel ratio in the combustion chamber. This approach is not compatible with the TWC working principle introduced above.

In naturally aspirated Diesel engines, the amount of air available is approximately the same for all loads, and only the amount of fuel injected changes in accordance with the driver's request. In modern CI engines the situation is even more complex as almost all engines are turbocharged. This turbocharger introduces additional feedback paths, which considerably complicate the dynamic behavior of the entire engine system. Another trend is to replace prechamber injection by direct-injection systems. The injection is realized either using so-called common-rail or integrated pump-injector systems.



CAT	oxidation catalytic converter	$u_{\varepsilon 1,2}$	EGR valve(s) command	p_c	pressure after COM
COM	compressor	u_{cr}	CR pump command	p_2	intake pressure
CR	common-rail system	u_i	injection command	p_{cr}	CR injection pressure
IR	intake receiver	u_{vnt}	turbine nozzle command	$\dot{m}_c$	intake air mass flow
OR	outlet receiver	u_{WG}	WG command	ϑ_1	intake air temperature
IC	intercooler	T_l	load torque at the flywheel	ϑ_{cw}	cooling water temperature
VNT	variable nozzle turbine	ω_{tc}	turbocharger speed	ω_e	engine speed
WG	waste-gate (alternative to VNT)	$\dot{m}_e$	total engine-in mass flow	$\dot{m}_{egr}$	exhaust gas recirculation

Fig. 1.8. Overview of a typical Diesel engine system structure.

Compression ignition, or Diesel engines, have been traditionally less advanced in electronic controller utilization due to cost, reliability and image

problems in the past. This situation has changed recently, and it is expected that electronic control systems will help to substantially improve the total system behavior (especially the pollutant emission) of Diesel engines [65].

Figure 1.8 shows an overview of a typical modern Diesel passenger car engine. The main objective for electronic Diesel engine control systems is to provide the required engine torque with minimal fuel consumption under the constraint of meeting the given exhaust gas and noise emission regulations. This requires an optimal coordination of injection, turbocharger and exhaust gas recirculation (EGR) systems in stationary and dynamic operating conditions.

From a control engineering point of view, there are three important paths which have to be considered, *viz.*:

- the fuel path;
- the air path; and
- the EGR path.

Figure 1.9 gives a schematic overview of the basic structure of a typical Diesel engine control system (for more details on the inner structure of the Diesel engine, see Sect. 2.1). Notice that a speed controller is standard in Diesel engines; at least the top speed must be actively controlled in order to prevent engine damage.

The fuel path with the outputs torque, speed and exhaust gas emission obtains its inputs from an engine speed feedback controller, see Fig. 1.9. The control inputs to the fuel path are start of injection, injection duration, and injection pressure. With common-rail systems, new degrees of freedom, such as the choice of a pilot injection, main and after-injection quantities with different dwell times in between, are added. Using the measurement of the air mass flow into the engine, the maximum quantity of injected fuel is limited in order to avoid visible smoke.

The turbocharger dominates the air path. Especially in applications with heavy transient operations, turbocharger designs with small A/R ratios (nozzle area over diameter of the turbine wheel) are chosen to get a good acceleration performance of the supercharging device. Unfortunately, at large loads the small-sized turbocharger creates boost pressures that, after compression in the engine, would exceed the engine's peak pressure limits. Therefore, a substantial fraction of the exhaust gas has to bypass the turbine through a waste gate. The side effect of this bypass is a considerable loss in efficiency. With the expectations of good acceleration performance and high turbocharger efficiency over the whole range of operation, instead of waste-gate systems variable nozzle turbochargers (VNT) are used. VNT are typically controlled using a closed-loop approach where the measured output is the boost pressure. Moreover, special care has to be taken to avoid VNT to reach dangerously high rotation speeds.

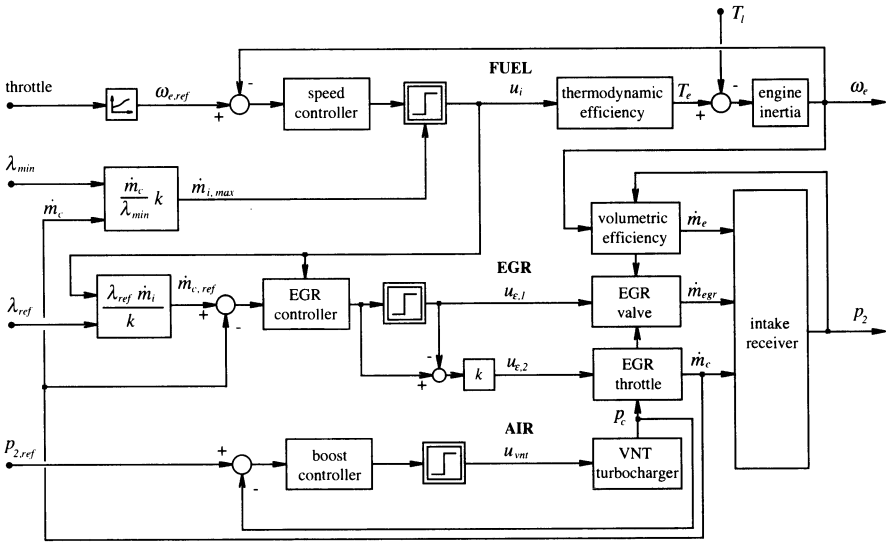


Fig. 1.9. Basic Diesel engine control system structure, variables defined in Fig. 1.8.

1.3.2 Main Control Loops in Diesel Engines

With exhaust emission laws becoming ever more stringent, exhaust gas recirculation is needed for NO_x reduction. A closed-loop control system takes care of the EGR path. Even if the objective is EGR control, the closed loop takes the measured air mass flow as the feedback variable. A Diesel engine starts smoking if the air/fuel ratio falls below a certain value. The air/fuel ratio with the best NO_x reduction under the boundary condition of no increase in smoke generation is mapped over the quantity of fuel injected and the engine speed. Together with the quantity of fuel injected, which is known from the injection table, the reference air mass flow can be derived. The EGR valve position is determined from the difference between reference and measured air mass flows. The EGR flow heavily influences the air mass flow through the states of the intake and outlet receiver. Additional devices, such as throttles, must be used to maintain a pressure difference over the EGR valve.

While engine performance, dynamic behavior, and bsfc are critical factors for the manufacturers of an engine, emission laws must be met by all engine systems. Figure 1.10 summarizes the main input-output relationship and attempts to illustrate the complex network of the underlying physics. In most cases, a single control input affects several outputs.

Three main phenomena are especially important to understanding the key issues of Diesel emission control:

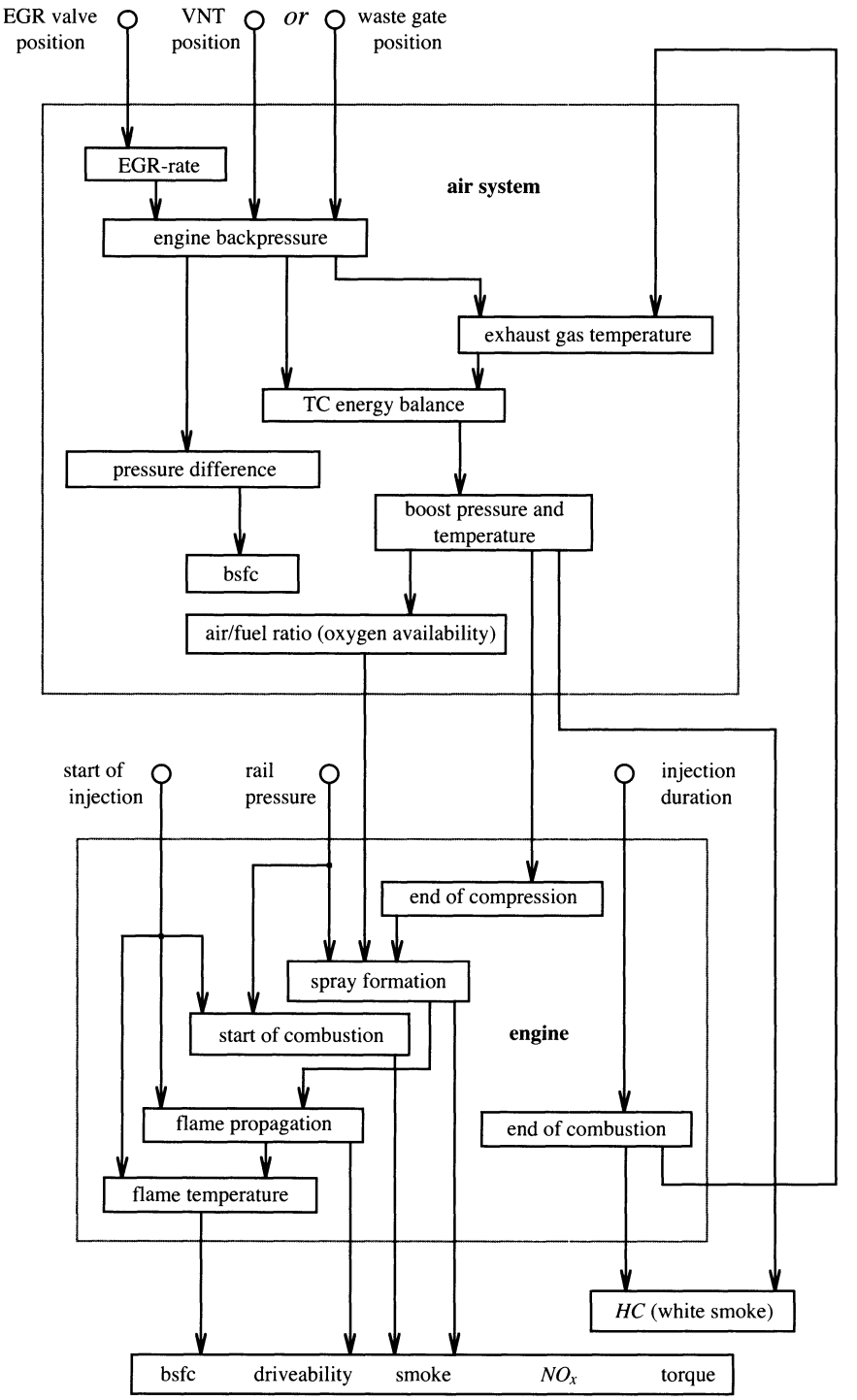


Fig. 1.10. Qualitative input-output relations in a Diesel engine.

- The thermal efficiency of a combustion process η_{Carnot} depends on the mean combustion temperature ϑ_{com} (determined by the Carnot cycle)

$$\eta_{Carnot} = 1 - \frac{\vartheta_{exh}}{\vartheta_{com}} \tag{1.1}$$

  where ϑ_{exh} is the mean exhaust temperature. Obviously, the higher the combustion temperature with respect to the exhaust temperature, the better the thermal efficiency and, therefore, the brake-specific fuel consumption bsfc.
- The rate at which NO is produced can be approximated by [81]

$$\frac{d}{dt}[NO] = \frac{6 \cdot 10^{16}}{\sqrt{\vartheta}} \cdot e^{\frac{-69090}{\vartheta}} \cdot [O_2]^{1/2} \cdot [N_2] \tag{1.2}$$

  where [.] denotes equilibrium concentrations. The strong dependence of the NO formation on the temperature ϑ of the burned gas fraction in the exponential term is evident. High temperatures and oxygen concentrations, therefore, result in high rates of NO formation.
- Diesel particulate matter consists principally of combustion-generated soot on which some organic compounds have become adsorbed. Lubricating oil contributes to the formation of particulate matter. The amount of Diesel particulate matter depends upon oxygen availability during combustion, upon spray formation and upon oxidation conditions towards the end of the combustion process.

This raises the question of how the electronic control inputs affect the key parameters mentioned:

- Brake-specific fuel consumption: With a given injection rate, the main inputs to improve bsfc are start of injection, rail pressure and boost pressure. What is desired is an early start of injection which results in a fast heat release around top dead center (TDC). Because of the sinusoidal motion of the piston in this area, the volume of the combustion chamber remains almost constant, which results in high gas temperatures and, thus, a good Carnot efficiency. Increasing rail pressures leads to shorter ignition delays and faster burn rates. Therefore, for the same bsfc, the start of injection may be delayed.
- NO formation: This effect is related to the temperature that determines the Carnot efficiency. As a consequence, a late start of injection combined with EGR produces low NO emission. EGR reduces the peak temperatures and, thus, reduces the NO formation. At the same time, relatively high temperatures during the expansion stroke enhance the reduction of formed NO.
- Particulate matter: Early start of injection with its fast, hot and complete combustion, produces low amounts of particulate matter. Additionally, with an early start of injection, conditions for soot oxidation are good

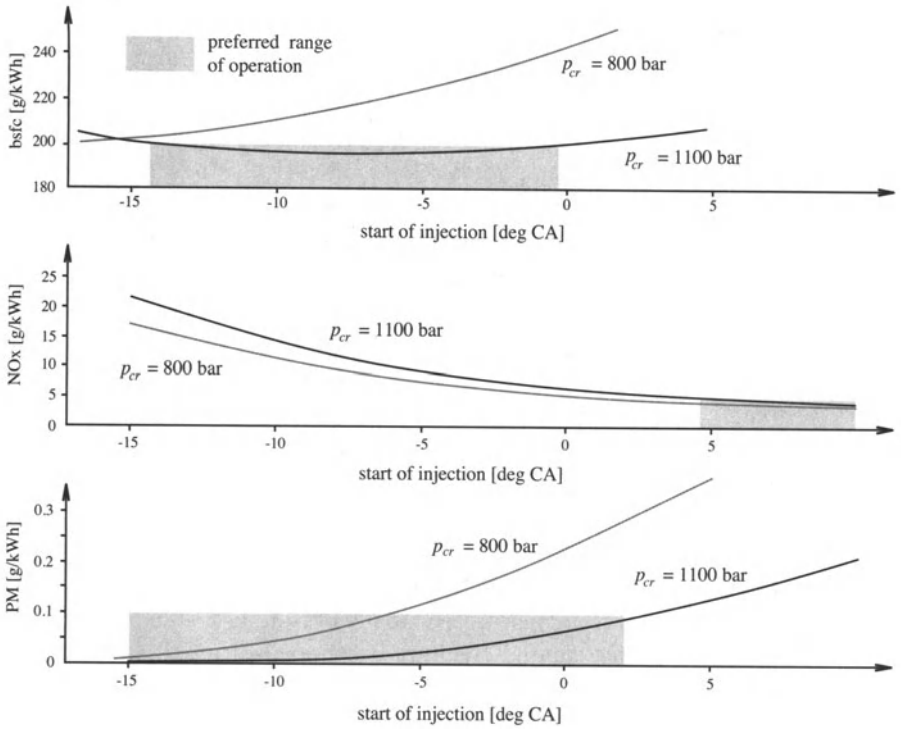


Fig. 1.11. Influence of start of injection on bsfc, and on the emission of PM and NO_x, with p_{cr} representing the injection pressure.

during the long period of the expansion stroke. Due to the good influence on spray formation, high injection pressures also are beneficial to obtain low amounts of PM.

The tendencies of various input-output relations are summarized in Table 1.1. It shows the difficulty of stating clear control objectives, since nearly every input has a good and a bad effect on the outputs of interest. A well-known method for dealing with the PM-NO_x trade-off is to vary the start of injection between early start of injection (e.g., 30 degrees before TDC) to late start of injection (e.g., eight degrees after TDC). The desired injection timing occurs where the trade-off curve crosses the acceptable emission limits as defined by the corresponding emission regulations. This approach, however, does not take into account that bsfc should be as low as possible.

The PM-NO_x trade-off is inherently connected to the principle of the Diesel cycle and creates an obstacle that is difficult to surmount. Figure 1.11 shows the influence of injection on bsfc and on the emission of NO_x and PM. The shaded areas indicate where the emission values are within the regulations and where bsfc is below 200 g/kWh. In this case, even with an injection pressure of 1100 bar, there is no starting point for the injection at which

Table 1.1. Tendencies in the influence of control inputs on fuel economy (bsfc) and emission quantity.

Control Input	Result
early start of injection	good bsfc low particulate matter high NO_x
high rail pressure	increased NO_x low PM slightly improved bsfc
pilot injection(s)	low noise
smaller VNT area	improved bsfc lower particulate higher NO_x
increased EGR	lower NO_x danger of higher PM improved noise equal or slightly increased bsfc

NO_x and PM satisfy current emission limits. Especially in Europe, selective catalytic reduction (SCR) technologies are being investigated to clean the emission in an after-treatment system, and to break the PM-NO_x trade-off.

For current applications, the problem is being tackled in the following order: sophisticated developments are pursued in the areas of the combustion chamber, the injection, air and EGR systems. The electronic control system must then guarantee the best use of the given hardware under stationary and transient conditions. In all of these approaches, model-based controller designs play an important role as an enabling technology.

1.3.3 Future Developments

Diesel engines clearly have a large potential to become much cleaner. On one side, progress in reducing engine-out emissions will continue. On the other side, several aftertreatment systems are ready to be introduced on a large scale.

In heavy-duty applications, where fuel economy is a top priority, lean de-NO_x systems using a selective catalytic reduction (SCR) approach are an interesting alternative. Such systems can reduce engine-out NO_x by approximately one order of magnitude. This permits the engine to be calibrated at the high-efficiency/low-PM boundary of the trade-off curve (see Fig. 1.11, early

injection angle). The drawback of this approach is, of course, the need for an additional fluid distribution infrastructure (most likely urea).

While such systems are feasible in heavy-duty applications, for passenger cars it is generally felt that a solution with Diesel particulate filters (DPF) is more likely to be successful on a large scale. These filters permit an engine calibration at the high-PM/low-NO_x boundary of the trade-off curve mentioned before. Using this approach, it is hoped that the engine-out NO_x emissions will meet the legislation limits without further aftertreatment systems. Of course, this will require high EGR rates. If this approach is not sufficient, the tailpipe NO_x emission can be further reduced using lean NO_x trap catalytic converters.

Radically new approaches, such as cold-flame combustion (see e.g., [159]) or homogeneous-charge compression ignition engines (HCCI, see [156], [157], [142] for control-oriented discussion) promise further reduction in engine-out emissions especially at part-load conditions.

In all of these approaches, feedforward and feedback control systems will play an important role as an enabling technology. Moreover, with ever increasing system complexity, model-based approaches will become even more important.

1.4 Structure of the Text

The main body of this text is organized as follows:

- Chapter 2 introduces mean-value[8] models of the most important phenomena in IC engines.
- Chapter 3 derives discrete-event or crank-angle models for those subsystems that will need such descriptions to be properly controlled.
- Chapter 4 discusses some important control problems by applying a model-based approach for the design of feedforward as well as feedback control systems.

In addition to these three chapters, the two appendices contain the following information:

- Appendix A summarizes, in a concise formulation, most control system analysis and synthesis ideas that are required to follow the main text.
- Appendix B illustrates the concepts introduced in the main text by showing the design of a simplified idle-speed controller. This includes some remarks on parameter identification and on model validation using experimental data.

[8] The term mean-value is used to designate models that do not reflect the engine's reciprocating and hence crank-angle sampled behavior, but which use a continuous-time lumped parameter description. Discrete-event models, on the other hand, explicitly take these effects into account.

1.5 Notation

The notation used in this text is fairly standard. The *derivative* of a variable $x(t)$, with respect to its independent variable t, is denoted by

$$\frac{d}{dt}x(t)$$

while the notation

$$\dot{x}(t)$$

is used to indicate a *flow* of mass, energy, etc. Both variables $\frac{d}{dt}x(t)$ and $\dot{x}(t)$ have the same units, but they are different objects. No special distinction is made between scalars, vectors and matrices. The dimensions of a variable, if not a scalar, are explicitly defined in the context. Input signals are usually denoted by $u_{...}$ and output signals by $y_{...}$ where the index ... specifies what physical quantity is actuated or measured.

Concentrations of chemical species C are denoted by $[C]$, with units mol/mol, with respect to the reference substance. The concentrations are therefore limited to the interval $[0, 1]$. The concentration of pollutant species are often shown in plots or tables using *ppm* units (part per million), i.e., by using a amplification factor of 10^6. For mass storage and transportation models it is advantageous to use mass fractions, which are denoted by ξ having units [kg/kg].

In general, all variables are defined at that place in the text where they are used for the first time. To facilitate the reading, some symbols have been reserved for special physical quantities:

α : heat transfer coefficient, units $W/(m^2\,K)$;
A : area, units m^2;
c_x : specific heat capacities ($x = p$ or v), units $J/(kg\,K)$;
η : efficiency, units $-$;
ϕ : crank angle degrees, units $^\circ$ or rad;
$\dot{H}$: enthalpy flows, units W;
κ : ratio of specific heats, units $-$.
λ : air/fuel ratio or volumetric efficiency, units $-$.
m : mass, units kg;
N : for a four-stroke engine $N = 2$, for a two-stroke engine $N = 1$;
p : pressure, units Pa or bar;
P : power, units W;
Π : pressure ratios, units $-$;
$\dot{Q}$: heat flows, units W;
ρ : densities, units kg/m^3;
R : specific gas constant, units $J/(kg\,K)$;
$\mathcal{R}$: universal gas constant, units $J/(mol\,K)$;
t : time as independent variable, units s;
τ : time as time constants, units s;

ϑ : temperature, units degrees K or C;
θ : occupancy, units $-$;
T : torque, units Nm;
Θ : inertia, units m^2kg;
V : volume, units m^3 or l;
ζ : ignition angle, units $^\circ$ or rad; and
ω : speed or angular frequency, units rad/s.

Similarly, some indices have been reserved for special use:

α, a, β for air or ambient;
c for compressor or cylinder;
e for engine;
egr or ε for exhaust gas recirculation;
m for manifold or mean values;
seg for segment;
t for turbine;
φ, ψ for fuel;
ξ for combustion; and
ζ for timing (ignition, injection, ...).

In a turbocharged engine system, the four most important locations are designated by the indices 1 for "before compressor," 2 for "after compressor," 3 for "after engine," and 4 for "after turbine."

In general, all numerical values listed in this text are shown in SI units. A few exceptions are made where non-SI units are widely accepted. These few cases are explicitly mentioned in the text.

The most commonly used abbreviations are:

me — (brake) mean-effective (pressure), some authors use bmep;
bsfc — brake-specific fuel consumption;
CI — compression-ignition (Diesel) engines;
CNG — compressed natural gas;
COM — control oriented model;
DEM — discrete event model;
ECU — electronic (engine) control unit;
IEG — induction-to-exhaust delay;
IPS — induction-to-powerstroke delay;
IVC — inlet valve closing;
EVC — exhaust valve closing;
MBT — maximum brake torque;
ODE — ordinary differential equation;
PDE — partial differential equation;
SCR — selective catalytic reduction;
SI — spark-ignition (gasoline, Otto) engines;
TDC — top dead center;
BDC — bottom dead center;

TPU — time-processing unit;
TWC — three way catalytic converter;
VNT — variable-nozzle turbine; and
WOT — wide open throttle.

2

Mean-Value Models

In this chapter, mean-value models *(MVM) of the most important subsystems of SI and Diesel engines are introduced. In this book, the notion of MVM*[1] *will be used for a specific set of models as defined below. First, a precise definition of the term MVM is given. This family of models is then compared to other models used in engine design and optimization. The main engine sub models are then discussed: the air system that defines how much air is inducted into the cylinder; the fuel system that defines how much fuel is inducted into the cylinder; the torque generation system that defines how much torque is produced by the air and fuel in the cylinder as defined by the first two parts; the engine inertial system that defines the engine speed; the engine thermal system that defines the dynamic thermal behavior of the engine; the pollution formation system that models the engine-out emission; and the pollution abatement system that models the behavior of the catalysts, the sensors and other relevant equipment in the exhaust pipe.*

All these models are control oriented models *(COM), i.e., they model the input-output behavior of the systems with reasonable precision but low computational complexity, and include, explicitly, all relevant transient (dynamic) effects. Typically, these COM will be represented by systems of nonlinear differential equations. Only physics-based COM will be discussed, i.e., models that are based on physical principles and few experiments necessary to identify some key parameters.*

[1] The terminology MVM was probably first introduced in [73]. One of the earliest papers proposing MVM for engine systems is [164]. A good overview of the first developments in the area of MVM of SI engine systems can be found in [137]. A more recent source of information on this topic is [37].

L. Guzzella et al., *Introduction to Modeling and Control of Internal Combustion Engine Systems*

2.1 Introduction

Reciprocating engines in passenger cars clearly differ in at least two aspects from continuously operating thermal engines like gas turbines:

- the combustion process itself is highly transient (Otto or Diesel cycle, with large and rapid temperature and pressure variations); and
- the thermodynamic boundary conditions, that govern the combustion process (intake pressure, composition of the air/fuel mixture, etc.), are not constant.

The thermodynamic and kinetic processes in the first class of phenomena are very fast (few milliseconds for a full Otto or Diesel cycle) and usually not accessible for control purposes. Moreover, the models necessary to describe these phenomena are rather complex and are not useful for the design of real-time control systems.

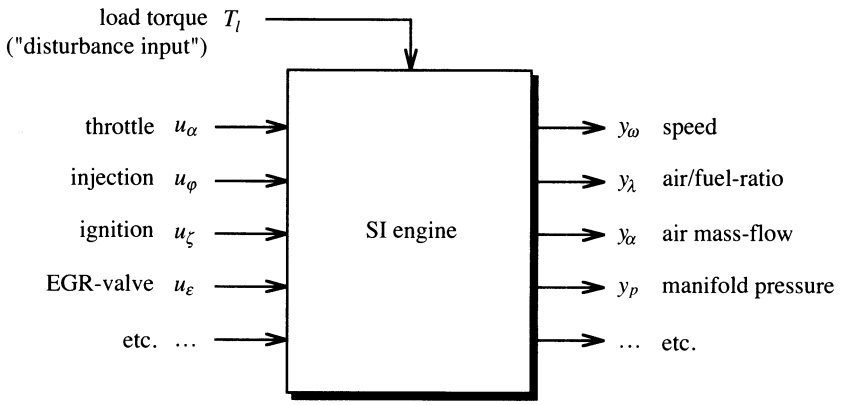


Fig. 2.1. Main system's input/output signals in a COM of an SI engine (similar for Diesel engines).

This text focuses on the second class of phenomena using control-oriented models, and simplifies the fast combustion characteristics as static effects. The underlying assumption is that, once all important thermodynamic boundary conditions at the start of an Otto or Diesel cycle are fixed, the combustion itself will evolve in an identical way each time the same initial starting conditions are imposed. Clearly, such models will not be able to reflect all phenomena (the random combustion pressure variations in SI engines, for example).

As shown in Fig. 2.1, in the COM paradigm, the engine is a "gray box" that has several input (command) signals, one main disturbance signal (the load torque) and several output signals. The inputs are signals, i.e., quantities

that can be arbitrarily chosen.[2] The outputs also are signals (not physical quantities) that can be used by the controller without the system behavior being affected. The only physical link of the engine to the rest of the power train is the load torque, which in this text will be assumed to be known.

The reciprocating behavior of the engine induces another dichotomy in the COM used to describe the engine dynamics:

- *Mean value models* (MVM), i.e., continuous COM, which neglect the discrete cycles of the engine and assume all processes and effects are spread out over the engine cycle;[3] and
- *Discrete event models* (DEM), i.e., COM that explicitly take into account the reciprocating behavior of the engine.

In MVM, the time t is the independent variable, while in DEM, the crankshaft angle ϕ is the independent variable. Often, DEM are formulated assuming a constant engine speed. In this case, they coincide with classical sampled data systems, for which a rich and (at least for linear systems) complete theoretical background exists. These aspects will be treated in detail in Chapter 3.

In MVM, the reciprocating behavior is captured by introducing delays between cylinder-in and cylinder-out effects (see Fig. 2.2). For example, the torque produced by the engine does not respond immediately to an increase in the manifold pressure. Only after the induction-to-power-stroke (IPS) delay [136]

$$\tau_{IPS} \approx \frac{2\pi}{\omega_e} \tag{2.1}$$

has elapsed will the new engine torque be active.[4]

Similarly, any changes of the cylinder-in gas composition (air/fuel ratio, EGR ratio, ...) will be perceived at the cylinder exhaust only after the exhaust gas (IEG) delay

$$\tau_{IEG} \approx \frac{3\pi}{\omega_e} \tag{2.2}$$

The proper choice of model class depends upon the problem to be solved. For example, MVM are well suited to relatively slow processes in the engine periphery, constant engine speed DEM are useful for air/fuel ratio feedforward control, and crank-angle DEM are needed for misfire detection algorithms based on crankshaft speed measurement.

[2] In order to allow full control of the engine, usually these will be electric signals, e.g., the throttle plate will be assumed to be "drive-by-wire."

[3] In MVM, the finite swept volume of the engine can be viewed as being one that is distributed over an infinite number of infinitely small cylinders.

[4] The expression (2.1) is valid for four-stroke engines. Two-stroke engines have half of that IPS delay. As shown in Chapter 3, additional delays are introduced by the electronic control hardware.

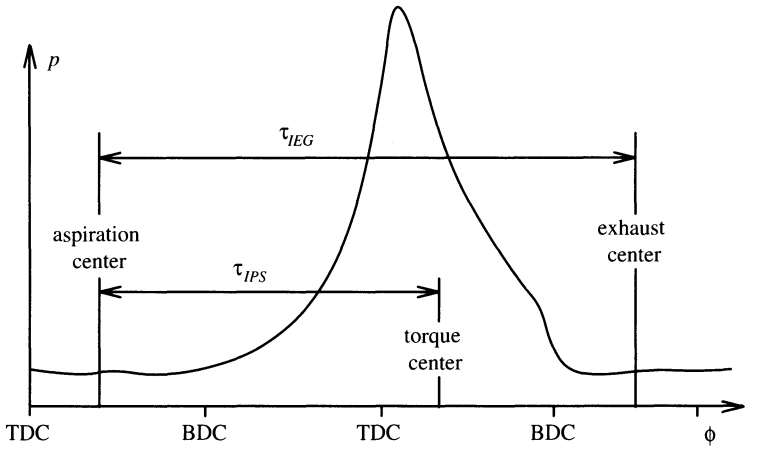


Fig. 2.2. Definition of IPS and IEG delays for MVM using a pressure/crank angle diagram.

Most MVM are *lumped parameter* models, i.e., system descriptions that have no spatially varying variables and that are represented by ordinary differential equations (ODE). If not only time but location must be used as an independent variable, *distributed parameter models* result that are described by partial differential equations (PDE). Such models usually are computationally too demanding to be useful for real-time applications[5] such that a spacial discretization is necessary (see e.g., Sect. 2.8).

2.2 Cause and Effect Diagrams

In this section, the internal structure of MVM for SI and CI engines will be analyzed in detail. When modeling any physical system there are two main classes of objects that must be taken into account:

- *reservoirs*, e.g., of thermal or kinetic energy, of mass or of information (there is an associated *level* variable to each reservoir that depends directly on the reservoir's content); and
- *flows*, e.g., energy, mass, etc. flowing between the reservoirs (typically driven by differences in reservoir levels).

A diagram containing all relevant reservoirs and flows between these reservoirs will be called a *cause and effect diagram* (see, for instance, Fig. 2.5). Since such a diagram shows the driving and driven variables, the cause and effect relations become clearly visible.

[5] There are publications which propose PDE-based models for control applications, see for instance [36].

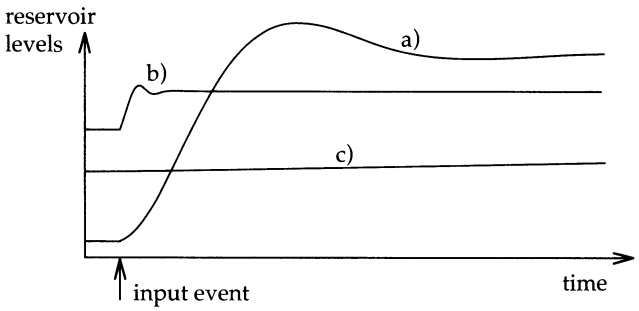


Fig. 2.3. Relevant reservoirs: a) variable of main interest, b) very fast and c) very slow variables.

A good MVM will contain only the *relevant* reservoirs (otherwise "stiff" systems will be obtained). To define relevant more precisely, the three signals shown in Fig. 2.3 can be useful. Signal a) is the variable of main interest (say, the manifold pressure). Signal b) is very fast compared to a) (say, the throttle plate angle dynamics) and must be modeled as a purely static variable which can depend in an algebraic way on the main variable a) and the input signals. Signal c) is very slow compared to a) (say, the temperature of the manifold walls) and must be modeled as a constant (which may be adapted after a longer period). Only in this way will a useful COM be obtained.

Unfortunately, there are no simple and systematic rules of how to decide a priori which reservoirs can be modeled in what way. Here, experience and/or iteration will be necessary, making system modeling partially an "engineering art."

2.2.1 Spark-Ignited Engines

Port-Injection SI Engines

A typical *port-injected* SI engine system has the structure shown in Fig. 2.4. In a mean value approach, the reciprocating behavior of the cylinders is replaced by a continuously working volumetric pump that produces exhaust gases and torque. The resulting main engine components are shown in Fig. 2.4. The different phenomena will be explained in detail in the following sections.

However, the main reservoir effects can be identified at the outset:

- gas mass in the intake and exhaust manifold;
- internal energy in the intake and exhaust manifold;
- fuel mass on the intake manifold walls (wall-wetting effect);
- kinetic energy in the engine's crankshaft and flywheel;
- induction-to-power-stroke delay in the combustion process (essentially an information delay); and
- various delays in the exhaust manifold (including transport phenomena).

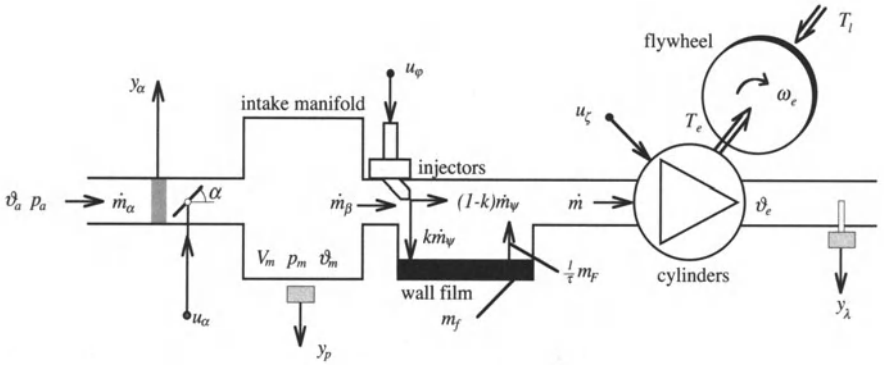


Fig. 2.4. Abstract mean-value SI engine structure.

Figure 2.5 shows the resulting simplified cause and effect diagram of an SI engine (assuming isothermal conditions in the intake manifold and modeling the exhaust manifold as a pure delay system). In the cause and effect diagram, the reservoirs mentioned appear as blocks with black shades. Between these reservoir blocks, flows are defined by static blocks (gray shades). The levels of the reservoirs define the size of these flows.

Each of these blocks is subdivided into several other parts which will be discussed in the sections indicated in the corresponding square brackets.[6] However, the most important connections are already visible in this representation. Both air and fuel paths affect the combustion through some delaying blocks while the ignition affects the combustion (almost) directly. The main output variables of the combustion process are the engine torque T_e, the exhaust gas temperature ϑ_e and the air/fuel ratio λ_e.

The following signal definitions have been used in Figs. 2.4 and 2.5:

$\dot{m}_\alpha$ air mass flow entering the intake manifold through the throttle;
$\dot{m}_\beta$ air mass flow entering the cylinder;
p_m pressure in the intake manifold;
$\dot{m}_\psi$ fuel mass flow injected by the injectors;
$\dot{m}_\varphi$ fuel mass flow entering the cylinder;
$\dot{m}$ mixture mass flow entering the cylinder with $\dot{m} = \dot{m}_\beta + \dot{m}_\varphi$;
T_e engine torque;
ω_e engine speed;
ϑ_e engine exhaust gas temperature; and
λ_e normalized air/fuel ratio.

[6] The block [x] will be discussed in Sect. 2.x.

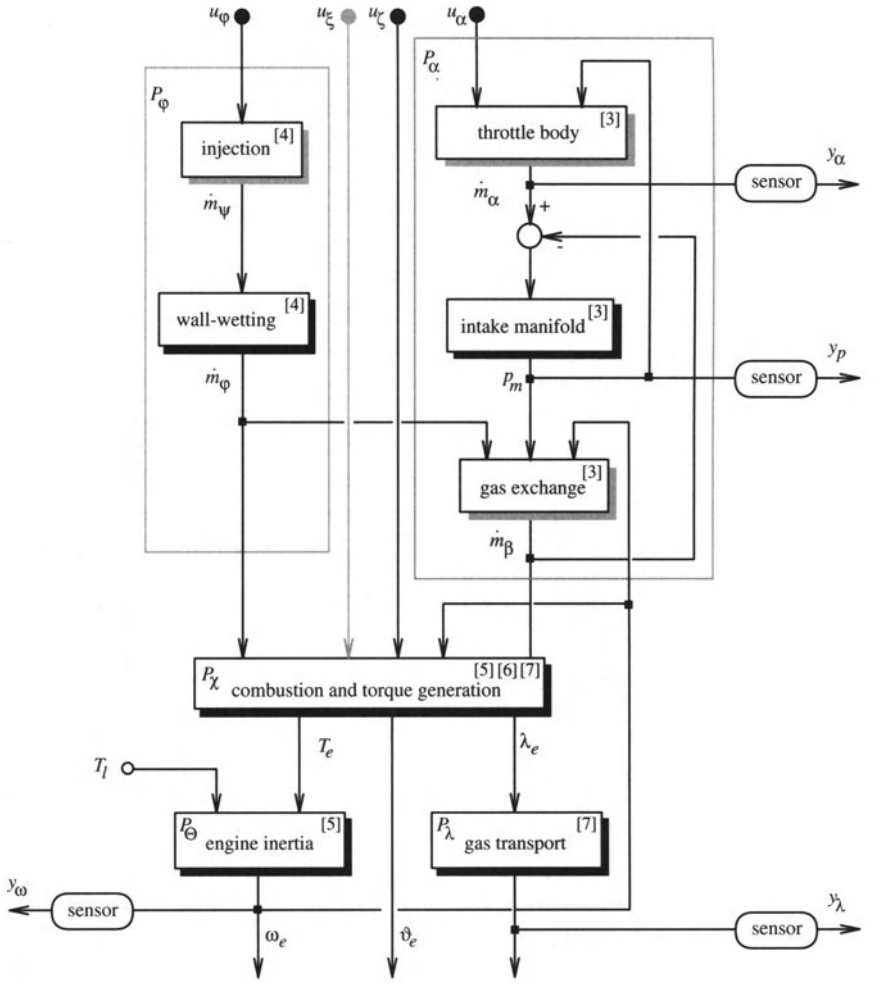


Fig. 2.5. Cause and effect diagram of an SI engine system (numbers in brackets indicate corresponding sections, gray input channel only for GDI engines, see text).

Direct-Injection SI Engines

Direct-injection SI engines (often abbreviated as GDI engines — for gasoline direct injection) are very similar to port-injected SI engines. The distinctive feature of GDI engines is their ability to operate in two different modes:

- Homogeneous charge mode (typically at higher loads or speeds), with injection starting during air intake, and burning stoichiometric air/fuel mixtures.
- Stratified charge mode (at low to medium loads and low to medium speeds), with late injection and lean air/fuel mixtures.

The *static properties* of the GDI engine (gas exchange, torque generation, pollution formation, etc.) deviate substantially from those of a port injected engine as long as the GDI engine is in *stratified charge mode*. These aspects will be discussed in the corresponding sections below.

The main differences from a controls point of view are the additional control channel (input signal u_ξ in Fig. 2.5) and the missing wall-wetting block [4] (see [166]). The signal u_ξ controls the injection process in its timing and distribution (multiple pulses are often used in GDI engines) while the signal u_φ indicates the fuel quantity to be injected.

2.2.2 Diesel Engines

As with SI engines, in a mean value approach, CI engines are assumed to work continuously. The resulting schematic engine structure has a form similar to the one shown in Fig. 2.4. The cause/effect diagram of a supercharged direct-injection Diesel engine (no EGR) is shown in Fig. 2.6. Even without considering EGR and cooling of the compressed intake air, the cause and effect diagram is considerably more complex than that of an SI engine.

The main reason for this complexity is the turbocharger, which introduces a substantial coupling between the engine exhaust and the engine intake side. Moreover, both in the compressor and in the turbine, thermal effects play an important role. However, there are also some parts that are simpler than in SI engines: fuel injection determines both the quantity of fuel injected and the ignition timing, and, since the fuel is injected directly into the cylinder, no additional dynamic effects are to be modeled in the fuel path.[7]

The following *new* signal definitions have been used in Fig. 2.6:

$\dot{m}_c$ air mass flow through the compressor;
$\dot{m}_t$ exhaust mass flow through the turbine;
p_c pressure immediately after the compressor;
p_2 pressure in the intake manifold;
p_3 pressure in the exhaust manifold;
ϑ_c air temperature after the compressor;
ϑ_2 air temperature in the intake manifold;
ϑ_3 exhaust gas temperature in front of the turbine;
ω_{tc} turbocharger rotational speed;
T_t torque produced by the turbine; and
T_c torque absorbed by the compressor.

Compared to an SI engine, there are several additional reservoirs to be modeled in a Diesel engine system. In the intake and exhaust manifolds, for instance, not only masses, but thermal (internal) energy is important. Accordingly, two level variables (pressure and temperature) form the output of these

[7] For fluid dynamic and aerodynamic simulations, usually a high-bandwidth model of the rail dynamics is necessary, see [106] or [117].

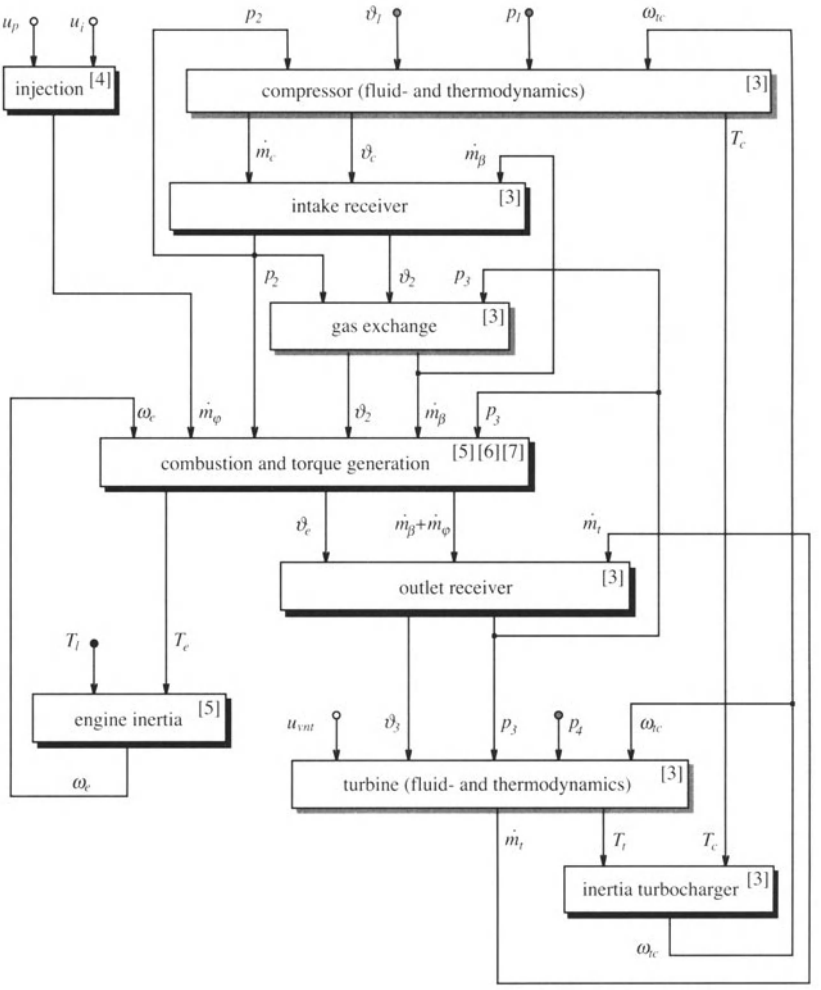


Fig. 2.6. Cause and effect diagram of a Diesel engine (EGR and intercooler not included).

blocks. The turbocharger's rotor, which stores kinetic energy, is an additional reservoir.

If EGR and intercooling is modeled as well, the cause and effect diagram has a similar, but even more complex, structure. The most important additional variable is the intake gas composition (the ratio between fresh air and burnt gases in the intake). If perfect mixing can be assumed, this leads to only one additional reservoir, see Sect. 2.3.4.

2.3 Air System

2.3.1 Receivers

The basic building block in the air intake system and also in the exhaust part is a *receiver*, i.e., a fixed volume for which the thermodynamic states (pressures, temperatures, composition, etc., as shown in Fig. 2.7) are assumed to be the same over the entire volume (lumped parameter system).

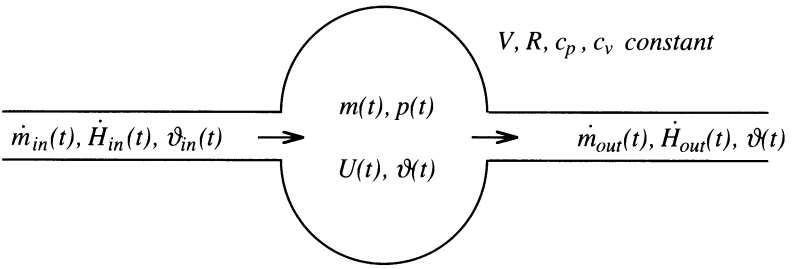


Fig. 2.7. Inputs, states and outputs of a receiver.

The inputs and outputs are the mass and energy flows,[8] the reservoirs store mass and thermal energy,[9] and the level variables are the pressure and temperature. If one assumes that no heat or mass transfer through the walls, and that no substantial changes in potential or kinetic energy in the flow occur, then the following two coupled differential equations describe such a receiver

$$\frac{d}{dt}m(t) = \dot{m}_{in}(t) - \dot{m}_{out}(t) \tag{2.3}$$

$$\frac{d}{dt}U(t) = \dot{H}_{in}(t) - \dot{H}_{out}(t) \tag{2.4}$$

Assuming that the fluids can be modeled as perfect gases, the coupling between these two equations is given by the ideal gas law

$$p(t) \cdot V = m(t) \cdot R \cdot \vartheta(t) \tag{2.5}$$

and by the caloric relations

$$\begin{aligned} U(t) &= c_v \cdot \vartheta(t) \cdot m(t) \\ \dot{H}_{in}(t) &= c_p \cdot \vartheta_{in}(t) \cdot \dot{m}_{in}(t) \\ \dot{H}_{out}(t) &= c_p \cdot \vartheta(t) \cdot \dot{m}_{out}(t) \end{aligned} \tag{2.6}$$

[8] Thermodynamically correct enthalpy flows, $\dot{H}(t)$.
[9] Thermodynamically correct internal energy, $U(t)$

Note that the temperature $\vartheta_{out}(t)$ of the out-flowing gas is assumed to be the same as the temperature $\vartheta(t)$ of the gas in the receiver (lumped parameter approach). The following well-known definitions and connections among the thermodynamic parameters will be used below:

c_p specific heat at constant pressure, units $J/(kg\,K)$;
c_v specific heat at constant volume, units $J/(kg\,K)$;
κ ratio of specific heats, i.e., $\kappa = c_p/c_v$; and
R gas constant (in the formulation of (2.5) this quantity depends on the specific gas), units $J/(kg\,K)$, that satisfies the equation $R = c_p - c_v$.

Substituting (2.5) and (2.6) into (2.3) and (2.4), the following two differential equations for the level variables pressure and temperature (which are the only measurable quantities) are obtained after some simple algebraic manipulations

$$\begin{aligned} \tfrac{d}{dt}p(t) &= \tfrac{\kappa \cdot R}{V} \cdot [\dot{m}_{in}(t) \cdot \vartheta_{in}(t) - \dot{m}_{out}(t) \cdot \vartheta(t)] \\ \tfrac{d}{dt}\vartheta &= \tfrac{\vartheta \cdot R}{p \cdot V \cdot c_v} \cdot [c_p \cdot \dot{m}_{in} \cdot \vartheta_{in} - c_p \cdot \dot{m}_{out} \cdot \vartheta - c_v \cdot (\dot{m}_{in} - \dot{m}_{out}) \cdot \vartheta] \end{aligned} \tag{2.7}$$

(for reasons of space here, the explicit time dependencies have been omitted in the second line of (2.7)).

The adiabatic formulation (2.7) is one extreme that well approximates the receiver's behavior when the dwell time of the gas in the receiver is small or when the surface to volume ratio of the receiver is small. If the inverse is true, then the isothermal assumption is a better approximation, i.e., such a large heat transfer is assumed to take place that the temperatures of the gas in the receiver ϑ and of the inlet gas ϑ_{in} are the same (e.g., equal to the engine compartment temperature). In this case, (2.7) can be simplified to

$$\begin{aligned} \tfrac{d}{dt}p(t) &= \tfrac{R \cdot \vartheta(t)}{V} \cdot [\dot{m}_{in}(t) - \dot{m}_{out}(t)] \\ \vartheta(t) &= \vartheta_{in}(t) \end{aligned} \tag{2.8}$$

In Sect. 2.6, some remarks will be made for the intermediate case where a substantial, but incomplete, heat transfer through the walls takes place (e.g., by convection or by radiation). A detailed analysis of the effects of the chosen model on the estimated air mass in the cylinder can be found in [78].

2.3.2 Valve Mass Flows

The flow of fluids between two reservoirs is determined by valves or orifices whose inputs are the pressures upstream and downstream. The difference between these two level variables drives the fluid in a nonlinear way through such restrictions. Of course, this problem is at the heart of fluid dynamics and a large amount of theoretical and practical knowledge exists on this topic. For the purposes pursued in this text, the two simplest formulations will suffice.

The following assumptions are made:

- no friction[10] in the flow;
- no inertial effects in the flow (the piping around the valves is small compared to the receivers to which they are attached);
- completely isolated conditions (no additional energy, mass ... enters the system); and
- all flow phenomenon zero dimensional, i.e., no spatial effects need be considered.

If, in addition, one can assume that the fluid is *incompressible*,[11] Bernoulli's law can be used to derive a valve equation

$$\dot{m}(t) = c_d \cdot A(t) \cdot \sqrt{2\rho} \cdot \sqrt{p_{in}(t) - p_{out}(t)} \tag{2.9}$$

where the the following definitions have been used:

$\dot{m}$	mass flow through the valve;
A	open area of the valve;
c_d	discharge coefficient;
ρ	density of the fluid, assumed to be constant;
p_{in}	pressure upstream of the valve; and
p_{out}	pressure downstream of the valve.

For *compressible* fluids, the most important and versatile flow control block is the *isothermal orifice*. When modeling this device, the key assumption is that the flow behavior may be separated as follows:

- No losses occur in the accelerating part (pressure decreases) up to the narrowest point. All the potential energy stored in the flow (with pressure as its level variable) is converted isentropically into kinetic energy.
- After the narrowest point, the flow is fully turbulent and all of the kinetic energy gained in the first part is dissipated into thermal energy. Moreover, no pressure recuperation takes place.

The consequences of this are that the pressure in the narrowest point of the valve is (approximately) equal to the downstream pressure and that the temperature of the flow before and after the orifice is (approximately) the same (hence the name, see the schematic flow model shown in Fig. 2.8).

Using the thermodynamic relationships for isentropic expansion [120] the following equation for the flow can be obtained

$$\dot{m}(t) = c_d \cdot A(t) \cdot \frac{p_{in}(t)}{\sqrt{R \cdot \vartheta_{in}(t)}} \cdot \Psi\left(\frac{p_{in}(t)}{p_{out}(t)}\right) \tag{2.10}$$

where the flow function $\Psi(.)$ is defined by

[10] Friction can be partially accounted for by discharge coefficients that must be experimentally validated.

[11] For liquid fluids, this is often a good approximation.

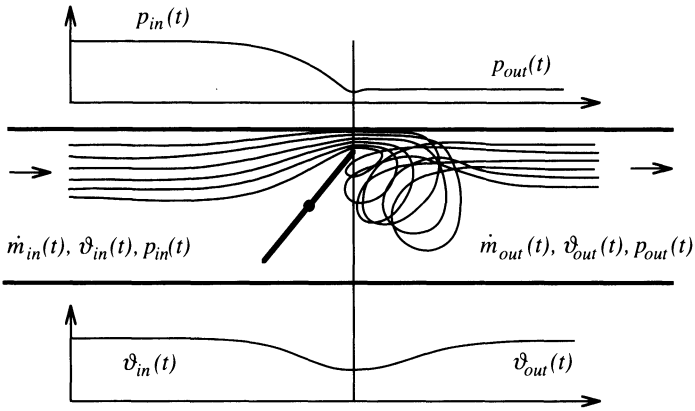


Fig. 2.8. Flow model in an isenthalpic throttle.

$$\Psi\left(\frac{p_{in}(t)}{p_{out}(t)}\right) = \begin{cases} \sqrt{\kappa\left[\frac{2}{\kappa+1}\right]^{\frac{\kappa+1}{\kappa-1}}} & \text{for} \quad p_{out} < p_{cr} \\ \left[\frac{p_{out}}{p_{in}}\right]^{1/\kappa} \cdot \sqrt{\frac{2\kappa}{\kappa-1} \cdot \left[1 - \frac{p_{out}}{p_{in}}\right]^{\frac{\kappa-1}{\kappa}}} & \text{for} \quad p_{out} \geq p_{cr} \end{cases} \tag{2.11}$$

and where

$$p_{cr} = \left[\frac{2}{\kappa+1}\right]^{\frac{\kappa}{\kappa-1}} \cdot p_{in} \tag{2.12}$$

is the critical pressure where the flow reaches sonic conditions in the narrowest part.

Equation (2.10) is sometimes rewritten in the following form

$$\frac{\sqrt{\vartheta_{in}}}{p_{in}} \cdot \dot{m} = \frac{c_d \cdot A}{\sqrt{R}} \cdot \Psi\left(\frac{p_{in}}{p_{out}}\right) = \widetilde{\Psi}\left(\frac{p_{in}}{p_{out}}\right) \tag{2.13}$$

For a fixed orifice area A, the function $\widetilde{\Psi}(p_{in}, p_{out})$ can be *measured* for some reference conditions $p_{in,0}$ and $\vartheta_{in,0}$. This approach permits the exact capture of the influence of the discharge coefficient c_d, which, in reality, is not constant. At a later stage, the actual mass flow $\dot{m}$ is computed using the corresponding conditions p_{in} and ϑ_{in} and the following relationship

$$\dot{m} = \frac{p_{in}}{\sqrt{\vartheta_{in}}} \cdot \widetilde{\Psi}\left(\frac{p_{in}}{p_{out}}\right) \tag{2.14}$$

For many working fluids (e.g., intake air, exhaust gas at lower temperatures, etc.) with $\kappa \approx 1.4$ (2.11) can be approximated quite well by

$$\Psi\left(\frac{p_{in}(t)}{p_{out}(t)}\right) \approx \begin{cases} 1/\sqrt{2} & \text{for} \quad p_{out} < \frac{1}{2}\, p_{in} \\ \sqrt{\frac{2\, p_{out}}{p_{in}}\left[1 - \frac{p_{out}}{p_{in}}\right]} & \text{for} \quad p_{out} \geq \frac{1}{2}\, p_{in} \end{cases} \tag{2.15}$$

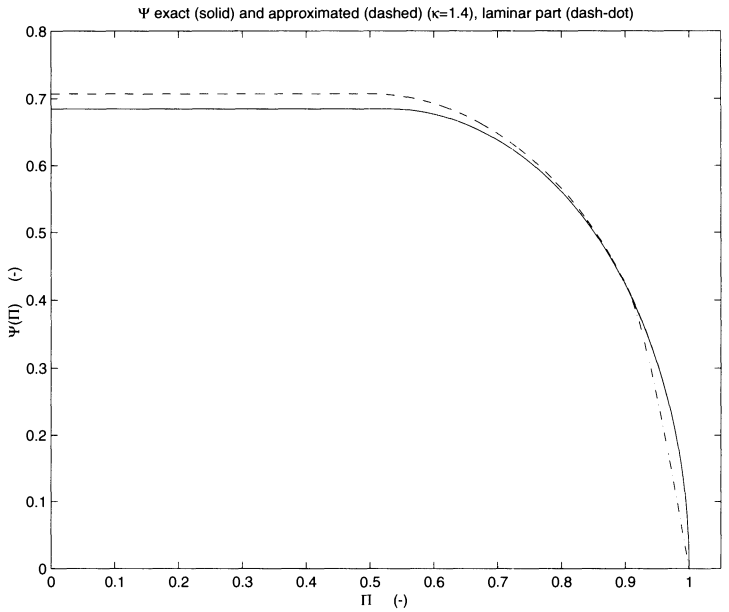


Fig. 2.9. Comparison of (2.11) (black) and (2.15) (dash) for the nonlinear function $\Psi(.)$. Also shown, approximation (2.17) (dash-dot) around $\Pi \approx 1$ for an unrealistic $\Pi_{tr} = 0.9$.

Figure 2.9 shows the exact curve (2.11) and the approximation (2.15).

As Fig. 2.9 also shows, the nonlinear function $\Psi(.)$ has an infinite gradient (in both formulations) at $p_{out} = p_{in}$. Inserting this relationship into the receiver equation, (2.7) or (2.8), yields a system that is not Lipschitz and which, therefore, will be difficult to integrate for values $p_{out} \approx p_{in}$ [178].

To overcome this problem, a laminar flow condition can be assumed for very small pressure ratios (see [47] for a formulation that is based on incompressible flows). This (physically quite reasonable) assumption yields a differentiable formulation at zero pressure difference and thus eliminates the mentioned problems.

An even more pragmatic approach to avoid that singularity is to assume that there exists a threshold

$$\Pi_{tr} = \left.\frac{p_{out}}{p_{in}}\right|_{tr} < 1 \tag{2.16}$$

under which (2.11) or (2.15) are valid. If larger pressure ratios occur, then a smooth approximation of the form

$$\Psi(\Pi) = a \cdot (\Pi - 1)^3 + b \cdot (\Pi - 1) \tag{2.17}$$

with

$$a = \frac{\Psi'_{tr} \cdot (\Pi_{tr} - 1) - \Psi_{tr}}{2\,(\Pi_{tr} - 1)^3}, \qquad b = \Psi'_{tr} - 3\,a \cdot (\Pi_{tr} - 1)^2 \tag{2.18}$$

is used (Ψ_{tr} is the value of $\Psi(.)$ and Ψ'_{tr} the value of the derivative of $\Psi(.)$ at the threshold Π_{tr}). The special form of (2.17) and (2.18) guarantees a symmetric behavior around the critical pressure ratio $\Pi = 1$ and a smooth transition at the departure point Π_{tr}. Figure 2.9 shows an example of this curve (for reasons of clarity, an unrealistically large threshold $\Pi_{tr} = 0.9$ has been chosen).

2.3.3 Engine Mass Flows

Regarding the air system, the engine itself can be approximated as a volumetric pump, i.e., a device that enforces a volume flow approximately proportional to its speed. A typical formulation for such a model is

$$\dot{m}(t) = \rho_{in}(t) \cdot \dot{V}(t) = \rho_{in}(t) \cdot \lambda_l(p_m, \omega_e) \cdot \frac{V_d}{N} \cdot \frac{\omega_e(t)}{2\pi} \tag{2.19}$$

where the new variables are defined as follows:

ρ_{in} density of the gas at the engine's intake (related to the intake pressure and temperature by the ideal gas law, (2.5));
λ_l *volumetric efficiency*, which describes how far the engine differs from a perfect volumetric device (see below);
V_d displaced volume; and
N number of revolutions per cycle ($N = 2$ for four-stroke and $N = 1$ for two-stroke engines, the formulations presented below will be for four-stroke engines).

An alternative definition of the volumetric efficiency is obtained by considering the *air* mass flow only

$$\dot{m}_\beta(t) = \rho_m(t) \cdot \dot{V}(t) = \rho_m(t) \cdot \tilde{\lambda}_l(p_m, \omega_e) \cdot \frac{V_d}{N} \cdot \frac{\omega_e(t)}{2\pi} \tag{2.20}$$

where $\dot{m}_\beta$ is the air mass flow aspirated by the engine and ρ_m the intake manifold density. The experimental identification of (2.20) is simpler than those of (2.19). The drawback of (2.20) is that it cannot account for varying fuel mass flows.

The volumetric efficiency determines the engine's ability to aspire mixture or air and is of special importance at full-load conditions. It is rather difficult to predict reliably since many effects influence it (internal exhaust gas recirculation, ram effects in the intake runner, resonance in the manifold, cross-coupling between the different cylinders, etc.). Precise models will therefore have to rely on measurements of this quantity (usually as a two-dimensional map over speed and load obtained in steady-state conditions) or on detailed simulations of the fluid dynamic processes in the intake system

(see e.g., [36]). Note also that, for gasoline engines, the evaporation of the fuel in the intake port causes a substantial reduction of the temperature of the inflowing mixture. This effect can increase volumetric efficiency [6].

As a first approximation, a multilinear formulation is sometimes useful, i.e.,

$$\lambda_l(p_m, \omega_e) = \lambda_{lp}(p_m) \cdot \lambda_{l\omega}(\omega_e) \tag{2.21}$$

Assuming perfect gases with constant κ and isentropic processes, the pressure dependent part (which essentially describes the effects caused by the trapped exhaust gas at TDC) for an ideal Otto or Diesel cycle can be approximated as

$$\lambda_{lp}(p_m) = \frac{V_c + V_d}{V_d} - \left(\frac{p_{out}}{p_m}\right)^{1/\kappa} \cdot \frac{V_c}{V_d} \tag{2.22}$$

where p_{out} is the pressure at the engine's exhaust side (assumed to be constant), and V_c is the compression volume, i.e., the volume in the cylinder at TDC.

The advantage of using (2.21) and (2.22) is that, in this case, the experimental validation has only to consider changing engine speeds, and thus a reduced number of experiments are sufficient. Other simplified formulations of the volumetric efficiency are in use (see e.g., [175] and [137] for an affine form).

The volumetric efficiency of port and direct injected SI engines is very similar (see [166]). The only obvious difference is the cooling effect of the gasoline evaporation taking place inside the cylinder in the GDI case which causes a small increase of the volumetric efficiency. The overall behavior, however, is not affected by that.

Equation (2.19) represents the total mass flow $\dot{m}$ that the engine aspirates into its cylinders. For port-injected engines (e.g., gasoline or especially natural gas fueled SI engines) a certain part of $\dot{m}$ is fuel[12] such that the air mass flow $\dot{m}_\beta$ leaving the intake manifold is smaller than $\dot{m}$. Assuming the fuel to be fully evaporated, the density in (2.19) must be changed in this case, as follows

$$\rho_{in}(t) = \frac{p_m}{R_m \cdot \vartheta_m} \tag{2.23}$$

where the temperature ϑ_m and the gas constant R_m of the mixture are found using the corresponding variables of the fuel (subscript φ) and the fresh air at the engine inlet valve (subscript β), i.e.,

$$R_m = \frac{\dot{m}_\beta \cdot R_\beta + \dot{m}_\varphi \cdot R_\varphi}{\dot{m}_\beta + \dot{m}_\varphi} \tag{2.24}$$

and

$$\vartheta_m = \frac{\dot{m}_\beta \cdot c_{p\alpha} \cdot \vartheta_\beta + \dot{m}_\varphi \cdot c_{p\varphi} \cdot \vartheta_\varphi}{\dot{m}_\beta \cdot c_{p\alpha} + \dot{m}_\varphi \cdot c_{p\varphi}} \tag{2.25}$$

[12] Additional exhaust gas recirculation is discussed in the next section.

Of course, these last two equations are based on the implicit assumption that perfect and adiabatic mixing takes place at the intake valve.

Using the mass balance

$$\dot{m} = \dot{m}_\beta + \dot{m}_\varphi \tag{2.26}$$

and assuming that all of the evaporated fuel is indeed aspirated into the cylinder, the air flowing out of the intake manifold can be computed by solving a quadratic equation which is derived combining (2.19) with (2.24) through (2.26)

$$\dot{m}_\beta = \frac{-b + \sqrt{b^2 - 4\,a \cdot c}}{2\,a} \tag{2.27}$$

where the following definitions have been used in (2.27)

$$\begin{aligned} a &= R_\beta \vartheta_\beta c_{p,\alpha} \\ b &= \left(R_\beta \vartheta_\varphi c_{p,\varphi} + R_\varphi \vartheta_\beta c_{p,\alpha}\right) \dot{m}_\varphi - p_m V_d \lambda_l \tfrac{\omega_e}{2N\pi} c_{p,\alpha} \\ c &= \left(\dot{m}_\varphi R_\varphi \vartheta_\varphi - p_m V_d \lambda_l \tfrac{\omega_e}{4\pi}\right) c_{p,\varphi} \dot{m}_\varphi \end{aligned} \tag{2.28}$$

Note that for physically meaningful parameters, the expression $b^2 - 4\,a \cdot c$ is positive and that the ambiguity in the solution of the quadratic equation is easily solved.

If the fuel's temperature, specific heat and gas constant are not known, or if the computations must be limited to a bare minimum, the following approximation of the air mass flow $\dot{m}_\beta$ can be used

$$\dot{m}_\beta = \frac{p_m}{\vartheta_\beta \cdot R_\beta} \cdot \lambda_l(p_m, \omega_e) \cdot \frac{V_d \cdot \omega_e}{N \cdot 4\pi} - \dot{m}_\varphi \tag{2.29}$$

where the properties of the mixture are approximated by the properties of the air. For gasoline, the error made with such an approximation is only ≈5%; for other fuels, however, it can be larger.

2.3.4 Exhaust Gas Recirculation

Exhaust gas recirculation (EGR) is either direct or cooled recirculation of the cylinder exhaust gases into the intake manifold. In the following, it is assumed that:

- the exhaust gas is collected immediately beyond the exhaust valve (no dynamic gas mixing or delays in the exhaust part);
- cooled EGR is realized (such that the inlet manifold can be modeled isothermically by (2.8));
- the fresh air and the exhaust gas are ideal gases with identical gas constants R;[13] and

[13] At ambient temperature, the actual values are $R_{air} = 287.1\ J/kg\,K$ and $R_{eg} = 286.6\ J/kg\,K$, and these values do not change substantially in the relevant temperature range.

- the mixture is stoichiometric or lean (but not rich).

Since EGR is usually not applied in SI engines at full load (otherwise power would be lost), the last assumption is not restricting. The first and third assumptions are usually satisfied. The second assumption is not always correct, but it is possible to extend the ideas presented below to the adiabatic case (using (2.7) instead of (2.8)).

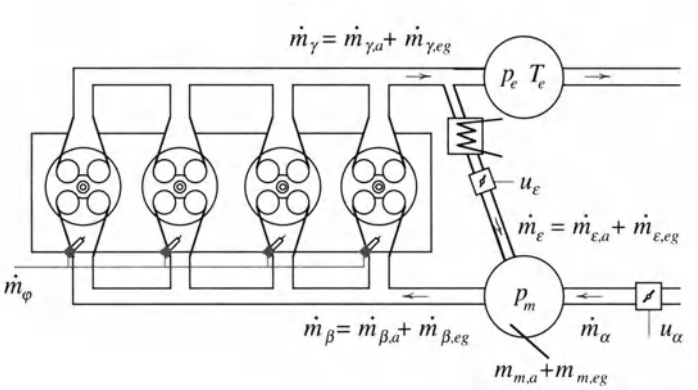


Fig. 2.10. Definitions of variables for the EGR model.

Figure 2.10 illustrates the definitions of the main variables of the EGR model. The index a represents air and the index eg exhaust gas. As usual, mass flows are labelled by $\dot{m}_x$ and masses by m_x where $x = \{a, eg\}$. The index m indicates a variable pertaining to the intake manifold, β to the engine intake, γ to the engine outlet, and ε to the exhaust gas recirculation duct. The three inputs are the fuel mass flow entering the cylinders $\dot{m}_\varphi$, the command signal to the fresh air throttle u_α, and the command signal to the exhaust gas recirculation valve u_ε.

In the following, the air/fuel ratio λ will be used which describes the ratio of air and fuel flow (or mass) into the cylinder in relation to the stoichiometric constant σ_0,[14] i.e.,

$$\lambda(t) = \frac{1}{\sigma_0} \cdot \dot{m}_{\beta,a}(t)/\dot{m}_\varphi(t) \tag{2.30}$$

According to the assumptions stated above $\lambda(t) \geq 1$ for all times t. A mass flow balance over the cylinders yields the mass flows

$$\dot{m}_{\gamma,eg}(t) = \dot{m}_{\beta,eg}(t - \tau_{IEG}) + \dot{m}_{\beta,a}(1 - \tau_{IEG}) \cdot \frac{1 + 1/\sigma_0}{\lambda(t - \tau_{IEG})} \tag{2.31}$$

[14] For typical RON 95 gasoline, σ_0 is approximately 14.67. This quantity can be calculated for any hydrocarbon fuel having b carbon and a hydrogen atoms (no oxygen) by $\sigma_0 \approx 137 \cdot (b + a/4)/(12 \cdot b + a)$. More details are given in Sect. 2.7.

and

$$\dot{m}_{\gamma,a}(t) = \dot{m}_{\beta,a}(t - \tau_{IEG}) \cdot \left[1 - \frac{1}{\lambda(t - \tau_{IEG})}\right] \tag{2.32}$$

In these equations the (relatively small) influence of incomplete combustion and pollutant formation has been neglected. Note also the influence of the time delays as discussed in the Sect. 2.1 above.

The air/fuel ratio is defined by the control variable u_φ and the fuel path dynamics (see Sect. 2.4.1 below). The total engine-in mass flow $\dot{m}$ is given by (2.19) such that $\dot{m}_{\beta,eg}$ and $\dot{m}_{\beta,a}$ are the only engine-in quantities not yet known. These variables are derived using the assumption of a perfect gas mixing in the inlet manifold, i.e., it is assumed that the EGR ratio of the out-flowing gas corresponds to the EGR ratio inside the intake manifold

$$x_{egr}(t) = \frac{m_{m,eg}(t)}{m_m(t)} = \frac{m_{m,eg}(t)}{m_{m,a}(t) + m_{m,eg}(t)} \tag{2.33}$$

Therefore,

$$\dot{m}_{\beta,eg}(t) = \dot{m}_\beta(t) \cdot x_{egr}(t), \qquad \dot{m}_{\beta,a}(t) = \dot{m}_\beta(t) \cdot (1 - x_{egr}(t)) \tag{2.34}$$

The next step is to formulate a first-order mass balance for the intake manifold which will comprise two species

$$\begin{aligned} \tfrac{d}{dt} m_{m,a}(t) &= \dot{m}_\alpha(t) - \dot{m}_{\beta,a}(t) + \dot{m}_{\varepsilon,a}(t) \\ \tfrac{d}{dt} m_{m,eg}(t) &= \dot{m}_{\varepsilon,eg}(t) - \dot{m}_{\beta,eg}(t) \end{aligned} \tag{2.35}$$

The manifold pressure is found using the ideal gas law (recall that both gas constants are assumed to be equal)

$$p_m(t) \cdot V_m = [m_{m,a}(t) + m_{m,eg}(t)] \cdot R \cdot \vartheta_m(t) \tag{2.36}$$

The total recirculated gas mass flow $\dot{m}_\varepsilon(t) = \dot{m}_{\varepsilon,a} + \dot{m}_{\varepsilon,eg}$ can be found using the exhaust and intake manifold pressures and temperatures (assumed to be known) and the orifice equation (2.10). Assuming again a perfect mixing behavior at the engine outlet, the exhaust gas and air fraction can be found by

$$\dot{m}_{\varepsilon,eg}(t) = \dot{m}_{\gamma,eg}(t) \cdot y_{egr}(t), \qquad \dot{m}_{\varepsilon,a}(t) = \dot{m}_{\gamma,a}(t) \cdot y_{egr}(t) \tag{2.37}$$

where y_{egr} is the ratio of the recirculated mass flow to the total exhaust gas mass flow, i.e.,

$$y_{egr}(t) = \frac{\dot{m}_\varepsilon(t)}{\dot{m}_\gamma(t)} \tag{2.38}$$

2.3.5 Superchargers

So far, the amount of air in the cylinder has been determined by the properties of the engine itself. These *naturally aspirated* engines (both SI and CI) are limited in their power by their volumetric efficiency. Consequently, in most modern engines the maximum torque is closely proportional to the displaced volume.

Supercharged[15] engines can produce much more power for a given displaced volume by forced air transport into the intake part. Three basically different devices are in use:

- turbochargers, i.e., *fluid dynamic* devices consisting of a compressor and a turbine on the same shaft (with rotation speeds of more than 200'000 rpm), utilizing some of the exhaust gas exergy to compress the intake air [181];
- pressure wave superchargers, i.e., *gas dynamic devices* consisting of a barrel whose narrow channels are in direct contact with both engine in and exhaust sides, also utilizing the exhaust gas exergy for the boosting process [68]; and
- mechanical superchargers (preferably with internal compression), which receive the power needed for boosting directly from the engine's crankshaft.

From a control engineering point of view, these devices substantially differ only in their static functions (blocks "compressor" and/or "turbine" in Fig. 2.6). The interconnections with the receivers and the other reservoir blocks remain the same. The key assumption is that the dynamic phenomena in the fluid or gas dynamic processes are much faster than the rate of change in the thermodynamic boundary conditions (taking place in the receivers to which the superchargers are connected). In this case, the fluid or gas dynamic processes can be modeled as quasi stationary phenomena (type b) events in Fig. 2.3) and static maps can be used to describe the behavior of these devices, which is usually highly nonlinear.

Fluid Dynamic Compressors

General Remarks

Typically, a map similar to the one shown in Fig. 2.11 is used to describe the main variables of a fluid dynamic compressor. The inputs (causes) being:

- the ratio of the pressure at the compressor inlet and outlet ports (starting point 1 in Fig. 2.11);
- the compressor rotational speed (curve 2 in Fig. 2.11); and

[15] The term supercharging will be used in this text to describe the forced induction of air or mixture into an engine by any device.

- the temperature at the compressor inlet port (needed to compute the corrected mass flow and speed, see below for details).

Note that all these variables are level variables and are associated with some reservoir blocks (two are mass, one is kinetic energy and one is internal energy). With this information, three outputs (effects) are calculated: the mass flow $\dot{m}_c$ through the compressor (point 3 in Fig. 2.11), the torque T_c needed to drive the compressor at the actual state and the temperature at the outlet of the compressor ϑ_c.

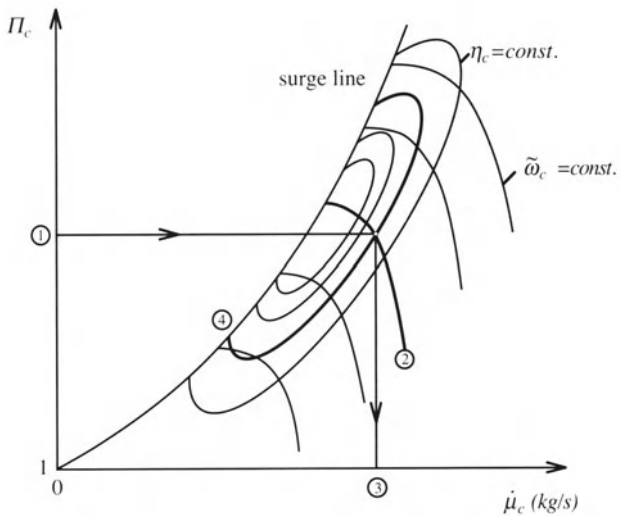


Fig. 2.11. Fluid dynamic (radial) compressor map (normalized mass flows and efficiencies, schematic representation).

The data used to draw the compressor maps are often measured using dedicated turbocharger test rigs. Similar to the approach described in Sect. 2.3.2 (see (2.13)), the maps obtained in this way are correct for a specified reference inlet pressure $p_{1,0}$ and reference inlet temperature $\vartheta_{1,0}$. Since the actual inlet conditions will often not be the same as the reference values, a correction must be made.[16] For this correction the following relation is used:

$$\dot{\mu}_c \cdot \frac{\sqrt{\vartheta_{1,0}}}{p_{1,0}} = \dot{m}_c \cdot \frac{\sqrt{\vartheta_1}}{p_1} \tag{2.39}$$

Accordingly, the following steps must be made to obtain the compressor massflow (see Fig. 2.11):

1. compute the pressure ratio over the compressor $\Pi_c = p_2/p_1$;

[16] The same arguments hold for the turbine.

2. compute the corrected compressor speed

$$\widetilde{\omega}_{tc} = \sqrt{\vartheta_{1,0}/\vartheta_1} \cdot \omega_{tc} \tag{2.40}$$

with $\vartheta_{1,0}$ being the inlet temperature which was used when deriving the map;

3. find the reference mass flow $\dot{\mu}_c$ either graphically or using any of the map fitting techniques described below; and
4. compute the actual compressor mass flow

$$\dot{m}_c = (p_1/p_{1,0}) \cdot \sqrt{\vartheta_{1,0}/\vartheta_1} \cdot \dot{\mu}_c \tag{2.41}$$

with $p_{1,0}$ being the inlet pressure which was used when deriving the map.

To compute the power (or torque) absorbed by the compressor for a given operating point, its isentropic efficiency must be known. Typically, this quantity is also plotted in the compressor maps as iso-level contours (see point 4 in Fig. 2.11).

The power absorbed by the compressor is the *isentropic power*[17] $P_{c,s}$ divided by the isentropic efficiency η_c, which takes into account all losses that the real compressor induces. To be more precise

$$P_c = P_{c,s} \cdot \frac{1}{\eta_c} = \dot{m}_c \cdot c_p \cdot \vartheta_1 \cdot \left[\Pi_c^{(\kappa-1)/\kappa} - 1 \right] \cdot \frac{1}{\eta_c} \tag{2.42}$$

Since the compressor speed is a level variable (i.e., an available output of a reservoir block) the torque absorbed by the compressor can be computed as follows

$$T_c = \frac{P_c}{\omega_{tc}} \tag{2.43}$$

Similarly, the temperature increase due to the compression is formulated as an isentropic (unavoidable) part and an additional increase due to technical imperfections in the compressor realization

$$\vartheta_c = \vartheta_1 + \left[\Pi_c^{(\kappa-1)/\kappa} - 1 \right] \cdot \frac{\vartheta_1}{\eta_c} \tag{2.44}$$

Note that adiabatic conditions have been assumed in (2.44). Additional heat losses through the compressor walls and in the compressor outlet will reduce the gas temperature before entering the intercooler or the intake manifold.

[17] The isentropic power $P_{c,s}$ is that mechanical power that the best possible compressor would need to achieve the actual compression ratio [120].

Limits of Operation

Every compressor is, of course, limited in its operating range. As Fig. 2.12 shows, there are at least four limiting conditions:

- the mechanical limit, i.e., the maximum rotational speed that is allowed before mechanical damage can occur due to the large centrifugal forces (this is especially critical when vehicles are operated at high altitude and therefore low air density);
- the choke limit, where the flow in the compressor reaches sonic conditions in its most narrow part;
- the blocking limit, i.e., the behavior of the compressor at zero (or very low) speed where it represents a blocking orifice; and
- the surge limit, where fluid dynamic instabilities destroy the regular flow patterns inside the compressor (see below for more details).

These limitations are very complex to model accurately, and are the topic of ongoing research efforts. For control purposes, however, simple models must be found that describe the system's behavior even outside the regular operating range.[18]

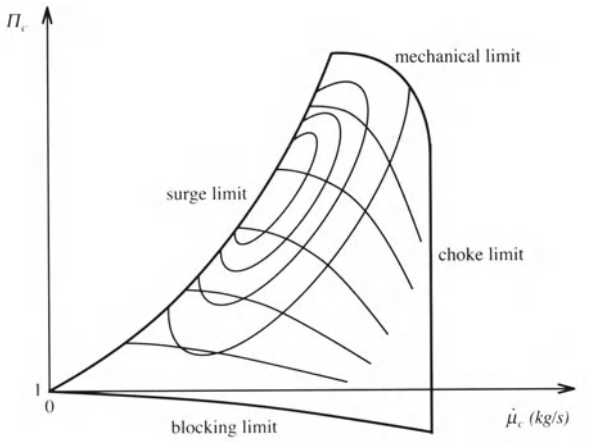


Fig. 2.12. Boundaries of operation of a compressor.

Maximum speed and choked flow usually do not pose any special problems. A simple approximation valid for the low-speed regime $\omega_{tc} < \omega_{tc,min}$ is to assume the compressor will behave like a fixed orifice as described by (2.10). In this regime there is no mass flow unless the pressure at the compressor's

[18] These conditions occur, for instance, when simulating driving cycles, e.g., engine start-up implies zero compressor speed, and quick throttle closing often induces compressor surge.

outlet is smaller than at its inlet port. No changes in the cause/effect structure are necessary, however. Note also that the minimum speed $\omega_{tc,min}$, at which the compressor changes its behavior, is, in general, larger than zero. A smooth transition between the two regimes (with a hysteresis) is recommended.

The behavior of the compressor in the surge (or pumping) mode[19] is much more complex. A simple model that permits qualitatively realistic simulations with few excursions into the surge regime will be presented below.

Fitting of Compressor Maps

Measured maps are the corner stones of the compressor models presented so far. Usually, the data obtained from turbocharger manufacturers or the data measured on dedicated test benches are not in a form that is directly useful in engine system simulations.

Figure 2.13 shows a typical example of a measured data set. The dashed lines in Fig. 2.13 are the actual measurements. The surge line is not measured separately, but rather it is constructed from the most left point of every speed line.

Often data points with rotational speeds below 100000 rpm are not measured because they represent only a small region in the compressor map, and measurements in this area are difficult. The mass flows and the pressure ratios are small, which leads to large relative errors. Moreover, heat flow from the turbine to the compressor becomes substantial in comparison to the enthalpy flow of the fluids. This results in large deviations of the efficiency calculated.

However, what looks like a small region in the compressor map is very relevant for SI engines systems. Especially while idling or under low engine load conditions, the turbocharger speed can fall even below 10000 rpm. Consequently, an extrapolation of the compressor map to lower speed regions is necessary. This extrapolation can be done either with a good physical model or with map fitting. Unfortunately, quite often the former approach does not result in sufficiently accurate models.

Several methods for map fitting are presented in [151] and [119]. Good results can be achieved by the inverted approach proposed in [87]. In this approach, the pressure ratio Π_c is expressed as a function of the mass flow $\dot{\mu}_c$ and the rotational speed $\widetilde{\omega}_{tc}$

$$\Pi_c = f(\dot{\mu}_c, \widetilde{\omega}_{tc}) \tag{2.45}$$

For the fitting, two dimensionless variables must be defined, namely the head parameter

$$\Psi_c = \frac{c_p \cdot \vartheta_a \cdot \left(\left(\frac{p_2}{p_1} \right)^{\frac{\kappa-1}{\kappa}} - 1 \right)}{\frac{1}{2} u_c^2} \qquad \text{with} \qquad u_c = r_c \cdot \omega_{tc} \tag{2.46}$$

[19] Compressor pumping occurs when the flow in the compressor is stalled, i.e., when the pressure gradient becomes too large to be sustained. Several forms of stalling (starting from local rotating stall to full rotor stall) are known [143].

and the normalized compressor flow rate

$$\Phi_c = \frac{\dot{\mu}_c}{\rho_a \cdot \pi \cdot r_c^2 \cdot u_c} \tag{2.47}$$

where r_c is the compressor wheel radius.

The head parameter is then approximated as

$$\Psi_c = \frac{k_1 + k_2 \cdot \Phi_c}{k_3 - \Phi_c} \qquad k_i = k_{i1} + k_{i2} \cdot Ma \tag{2.48}$$

where Ma is the Mach number at the ring orifice of the compressor

$$Ma = \frac{u_c}{\sqrt{\kappa \cdot R \cdot \vartheta_a}} \tag{2.49}$$

With a least-square fitting algorithm, the parameters k_{ij} can be identified on measurement data.

Since (2.48) is invertible, this compressor model can be used in mean value models to describe the compressor

$$\dot{\mu}_c = g(\phi_c, \widetilde{\omega}_{tc}) \quad \text{with} \quad \phi_c = \frac{k_3 \cdot \Psi_c - k_1}{k_2 + \Psi_c} \tag{2.50}$$

and Ψ_c according to (2.46).

A typical example of such a data fitting procedure can be seen in Fig. 2.13. As visible in that figure, the approach discussed above is able to capture the very flat region at low mass flows, as well as the very steep characteristic close to the choke line.

For very high speeds, the measured speed lines no longer have their maximum value on the surge line, but rather at higher mass flows. In [151], this phenomenon is explained with gas dynamic effects, which falsify the measurement. Note that this behavior cannot be included in the class of system models discussed here. In fact, such a speed-to-mass-flow function is not invertible and, therefore, the approach proposed above fails to produce a result.

Sometimes it is useful to use "analytic" (closed form) descriptions of the compressor characteristics (e.g., for numerical optimization where the gradients must be supplied to the optimization routines). The mass-flow-to-pressure ratio curves depend on the compressor speed. An implicit formulation (which has the advantage of being easily identifiable with the measured curves in the compressor map) and which is valid for $\dot{\mu}_c > \dot{\mu}_{c,0}$ (see Fig. 2.16) is

$$\Pi_c(\dot{\mu}_c, \widetilde{\omega}_{tc}) = \Pi_{c,max}(\widetilde{\omega}_{tc}) - c_{c,c}(\widetilde{\omega}_{tc}) \cdot [\dot{\mu}_c - \dot{\mu}_{c,0}(\widetilde{\omega}_{tc})]^2 \tag{2.51}$$

The parameters $\Pi_{c,max}(\omega_{tc})$, $c_{c,c}(\omega_{ct})$ and $\dot{\mu}_{c,0}(\omega_{tc})$ depend only on the compressor speed. Typically, they can be parameterized as low order polynomial in ω_{tc}. Note that (2.51) must be resolved for the mass flow $\dot{\mu}_c$ to fit into the cause and effect structure of the compressor model. This is why only second-order polynomials are used in (2.51).

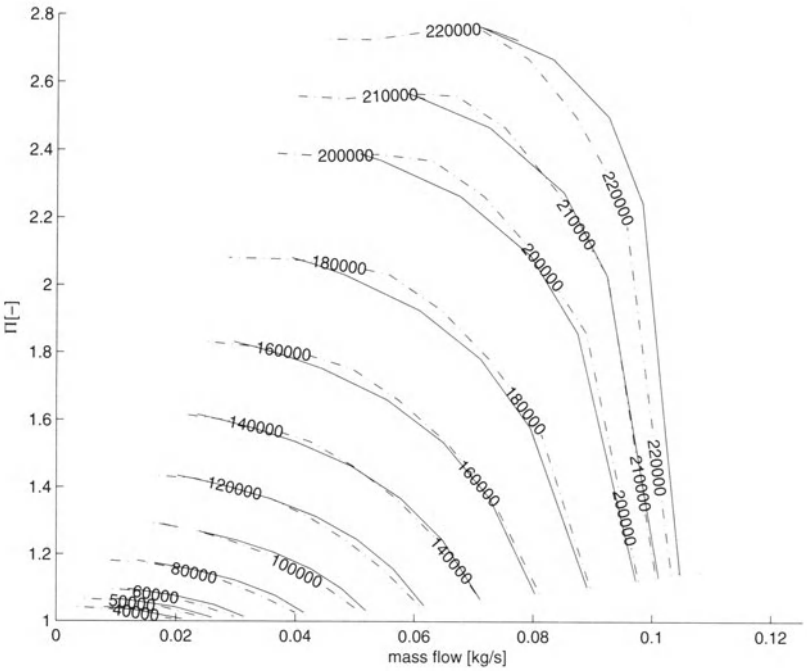


Fig. 2.13. Comparison of the measured and fitted compressor maps; numbers in plot indicate compressor speed in rpm.

Fortunately, fitting the mass flow/speed lines is the most difficult task in modeling a compressor. The following relatively simple parametrization is a possible choice for compressor efficiency

$$\eta_c(\Pi_c, \dot{\mu}_c) = \eta_{c,opt} - \chi^T(\Pi_c, \dot{\mu}_c) \cdot Q \cdot \chi(\Pi_c, \dot{\mu}_c) \tag{2.52}$$

where

$$\chi^T(\Pi_c, \dot{\mu}_c) = [\Pi_c - \Pi_{c,opt},\ \dot{\mu}_c - \dot{\mu}_{c,opt}] \tag{2.53}$$

These formulations require six parameters to be fitted to the compressor map at hand. The result of the fitting process can sometimes be improved by rescaling the ordinate

$$1 + \sqrt{\Pi_c - 1} \Rightarrow \Pi_c \tag{2.54}$$

Turbines

General Remarks

For control purposes, the mass flow behavior of turbines can be modeled well as orifices with variable area in the case of variable nozzle turbines and a blade-speed-dependent discharge coefficient c_d, as shown in Fig. 2.15. Analytic

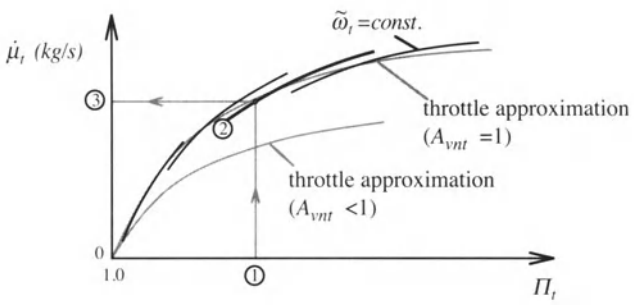


Fig. 2.14. Simplified turbine map with normalized mass flows.

descriptions similar to (2.10) or "maps" (as shown in Fig. 2.14) are commonly used.

In both cases, similar considerations hold as in the compressor case, i.e., turbine speed and pressure ratio are the input variables (points 1 and 2 in Figs. 2.14 and 2.15) and the corrected mass flow and the turbine efficiency are the system's output. The following definitions are used below:

- $\Pi_t = p_3/p_4$, the pressure ratio over the turbine ($p_4 \approx p_1$).
- $\tilde{\omega}_{tc} = \sqrt{\vartheta_{3,0}/\vartheta_3} \cdot \omega_{tc}$, the corrected turbine speed, with $\vartheta_{3,0}$ being the inlet temperature which was used deriving the map.
- $\dot{m}_t = \sqrt{\vartheta_{3,0}/\vartheta_3} \cdot p_3/p_{3,0} \cdot \dot{\mu}_t$, the corrected turbine mass flow, with $p_{3,0}$ being the turbine inlet pressure which was used deriving the map.

Note that, for turbines, the correction terms play a more important role since both inlet temperature and pressure do vary substantially. The temperature of the exhaust gas entering the turbine may be found using an approach proposed in [51] or the more detailed approaches shown in Sect. 2.6.

Again, the torque and the turbine-out temperature can be determined only when the turbine's isentropic efficiency is known. Similar to the compressor case, representations such as those shown in Fig. 2.14 are in use. Since turbine efficiency depends primarily on the angle of incidence of the in-flowing gas, the turbine-blade-speed ratio $\tilde{c}_{us}$ is used as the main independent variable

$$\tilde{c}_{us} = \frac{r_t \cdot \omega_{tc}}{c_{us}} \tag{2.55}$$

where r_t is the mean turbine radius and

$$c_{us} = \sqrt{2\, c_p \cdot \vartheta_3 \cdot \left[1 - \Pi_t^{(1-\kappa)/\kappa}\right]} \tag{2.56}$$

is the speed resulting from an isentropic expansion.

Once the turbine efficiency is known, the power produced by the turbine can be found by

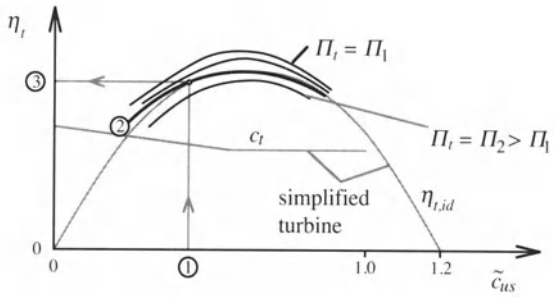


Fig. 2.15. Simplified turbine map with efficiencies and discharge coefficient.

$$P_t = P_{t,s} \cdot \eta_t = \dot{m}_t \cdot c_p \cdot \vartheta_3 \cdot \left[1 - \Pi_t^{(1-\kappa)/\kappa}\right] \cdot \eta_t \tag{2.57}$$

where $P_{t,s}$ is the (maximum) power which would be produced by a perfect (isentropic) turbine. Again, since the turbine speed is a level variable, the torque produced by the turbine can be directly derived from (2.57) to be

$$T_t = P_t / \omega_{tc} \tag{2.58}$$

Fitting of Turbine Maps

Usually, a closed-form description of the turbine's efficiency is quite accurate, i.e., η_t, as shown in Fig. 2.15, can be approximated well by

$$\eta_t(\tilde{c}_u s) = \eta_{t,max} \cdot \left[\frac{2\,\tilde{c}_u s}{\tilde{c}_{us,opt}} - \left(\frac{\tilde{c}_u s}{\tilde{c}_{us,opt}}\right)^2\right] \tag{2.59}$$

The only two parameters to be fitted are $\eta_{t,max}$ and $\tilde{c}_{us,opt}$. In automotive applications, typical values are $\eta_{t,max} \approx 0.65 \ldots 0.75$ and $c_{us,opt} \approx 0.55 \ldots 0.65$.

For turbines with a variable nozzle area, the value of $c_{us,opt}$ depends on the nozzle area, with larger areas yielding larger values.

If little measurements are available, the mass flow behavior of a turbine may be approximated using the approach described in Sect. (2.3.2) for a normalized orifice (2.13)

$$\dot{m}_t = \frac{p_3}{\sqrt{\vartheta_3}} \cdot c_t \cdot \sqrt{1 - \Pi_t^{k_t}} \tag{2.60}$$

In this case, only two parameters c_t and k_t must be fitted to the measured data using, for instance, a least-square algorithm. A typical value for the latter is $k_t = 2.2$.

Simplified Moore-Greitzer Surge Model

As mentioned above, surge imposes serious limitations on the operating range of automotive compressors.[20] In this subsection, a model is presented that

[20] In practice, ancillary devices, such as bleed valves, are used to reduce these restrictions.

can be used to approximately describe surge effects. This permits simulation of the dynamic behavior of a supercharged engine system, even for cases in which the compressor crosses the surge boundaries during heavy transients.

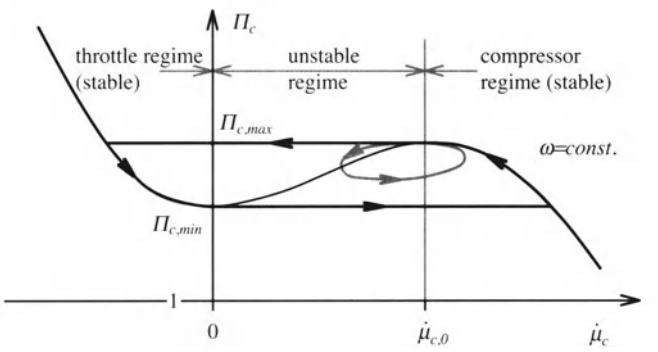


Fig. 2.16. Compressor behavior in "deep surge" (black) and "mild surge" (gray).

As Fig. 2.16 shows for one fixed rotor speed, the key idea is to extend the describing function between pressure ratio and mass flow from the stable compressor regime through the unstable compressor zone into the stable throttle regime. The compressor cannot operate in the inner unstable zone for longer time intervals.[21] A full deep-surge cycle can be visualized, as shown in Fig. 2.16, as a limit cycle whose frequency is defined by the receiver's volume after the compressor and the mass flow characteristics of the compressor in its two stable modes.

A simplified model of a compressor/manifold/throttle system that is able to reproduce (deep) surge phenomena has been proposed in [118] and extended in several subsequent publications (see e.g., [62]). The system structure is shown in Fig. 2.17. In the simplest case, the model neglects all thermal phenomena and assumes constant compressor speed and incompressible flows. As mentioned above, the starting point is to formulate an approximation of the pressure-ratio-to-mass-flow behavior that is valid beyond the surge limit. A third-order polynomial is the simplest possible approach

$$\widetilde{\Pi}_c(\dot{\mu}_c) = \Pi_{c,min} + \Pi_{c,1} \cdot \left[\frac{1}{2} + \frac{3}{2} \left(\frac{\dot{\mu}_c}{\dot{\mu}_{c,0}} - \frac{1}{2} \right) - 2 \left(\frac{\dot{\mu}_c}{\dot{\mu}_{c,0}} - \frac{1}{2} \right)^3 \right] \tag{2.61}$$

(see Fig. 2.18 for the definition of the parameters $\Pi_{c,min}$, $\Pi_{c,1}$ and $\dot{\mu}_{c,0}$).

The Moore-Greitzer model assumes that both the pressure in the manifold and the mass flow through the compressor are dynamic variables, i.e., that

[21] In this regime, a small negative pressure disturbance leads to smaller mass flow and, hence, to even smaller pressure after the compressor. This represents a dynamic system with positive feedback, hence this regime is rapidly crossed once the compressor reaches its surge limit.

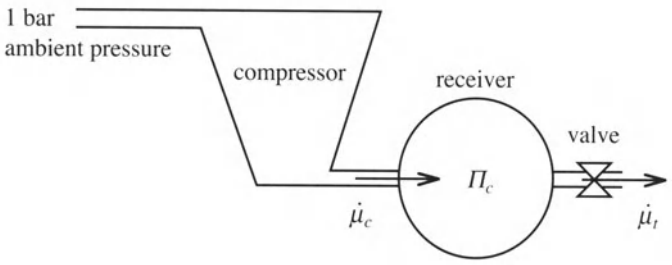


Fig. 2.17. Simplified Moore-Greitzer compressor/receiver/throttle system.

the system shown in Fig. 2.17 is described by a coupled system of differential equations

$$\begin{aligned} \frac{d}{dt}\dot{\mu}_c(t) &= \frac{1}{\tau_c} \cdot \left[\widetilde{\Pi}_c(\dot{\mu}_c(t)) - \Pi_c(t)\right] \\ \frac{d}{dt}\Pi_c(t) &= \frac{1}{\tau_m} \cdot [\dot{\mu}_c(t) - \dot{\mu}_t(t)] \end{aligned} \tag{2.62}$$

The mass flow $\dot{\mu}_t$, which represents the turbine mass flow, is approximated using (2.9).

The physical interpretation of these equations is that the compressor has some delays and time lags (inherent dynamics) which limit the rate at which the mass flow $\dot{\mu}_c(t)$ can be changed. These delays are modeled using a first-order lag with a time constant τ_c. A second time constant, τ_m, characterizes the usual dynamics of the manifold.

Note that if the system of (2.62) reaches a stable equilibrium, the pressure ratio satisfies the condition $\Pi_c = \widetilde{\Pi}_c(\dot{\mu}_c)$, which corresponds to the behavior of the quasi-static compressor model introduced above.

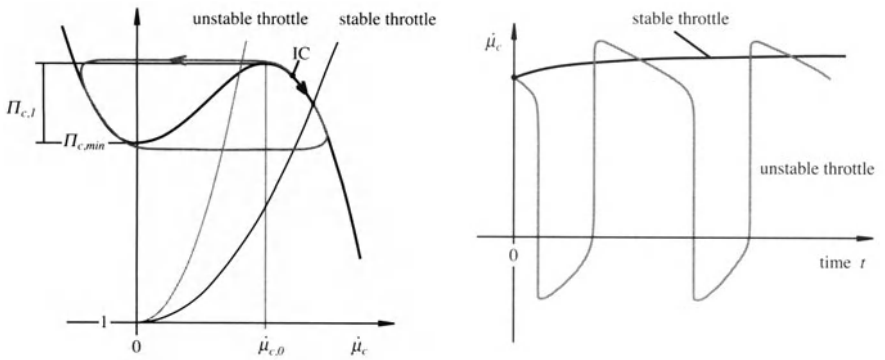


Fig. 2.18. Limit cycle behavior of the Moore-Greitzer system. Left as a trajectory in the compressor map, right as a function of time. IC is the initial condition in both cases.

In Fig. 2.18, two simulations of the Moore-Greitzer system are shown. In one case (black curves), the throttle has been chosen sufficiently large such that the system is stable. In the other case (gray curves), a smaller throttle has been chosen and the system never reaches a stable equilibrium, but rapidly reaches a limit-cycle orbit.

The dynamic properties of this orbit are defined by the two time constants τ_m and τ_c, and the function (2.61). The boundary between stable equilibrium operation and limit cycle behavior is *not* found simply by requiring the throttle to satisfy the condition

$$\dot{\mu}_t(\Pi_{c,min} + \Pi_{c,1}) = \dot{\mu}_{c,0} \tag{2.63}$$

Mild surge and even more complex behavior can be observed for turbine throttle characteristics close to that critical value.

Mechanical Behavior of Turbochargers

This part of a turbocharger system is the most simple to model. The mechanical behavior of a combined turbocharger group (compressor and turbine on the same shaft) is described by

$$\frac{d}{dt}\omega_{tc}(t) = \frac{1}{\Theta_{tc}} \cdot [T_t(t) - T_c(t) - T_f(t) + T_a(t)] \tag{2.64}$$

where Θ_{tc} is the turbocharger's rotational inertia and T_i represents the torques acting on the shaft.[22] The first two torques have been introduced above (see (2.43) and (2.58)). The remaining two torques model additional friction losses and possible external auxiliary torques, e.g., as produced by an electric motor to improve the acceleration behavior [99].

Other Supercharger Systems

Pressure wave or mechanical superchargers have maps similar to turbochargers. In particular the input/output relations are identical, such that the formulations presented above can be applied in that case, too.

For example, a mechanical compressor (connected to the crankshaft with a clutch and a 1:2 gear box) will have a map similar to the one shown in Fig. 2.19. Again, knowing the pressure ratio (point 1) and the compressor speed (point 2), the volume flow can be found in this map. The air mass flow at input conditions can be found easily once the ambient air density is known. The efficiency (in the case of Fig. 2.19 the total efficiency, i.e., including the mechanical losses) can then be found in the compressor map as well.

Mechanical superchargers do not have the surge problems present in fluid dynamic compressors and they can be designed to provide excellent dynamic

[22] Of course, it is also possible to model the rotor speed dynamics utilizing a *power* balance instead of the torque balance used in (2.64).

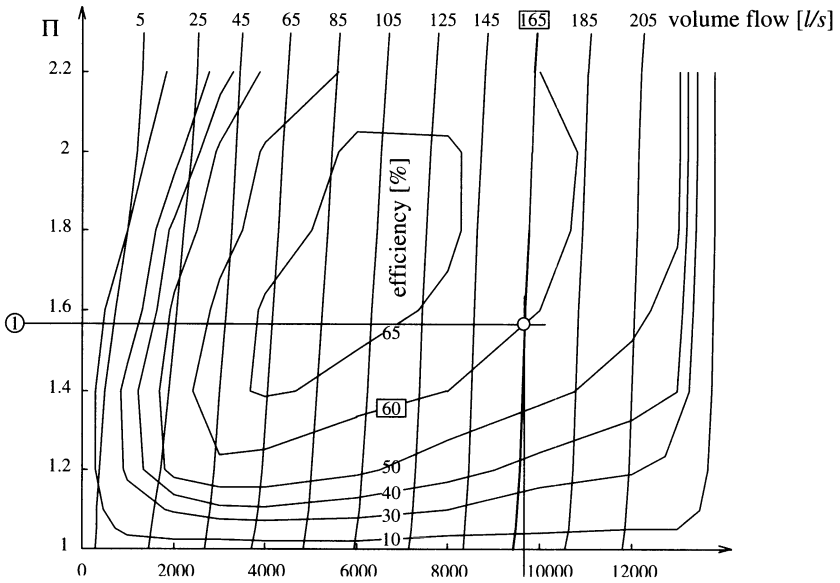


Fig. 2.19. Mechanical supercharger "map" (input volume flow and mechanical efficiencies).

response times. Problems arise if the compressor is used as a fuel economy device in a downsizing supercharging approach. In this case, the compressor is linked to the engine crankshaft only in high-power phases, and the control of the compressor clutch requires a careful design in order to avoid large torque disturbances.

Pressure wave superchargers (PWS) represent another class of devices that can be used to boost the in-cylinder pressure. Similar to turbochargers, they utilize the enthalpy of the exhaust gas to drive that process. The main advantage of PWS is that they transform the exhaust gas enthalpy directly to boost pressure by a tuned pressure-wave mechanism. This approach avoids additional time delays and yields high boost pressures even at low engine speeds. For more information on the basic principles of PWS, see [68]; for a modern application of PWS to realize fuel efficient engines, see [67]; and for the modeling and control of PWS, see [182] and [160].

2.4 Fuel System

2.4.1 Introduction

The fuel path, i.e., the subsystem that provides the necessary fuel for the combustion process to the cylinder, has its own dynamics. Both indirect and direct

fuel injection systems are in use.[23] Direct injection is the direct transport of fuel into the (main) combustion chamber (both in Diesel and SI engines), while indirect injection can be realized as port injection (i.e., upstream of the intake valve) in SI engines or as pre-chamber injection in Diesel engines. Modern Diesel engines almost exclusively use direct-injection fuel systems while most SI engines use a port-injection configuration.[24]

For direct injection systems, the dynamic effects described below are not present, i.e., the command inputs to the fuel system produce "immediate" corresponding fuel quantity changes in the next combustion process (delays due to the engine's reciprocating behavior will be discussed in the next chapter). The higher injection pressures induce complex dynamics in the fueling system which, in some situations, can have an influence on the thermodynamic processes inside the cylinder (see [117] or [106]).

In this section, only port-injected SI gasoline engine systems will be discussed. In such systems, the liquid fuel is injected through a solenoid on-off valve in the intake port, usually located directly in front of the intake valve. Since the injection pressure difference to the intake manifold is, in general, kept constant by a dedicated control loop, the injected fuel mass is (approximately) proportional to injection times (typically 5-20 ms). In addition, synchronization with the crank angle domain and other discrete event constraints (say, in communication with the electronic engine controller) cause delays in the fuel injection transmission path (for more details on this topic see Chapter 3).

2.4.2 Wall-Wetting Dynamics

Phenomenological Models

One of the most important dynamic effects in the fuel path is caused by the *wall-wetting* phenomena. The liquid fuel injected into the intake port only partially enters the cylinder in the next intake stroke. Some of it is stored in fuel puddles of the mass $m_f(t)$ at the intake port walls and at the back face of the intake valve.[25] Of course, fuel also evaporates from these puddles such that a mass balance can be expressed as follows (see [9])

[23] Carburetors are only used in special applications and will therefore not be treated in this text.

[24] Direct-injection SI engines can only reach substantial improvements in fuel economy by running in lean conditions ($\lambda >> 1$) at low loads. However, in this case the proven and cost-effective TWC technology cannot be used to achieve the required very low tail-pipe emission levels (see Sect. 1.2.2). More complex aftertreatment systems that are able to deal with a lean and non-homogeneous combustion are required in this case.

[25] Gaseous fuels, especially compressed natural gas, have, of course, no wall-wetting problems. Nevertheless, dynamic phenomena occur due to substantial back flow of aspirated mixture from the cylinder to the intake manifold. Since these effects are best described in a discrete-event setting, the corresponding models will be introduced in Sect. 3.2.4.

$$\dot{m}_\varphi(t) = (1 - \kappa(\omega_e, p_m, \vartheta_f, \ldots)) \cdot \dot{m}_\psi(t) + \frac{m_f(t)}{\tau(\omega_e, p_m, \vartheta_f, \ldots)} \tag{2.65}$$

and

$$\frac{d}{dt} m_f(t) = \kappa(\omega_e, p_m, \vartheta_f, \ldots) \cdot \dot{m}_\psi(t) - \frac{m_f(t)}{\tau(\omega_e, p_m, \vartheta_f, \ldots)} \tag{2.66}$$

where $\dot{m}_\psi$ is the fuel mass injected, $\dot{m}_\varphi$ the fuel mass aspirated into the cylinder, and m_f the mass of fuel stored in the fuel film. The coefficients $\kappa(.)$ and $\tau(.)$ depend on the engine speed and load, and on many other variables (mean fuel temperature ϑ_f, etc.).

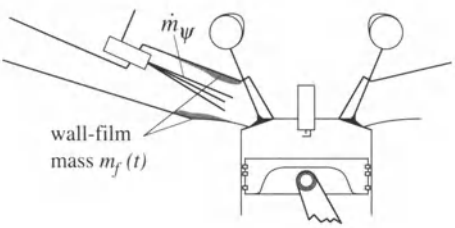


Fig. 2.20. Fuel wall-film model.

The usual approach to cope with this variability is to use only a pair of *constant* coefficients k and τ in the neighborhood of the operating point (engine speed and load) for which they were experimentally found. Of course, these coefficients must be updated when the engine moves away from the original operating point. Additionally, a correction factor is used to compensate for the changing engine temperature.

Another approach is to automatically identify the time-varying numerical values of the k and τ parameters (see e.g., [112], [171]) or to apply adaptive control methods (see e.g., [173]).

Notice that (2.65) and (2.66) inherently satisfy the structural constraint imposed by the mass conservation principle. Other formulations, which assume two (or more) separate wall films (with different time constants τ modeling fast and slow response contributions) or which include the influence of the back flow of exhaust gas, have been proposed as well (see e.g., [172]).

Notice also that the model represented by (2.65) and (2.66) (excluding the non-generic case $\kappa(.) = 1$) is *invertible*, i.e., if the actual wall film mass m_f is known, the fuel mass $\dot{m}_\psi$ which needs to be injected to guarantee a desired fuel mass flow $\dot{m}_{\varphi,des}$ to enter the cylinder can be computed by

$$\dot{m}_\psi(t) = \frac{1}{1 - \kappa(\omega_e, p_m, \vartheta_f, \ldots)} \cdot \left(\dot{m}_{\varphi,des}(t) - \frac{m_f(t)}{\tau(\omega_e, p_m, \vartheta_f, \ldots)} \right) \tag{2.67}$$

This property will be important in the air/fuel ratio control problem (see Chapter 4).

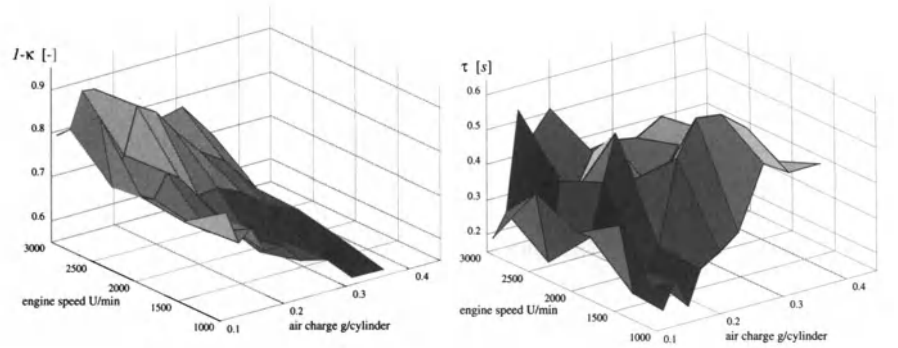


Fig. 2.21. Example of experimentally obtained maps of the coefficients k, τ used in (2.65) and (2.66); fully warmed-up engine (1.8-liter, 4 cylinders, 16V).

First-Principle Models

Since the experimental identification of the complete set of parameters k and τ is time consuming, approaches based on physical first principles that promise to shorten that time are of great interest. Several publications discuss this problem, ranging from complex simulations [97] to less detailed formulations [69].

In the following, a model proposed in [110], that is based on thermodynamic analogies, is presented. This model will be shown to have the same model structure as the model described by (2.65) and (2.66) while requiring much less experimental effort for the identification of the full set of parameters.

The wall-wetting phenomenon is split into three subproblems:

- the evaporation of droplets while they are airborne;
- the reduction of the number of droplets by collisions with the wall; and
- the evaporation of the liquid fuel on the manifold walls.

The change of the mass of fuel on the manifold walls can be expressed as the amount of fuel impinging on the walls minus the amount of fuel evaporating from the walls $\dot{m}_{EVF}$. The first contribution is the difference between the injected mass, $\dot{m}_{\psi}$, and the mass flow $\dot{m}_{EVD}$ that evaporates from the droplet while airborne. The two evaporation processes are modeled by using well-known thermodynamic analogies [84], i.e., the fuel mass flow resulting from the evaporation process is given by

$$\dot{m}_{EV} = h_m \cdot \rho_f \cdot A_f \cdot \ln(1 + B) \tag{2.68}$$

where h_m is the mass transfer coefficient, ρ_f is the density of the liquid fuel, A_f is the area of the surface on which the evaporation takes place, and B is the *Spalding* number which describes the evaporation properties of the liquid. This last parameter is defined as

$$B = \frac{x_{vf}(\vartheta_f) - x_{vg}}{1 - x_{vf}(\vartheta_f)} \tag{2.69}$$

where $x_{vf}(\vartheta_f)$ is the mass fraction of fuel that has vaporized at the actual temperature of the liquid fuel ϑ_f and x_{vg} is the mass fraction of fuel in the surrounding gas flow. Below, it is assumed that the vaporized fuel is rapidly convected past the liquid fuel zone such that x_{vg} is assumed to be zero. The function $x_{vf}(\vartheta_f)$ can be experimentally determined. Figure 2.22 shows such an evaporation curve for a standard unleaded RON98 type of gasoline.

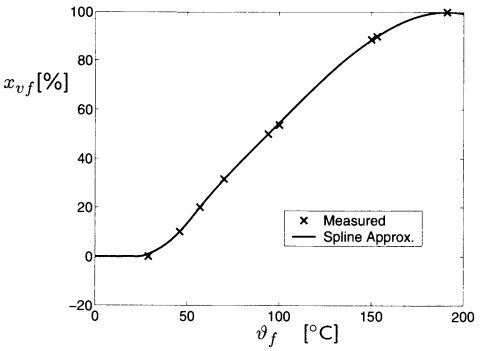


Fig. 2.22. Example of a fuel evaporation curve.

The mass transfer coefficient h_m must be calculated with thermodynamic analogies as well

$$h_m = Sh \cdot \frac{D_{AB}}{d_x} \tag{2.70}$$

where Sh is the *Sherwood number* (2.73) defined below and D_{AB} is the diffusivity coefficient, which may be taken from tables in literature (see e.g., [84]). Assuming ideal gas behavior, the pressure and temperature dependency of the diffusion coefficient for a binary mixture of gases, such as air and fuel, can be taken into account with the relation

$$D_{AB} \propto \frac{\vartheta_f^{3/2}}{p_m} \tag{2.71}$$

The characteristic length in the thermodynamic analogy of an evaporation problem is given by d_x where x indicates which specific geometry is relevant ($x = D$ for droplet and $x = F$ for wall film evaporation). The evaporated mass is

$$\dot{m}_{EV} = \frac{\rho_{fvs} \cdot D_{AB} \cdot A_f}{d_x} \cdot Sh \cdot \ln(1 + B) \tag{2.72}$$

The density ρ_{fvs} refers to the fuel vapor at the surface of the droplet. It can be approximated as half of the density of the liquid fuel.

The Reynolds analogy allows for the calculation of the Sherwood number based on the conditions of the air flux, represented by the Reynolds number Re, and on the characteristics of the fluid, represented by the Schmidt number Sc, [84]. To be more specific, the Sherwood number is calculated with the following analogy

$$Sh = c_r \cdot Re^{m_r} \cdot Sc^{n_r} \tag{2.73}$$

and the Reynolds and the Schmidt numbers are defined by

$$Re = \frac{u_\infty \cdot d}{\nu} \qquad Sc = \frac{\nu}{D_{AB}} \tag{2.74}$$

The kinematic viscosity ν depends on the temperature of the fluid and can be taken from tables as well. The parameters c_r, m_r, and n_r depend on the specific evaporation problem.

The relevant temperatures for the two evaporation problems cannot be measured. The only temperatures that are accessible in conventional engine management systems are the intake air temperature and the engine coolant temperature. Using these two measurements, the temperatures of the droplets and the fuel film can be estimated using simple models, which will be introduced in Sect. 2.6. Relations for the relevant heat fluxes can be found in [42], [2], [40], [3], and [49].

Droplet Evaporation

For the case of droplet evaporation, [139] provides the following correlations

$$c_r = 2 \cdot 0.276 \qquad m_r = 0.50 \qquad n_r = 0.33$$

For this problem, the droplet diameter d_D is chosen for the characteristic length. For the calculation of the Reynolds number, the velocity of the droplets must be calculated. The Bernoulli approach, introduced in (2.9), is used for that purpose.

The mass balance for one single droplet yields

$$\frac{d}{dt} m_D(t) = \rho_f \frac{d}{dt} V_D(t) = \rho_f \frac{1}{2} \pi \, d_D^2 \frac{d}{dt} d_D(t) = -\dot{m}_{EVD}(t) \tag{2.75}$$

Combining (2.75) and (2.72), and considering $A_{fD} = \pi \, d_D^2$, the dynamic behavior of the droplet diameter is described by the following first-order differential equation

$$\frac{d}{dt} d_D(t) = -\frac{1}{d_D} \cdot D_{AB} \cdot \ln(1 + B) \cdot Sh_D \tag{2.76}$$

Assuming all thermodynamic parameters (diffusivity coefficient, Sherwood number, Spalding Number, etc.) to remain constant, this equation can be solved in closed form

$$d_D(t) = \sqrt{d_0^2 - 2 \cdot D_{AB} \cdot \ln(1+B) \cdot Sh_D \cdot t} \tag{2.77}$$

Using this expression, the mass of evaporated fuel can easily be calculated

$$\dot{m}_{EVD}(t) = \frac{1}{2} \cdot \rho_f \cdot \pi \cdot d_D \cdot D_{AB} \cdot \ln(1+B) \cdot Sh_D \tag{2.78}$$

For the calculation of the total mass resulting from droplet evaporation the number of droplets must be assumed. In [110], an exponential decay for the number of airborne droplets is introduced. It is assumed that all the fuel is injected instantaneously and that all droplets are identical at the beginning. Then, due to the geometry of the intake runner, some droplets will hit the manifold walls. The time constant of the exponential decay, τ_{DA}, thus depends on the geometry of the engine under investigation and can be identified once and for all. The number of droplets at the beginning of the evaporation process can be calculated using a starting diameter d_0 for the droplets of approximately 100 μm, [113]. The number of airborne droplets $N(t)$ can then be calculated as follows.

$$N(t) = \frac{m_\psi}{\frac{\pi}{6} \cdot d_0^3 \cdot \rho_f} \cdot e^{-t/\tau_{DA}} \tag{2.79}$$

where m_ψ is the mass of fuel injected into one intake runner during one engine cycle.

To calculate the total mass of evaporated fuel due to droplet evaporation for one injection pulse, the differential equation for the evaporation must be integrated over an interval which starts with the beginning of the injection pulse, t_{is}, and ends with the closing of the intake valve, t_{IVC}). Combining (2.79) with (2.75) yields

$$m_{EVD} = \int_{t_{is}}^{t_{IVC}} N(t) \cdot \dot{m}_{EVD} \cdot dt \tag{2.80}$$

Usually, the time until a droplet is completely evaporated, t_{ev}, is smaller than the length of this interval. Thus, for a fully warmed up engine it is sufficient to integrate only up to this time.

$$t_{ev} = \frac{d_0^2}{2 \cdot C_D} \tag{2.81}$$

In this expression, all operating-point-dependent factors have been combined into one parameter

$$C_D = D_{AB} \cdot \ln(1+B) \cdot Sh_D \tag{2.82}$$

Since it was assumed that all droplets have the same diameter at $t = 0$, the time when all droplets have evaporated depends only upon the starting diameter and not on the amount of fuel injected. Consequently, the mass of fuel that evaporates from all droplets can be expressed by

$$m_{EVD} = (1 - \kappa(\pi_{op})) \cdot m_{\psi} \tag{2.83}$$

where the parameter $\kappa(\pi_{op})$ defined by

$$1 - \kappa(\pi_{op}) = \frac{3}{d_0^3} \cdot C_D \cdot \int_0^{t_{IVC} - t_{is}} e^{-\frac{t}{\tau_{DA}}} \cdot \sqrt{d_0^2 - 2 \cdot C_D \cdot t} \cdot dt \tag{2.84}$$

summarizes all operating parameters (intake manifold pressure and temperature, engine speed, fuel properties, engine geometry). For a given set of thermodynamic properties, i.e., temperature and pressure, $\kappa(\pi_{op})$ is constant and thus m_{EVD} is proportional to the amount of fuel injected.

Wall Film Evaporation

The mass $m_{\psi} - m_{EVD}$ of fuel droplets hitting the manifold wall, represent the mass increase for the wall film. This input is therefore dependent on the droplet evaporation model discussed before.

Experiments on a wetted tube, as described in [42] and [170], provide the analogy parameters used in (2.73) for this type of evaporation

$$c_r = 0.023 \quad , \quad m_r = 0.83 \quad , \quad \text{and} \quad n_r = 0.44$$

The fuel film is assumed not to be moving. The evaporation is, therefore, directly dependent upon the convection on the film surface caused by the air mass flow through the duct. Since it is assumed that the wall film forms directly around the intake valve(s), it is the air flow velocity through these valves that is relevant for the convection. For each cylinder, the average air flow velocity $u_{F\infty}$ through the valves can be approximated well by

$$u_{F\infty} = \frac{\dot{m}_{\beta}}{n_{cyl}} \cdot \frac{1}{\rho_{in} \cdot \frac{\pi}{4} d_F^2} \tag{2.85}$$

where d_F represents the diameter of the intake valve. This diameter is used as the characteristic length in the Reynolds analogy as well.

There are many approaches to describe the geometry of the wall film. One possibility is to represent it by a cylindric ring of fuel around the intake valve with thickness δ_{th} and height h_F as proposed in [110]. The height is the actual state variable of the wall film, whereas the thickness is a parameter which is slowly varying depending on the wall temperature. The thickness can either be identified in advance for a specific engine, or (preferably) identified on-line with an extended Kalman Filter.

Assuming $\delta_{th} \ll d_F$, this results in a relevant area A_F for the evaporation of the wall film of

$$A_F = \pi \cdot d_F \cdot h_F \tag{2.86}$$

Thus the mass of fuel evaporating from the wall film is

$$\dot{m}_{EVF} = \frac{\rho_{fvs} \cdot A_F}{d_F} \cdot D_{AB} \cdot Sh_F \cdot \ln(1+B) \tag{2.87}$$

$$= \rho_{fvs} \cdot \pi \cdot h_F \cdot D_{AB} \cdot Sh_F \cdot \ln(1+B) \tag{2.88}$$

The mass of fuel in the wall film is described by

$$m_f = \rho_f \cdot \pi \cdot d_F \cdot \delta_{th} \cdot h_F \tag{2.89}$$

The change in the mass of the wall film is then the difference between the mass of fuel of the injection which has not evaporated in the droplet phase and the mass of fuel evaporated from the wall film. The interconnections between droplet model and wall film model are illustrated in Fig. 2.23.

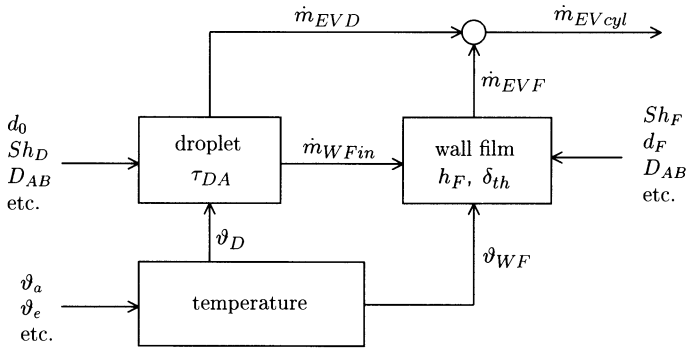


Fig. 2.23. Interconnections between droplet model and wall film model

Notice that the fuel injection process takes place at discrete points in time, and that the droplet evaporation processes start after that event. This is an inherently discrete-event formulation and, as such, not compatible with the mean-value paradigm used in this chapter. Nevertheless, once the corresponding fuel mass (2.83) has been computed, it is possible to spread out this variable over the entire engine cycle, neglecting its exact location on the time axis. By that approximation, a mean-value formulation is obtained.

For instance, the mass balance for the fuel in the wall film can be formulated as follows

$$\frac{d}{dt} m_f(t) = \kappa(\pi_{op}) \cdot \dot{m}_\psi(t) - \rho_{fvs} \cdot \pi \cdot h_F \cdot D_{AB} \cdot Sh_F \cdot \ln(1+B)$$

$$= \kappa(\pi_{op}) \cdot \dot{m}_\psi(t) - \dot{m}_{EVF}(t) \tag{2.90}$$

This equation can be transformed to the Aquino structure (2.65) and (2.66). Again utilizing the assumption $\rho_{fvs} = \rho_f/2$, the time constant of the evaporation is found to be

$$\frac{1}{\tau(\pi_{op})} = \frac{D_{AB} \cdot Sh_F \cdot \ln(1+B)}{2 \cdot d_F \cdot \delta_{th}} \tag{2.91}$$

Figures 2.24 and 2.25 show the results obtained with the approach presented here for two very different engine temperatures. Table 2.1 lists all relevant parameters of the two experiments and the simulations shown in these figures. Case 1 represents a fully warmed up engine in an operating point with low engine speed and load. For the measurements of case 2, the engine's thermostat has been removed. This results in a large coolant flow through the engine, which prevents the engine from warming up.

Table 2.1. Parameters for the wall film model.

Parameter	Case 1	Case 2
ω_e [rad/s]	157	157
m_β [g/cyl]	0.2	0.2
m_ψ [g/cyl]	$0.0136 \pm 5\%$	$0.0138 \pm 5\%$
p_m [bar]	0.578	0.586
ϑ_e [$^\circ K$]	335	303
ϑ_f [$^\circ K$]	430	390
ρ_f [kg/m^3]	747	747
Droplets		
c_r [-]	0.552	0.552
m_r [-]	0.5	0.5
n_r [-]	0.33	0.33
D_{AB} [m^2/s]	$2.2e^{-5}$	$1.8e^{-5}$
B [-]	0.35	0.0067
Sh_D [-]	1.30	1.32
$\tau_{DA}[s]$	0.012	0.012
$t_{ev}[s]$	0.00029	0.0157
Film		
c_r [-]	0.023	0.023
m_r [-]	0.83	0.83
n_r [-]	0.44	0.44
D_{AB} [m^2/s]	$2.8e^{-5}$	$2.6e^{-5}$
B [-]	8.5	1.5
Sh_F [-]	1.95	2.10
d_F [m]	0.052	0.052
$\delta_{th}[m]$	0.00085	0.00137
Aquino Parameters		
k [-]	0.81	0.53
τ [s]	0.36	1.42

Direct Injection Engines

Obviously, the wall-wetting effect is not present in direct injection engines. The mass of fuel entering the cylinder $\dot{m}_\varphi$ for this class of engines is an immediate control input. Moreover, not only can the mass of fuel be assigned for each

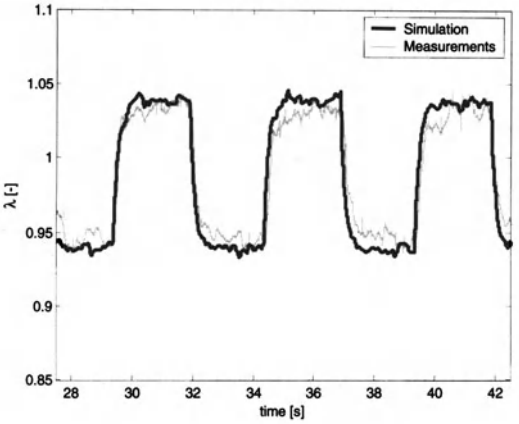


Fig. 2.24. Air/fuel ratio for warmed engine (85°C), fuel injection modulation.

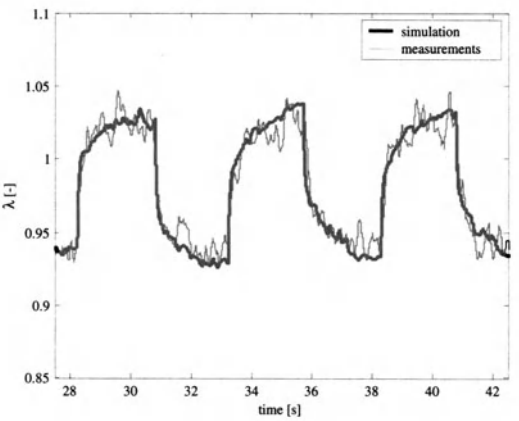


Fig. 2.25. Air/fuel ratio for cold engine (35°C), fuel injection modulation.

cycle individually, but the timing of the different fuel injections pulses can also be used to control the engine's behavior (torque, pollutant formation, etc., see below).

The injection pressure (fuel rail pressure) is another control variable that influences the mixture formation inside the cylinder and, hence, the engine behavior. Simple models of the fuel rail pressure dynamics are easy to derive (lumped parameter gasoline reservoir, gasoline as incompressible fluid, etc.). Precise fuel line dynamics (including compressibility effects and fluid dynamic oscillation phenomena) have also been developed (see for instance [23]). Such models, however, are not applicable to real-time control objectives.

2.4.3 Gas Mixing and Transport Delays

In the last sections, the air and fuel paths up to the cylinder entrance have been introduced. The next dynamic block will be the air/fuel ratio formation and propagation to the oxygen sensor. Several dynamic effects must be considered in that part (see Fig. 2.26 for some definitions used below) that will be modeled in this section using a MVM approach. Obviously, such a formulation is not able to correctly model single-cylinder effects. These aspects will be discussed in Sect. 3.2.6.

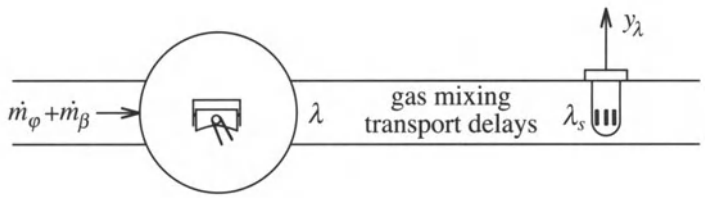


Fig. 2.26. Definitions of variables in the air/fuel ratio path.

The discrete engine operation introduces the induction to exhaust gas delay τ_{IEG} as defined in (2.2) such that, neglecting all other phenomena, immediately after the cylinder the air/fuel ratio should be

$$\widetilde{\lambda}(t) = \frac{\dot{m}_\beta(t - \tau_{IEG})}{\dot{m}_\varphi(t - \tau_{IEG})} \cdot \frac{1}{\sigma_0} \tag{2.92}$$

However, additional dynamics are introduced by the engine's gas exchange behavior. As shown above (see for instance (2.22)), a substantial part of the burnt gases remains inside the cylinder, especially at low loads. This gas fraction will carry on the air/fuel ratio of the present cycle to the next. A possible formulation[26] of an MVM for this phenomenon is

$$\lambda(t) = \frac{\widetilde{\lambda}(t) \cdot m(t) + \lambda(t - \tau_{IEG}) \cdot m_{eg}(t)}{m(t) + m_{eg}(t)} \tag{2.93}$$

where $m(t)$ is the mass of fresh charge that enters the cylinder and $m_{eg}(t)$ is the mass of burnt gas that remains in the cylinder. Note that (2.93) is a recursive formulation. The fresh charge is obtained by (Euler) integration over one segment, i.e.,

$$m(t) = \dot{m}(t - \tau_{IEG}) \cdot \frac{4\,\pi}{N \cdot \omega_e} \tag{2.94}$$

[26] In Sect. 3.2.5 a more detailed analysis of this problem will be made using a discrete-event approach. The formulation (2.93) will be shown to be a good approximation of the EGR dynamics.

(the engine speed is assumed to be constant over this short period). An estimation of $m_{eg}(t)$ is possible when the cylinder pressure p_z and burnt gas temperature ϑ_z at exhaust valve closing are known

$$m_{eg} = V_c \cdot \frac{p_z}{R_{eg} \cdot \vartheta_z} \tag{2.95}$$

(see Sect. 2.6 for details on ϑ_z). After exiting the cylinder, the exhaust gases must reach the oxygen sensor, and this sensor has its own dynamics which must be considered in an air/fuel ratio loop model. The exhaust gas transport can be approximated by a combination of delay and gas mixing, i.e., using a frequency domain formulation

$$\lambda_s(s) = e^{-\tau_1 \cdot s} \cdot \frac{1}{\tau_2 \cdot s + 1} \cdot \lambda(s) \tag{2.96}$$

The time constants τ_1 and τ_2 depend upon the engine system's geometry (volume and length of exhaust pipes) and upon the actual operating point (especially exhaust gas mass flow and temperature), as will be discussed in Sect. 2.6.

The air/fuel ratio sensor, provided it is actively heated and built using state-of-the-art planar technology, is rather fast. A first-order lag is usually a good approximation of its dynamics [58]

$$x_\lambda(s) = \frac{1}{\tau_s \cdot s + 1} \cdot \lambda_s(s) \tag{2.97}$$

where the time constant τ_s is in the order of $10 - 20\ ms$ (depending on the type of protection cap used). The output voltage y_λ depends upon the fictitious intermediate signal x_λ in a nonlinear but static way. Both switching and continuous sensors are in use.

2.5 Mechanical System

2.5.1 Torque Generation

Introduction

The primary objective of an engine is to produce mechanical power. Its speed is a level variable, i.e., it is not arbitrarily assignable. The torque, however, can be changed at will, provided that the amount of mixture in the cylinder and/or its composition, for instance, can be changed arbitrarily. The mean-value engine torque is a nonlinear function of many variables (fuel mass in cylinder, air/fuel ratio, engine speed, ignition or injection timing, EGR rate, etc.)

$$T_e = f(\dot{m}_\varphi, \lambda, \omega_e, \zeta, x_{egr}, \ldots) \tag{2.98}$$

Detailed thermodynamic simulations are necessary to correctly predict the engine torque. For control purposes, however, such simulations are too time-consuming. Thus, alternative approaches have been investigated. An obvious (but not feasible) solution would be to grid the space of all possible operating conditions, to measure or compute the engine torque in each vertex, and to store that information for later on-line use in appropriate maps.[27] Another, more promising approach is to use some physical insight to separate the different influencing variables and to divide the modeling task into several low-dimensional problems.

Before entering into a detailed discussion of this approach, two definitions are introduced. Since engine torque mainly depends on engine size, i.e., its displaced volume V_d, it is advantageous to use a normalized formulation introducing the *brake mean effective pressure*

$$p_{me} = \frac{T_e \cdot 4\,\pi}{V_d} \tag{2.99}$$

and the *fuel mean effective pressure*

$$p_{m\varphi} = \frac{H_l \cdot m_\varphi}{V_d} \tag{2.100}$$

where the mass of fuel, with lower heating value H_l, burnt per combustion cycle m_φ in (2.100) is related to the mean-value fuel mass flow by

$$\dot{m}_\varphi(t) = m_\varphi(t) \cdot \frac{\omega_e(t)}{4\,\pi} \tag{2.101}$$

The brake mean effective pressure p_{me} is the pressure that has to act on the piston during one full expansion stroke to produce the same amount of work as the real engine does in two engine revolutions (assuming a four-stroke engine). The fuel mean effective pressure $p_{m\varphi}$ is that brake mean effective pressure that an engine with an efficiency of 1 would produce with the fuel mass m_φ burnt per engine cycle (perfect conversion of the fuel's thermal energy into mechanical energy).

Obviously, the engine's efficiency can then be written as

$$\eta_e = \frac{p_{me}}{p_{m\varphi}} = \frac{T_e \cdot 4\,\pi}{m_\varphi \cdot H_l} \tag{2.102}$$

With this definition (2.99) can be rewritten as

$$p_{me} = \eta_e(m_\varphi, \omega_e, \lambda, \zeta, x_{egr}, \ldots) \cdot p_{m\varphi} \tag{2.103}$$

and an approximation of the following form

[27] Taking into consideration only the five independent variables mentioned in (2.98), using for each variable say, 20 support points, and assuming a measurement duration of one minute per vertex results in six years of consecutive experimentation.

$$\eta_e \approx [\eta_{e0}(m_\varphi, \omega_e) \cdot \eta_\lambda(\lambda) \cdot \eta_\zeta(\zeta) \cdot \eta_{egr}(x_{egr}) \cdot \ldots] \cdot p_{m\varphi} \tag{2.104}$$

of the efficiency can be used. The first factor $\eta_{e0}(m_\varphi, \omega_e)$ is the engine's nominal efficiency map and the remaining factors represent the changes occurring when the respective inputs deviate from their nominal values.

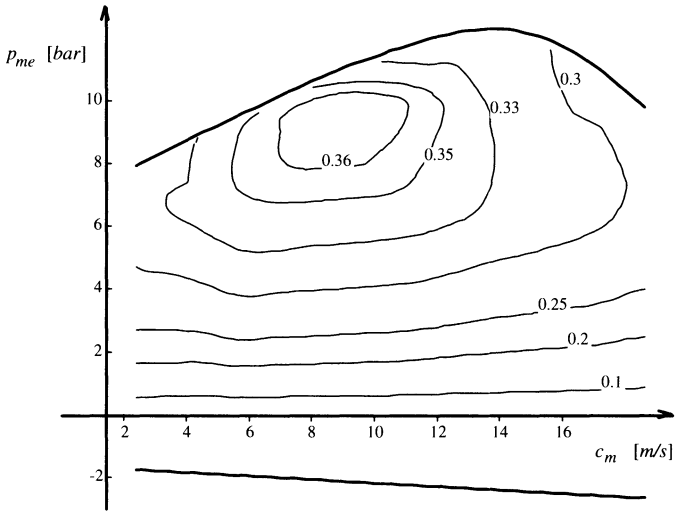


Fig. 2.27. Engine map of a modern naturally aspirated 2-liter SI engine.

For existing engines, the nominal engine efficiency $\eta_{e0}(m_\varphi, \omega_e)$ is often available in the form of a measured engine map (see Fig. 2.27 for a typical example). However, in control applications this representation has several drawbacks:

- At low loads efficiency is not precisely measurable since unavoidable small errors in measured torque produce relatively large errors in estimated brake power.
- The efficiency is displayed as a function of mean piston speed c_m[28] (this poses no problems) and mean effective pressure p_{me}, which, according to Fig. 2.5, is not a level variable but the *output* of the combustion block.

Other forms are therefore necessary to display the engine's nominal torque production properties. Figure 2.28 shows different possibilities, starting with the usual engine map in the top left corner. The form shown in the bottom right corner is especially useful for mean-value engine models since it interfaces directly with the main inputs (see Fig. 2.5).

[28] Mean piston speed is defined by $c_m = S \cdot \omega_e / \pi$ where S is the engine's stroke.

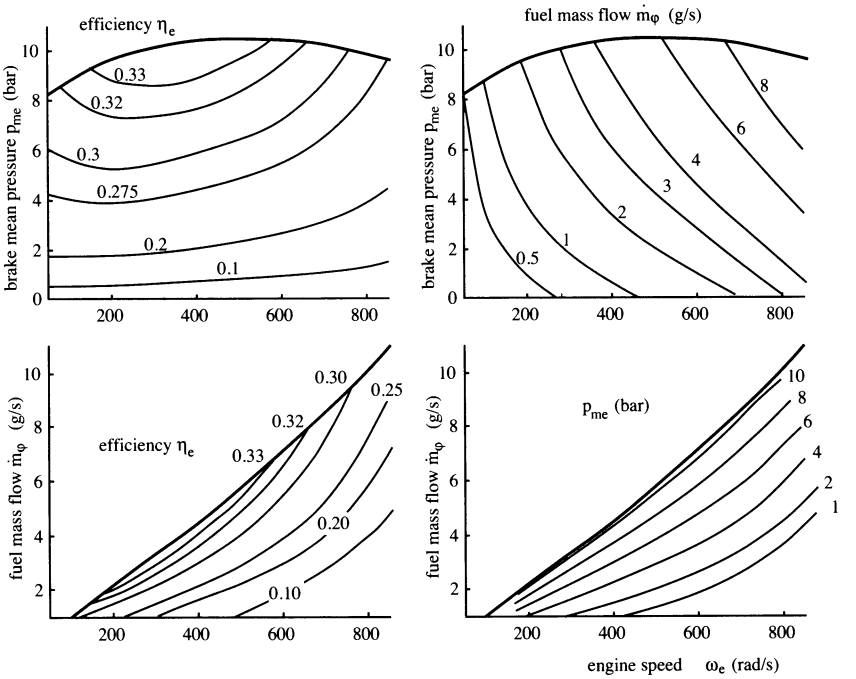


Fig. 2.28. Various alternative representations of an engine map.

Willans Approximation

A useful simplification of the engine's torque and efficiency characteristics can often be made, especially for Diesel engines. In many measurements and process simulations, it has been observed [63] that a simple relationship between mean fuel pressure and mean effective pressure approximates real engine behavior astonishingly well

$$p_{me} = e(m_\varphi, \omega_e, \lambda, \zeta, x_{egr}, \ldots) \cdot p_{m\varphi} - p_{me0}(\omega_e, m_\beta, \vartheta_e, \ldots) \tag{2.105}$$

The coefficient $e(.)$ represents the thermodynamic properties of the engine (related to the indicated or inner mean pressure) and the second part $p_{me0}(.)$ represents the engine's friction and gas exchange losses. Quite often the simplification

$$\frac{\partial e(m_\varphi, \omega_e, \lambda, \zeta, x_{egr}, \ldots)}{\partial m_\varphi} = 0 \tag{2.106}$$

is possible (especially when excluding full-load conditions), in which case (2.105) becomes affine in the injected fuel mass (see Fig. 2.29).

Aside from its simplicity, (2.105) has another important advantage: it clearly distinguishes between the contribution of the internal (thermodynamic) and external effects (friction, gas exchange, etc.) on engine efficiency.

This separation simplifies the modeling of the different torque correction factors as introduced in (2.104).

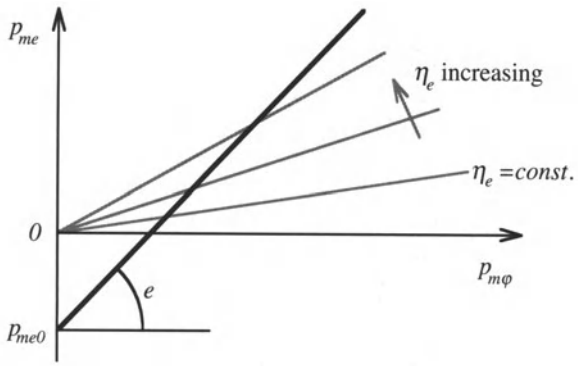


Fig. 2.29. Affine Willans model of the engine efficiency (constant speed, air/fuel ratio, ...).

Gas Exchange and Friction Losses

The losses represented by $p_{me0}(\ldots)$ are discussed first. As mentioned above, these losses comprise a friction and a gas exchange component

$$p_{me0}(\omega_e, \dot{m}_\beta, \vartheta_e, \ldots) = p_{me0f}(\omega_e, \vartheta_e, \ldots) + p_{me0g}(\dot{m}_\beta) \tag{2.107}$$

The parameter $p_{me0g}(\dot{m}_\beta)$ is the cycle-averaged pressure difference between the intake and exhaust port of the engine. When the engine is motored (with no fuel injection and completely closed throttle valve for SI engines and no boost for Diesel engines), the following values are reasonable estimations

$$\text{Otto} : p_{me0g}(0) \approx 0.9\ bar, \qquad \text{Diesel} : p_{me0g}(0) \approx 0.1 \ldots 0.2\ bar \tag{2.108}$$

In Diesel engines, $p_{me0g}(\dot{m}_\beta)$ is primarily a function of the turbocharger characteristics, i.e., the total gas exchange losses[29] are the pressure difference between the engine's exhaust and intake side plus the 0.1 ... 0.2 bar of $p_{me0g}(0)$ which model flow-friction losses (intake and exhaust valves, intercooler, etc.).

In naturally aspirated SI engines, the exhaust pressure varies less and an approximation of the gas-exchange losses can be derived *assuming* a linear dependency between intake pressure and mass of aspirated air

$$p_{me0g}(\dot{m}_\beta) = p_{me0g}(0) \cdot \left[1 - \frac{m_\beta}{1.1 \ldots 1.2 \cdot m_{\beta,max}}\right]$$

[29] Of course, in some operating points the pressure difference can be positive, i.e., positive gas-exchange loops can occur in turbocharged systems.

$$= p_{me0g}(0) \cdot \left[1 - \frac{\lambda \cdot \sigma_0 \cdot V_d/H_l}{1.1 \ldots 1.2 \cdot m_{\beta,max}} \cdot p_{m\varphi}\right] \tag{2.109}$$

where $m_{\beta,max}$ is the maximum amount of air mass the engine can aspirate in one cycle.

Note that for both Diesel and SI engines, the intake and exhaust pressures are readily available if a model according to Fig. 2.5 or Fig. 2.6 is implemented and, therefore, the relation

$$p_{me0g} = p_3 - p_2 \tag{2.110}$$

can be used directly. The approximation (2.109) is only used in less detailed simulations, for instance in quasi-static estimations of the fuel consumption [66].

Engine friction, as represented by the $p_{me0f}(\omega_e, \vartheta_e, \ldots)$ term in (2.107), is discussed now. Several models have been proposed for this part (see e.g., [132]). The friction model developed in [162] is used in this text. It has the following form

$$p_{me0f}(\omega_e, \vartheta_e, \ldots) = k_1(\vartheta_e) \cdot \left(k_2 + k_3 \cdot S^2 \cdot \omega_e^2\right) \cdot \Pi_{e,max} \cdot \sqrt{\frac{k_4}{B}} \tag{2.111}$$

Typical values for the parameters are given in Table 2.2. The variable $\Pi_{e,max}$ is the maximum boost ratio for which the engine is designed to operate at low speeds,[30] B is the engine's bore diameter and S its stroke. The temperature behavior of k_1 is assumed to vary as shown in Fig. 2.30 where the engine temperature ϑ_e is an average value that describes the thermal state of the engine (see Sect. 2.6).

Table 2.2. Parameters of ETH friction model.

	SI	Diesel
$k_1(\vartheta_\infty)$	$1.44 \cdot 10^5$ (Pa)	$1.44 \cdot 10^5$ (Pa)
k_2	0.46 $(-)$	0.50 $(-)$
k_3	$9.1 \cdot 10^{-4}$ (s^2/m^2)	$1.1 \cdot 10^{-3} (s^2/m^2)$
k_4	0.075 (m)	0.075 (m)

[30] Higher boost ratios require larger crankshaft bearings, etc., thus leading to increased friction.

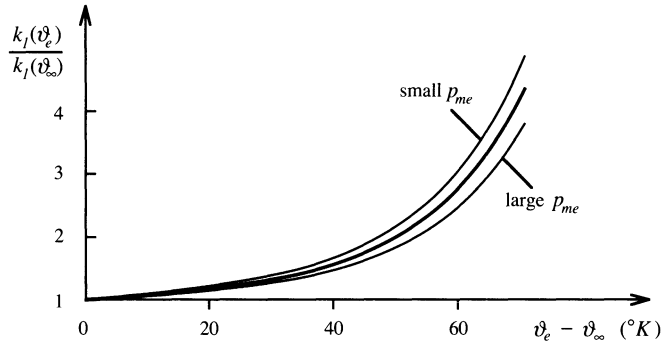


Fig. 2.30. Temperature dependency of mechanical friction. Mean engine temperature ϑ_e, with ϑ_∞ the temperature in fully warmed conditions.

Thermodynamic Efficiencies

According to the approach introduced in (2.104), the internal (thermodynamic) efficiency is separated as follows

$$e(\omega_e, \lambda, \zeta, x_{egr}, \ldots) = e_\omega(\omega_e) \cdot e_\lambda(\lambda) \cdot e_\zeta(\zeta) \cdot e_{egr}(x_{egr}) \cdot \ldots \qquad (2.112)$$

where the first factor $e_\omega(\omega_e)$ is centered at the optimum point of the other variables $\lambda, \zeta, x_{egr}, \ldots$

Engine Speed

Typically, the factor $e_\omega(\omega_e)$ has a parabolic form, as shown in Fig. 2.31. At very low speeds, the relatively large heat losses through the wall reduce engine efficiency, while at very high speeds, the combustion times become unfavorably large compared to the available interval in the expansion stroke. Note that $e_\omega(\omega_e)$ has a magnitude substantially smaller than 1 since it incorporates the basic thermodynamic efficiency mechanisms, while the other factors by design are around 1.

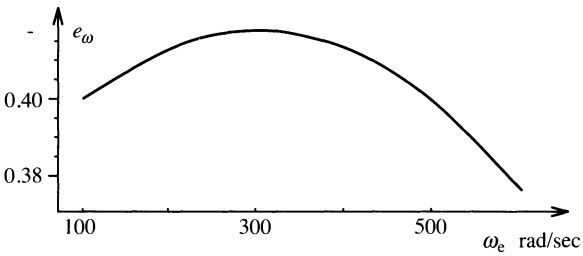


Fig. 2.31. Engine speed influence on efficiency for a SI engine (for a Diesel engine, similar form but maximum value around 0.46).

Air/Fuel Ratio

The factor $e_\lambda(\lambda)$ models the influence of the changing air/fuel ratio on the thermodynamic efficiency. This influence is, of course, very different for SI and CI engines.

In SI engines, rich to lean mixtures are possible. For rich mixtures, incomplete combustion and water/gas shift reactions[31] substantially reduce the thermodynamic efficiency. For mixtures more than slightly lean, sufficient oxygen is available for complete combustion and efficiency is not affected by changing λ. For intermediate $\lambda_1 \leq \lambda \leq \lambda_2$, a smooth transition is observed. For $\lambda > \lambda_{max}$, finally, misfires start to interfere such that these operating points are not modeled.

This behavior can be expressed quantitatively as

$$e_\lambda(\lambda) = \begin{cases} \gamma_1 \cdot \lambda - \gamma_0 & \text{for } \lambda_{min} < \lambda < \lambda_1 \\ e_{\lambda,1} + (1 - e_{\lambda,1}) \cdot \sin\left(\frac{\lambda - \lambda_1}{1 - \lambda_1}\right) & \text{for } \lambda_1 < \lambda < \lambda_2 \\ 1 & \text{for } \lambda_2 < \lambda < \lambda_{max} \end{cases} \tag{2.113}$$

where $e_{\lambda,1} = \gamma_1 \cdot \lambda_1 - \gamma_0$. Moreover, for continuity reasons, the following equation must be satisfied

$$\lambda_2 = \lambda_1 + \frac{\pi}{2} \cdot (1 - \lambda_1) \tag{2.114}$$

Figure 2.32 shows the form of $e_\lambda(\lambda)$ for realistic parameter values.

In Diesel engines, which are always operated in lean conditions, the air/fuel ratio has little influence on efficiency with the exception of the extreme levels. Both at very rich ($\lambda < 1.4$) and very lean conditions, the efficiency slightly drops. However, for most control purposes these extreme operating regions can be neglected.

Ignition/Injection Timing

The timing of the ignition in SI and of the injection in CI engines has a similar influence on torque production:

- There is an optimal (maximum break torque, MBT) ignition angle and start of injection angle denoted as $\zeta_0(\omega_e, p_{me})$;
- This MBT angle varies strongly with engine operating conditions (especially speed and load);
- For ignition or injection angles other than MBT the torque drops almost parabolically; and

[31] The conversion of fuel (hydrocarbon) into H_2 and CO consumes some of the fuel's enthalpy, such that it is not available for the thermodynamic processes.

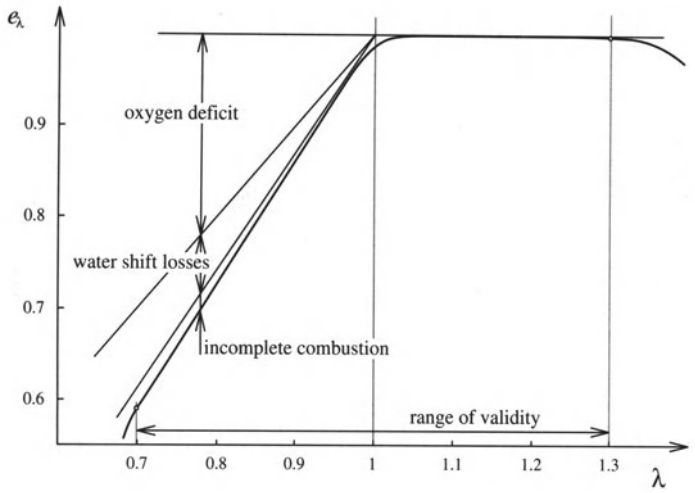


Fig. 2.32. Air/fuel ratio influence on engine efficiency. Parameters $\lambda_{min} = 0.7$, $\lambda_1 = 0.95$, $\lambda_2 = 1.0285$, $\lambda_{max} = 1.3$, $\gamma_0 = 0.373$, $\gamma_1 = 1.373$.

- The form of this reduction is, in first approximation, independent of the operating point and of the engine type [57]. An example of that reduction for an SI engine is shown Fig. 2.33.

  The factor $e_\zeta(\zeta)$ in (2.112) can thus be written as

$$e_\zeta(\zeta) = 1 - k_\zeta \cdot (\zeta - \zeta_0(\omega_e, p_{me}))^2 \tag{2.115}$$

Note that in both engine types, MBT settings might not be attainable due to knocking phenomena or pollutant emission restrictions (see Sect. 2.7).

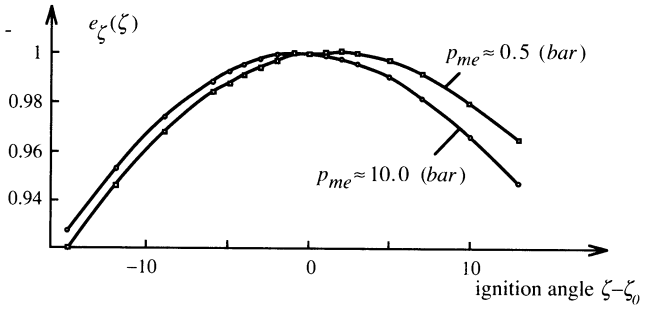


Fig. 2.33. Influence of ignition angle deviation from MBT settings on engine torque at constant engine speed for almost idle to full-load conditions.

Burnt Gas Fraction

The factor $e_{egr}(x_{egr})$ models the influence of the changing burnt gas fraction on the thermodynamic efficiency. This influence is, of course, very much different for SI and Diesel engines. In both engine types, burnt gases are always present (internal EGR). Their effect on the thermodynamic efficiency is already included in the formulations above.

Additional (external or internal) EGR reduces the formation of nitrogen oxide, but negatively affects the thermodynamic efficiency of SI engines. In fact, increased EGR rates lead to slower burning speeds and, therefore, to reduced factors $e_{egr}(x_{egr})$. This effect will have a qualitative form as shown in Fig. 2.34. Note that the reduced burning speed has a more negative influence at high engine speeds, where relatively less time for charge burning is available. Therefore, a speed-dependent formulation of the EGR term of the following form is recommended

$$e_{egr}(x_{egr}) = 1 - k_{egr,1} \cdot (1 + k_{egr,2} \cdot \omega_e) \cdot x_{egr}^2 \tag{2.116}$$

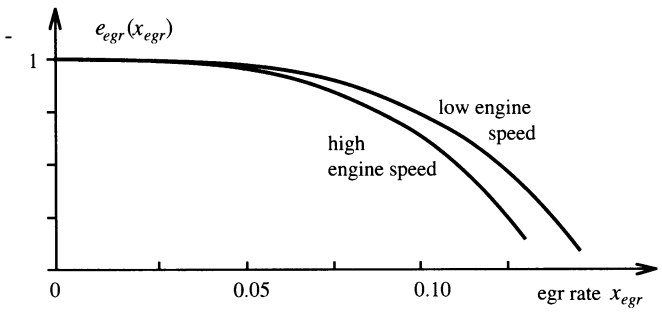


Fig. 2.34. Influence of EGR rate on SI engine efficiency.

In Diesel engines the EGR rate has — with the exception of very large ratios — almost no effect on efficiency. The Diesel-specific diffusion-controlled burning mechanism and the always abundant oxygen render this type of combustion much more tolerant to high EGR, which is one of the reasons why EGR in modern Diesel engines is a standard feature. However, as stated in the Introduction, higher EGR rates lead to increased particulate matter formation, such that this trade-off is often an important limiting factor in Diesel engines.

Direct-Injection SI Engines

All the assertions made above are valid, in principle, for direct-injection gasoline engines operating in lean and stratified mode alike (in homogeneous mode there are almost no differences at all). The differences that can be observed

concern the specific shape of the functions relating torque and input parameters, but not the structural properties. Of course, for GDI engines, more degrees of freedom are available.

Injection timing ξ, together with ignition timing ζ, substantially influences engine torque in the stratified mode. This can be captured, for instance, by adding an additional term in (2.112)

$$e(\omega_e, \lambda, \zeta, x_{egr}, \xi, \ldots) = e_\omega(\omega_e) \cdot e_\lambda(\lambda) \cdot e_\zeta(\zeta) \cdot e_{egr}(x_{egr}) \cdot e_\xi(\xi) \cdot \ldots \quad (2.117)$$

or by modifying the ignition factor (2.115) to include also the injection timing effects

$$e_\zeta(\zeta, \xi) = 1 - k_\zeta(\xi) \cdot (\zeta - \zeta_0(\omega_e, p_{me}, \xi))^2 \quad (2.118)$$

2.5.2 Engine Speed

System Dynamics

In a mean value setting, modeling the engine speed behavior is a straightforward process. In fact, the engine inertia is assumed to be constant and all engine friction phenomena are already included in the torque model described by (2.105).

The only relevant reservoir is the engine flywheel. It stores kinetic energy and the differential equation for the corresponding level variable $\omega_e(t)$ is

$$\Theta_e \cdot \frac{d}{dt}\omega_e(t) = T_e(t) - T_l(t) \quad (2.119)$$

The load torque T_l is assumed to be known. In a more general setting this variable would depend on engine speed, drive train properties and external loads like road slope or wind disturbances. However, as mentioned in the Introduction, these aspects are not analyzed in this text.

Sometimes a crank-angle-based formulation of (2.119) is preferred, which is obtained using the transformation $t \rightarrow \phi$ where ϕ is the engine's crank angle

$$\Theta_e \cdot \frac{d}{d\phi}\omega_e(\phi) = \frac{1}{\omega_e(\phi)} \cdot [T_e(\phi) - T_l(\phi)] \quad (2.120)$$

Note that for torques that depend linearly on engine speed, (2.119) is linear, while (2.120) — describing the same system — is nonlinear. On the other hand, many engine phenomena (e.g., air intake) are easier to describe with the crank angle as the independent variable, since the influence of the engine speed on the dynamic phenomena is then substantially reduced. More details on these aspects will be given in Chapter 3.

Note also that, for diagnostic purposes (e.g., in on-board diagnosis systems for misfire detection), much more detailed models are in use. In these models, the reciprocating structure (both of the combustion and of the crank mechanism) is correctly modeled such that the engine speed variations during one engine cycle are predicted (see the original papers [144] and [134], or [93] and [31] for more recent publications).

Constant Speed Approximation

In discrete-event models, the mean engine speed is often assumed to be almost constant during the next few (say 10) cycles such that the discrete-*event* model can be approximated by a discrete-*time* model in which the crank angle and the time evolve in parallel.[32] For this assumption to be realistic, the speed dynamics described by (2.119) must be much slower than the other dynamic effects of interest. For naturally aspirated SI engines, one of the most important of these other variables is the torque, which depends, according to (2.105), mainly on manifold pressure. All other variables have much faster time constants or are linked to the manifold pressure (air/fuel ratio, EGR ratio, etc.). In the following, only SI engines will be analyzed (see [176] for a similar problem formulation arising in Diesel engines).

As a first step, the slowest part of the manifold pressure dynamics has to be identified, for which a constant engine speed $\omega_e(t) = \omega_{e0}$ is assumed temporarily. The mean-value intake manifold dynamics for sonic conditions and isothermal[33] gas behavior in the manifold can be derived from (2.8), (2.10), (2.15) and (2.19) to be

$$\frac{d}{dt}p(t) = \frac{R \cdot \vartheta_m}{V_m} \cdot \left(A(t) \frac{p_a}{\sqrt{R \cdot \vartheta_m}} \cdot \frac{1}{\sqrt{2}} - \frac{p(t)}{R \cdot \vartheta_m} \cdot \frac{V_d \cdot \omega_{e0}}{4\pi} \right) \tag{2.121}$$

(for the sake of simplicity, discharge coefficient and volumetric efficiency have been set to 1), or for subsonic conditions

$$\frac{d}{dt}p(t) = \frac{R\vartheta_m}{V_m} \left(A(t) \frac{p_a}{\sqrt{R\vartheta_m}} \sqrt{\frac{2p(t)}{p_a}[1 - \frac{p(t)}{p_a}]} - \frac{p(t)}{R\vartheta_m} \frac{V_d \omega_{e0}}{4\pi} \right) \tag{2.122}$$

For constant engine speeds, $\omega_e(t) = \omega_{e0}$, (2.121) is linear, with the time constant

$$\tau_l = \frac{4\,\pi \cdot V_m}{\omega_{e0} \cdot V_d} \tag{2.123}$$

Equation (2.122) is nonlinear. It is linearized around the equilibrium operating point $\{A_0, p_0\}$

$$A_0 = \frac{V_d \cdot \omega_e \cdot p_0}{\sqrt{R \cdot \vartheta_m} \cdot 4\,\pi \cdot \sqrt{2\,p_0 \cdot [p_a - p_0]}} \tag{2.124}$$

Linearizing (2.122) around this operating point yields the following time constant

$$\tau_{nl} = \frac{4\,\pi \cdot V_m}{\omega_{e0} \cdot V_d} \cdot \frac{2\,(p_a - p_0)}{p_a} \tag{2.125}$$

[32] Analyzing an engine system as a truly discrete-event system requires sophisticated mathematical tools, see, for instance, [19] or [20].

[33] The same conclusions are obtained for the adiabatic formulation.

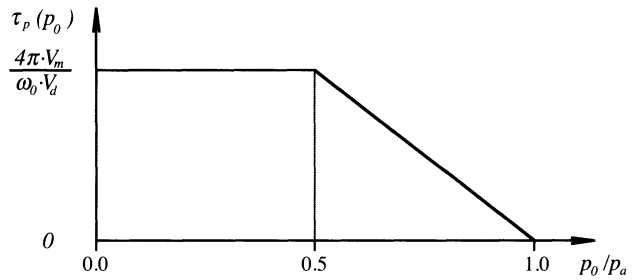


Fig. 2.35. Time constants of linearized manifold system for different nominal points.

If these time constants are much faster than the speed dynamics for all operating points possible, then the postulated simplification of the discrete-event system to a discrete-time system is indeed valid. As Fig. 2.35 shows, the critical case (the largest time constant) occurs at very low loads or manifold pressures.

Of course, it could be argued that the linear system behavior is not relevant since large pressure swings do occur often. However, as Fig. 2.36 shows, in the nonlinear case (dark curves in the middle plot), the manifold pressure converges even faster to its steady-state value than in the linear case (gray curve). In fact, the switch between the sonic and the subsonic conditions accelerates the convergence of the manifold pressure because inlet air mass flow is reduced more than in the linear case where downstream pressure is not influencing throttle air mass flow (see (2.121) and (2.122)).

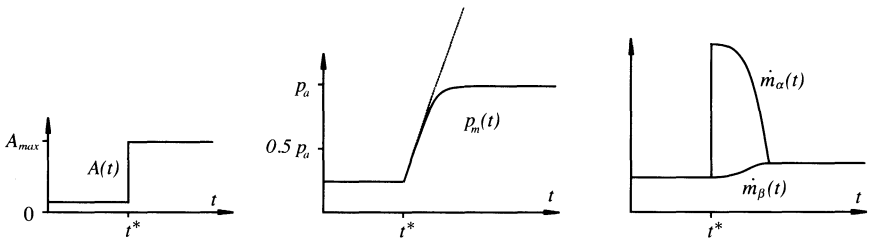


Fig. 2.36. Large-signal system behavior of the manifold pressure dynamics.

In summary, the worst case occurs at:

- low loads and small deviations in manifold pressure from equilibrium (slowest manifold pressure dynamics); and
- open clutch, i.e., zero load torque, since in that case only the engine inertia is active and the vehicle's power train inertia and mass do not delay the speed changes (fastest engine speed dynamics).

If in those situations the engine speed dynamics are much slower than the torque dynamics, the simplification of the engine's discrete-event behavior as a discrete-time system is indeed justified. This will be analyzed in the rest of this section.

The starting point is the combined and simplified manifold pressure and speed dynamics

$$\Theta_e \cdot \frac{d}{dt}\omega_e(t) = \frac{e \cdot H_l}{4\pi} \cdot \frac{1}{1+\sigma_0} \cdot \frac{V_d}{R \cdot \vartheta_m} \cdot p_m(t) - \frac{V_d}{4\,\pi} \cdot p_{me0} \tag{2.126}$$

and

$$\frac{d}{dt}p_m(t) = \frac{R \cdot \vartheta_m}{V_m} \cdot \left(A(t) \cdot \frac{p_a}{\sqrt{R \cdot \vartheta_m}} \cdot \frac{1}{\sqrt{2}} - \frac{p_m(t)}{R \cdot \vartheta_m} \cdot \frac{V_d}{2} \cdot \frac{\omega_e(t)}{2\pi} \right) \tag{2.127}$$

To simplify the subsequent computations, the Willans coefficients are assumed to be constant, volumetric efficiency and air/fuel ratio are set to 1, the IPS delay is neglected,[34] and the gas properties of mixture and air are assumed to be identical.

To simplify the notation, the following abbreviations are introduced

$$\begin{aligned} k_1 &= \sqrt{R \cdot \vartheta_m/2} \cdot p_a/V_m \\ k_2 &= V_d/(4\,\pi \cdot V_m) \\ k_3 &= e \cdot V_d \cdot H_l/(4\pi \cdot (1+\sigma_0) \cdot R \cdot \vartheta_m \cdot \Theta_e) \\ k_4 &= V_d \cdot p_{me0}/(4\,\pi \cdot \Theta_e) \end{aligned} \tag{2.128}$$

Using these definitions the combined system can be written compactly as

$$\begin{aligned} \tfrac{d}{dt}p_m(t) &= k_1 \cdot A(t) - k_2 \cdot \omega_e(t) \cdot p_m(t) \\ \tfrac{d}{dt}\omega_e(t) &= k_3 \cdot p_m(t) - k_4 \end{aligned} \tag{2.129}$$

If, for the moment, the assumption holds that the speed dynamics (the second equation) are much slower than the torque dynamics (the validity of this assumption will be verified later), then a separation[35] of the two equations is possible as follows:

- Case a) assuming a constant engine speed $\omega_e(t) = \omega_{e0}$ and a dynamic behavior of the manifold pressure.
- Case b) assuming an algebraic relationship between engine speed and manifold pressure, i.e., the derivative of $p_m(t)$ is set to zero and a corresponding dynamic behavior of the engine speed is derived.

[34] Even at idle speed the IPS delay is smaller than the manifold time constant. However, as discussed in detail in the Appendix B, the delays present in the system have a substantial influence on the system behavior.

[35] In system theory, this approach is known as singular perturbation, [178].

In case a), the manifold pressure satisfies the linear first-order ODE

$$\frac{d}{dt}\widetilde{p}_m(t) = k_1 \cdot A(t) - k_2 \cdot \omega_{e0} \cdot \widetilde{p}_m(t) \tag{2.130}$$

and in case b), the engine speed satisfies the nonlinear first-order ODE

$$\frac{d}{dt}\widetilde{\omega}_e(t) = \frac{k_1 \cdot k_3}{k_2} \cdot \frac{A(t)}{\widetilde{\omega}_e(t)} - k_4 \tag{2.131}$$

where the tilde is used to denote the corresponding variables in the fully separated limit case. Using the two equilibrium values

$$p_{m,\infty} = \frac{k_4}{k_3}, \qquad A_\infty = \frac{k_2 \cdot k_4}{k_1 \cdot k_3} \cdot \omega_{e0} \tag{2.132}$$

to linearize (2.131) — the system (2.130) is always linear — yields the two separated time constants

$$\widetilde{\tau}_p = \frac{1}{k_2 \cdot \omega_{e0}}, \qquad \widetilde{\tau}_\omega = \frac{\omega_{e0}}{k_4} \tag{2.133}$$

Inserting the definitions (2.128), the ratio of these two time constants is then found to be

$$\chi = \frac{\widetilde{\tau}_p}{\widetilde{\tau}_\omega} = \frac{p_{me0} \cdot V_m}{\Theta_e \cdot \omega_{e0}^2} \tag{2.134}$$

Inserting typical numerical values[36] for the critical case, i.e., at idle speed, values of χ in the order of 0.1 are obtained for typical naturally aspirated SI engines. In other words, in the *worst case*, the manifold dynamics are faster than the speed dynamics; in all other cases (higher engine speeds or loads, large transients, etc.), the ratio χ is *much* smaller. This justifies in most situations the time scale separation introduced above and, consequently, the assumption that in most situations engines can be well approximated as discrete-time dynamic systems.

Note that the non-dimensional parameter χ is the ratio of two types of energies. The numerator is proportional to a fictitious pneumatic energy stored in the manifold, and the numerator is proportional to the kinetic energy of the engine's flywheel. Therefore, the simplification of the engine as a discrete-time system is based on the fact that, in most situations, much more energy is stored in the engine's flywheel than in its intake manifold.

If the parameters of the engine system are such that the value of χ is not substantially smaller than 1, the system (2.129) has to be analyzed without adopting the simplifying assumption used in deriving (2.130) and (2.131). Linearizing the equation (2.129) at the equilibrium $\{p_0, A_0(\omega_0)\}$ defined by

[36] For a 50 kW SI engine the following values are reasonable estimates $p_{me0} \approx 10^5$ Pa, $V_m \approx 2 \cdot 10^{-3}$ m^3, $\Theta_e \approx 0.2$ kg/m^2 and $\omega_{e0} \approx 100$ rad/s.

$$p_0 = \frac{k_4}{k_3}, \quad A_0 = \frac{k_2 \cdot k_4}{k_1 \cdot k_3} \cdot \omega_0 \tag{2.135}$$

yields a second-order linear system whose state-space description (A.34) includes the following matrix A

$$A = \begin{bmatrix} -k_2 \cdot \omega_0 & -k_2 \cdot k_4/k_3 \\ k_3 & 0 \end{bmatrix} \tag{2.136}$$

The two eigenvalues of this matrix are

$$\lambda_{1,2} = \frac{k_2 \cdot \omega_0}{2} \cdot \left(-1 \pm \sqrt{1 - \frac{4 \cdot k_4}{k_2 \cdot \omega_0^2}} \right) \tag{2.137}$$

For a given engine, i.e., for fixed k_2 and k_4, and for increasing nominal speeds ω_0 these two eigenvalues will converge to the negative inverse of the time constants defined in (2.133)

$$\lambda_{1,2} \rightarrow \{-1/\widetilde{\tau}_p, \ -1/\widetilde{\tau}_\omega\} \tag{2.138}$$

Note that the eigenvalues (2.137) are real only if the following condition is satisfied

$$\chi < \frac{1}{4} \tag{2.139}$$

If the parameters $\{p_{me0}, V_m, \Theta_e, \omega_{e0}\}$ are such that χ is larger than 0.25, oscillations in the engine speed and manifold pressure must be expected.[37] Note that in this case the constant speed assumption is not well satisfied. In such situations, approaches similar to the one proposed in [19] will be useful.

2.6 Thermal Systems

2.6.1 Introduction

The thermal behavior inside the cylinder is, of course, of paramount importance for the engine's performance. In steady state, these effects are already taken into consideration in the engine maps discussed above. The thermodynamic boundary conditions, however, evolve much slower such that, from an engine system point of view, these effects must be modeled as dynamic processes [10]. Only with such a model will a precise control of the engine's thermal behavior be possible (see Chapter 4 and [114], [103]).

To be more specific, in this section, the following effects are discussed in more detail:

[37] Due to the several simplifications adopted when deriving the equations (2.129), the onset of an oscillatory behavior can be observed for values of χ that are smaller than 0.25.

- enthalpy present in the engine's exhaust gases;
- thermal behavior of the exhaust manifold;
- simplified engine thermal behavior (one reservoir); and
- detailed engine thermal behavior (several reservoirs and delays).

2.6.2 Engine Exhaust Gas Enthalpy

The engine's exhaust gases carry a substantial part of the fuel's energy. This enthalpy is used in turbocharged engines for boosting and in catalytic converters to maintain the necessary temperature levels. In future engine systems this enthalpy will become even more valuable such that heat exchangers will be used to partially recover some of this enthalpy for various purposes [98].

The temperature of the exhaust gas is primarily a function of the air/fuel ratio and of the operating point's indicated efficiency. Diesel engine exhausts are, therefore, usually much colder (less than 900 K) than those of stoichiometrically operated SI engines (up to 1300 K). Since several effects (blowdown at exhaust valve opening, heat transfer to the engine block, etc.) influence the exhaust gas temperature, it is quite difficult to accurately predict this variable. Therefore, a combination of measurements (made with standard control settings, say MBT ignition in SI engines) and correcting functions (that model the influence of deviations in control settings) are often used. A different approach would be to carry out a complete thermodynamic cycle simulation which, of course, would be computationally much more demanding.

Coming back to the description of the nominal engine-out temperature, for an SI engine a measured map has a form similar to that shown in Fig. 2.37. At very high power levels, cooling is achieved by running the engine in rich conditions, as can be seen in Fig. 2.37 in the upper right-hand corner (close to wide-open throttle). Neglecting these deviations, a qualitative approach would be to assume the engine-out temperature to be an affine function of the engine's mechanical power output.

More precise modeling can be achieved by analyzing the engine-out enthalpy. Knowing the temperature and the mass flow of the exhaust gas, the engine-out enthalpy can be computed since in a first approximation, despite the relatively high temperatures, the exhaust gases can be assumed to be a perfect gas, i.e., that the enthalpy flow is well approximated by

$$\dot{H}_{eg}(t) = c_p \cdot \dot{m}_e(t) \cdot \vartheta_{eg}(t) \tag{2.140}$$

with a constant specific heat c_p.

In order to formulate an engine-out enthalpy model, it is useful to compute the specific per-cycle enthalpy difference over the engine divided by the displacement volume V_d, $\frac{\Delta h_{eg}(n,p_{me})}{V_d}$, in each operating point. Figure 2.38 shows this variable for the same engine of Fig. 2.37. In this example, $\Delta h_{eg}(n, p_{me})$ can be approximated accurately (the maximum errors in the stoichiometric part are less than 10%) by an affine approach, i.e.,

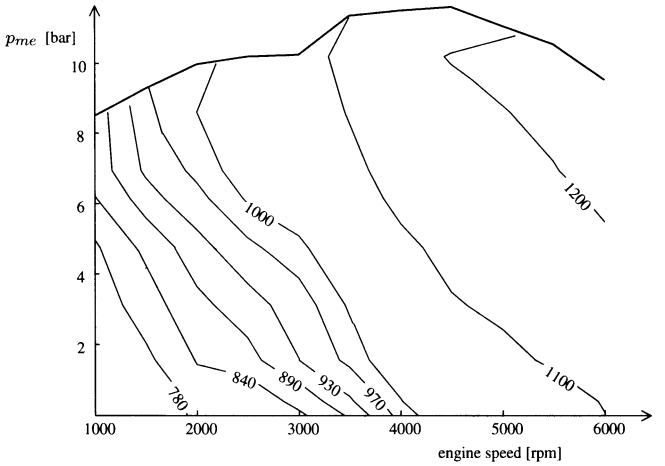


Fig. 2.37. Exhaust gas temperatures ϑ_{eg} (degrees K) for a 3-liter SI engine.

$$\Delta h_{eg}(t) = h_{eg,\omega} \cdot \omega_e(t) + h_{eg,p} \cdot p_{me}(t) + h_{eg,0} \tag{2.141}$$

where only three constants ($h_{eg,\omega}$, $h_{eg,p}$, $h_{eg,0}$) must be fitted to the engine to be modeled. This simplification is applicable to other engine classes (see for instance Fig. 2.43), and simply reflects the observation that the heat transfer to the cylinder wall is essentially a function of engine load [81], [70].

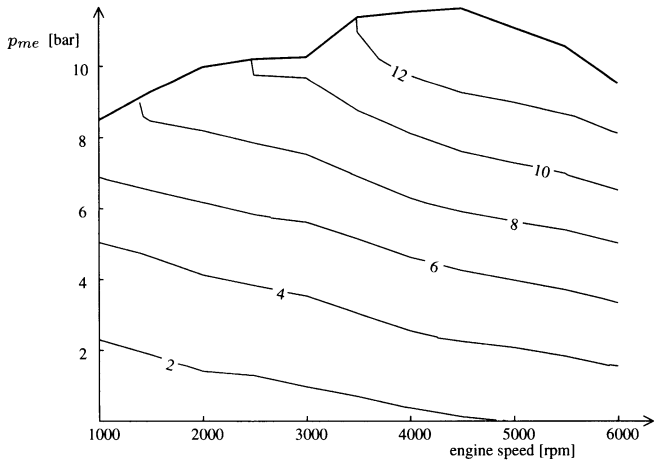


Fig. 2.38. Exhaust gas specific enthalpy $\frac{\Delta h_{eg}(n,p_{me})}{V_d}$ in [bar] for a 3-liter SI engine.

As mentioned above, deviations from the nominal behavior caused by non-nominal control settings can often be modeled by algebraic correction functions. An example of such an approach is the modeling of the influence of

the ignition angle on the exhaust gas temperature. The nominal engine-out temperature is described by a map $\vartheta_{eg,0}(\omega_e, p_m)$ similar to that shown in Fig. 2.37. The temperature increase $\Delta\vartheta_{eg}(\zeta)$ is related to (2.115). In fact, using the first law of thermodynamics and assuming the engine to be a Willans machine, the following approximation can be derived

$$\Delta\vartheta_{eg}(\zeta) \approx k_e \cdot \frac{H_m}{c_p} \cdot e_\omega(\omega_e) \cdot k_\zeta \cdot [\zeta - \zeta_0(\omega_e, p_{me})]^2 \tag{2.142}$$

Here, H_m is the mixture's lower heating value,[38] c_p is the specific heat capacity of the exhaust gas, $k_e \approx 0.5$ is a factor that quantifies how much of the available heat flux is diverted to the exhaust gas,[39] and ζ_0 and k_ζ are parameters that have been defined in (2.115).

2.6.3 Thermal Model of the Exhaust Manifold

The temperatures of the gases in the exhaust manifold can be computed using an approach similar to the one introduced in (2.3) and (2.4). The system to be analyzed is shown in Fig. 2.39. The in- and out-flowing exhaust gas masses are given by the engine (see (2.19)) and the downstream elements (modeled for instance as throttles (2.10)), respectively. The only change needed is the addition of an extra term in the enthalpy losses, i.e.,

$$\dot{H}_{out}(t) = \dot{m}_o(t) \cdot c_p \cdot \vartheta_o(t) + \dot{Q}_{in}(t) \tag{2.143}$$

The heat flow term $\dot{Q}_{in}(t)$ models the heat transfer from the exhaust gases to the exhaust manifold walls. It is assumed that this manifold has a mass m_w, a homogenous temperature ϑ_w, and a constant specific heat c_w, such that a simple enthalpy balance yields the differential equation for the mean wall temperature

$$m_w \cdot c_w \cdot \frac{d}{dt}\vartheta_w(t) = \dot{Q}_{in}(t) - \dot{Q}_{out}(t) \tag{2.144}$$

The heat flows $\dot{Q}_{in}(t)$ to and $\dot{Q}_{out}(t)$ from the manifold walls contain convection and radiation terms. A detailed modeling of all of these heat transfer terms is beyond the scope of this text. Interested readers can find more information in [22] or in [135]. In the following, just the most basic ideas will be introduced briefly such that the subsequent modeling examples can be followed.

The radiation terms are proportional to the difference in the fourth power of the temperatures. Assuming the ambient of the exhaust manifold to be

[38] The heating value of the mixture is equal to $H_m = H_l/(\lambda\sigma_0 + 1)$ where H_l is the fuel's lower heating value and σ_0 the fuel's stoichiometric constant. For a gasoline/air mixture at $\lambda = 1$ the heating value of the mixture H_m is equal to 2.7 MJ/kg.

[39] The remaining heat flux is carried away by the cooling water.

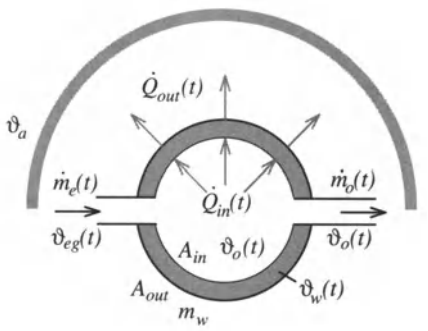


Fig. 2.39. Lumped parameter model of the thermal behavior of a receiver.

at a temperature ϑ_a that is much smaller than the exhaust manifold wall temperature ϑ_w, this heat flux can be approximated by

$$\dot{Q}_{out,rad}(t) \approx k_{out} \cdot A_{out} \cdot \vartheta_w^4(t) \tag{2.145}$$

The radiation from the ambient to the exhaust manifold and the radiation from the exhaust gas to the manifold walls and vice versa are neglected in this approximation.

The convection terms are linear in the temperature difference, i.e.,

$$\begin{aligned} \dot{Q}_{in,con}(t) &= \alpha_{in} \cdot A_{in} \cdot (\vartheta_o(t) - \vartheta_w(t)) \\ \dot{Q}_{out,con}(t) &= \alpha_{out} \cdot A_{out} \cdot (\vartheta_w(t) - \vartheta_a(t)) \end{aligned} \tag{2.146}$$

where the heat transfer coefficients depend upon the flow conditions. A simple approximation of the heat transfer coefficients valid for those conditions encountered in engine systems is given by the following expression

$$\alpha_{in,out} = \begin{cases} 28.6 + 4 \cdot v_g & \text{for } v_g < 5\ m/s \\ 21 \cdot v_g^{0.52} & \text{for } v_g \geq 5\ m/s \end{cases} \tag{2.147}$$

where v_g is a representative flow velocity at the interfaces between the exhaust gas and the inner manifold walls and the outer manifold walls and the ambient, respectively. This approximation has been derived from published approaches (see e.g. [83], [111]). The numerical values of the parameters have been adapted to the standard SI engine exhaust manifold conditions.

2.6.4 Simplified Thermal Model

In this subsection, a first, simple model of the engine's thermal behavior is introduced, which can be applied, for instance, in combination with the curve

shown in Fig. 2.30 to model temperature-dependent engine friction. The starting point is to treat the engine (engine block, cooling water, oil, etc.) in a simplified approach as one single thermal lumped-parameter element. An enthalpy balance over the engine then yields

$$m_e \cdot c_e \cdot \frac{d}{dt}\vartheta_e(t) = \dot{H}_w(t) - \alpha \cdot A \cdot [\vartheta_e(t) - \vartheta_a] \tag{2.148}$$

where m_e and c_e are the engine's mass and average specific heat capacity, $\dot{H}_w(t)$ represents the heat transfer to the engine block, α the (convective) heat transfer coefficient, A the active heat-exchange area and ϑ_a the ambient air temperature.

The variable $\dot{H}_w(t)$ can be derived using a first-law argument. In fact, once the total fuel enthalpy entering the system (2.65), the engine's actual efficiency $\eta_e(t)$ (using one of the several approaches introduced in Sect. 2.5.1), and the enthalpy exiting the system with the exhaust gases (2.140) are known, the missing information is derived as follows

$$\dot{H}_w(t) = (1 - \eta_e(t)) \cdot H_l \cdot \dot{m}_\varphi(t) - \dot{H}_{eg}(t) \tag{2.149}$$

Although this model is, of course, quite crude (simplified heat transfer to the environment, neglected temperature influence on wall heat transfer, etc.), it yields a useful first estimate of the main thermal effects. For instance, it can be used to estimate the engine's thermal time constant

$$\tau_e = \frac{m_e \cdot c_e}{\alpha \cdot A} \tag{2.150}$$

that is in the order of several thousand seconds.

An active cooling circuit (radiator, water pump, by-pass valve, see below) can be included quite easily in this formulation by modeling the radiating surface to be a function $A(u)$ of a control signal u that determines how much of the engine cooling liquid flows through the radiator.

2.6.5 Detailed Thermal Model

Introduction

Following the work described in [38], a more detailed model of the engine's thermal behavior including the external heat radiation equipment is developed in this section. Several other publications analyze this problem, see, for instance, [10] or [158]. In this section, both the external circuit bypass valve and the water pump speed can be used as control quantities, and the engine-in and engine-out water temperature are measured variables (see Fig. 2.40). The system shown in Fig. 2.40 may be subdivided into two main blocks for analysis, as shown in Fig. 2.41. The inner structure of the block, termed *internal cooling circuit* in Fig. 2.41, is detailed in Fig. 2.42. The system is approximated using

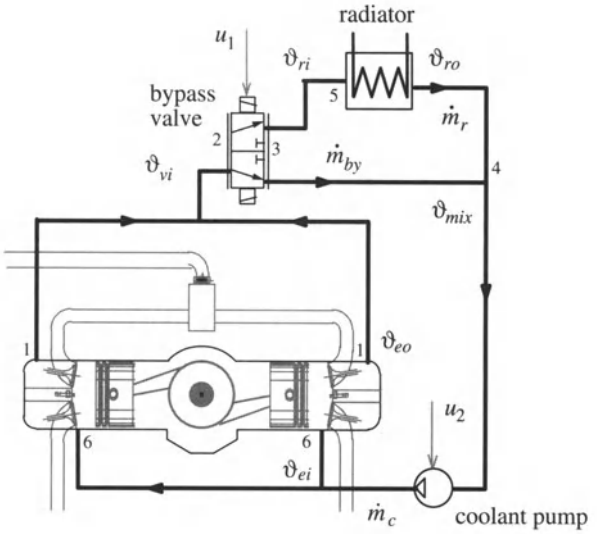


Fig. 2.40. Engine cooling system layout.

a lumped parameter approach, i.e., the various temperature distributions are assumed to be represented by a finite number of characteristic temperatures.

The following variables have been used in Fig. 2.40 and Fig. 2.41:

u_1 The ratio between the total coolant mass flow $\dot{m}_c$ and that mass flow $\dot{m}_{by}$ that bypasses the radiator.

$\dot{m}_r$ the mass flow through the radiator, i.e., $\dot{m}_r = \dot{m}_c - \dot{m}_{by}$.

u_2 The water pump speed, assumed to be proportional to the cooling $\dot{m}_c$.

z_1 The heat flux $\dot{Q}_{g,w}$ from the combustion chamber to the cylinder walls (see Fig. 2.42).

z_2 The heat flux to the engine block that is caused by engine friction $\dot{Q}_{if}$ minus the heat flux $\dot{Q}_{eb,a}$, which models the heat losses of the engine block to its environment.

z_3 The heat flux from the radiator to the environment that depends on vehicle speed, ambient temperature, etc.

$\vartheta_{...}$ Several cooling water temperatures.

Using these definitions, the mass flows are static multilinear functions of the inputs, i.e.,

$$\begin{aligned} \dot{m}_r(t) &= [1 - u_1(t)] \cdot u_2(t) \\ \dot{m}_{by}(t) &= u_1(t) \cdot u_2(t)\,. \end{aligned} \tag{2.151}$$

Modeling of the Internal Component

The model as illustrated in Fig. 2.42 comprises three state variables, namely:

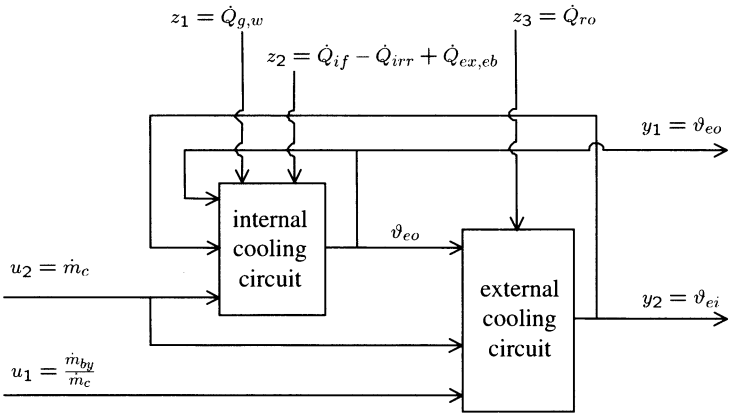


Fig. 2.41. Inputs, outputs and main blocks of the thermal engine model.

ϑ_w the cylinder wall temperature;
ϑ_c the cooling water temperature; and
ϑ_{eb} the engine block temperature (including the engine oil).

For each state, an enthalpy balance can be formulated:[40] for the wall

$$\frac{d}{dt}\vartheta_w(t) = \frac{1}{c_w \cdot m_w} \cdot [z_1(t) - \dot{Q}_{w,c}(t)] \tag{2.152}$$

for the coolant[41]

$$\frac{d}{dt}\vartheta_{eo}(t) = \frac{1}{c_c m_c}\left[c_c \dot{m}_c\{\vartheta_{ei}(t) - \vartheta_{eo}(t)\} + \dot{Q}_{w,c}(t) - \dot{Q}_{c,eb}(t)\right] \tag{2.153}$$

and, assuming $\vartheta_{eb} \approx \vartheta_{oil}$, for the engine block

$$\frac{d}{dt}\vartheta_{eb}(t) = \frac{1}{c_{eb} \cdot m_{eb} + c_{oil} \cdot m_{oil}} \cdot \left[z_2(t) + \dot{Q}_{c,eb}(t)\right] \tag{2.154}$$

The specific heat capacities and masses used in (2.152) to (2.154) are attributed to the following sub blocks:

$c_w \cdot m_w$: the cylinder walls;
$c_c \cdot m_c$: the cooling water inside the engine;
$c_{eb} \cdot m_{eb}$: the engine block; and
$c_{oil} \cdot m_{oil}$: the engine oil.

These parameters must be experimentally identified in order for the engine to be analyzed.

[40] There are no mass storage effects in this case, i.e., all fluids are assumed to be incompressible.

[41] As usual, the cooling water is assumed to be perfectly mixed, and, therefore, the output temperature coincides with the internal temperature.

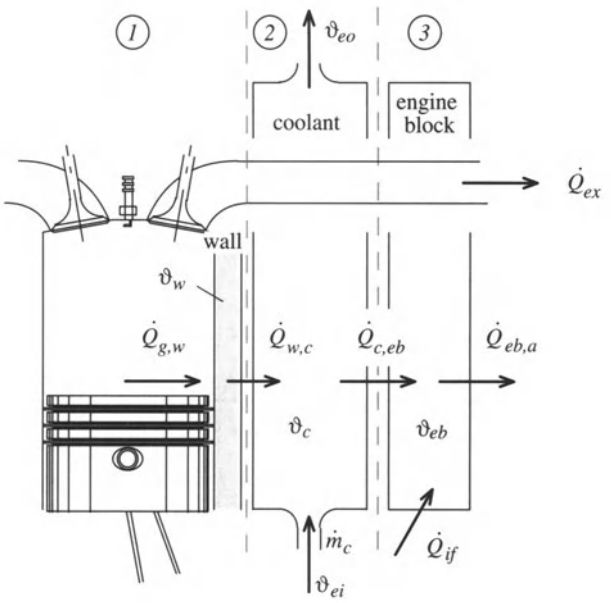


Fig. 2.42. Main heat flows inside the engine.

The heat flux $\dot{Q}_{g,w}$ from the combustion chamber to the cylinder wall is a function of several engine parameters, the most important being load p_{me}, speed ω_e and temperature of the cylinder wall ϑ_w. In general, this heat flux is very difficult to measure, but has to be estimated using a thermodynamic process simulation. For the engine described in [67], the results shown in Figs. 2.43 and 2.44 were obtained in [38]. These calculations can be carried out beforehand and the results can be stored in appropriate maps for later use in system simulation and controller design.

A further simplification is indicated in Fig. 2.44. In this diagram, it can be observed that the cylinder wall temperature influences the heat transfer from the cylinder to the wall almost linearly. Thus, this influence can be included easily as a correcting factor to the nominal case shown in Fig. 2.43.

The two internal heat fluxes $\dot{Q}_{w,c}$ and $\dot{Q}_{c,eb}$ are modeled, as usual, using the following formulation

$$\dot{Q}_{w,c}(t) = \alpha_c \cdot A_c \cdot \left[\vartheta_w(t) - \frac{\vartheta_{eo}(t) + \vartheta_{ei}(t)}{2}\right] \tag{2.155}$$

and

$$\dot{Q}_{c,eb}(t) = \alpha_{eb} \cdot A_{eb} \cdot \left[\frac{\vartheta_{eo}(t) + \vartheta_{ei}(t)}{2} - \vartheta_{eb}(t)\right] \tag{2.156}$$

where the heat transfer coefficients α_c and α_{eb} can be estimated using (2.147) (for other approaches see [38]) or experimental data. The contact areas, A_c

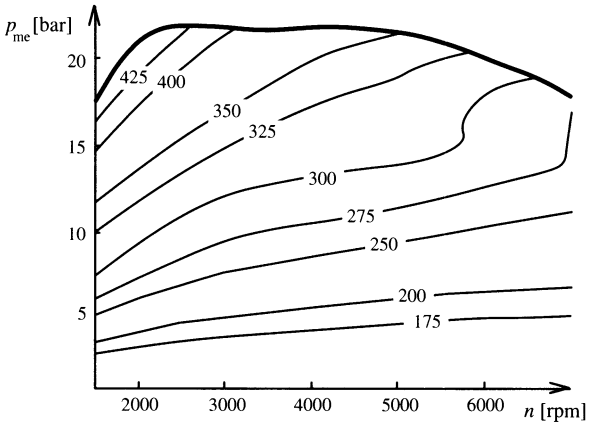


Fig. 2.43. Heat flux from the cylinder to the wall (J/cycle) at nominal wall temperature of the engine described in [38]. Engine at 2000 *rpm*.

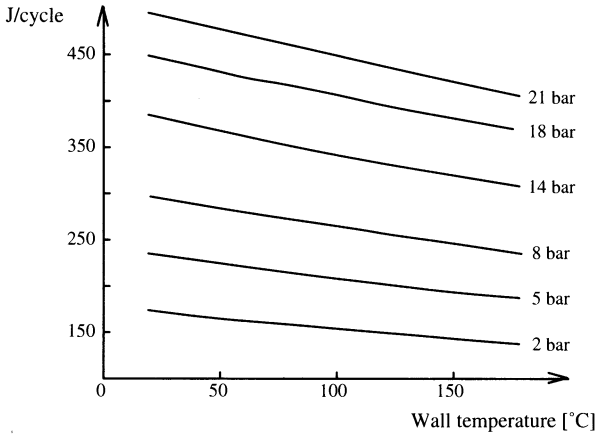


Fig. 2.44. Influence of the cylinder wall temperature on the heat flux from the cylinder to the wall (J/cycle).

between the cylinders and the cooling water, and A_{eb} between the cooling water and the engine block, respectively, can be inferred from the geometrical data of the engine.

The heat fluxes that are due to engine friction and to losses to the ambient are the last two unknown variables. The friction term can be derived using (2.111) as

$$\dot{Q}_{if}(t) = p_{me0f}(t) \cdot V_d \cdot \frac{\omega_e(t)}{4\,\pi} \tag{2.157}$$

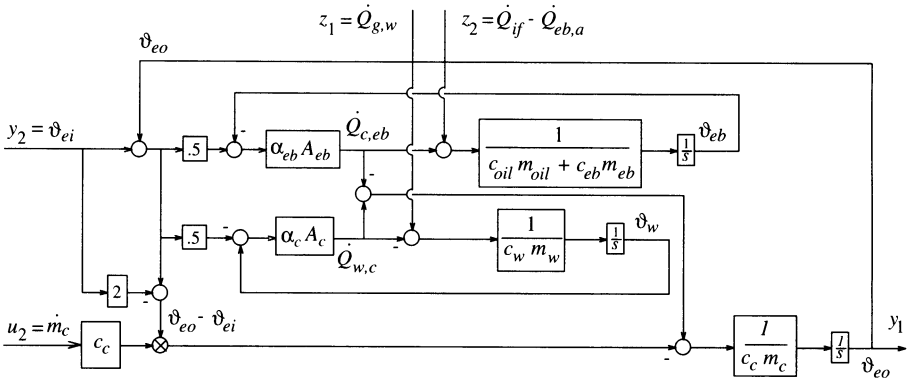


Fig. 2.45. Detailed structure of the engine thermal model.

In (2.157), it is assumed that all of the generated friction heat enters the engine block.

The heat losses to the environment can be modeled using the approach (2.146)

$$\dot{Q}_{eb,a}(t) = A_{eb} \cdot \alpha_{eb,a} \cdot \{\vartheta_{eb}(t) - \vartheta_a(t)\} \tag{2.158}$$

The heat transfer coefficient $\alpha_{eb,a}$ can be estimated using (2.147) or using experimental data. This concludes the modeling of the engine part. The resulting block diagram is shown in Fig. 2.45.

Model of the External Components

The external components (radiator, piping, etc.) has a simpler structure and its dynamics are mainly determined by *time delays*. All connecting pipes are assumed to be adiabatic and only the radiator exchanges heat with the environment, leading to

$$\frac{d}{dt}\vartheta_{ro}(t) = \frac{1}{c_c \cdot m_r} \cdot [\dot{Q}_{ri}(t) - \dot{Q}_{ro}(t)] \tag{2.159}$$

The heat flux to the radiator is given by

$$\dot{Q}_{ri}(t) = c_c \cdot \dot{m}_r(t) \cdot [\vartheta_{ri}(t) - \vartheta_{ro}(t)] \tag{2.160}$$

where the mass flow $\dot{m}_r$ is a function of the inputs, see (2.151). The heat flux from the radiator to the ambient strongly depends upon vehicle speed v_v and ambient temperature ϑ_a

$$\dot{Q}_{ro}(t) = \alpha_r(v_v(t)) \cdot A_r \cdot \left[\frac{1}{2}(\vartheta_{ri}(t) + \vartheta_{ro}(t)) - \vartheta_a(t)\right] \tag{2.161}$$

The heat transfer coefficient can be modeled using (2.147) or with more detailed approaches, see e.g., [38] or [21].

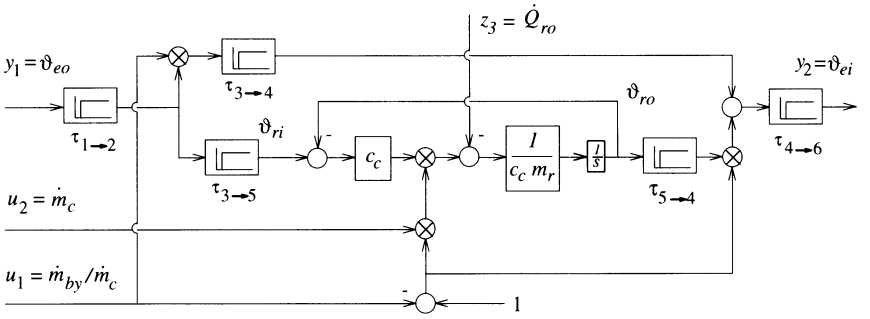


Fig. 2.46. Detailed structure of the model of the external cooling circuit.

From a control engineering point of view, the main dynamic effects are caused by the finite flow velocities in the system's pipes. The flow is assumed to be a perfect *plug flow*, i.e., the time delay $\tau_{i\to j}$ between the junctions i and j is defined by the implicit equation

$$\int_{t-\tau_{i\to j}}^{t} v_{i\to j}(\sigma)\, d\sigma \;=\; l_{i\to j} \tag{2.162}$$

where $v_{i\to j}$ is the flow velocity (assuming constant densities it is proportional to the mass flow) and $l_{i\to j}$ is the length of the piping between the junctions i and j as shown in Fig. 2.40.

Assuming that the velocities (control inputs) change slowly compared to the magnitude of the time delays, the variable $\tau_{i\to j}$ can be approximated by

$$\tau_{i\to j} \approx \frac{l_{i\to j}}{v_{i\to j}} \tag{2.163}$$

Perfect mixing is assumed at junction 4; therefore, the temperature of the water flowing out of that junction can be found using the following relation

$$\vartheta_{mix}(t) = \frac{1}{\dot{m}_c(t)} \cdot [\dot{m}_r(t) \cdot \vartheta_{ro}(t - \tau_{5\to4}) + \dot{m}_{by}(t) \cdot \vartheta_{eo}(t - \tau_{1\to4})] \tag{2.164}$$

The resulting detailed block diagram, which corresponds to the *external cooling circuit* block in Fig. 2.41, is shown in Fig. 2.46 (the junction labels used in this illustration were introduced in Fig. 2.40).

Figure 2.47 shows a comparison of a measured and a simulated step response. As can be seen in that illustration, the model presented above is able to describe the dynamic effects that are relevant for design and optimization of an engine-in and engine-out temperature controller.

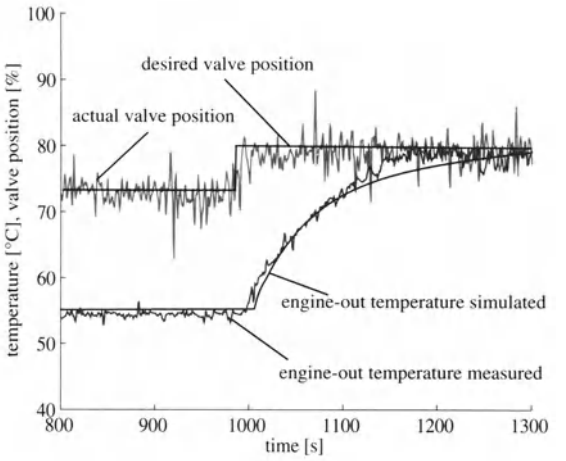


Fig. 2.47. System and model response to a step change in by-pass valve position.

2.7 Pollutant Formation

2.7.1 Introduction

Pollutant formation in internal combustion engines is rather difficult to predict theoretically or by numerical simulation. This is primarily because these phenomena are governed by the detailed spatial and temporal distribution of the mixture composition, temperature and pressure inside the combustion chamber.

Control-oriented models, therefore, often rely on the results of experiments which are summarized in appropriate maps. Once these maps are available, slowly varying state variables (engine speed, manifold pressure, EGR rate, etc.) and accessible control variables (injection pressure, ignition timing, etc.) are used to derive the corresponding engine-out pollution values.

The specific form of such maps is not unique. Starting with two-dimensional graphic representations, other forms like polynomial regressions, neural nets, etc. have been proposed in the literature, see, for instance, [11], [14], [186].

All these approaches rely on the fundamental assumption that the pollution formation process depends, in a deterministic way, on the control inputs and on the thermodynamic boundary conditions mentioned above. Of course, this determinism is not entirely true in real situations, such that only average estimations of engine-out pollutant concentration levels are possible. Moreover, aging and other, not easily modeled effects, cause large deviations in pollutant formation such that, in general, prediction errors are quite large. Control-oriented engine-out pollution models are, therefore, often used as a means to predict the *relative* impact of a specific controller structure or algorithm on pollutant emission.

In this section, the most important pollutant formation mechanisms are discussed using *qualitative* arguments. Both SI and Diesel engines are analyzed starting with a discussion of the air/fuel ratio needed for a stoichiometric combustion. As an example the last part of this section shows a *quantitative* COM of the NO_x formation in homogeneous-charge SI engines.

2.7.2 Stoichiometric Combustion

While the actual reactions are much more complicated and involve a large number of intermediate species, the combustion of hydrocarbon H_aC_b occurs according to the following overall chemical reaction:

$$1 \text{ mol } H_a\,C_b + c \text{ mol air} \rightarrow d \text{ mol } H_2O + e \text{ mol } CO_2 + f \text{ mol } N_2 \quad (2.165)$$

Assuming the ambient air to consist of 79% N_2 and 21% O_2, the following relations can be derived for a stoichiometric combustion

$$d = \frac{a}{2}, \quad e = b, \quad f = 3.76\,(b + \frac{a}{4}), \quad c = 4.76\,(b + \frac{a}{4}) \quad (2.166)$$

Accordingly, the stoichiometric mass of air needed to burn 1 kg of a hydrocarbon fuel $H_a\,C_b$ is given by the expression

$$m_{air,stoich} = \frac{137\ (1 + y/4)}{12 + y}\ kg \quad (2.167)$$

where $y = a/b$ is the fuel's hydrogen-to-carbon ratio.

Figure 2.48 shows the form of function (2.167) and the specific values for three important fuels (the limit values are 11.4 kg of air for pure carbon and 34.3 kg of air for pure hydrogen).

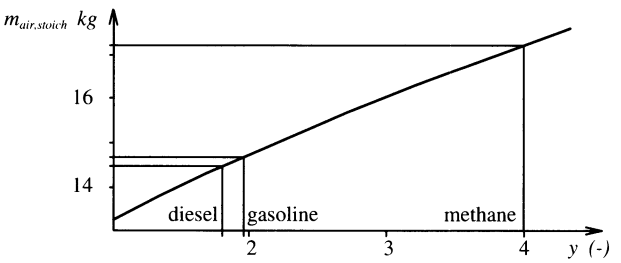


Fig. 2.48. Stoichiometric mass of air required to burn 1 kg of hydrocarbon fuel $H_a\,C_b$, with $y = a/b$ being the fuel's hydrogen-to-carbon ratio.

The complete combustion assumed in (2.165) is not attained even in perfect thermodynamic equilibrium conditions. In reality, traces (a few hundred to several tens of thousands of ppm) of pollutant species are always emitted. The main pollutant species are:

- products of an incomplete oxidation: carbon monoxide CO (with a concentration of typically one order of magnitude larger than those of any other pollutant in the exhaust gases) and hydrocarbon (either unburnt fuel or intermediate species) which — depending on the fuel — can occur as a simple or complex molecular species including cancer-causing components; and
- products of an unwanted oxidation: mostly nitric oxide, NO, but also some NO_2.

In diffusion-flame combustion systems (Diesel and stratified charge GDI engines) a substantial number of particulates is generated as well (sized from a few tens of nanometers to several hundred microns in size). In order to better understand the structural relations in pollution formation, the basic mechanisms are briefly reviewed below. For a more detailed discussion, see [81].

2.7.3 Pollutant Formation in SI Engines

Figure 2.49 shows typical pollutant concentrations in the exhaust gas of a port-injected SI engine as a function of the normalized air/fuel ratio λ. *Qualitative* explanations for the shape of the curves shown in this figure are given below.

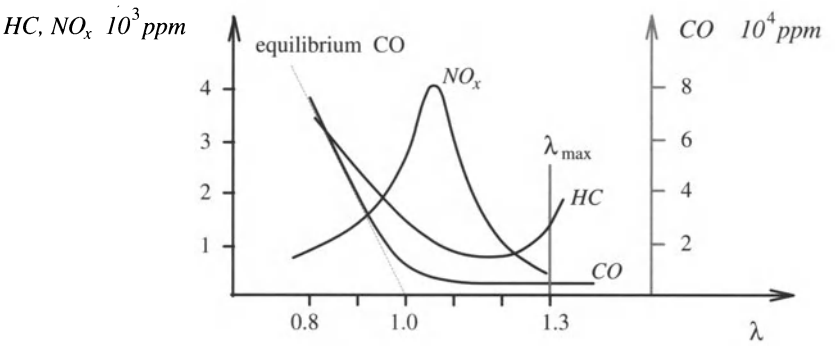


Fig. 2.49. Typical engine-out pollutant emission of a port-injected SI engine, λ_{max} is the lean combustion limit where misfires start (notice the different scales for CO and the other species).

Hydrocarbon and Carbon Monoxide

In rich conditions, the lack of oxygen produces very high HC and CO concentrations. When the engine is operated under extremely lean conditions, misfires start to occur which, of course, cause increased HC emission. The misfires also reduce in-cylinder and exhaust temperatures, which lead to smaller

post-combustion oxidation. Of course, CO does not increase due to misfires (when the mixture fails to ignite, no CO can be formed).

The reason that at intermediate air/fuel ratios the HC and CO concentrations do not reach the (almost) zero equilibrium levels can be explained as follows:

- Due to the cylinder's charge quick cooling during the expansion stroke, the necessary time required to reach chemical equilibrium is not available (high CO and HC concentrations are "frozen" shortly after their formation during combustion).
- During the compression stroke, the mixture is forced into several crevices present in the cylinder, and part of the mixture is absorbed into the oil film covering the cylinder walls (Fig. 2.50). As soon as the pressure decreases during the expansion stroke (and after combustion has terminated), these HC molecules are released into the cylinder. Since cylinder charge temperatures have already fallen substantially, a non-negligible portion of these HC molecules escapes post-flame oxidation.
- When the flame front reaches the proximity of the cylinder walls (a few 0.1 mm distance), large heat losses to the wall quench the flame before all the mixture has been oxidized.

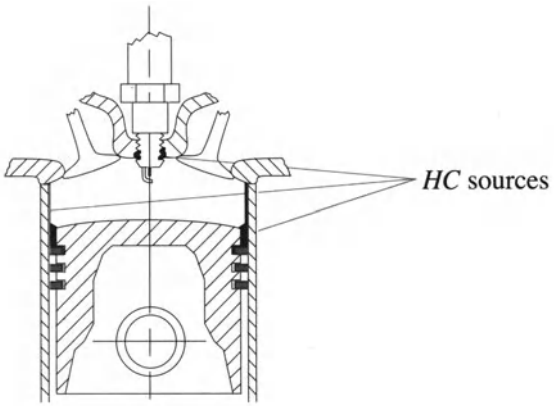


Fig. 2.50. Crevices and other sources of engine HC emissions.

Since a substantial part of the HC and CO is always oxidized in the post-flame phase inside the cylinder, and even during blow-down in the exhaust manifold, all effects that influence the conditions for this phenomenon (available oxygen, temperature, time, etc.) also affect the engine-out HC and CO emission.

The pollutant formation in direct-injection SI engines in homogeneous mode is essentially the same as in port-injection engines. In stratified mode, however, substantial differences appear (for instance, the late injection may

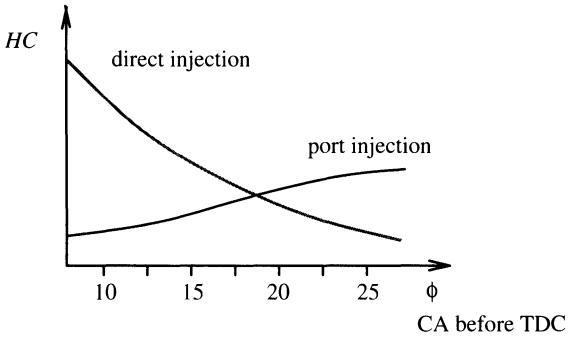


Fig. 2.51. Hydrocarbon emission of a port-injection and a direct-injection stratified-charge SI engine. The first engine operates at $\lambda = 1$ and at MBT ignition of 25 degrees BTDC. The second engine operates at $\lambda = 2$, see [166].

eliminate the crevice losses). Figure 2.51 shows that this does not necessarily lead to reduced pollutant levels. In this figure, HC emission are plotted as a function of ignition timing. While the port-injected engine has decreasing HC levels due to later ignition (the thermodynamic efficiency decreases, hence the burnt gases are hotter, which improves post-combustion oxidation), the direct-injection engine displays the inverse behavior. This shows that in direct-injection engines the ignition point may not be shifted without an appropriate change in injection parameters.

Note that in modern engines "blow-by" gases, i.e., that part of the mixture that enters the crank shaft compartment due to leakages of the piston rings, are collected and fed back to the combustion process.

Nitrogen Oxide

Engine-out NO_x emission (mostly NO, but also some NO_2) are formed during the peak-temperature phases, i.e., when the the burnt gas in the cylinder reaches temperatures of more than 2000 K. The three variables influencing that process most are *oxygen availability*, *cylinder pressure* and, most importantly, *burnt gas temperature*.

All of these parameters define the *equilibrium* concentration of NO shown in Fig. 2.52.[42] This figure clearly indicates that the gas temperature has a dominant influence on the NO concentration. The influence of the oxygen availability also seems strong, but Fig. 2.52 is somewhat misleading in that aspect. In fact, for a *constant* amount of fuel, a larger air/fuel ratio implies cooler burnt-gas temperatures and, hence, smaller NO concentrations. Finally, the influence of the gas pressure is noticeable, but clearly less important than the contributions of the previous two variables.

[42] The data and figures shown in this section have been taken from [27]

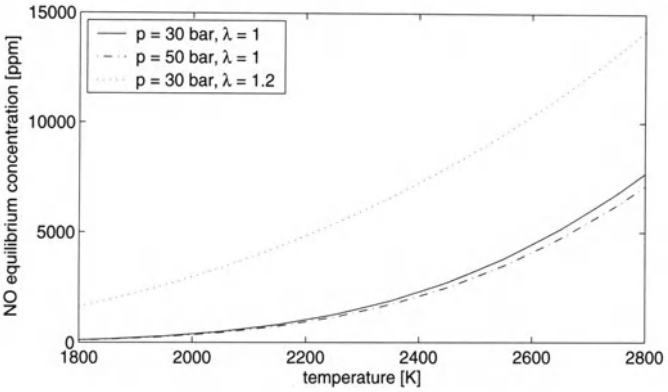


Fig. 2.52. Equilibrium NO concentration as a function of burnt-gas temperature (main parameter), oxygen availability and cylinder pressure.

The NO concentrations shown in Fig. 2.52 are reached under equilibrium conditions, i.e., after an infinitely long period. However, the reciprocating behavior of the engine limits the time available for these reactions to take place. Therefore, the *rate of formation* of NO has to be taken into account. The equilibrium conditions are reached only if the associated characteristic time constants τ_r of these reaction kinetics are much smaller than the time available during the expansion stroke (for a precise definition of τ_r see [81]). As shown in Fig. 2.53, this is only the case for temperatures much greater 2000 K, irrespective of the actual air/fuel ratios, cylinder pressures and engine speeds.

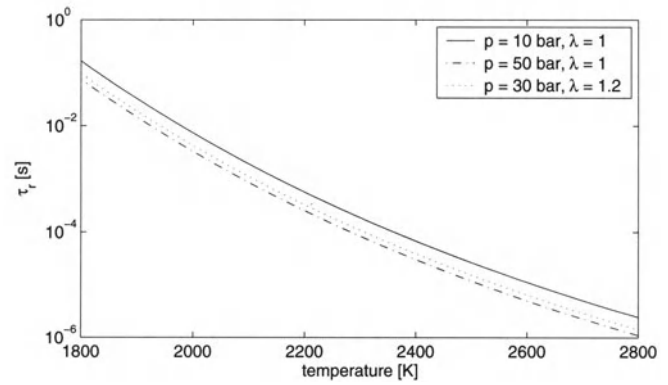


Fig. 2.53. Characteristic time constants of the NO formation rate in burnt gases.

This dependency of the reaction rate on the gas temperature is the main cause for the large NO concentrations in the exhaust gas. In fact, due to the short interval available in the expansion phase, the NO rapidly formed in

the high-temperature phases around TDC are not reduced to the equilibrium values that are expected at the cooler temperatures that prevail before the exhaust valve opens. As Fig. 2.53 shows, all reactions are essentially "frozen" below 2000 K.

An exact modeling of these phenomena requires a thermodynamic process simulation based on the energy conservation laws. The heat release due to the combustion, the piston work and the heat transfer through the cylinder walls has to be considered together with the NO formation laws in a crank-angle based formulation [81]. The results of such a model will be the transient temperature and NO concentration profiles as shown in Fig. 2.54. Unfortunately, the numerical solution of the system of ODE that describes that process requires substantial computational power such that real-time applications are not yet possible with standard ECUs.[43]

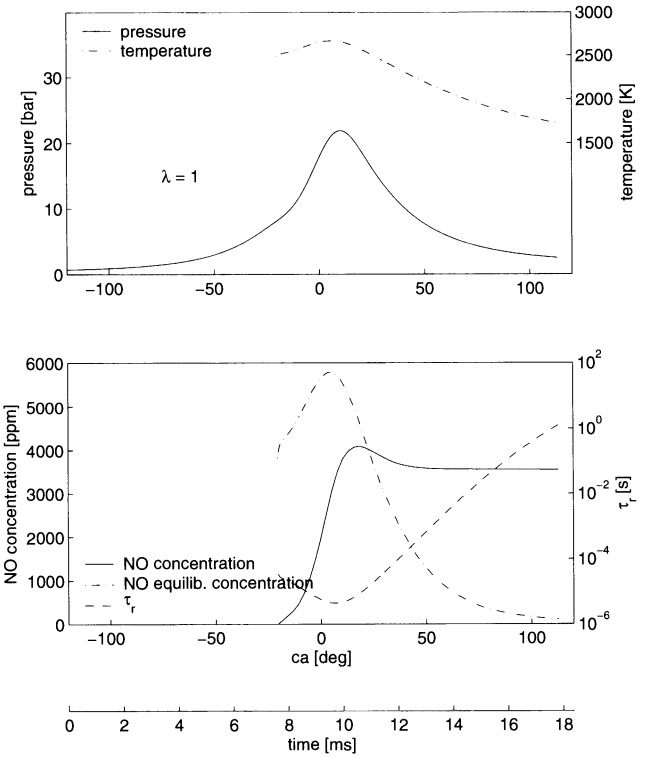


Fig. 2.54. Transient temperature and NO concentration in the burnt-gas zone. SI engine at $n = 2000$ rpm speed and $p_m = 0.5$ bar load.

[43] In [4], a simplified approach is presented that permits to avoid the explicit solution of the ODE.

The basic NO formation mechanisms explained above indicate that the engine-out NO emissions depend on almost all engine-system variables such as engine speed and load, air/fuel ratio, spark timing, etc. Since reaction rates depend exponentially on the temperature of the burnt gas, small changes in the engine control variables that affect this signal will have dramatic influences on the NO engine-out levels. The temperature of the burnt gases depends on many parameters. The highest temperatures are obtained for close-to-stoichiometric mixtures, but since oxygen availability is also important for NO formation, maximum NO concentrations values are usually obtained in slightly lean conditions (see Fig. 2.49).

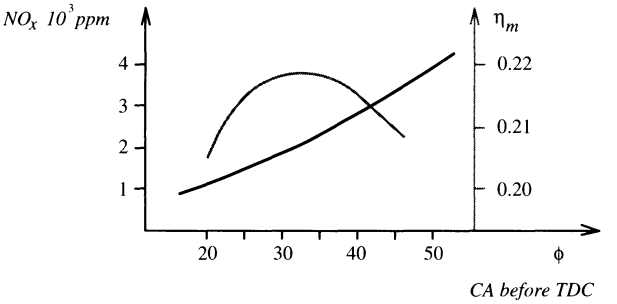


Fig. 2.55. Example of the engine-out NO emission and of the engine efficiency as a function of ignition timing (port-injected SI engine, medium load and speed).

Another important parameter is ignition timing. Shifting ignition timing toward early ignition typically improves engine efficiency. Unfortunately, this comes at the price of an increased NO formation rate, as shown in Fig. 2.55. Again, this is easy to see from the previous discussion. In fact, earlier ignition causes higher peak pressures and, hence, higher burnt-gas temperatures.

Finally, exhaust gas recirculation (EGR) has an important impact on the formation of NO as well. External EGR increases the mass of the cylinder charge and therefore acts as "thermal inertia." This additional mass reduces peak temperatures (especially if the external EGR has been cooled) and, hence, the formation of NO. External EGR is limited, however. In port-injected SI engines, this limit is approximately 15%; otherwise misfiring and unacceptably high levels of HC emission occur.

Summary

In summary, at least the following engine and control parameters affect the pollution formation of port-injected SI engines:

- Air/fuel ratio — most important parameter, strongly affects all pollutants.
- Engine speed — determines available reaction times and, hence, the formation and reduction of all pollutants.

- Engine load — determines peak and post-combustion oxidation temperatures and, hence, all pollutant species.
- External EGR rate — affects peak temperatures and, hence, especially NO_x formation.
- Ignition timing — determines peak temperatures and hence especially NO_x, but also has an impact on post-combustion temperatures and, hence, on HC.
- Engine temperature — influences the post-flame oxidation of CO, but also NO_x formation (this temperature is not the oil or the cooling water temperature, but a fictitious characteristic mean cylinder wall and head temperature).

In direct-injection engines operated in stratified charge mode, several other influencing parameters must be included, the most important being injection timing (start and end of the multiple-injection events). Note that in stratified mode the engines are operated at very lean air/fuel ratios — well beyond the ignition limits of homogeneous charge engines ($\lambda = 2$ and higher are possible). This, of course, leads to lower engine-out NO_x levels (the cooling effect dominates the higher availability of oxygen).

HC levels, however, can be much larger (by a factor of 2-3) and also particulate matter, which in homogeneous-charge SI engines is not an issue, is a problem in stratified-charge operation. Both effects are caused by the unstable flame structure at the very lean borders of the spray.

So far, warmed engines which are operated under stationary conditions have been assumed. Unfortunately, during cold start, engine-out emission may be much higher. This is especially the case for HC, which, due to the not-yet warmed cylinder and exhaust ports, are not oxidized after formation.

Additionally, deterioration of engine-out pollution levels is observed during transients. Without precise compensation of the involved dynamic variables (wall film mass, temperatures, speeds, ignition timing, etc.) unfavorable operating conditions can be imposed on the engine with the corresponding increase in pollution levels. These problems will be discussed in Chapter 3.

2.7.4 Pollutant Formation in Diesel Engines

Diesel engines have relatively low engine-out emissions. In particular, hydrocarbon and carbon monoxide can usually be neglected. The main pollutant species in the exhaust gas are nitrogen oxide NO_x and particulate matter (PM), see Fig. 2.56. As in SI engines, the air/fuel ratio is one primary factor that affects pollution formation. Since Diesel engines are load-controlled by air/fuel ratio variations, this parameter plays an even more important role here than for homogeneous-charge SI engines. Other important parameters are injection timing and pressure, as well as the EGR rate.

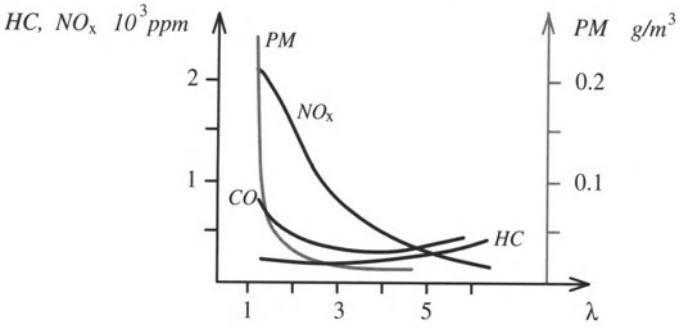


Fig. 2.56. Nitrogen oxide NO_x, hydrocarbon HC and particulate matter PM engine-out emission for a direct injection Diesel engine as a function of air/fuel ratio.

Figure 2.56 shows the typical relation between air/fuel ratio and engine-out pollutants. As Diesel engines are always operated under lean conditions,[44] CO and HC emission are relatively low. At very high air/fuel ratios, i.e., at very low loads, combustion is no longer complete (low temperatures, too lean local air/fuel mixture composition, etc.) and part of the fuel enters the cylinder too late to be burnt. At very high loads, regions with insufficient oxygen supply exist, which leads to increased CO emission.

At a first glance, Diesel engine NO_x emission levels are surprisingly high. In fact, the extrapolation of the NO_x levels of a homogeneous-charge SI engine to $\lambda \approx 2$ (see Fig. 2.49) predicts levels of less than 100 ppm. As Fig. 2.56 shows, the measured values are one order of magnitude higher than that.

This "contradiction" can be resolved by the following argument. As Fig. 2.57 shows, the diffusion combustion process in a direct-injection Diesel engine is, by its nature, inhomogeneous. Global air/fuel ratios are not representative of the local air/fuel ratio at which the actual combustion process takes place. Locally, very rich to pure air zones can be found. As a consequence, during inhomogeneous combustion, there will always be some regions where combustion takes place at conditions favorable for NO_x formation. Once this species has been formed, the charge cools down so rapidly that the reverse reaction becomes "frozen."

The inhomogeneity of the air/fuel mixture is also the primary cause of the formation o particulate matter. When the outer layer of a fuel droplet begins to burn, its core becomes so hot that cracking processes start. The shorter hydrocarbon chains produced by this cracking are not easily inflammable such that some escape combustion and appear at the exhaust port as soot particles. Typically, some hydrocarbon compounds are absorbed in these particles. The

[44] For the regeneration of lean NOx-trap catalytic converters, a stoichiometric-to-rich operation has been proposed, realized, for instance, by a late additional injection in common-rail systems.

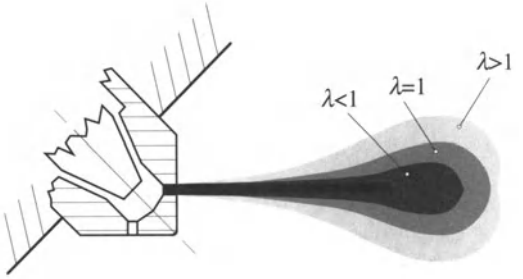


Fig. 2.57. Schematic representation of local air/fuel ratio distribution in a Diesel engine combustion chamber.

prediction of the actual PM emission is very difficult because a large amount is first formed and later burnt during the combustion process, see [150]. The actual engine-out emission is, therefore, the difference between two large numbers, such that small changes in these numbers can have substantial effects on the net difference.

2.7.5 Control-Oriented *NO* Model

In this section, an example of a COM of the NO emissions of an SI engine is described in some detail. The emphasis is not on the well-known basic NO formation laws that have been briefly described above, but on models useful for real-time applications where the computational burden must be limited to the bare minimum.

In this example, it is assumed that only four influencing parameters must be included in the COM, *viz.*, the manifold pressure p_m, the engine speed ω_e, the air/fuel ratio λ, and the ignition angle ζ. These four variables are mutually independent, and — choosing 20 support points for each variable results in $20^4 = 1.6 \cdot 10^5$ measurements necessary to find the complete four-dimensional "map" that describes the NO engine-out emissions. Assuming one minute per measurement, 110 consecutive days of engine testing would be required to obtain this data. Clearly, this straightforward approach is not feasible in practice.[45]

Instead, an approximation of the following form can be used to estimate the concentrations of the engine-out NO emissions

$$[NO](\omega_e, p_m, \lambda, \zeta) \approx [NO]_0(\omega_e, p_m, \lambda_0, \zeta_0) \cdot \tag{2.168}$$
$$(1 + \Delta_\lambda(\omega_e, p_m, \lambda - \lambda_0) + \Delta_\zeta(\omega_e, p_m, \zeta - \zeta_0))$$

where the *base map* $[NO]_0(\omega_e, p_m, \lambda_0, \zeta_0)$ is measured using the nominal values λ_0 and ζ_0. The two correcting functions $\Delta_\lambda(\omega_e, p_m, \lambda - \lambda_0)$ and

[45] In reality, more parameters such as engine temperature, EGR ratio, etc., must be included making such an approach even less likely to succeed.

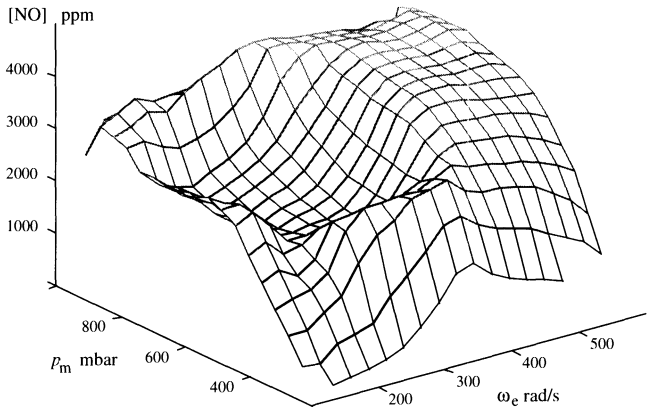


Fig. 2.58. Measured base map of engine-out NO emissions (3.2-liter V6 SI engine).

$\Delta_\zeta(\omega_e, p_m, \zeta - \zeta_0)$ are determined using process simulations as described above.[46] The advantage of measuring such a base map, an example of which is shown in Fig. 2.58, is that all subsequent computations can be validated at nominal operating points.

The structure of $[NO]_\lambda(\omega_e, p_m, \lambda - \lambda_0)$ and $[NO]_\zeta(\omega_e, p_m, \zeta - \zeta_0)$ can be derived using the underlying first laws of physics. For the air/fuel ratio, a parabolic function with its maximum at slightly lean conditions follows from the equilibrium values shown in Fig. 2.52 and the comments made in the adjacent text (see Fig. 2.59).

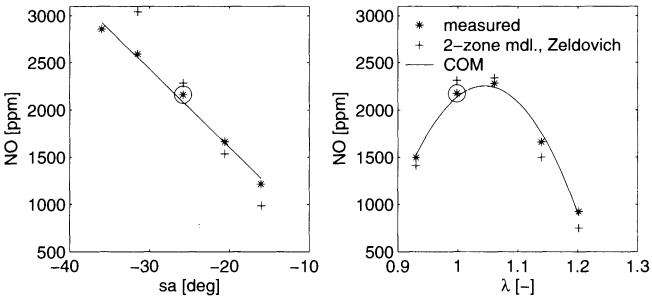


Fig. 2.59. Example of the the engine-out NO emission for variations in ignition timing (nominal value $\zeta_0 = -25°$) and air/fuel ratio (nominal value $\lambda_0 = 1$).

[46] Only 400 measurements are needed in this case. If, instead of a process simulation actual measurements are made, the number of experiments is $2 \cdot 20^3 = 1.6 \cdot 10^4$, i.e., one order of magnitude less than the full mapping approach described above.

Placing the ignition at an early angle increases the temperature during the combustion and, hence, the NO engine-out levels. Interestingly, as Fig. 2.59 shows, this increase is approximately linear rather than exponential as would be expected from the equilibrium values of Fig. 2.52 considering *peak* temperatures only. This apparent contradiction can be resolved with the illustration shown in Fig. 2.53. In fact, at sufficiently high temperatures the reaction time constants are so small that the NO levels in the burnt gases are close to their equilibrium value even during the first part of the expansion phase. However, in this phase the temperature of the burnt gases decreases and the equilibrium levels of NO follow that trend. Once the limit of approximately 2000 K is reached the reactions become so slow that almost no further NO reduction takes place. This results in the engine-out NO levels observed during the experiments.

These observations form the basis for the formulation of the structure of the correcting functions in (2.168)

$$\Delta_\lambda(\omega_e, p_m, \lambda - \lambda_0) = k_{\lambda,2} \cdot (\lambda - \lambda_0)^2 + k_{\lambda,1} \cdot (\lambda - \lambda_0) \tag{2.169}$$

$$\Delta_\zeta(\omega_e, p_m, \zeta - \zeta_0) = k_{\zeta,1} \cdot (\zeta - \zeta_0) \tag{2.170}$$

The coefficients $k_{\lambda,i}, k_{\zeta,i}$, which depend on the engine operating point, are chosen such that they minimize the error between the data obtained with the thermodynamic process simulation or measurements and the prediction of the COM (2.168). This minimization can be done using a least-squares approach, as shown in Appendix A.

Table 2.3 shows an example of the NO concentrations resulting from air/fuel ratio and spark advance variations. The symbol λ_- designates a richer mixture than λ_0, while λ_+ indicates a leaner mixture. Similarly, the symbol $\theta_{SA,-}$ indicates an earlier spark angle than $\theta_{SA,0}$, while $\theta_{SA,+}$ denotes a later spark angle. The concentrations in each field of Table 2.3 are shown in *ppm*. The first value in each field is the actual measured value. The second value is the NO concentration obtained using the detailed, but computationally demanding two-zone process simulation. Finally, the third value is calculated using the COM proposed in this section.

Since the measured value in the center field (2162 *ppm*) is used to calibrate both the detailed and the control-oriented model, the agreement shown in that field is to be expected. The values shown in the second row of the other eight fields are calculated using the two-zone model, without changing any of its calibration parameters. The measurements shown in the fields (1,2), (2,1), (2,3), and (3,2) are used to calibrate the additional parameters of the COM. Finally, the fields (1,1), (1,3), (3,1) and (3,3) contain the predictions obtained without changing the parameters of the COM model.

As Table 2.3 shows, neither the detailed model nor the COM is able to precisely predict the measured NO concentrations. In most cases, however, the simplified model yields the correct trend. As mentioned in the introduction to this section, this is all that can be expected from COM of the pollutant formation.

Table 2.3. Varying NO concentrations for air/fuel ratio and spark advance variations at engine speed of $\omega_e = 200\ rad/s$ and inlet manifold pressure of $p_m = 0.5\ bar$. Variations $\lambda_- \approx 0.9$, $\lambda_0 \approx 1$, $\lambda_+ \approx 1.1$, $\theta_{SA,-} \approx -34°$, $\theta_{SA,0} \approx -26°$, $\theta_{SA,+} = -18°$. The first value in each field is measured; the second value is calculated using a two-zone process simulation and the Zeldovich Mechanism for NO formation; and the third value is calculated using the control-oriented model (2.168).

	$\theta_{SA,-}$	$\theta_{SA,0}$	$\theta_{SA,+}$
λ_-	1378 ppm	1495 ppm	884 ppm
	1396 ppm	1409 ppm	620 ppm
	1677 ppm	1591 ppm	584 ppm
λ_0	2857 ppm	2162 ppm	1217 ppm
	3836 ppm	2285 ppm	985 ppm
	3018 ppm	2162 ppm	1362 ppm
λ_+	2773 ppm	1661 ppm	1044 ppm
	2948 ppm	1500 ppm	872 ppm
	2629 ppm	1830 ppm	1487 ppm

Note that (2.168) quantifies the *concentration* of the pollutants in the exhaust gas, i.e., the number of NO molecules relative to the total number of parts in the exhaust (usually mol/mol). However, most legislations limit the total *mass* of each pollutant species emitted during the corresponding test cycles.

To obtain the latter the total engine mass flow, (2.19) has to be combined wit the NO_x concentration. For small NO_x concentrations, the following approximation can be used to compute the total engine-out pollutant mass flow

$$\dot{m}_{NO_x}(t) = [NO_x](t) \cdot \frac{M_{NO_x}}{M_{eg}} \cdot \dot{m}(t) \tag{2.171}$$

where $\dot{m}(t)$ is the total engine mass flow (2.19), $[NO_x](t)$ the concentration of NO_x in the exhaust gas, and M_{eg} and M_{NO_x} are the molar mass of the exhaust gas and of the NO_x gas. Since more than 90% of the NO_x is NO, a value of $M_{NO_x} \approx 30\ kg/kMol$ is often used and, for stoichiometric gasoline mixtures, a value of $M_{eg} \approx 29\ kg/kMol$ is a reasonable estimate. Integrating the variable (2.171) during the entire test cycle yields the total mass of NO_x emitted.

Note that certification measurements must follow precise procedures that often utilize dilution tunnels and fill bags to collect all tail pipe emissions. These procedures are described in detail in [8].

2.8 Pollutant Abatement Systems

2.8.1 Introduction

Since 1980, the permitted levels for the concentrations of carbon monoxide CO, hydrocarbons HC, and NO_x (NO, NO_2) have been reduced significantly. A continuation of this trend is to be expected. Table 2.4 shows the development of the limits in Europe as well as of the standards for California. Similar legislations are imposed in Japan and other countries. In contrast to the European regulations, which limit the total hydrocarbon emissions, the US regulations consider only the non-methane hydrocarbons. This choice reflects the fact that the methane emission by passenger cars are very low compared to the agricultural emissions. However, the US standards regulate formaldehyde ($HCHO$) levels, which are not explicitly limited in the European Union.

Common to all of these emission limits is that they may not be exceeded during a predefined driving cycle. This includes a warm-up phase, transients and idle periods with defined boundary conditions such as ambient temperatures, etc.

As discussed in the last section, today's stringent emission regulations and those in the future will be difficult to meet. Only homogeneous charge SI engines operated with a stoichiometric air/fuel ratio can easily meet these limits by using three-way catalytic converters. All other engines (Diesel, direct-injection SI engines, etc.) must be combined with more complex exhaust gas aftertreatment systems such as lean NO_x traps, selective catalytic reduction converters and particulate filters.

In this section, three-way catalytic converters and selective catalytic reduction converters are analyzed from a control engineering point of view. Readers interested in lean-NO_x traps are referred to [105], [179], [92] [168], and [95] and for particulate filters to [101], [102] and [96].

2.8.2 Pollution Abatement Systems for SI Engines

The most common pollution abatement system for SI port-injection engines is the three-way catalytic converter, TWC. It derives its name from its ability to simultaneously reduce NO_x and oxidize CO and HC. State-of-the-art systems are capable of removing more than 98% of the pollutants. This can only be achieved by operating the engine within very narrow air/fuel ratio limits, as will be shown in the following sections.

Three-Way Catalytic Converters

A catalyst is a substance that promotes chemical reactions without being dissipated. An ideal TWC promotes the following three primary chemical reactions

$$NO + CO \rightarrow \frac{1}{2}N_2 + CO_2 \tag{2.172}$$

Table 2.4. European (top) and Californian (bottom) emission regulations.

	Euro I 1992	Euro II 1996	Euro III 2000	Euro IV 2005
SI Engines				
CO (g/km)	2.72	2.2	2.3	1
HC (g/km)			0.2	0.1
NO_x (g/km)			0.15	0.08
$HC + NO_x$ (g/km)	0.97	0.5		
PM (g/km)	0.14			
Diesel Engines				
CO (g/km)	3.16	1	0.64	0.5
HC (g/km)				
NO_x (g/km)			0.5	0.25
$HC + NO_x$ (g/km)	1.13	0.7	0.56	0.3
PM (g/km)	0.18	0.08	0.05	0.025

	TLEV	LEV	ULEV	SULEV
Passenger Cars				
CO (g/mi)	4.2	4.2	2.1	1
$NMOG$ (g/mi)	0.156	0.09	0.055	0.01
$HCOH$ (g/mi)	0.018	0.018	0.011	0.004
NO_x (g/mi)	0.6	0.07	0.07	0.02
PM (g/mi)	0.08	0.01	0.01	0.01

$$CO + \frac{1}{2} O_2 \rightarrow CO_2 \tag{2.173}$$

$$H_a C_b + (b + \frac{a}{4}) O_2 \rightarrow b\, CO_2 + \frac{a}{2} H_2O \tag{2.174}$$

In reality, the system of reactions is far more complex [88]. Several dozen reactions are involved, and other species, apart from those mentioned, are produced. For example, the concentration of N_2O is often increased by the reactions taking place inside the TWC.

Generally, the reactants are first adsorbed on the catalytic surface. Catalytic substances, such as platinum (Pt), rhodium (Rh) or palladium (Pd), weaken the bonds of the adsorbed species and, thus, allow for the formation of the desired products, according to the law of minimizing the (chemical) potential energy. The products are then desorbed and thus released to the gas phase.

Reactions (2.172) to (2.174) clearly show that optimal conversion rates can only be achieved if exactly as much oxygen is available as is used for the

oxidation of CO and HC. It would be favorable if (2.172) were dominant. In fact, in the presence of excess oxygen (at air/fuel ratios $\lambda > 1$), the reaction (2.173) always occurs in parallel to the other reactions and, therefore, under lean conditions usually inhibits the reduction of NO. With a shortage of oxygen at $\lambda < 1$, all NO is converted, but the removal of HC and CO is incomplete. Therefore, the TWC can be operated only within the very narrow band of air/fuel ratios around $\lambda = 1$.

Figure 2.60 illustrates these concepts by showing the engine-out emissions of a TWC as a function of air/fuel ratio (the upstream gas concentrations in Fig. 2.60 correspond to those shown in Fig. 2.49, with the region around $\lambda = 1$ being substantially enlarged).

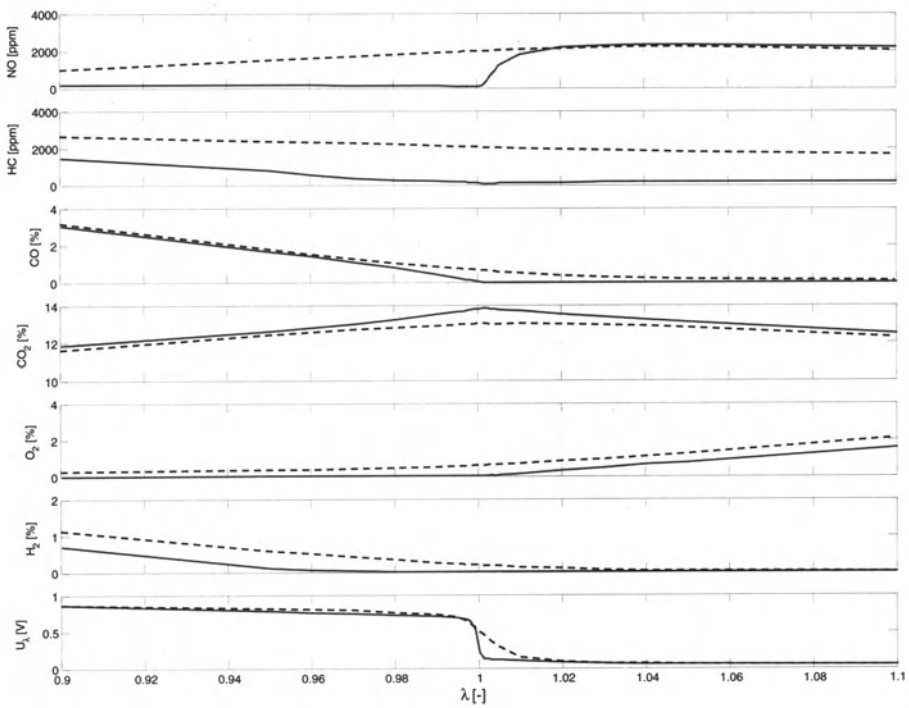


Fig. 2.60. Concentrations of NO, HC, CO, CO_2, O_2, H_2, and the switch-type lambda sensor voltage upstream (dashed) and downstream (solid) of an aged three-way catalytic converter.

Notice that downstream of the TWC a significant amount of hydrogen is still present. The ratio of H_2 to CO produced by new TWC is even higher than that of aged ones. Since air/fuel ratio sensors show a significant cross-sensitivity to hydrogen, this effect has to be taken into account for a correct interpretation of the downstream air/fuel ratio sensor signal. In fact, for a

rich mixture, the air/fuel ratio sensor can be employed as a hydrogen sensor. Further details on this topic may be found in [17].

Especially during transients, when the wall-wetting dynamics and other disturbances render a complete compensation of the varying air mass flow difficult, the air-to-fuel ratio cannot always be kept within the narrow band that is necessary for the TWC to work efficiently.

Therefore, the catalyst must be able to cope with temporary excursions to the lean or to the rich side. This is achieved by incorporating a reservoir for oxygen, which lies beneath the catalytically active surface and which consists of special materials, e.g., cerium. Excess oxygen can then temporarily be stored in these materials as well as on the catalytic surface. The oxygen storage within cerium can be described chemically by the following reaction

$$Ce_2\,O_3 + O^* \leftrightarrow 2\,Ce\,O_2 \tag{2.175}$$

The adsorbed oxygen O^* combines with $Ce_2\,O_3$ to form a new grid structure. The ratio of the catalytically active surfaces of the noble metal (Pt, Pd, Rh) and the cerium considerably influence the water-gas shift reaction

$$CO + H_2O \rightarrow CO_2 + H_2 \tag{2.176}$$

This results in a changing H_2/CO ratio during the lifetime of the TWC, since the catalytically active surface decreases with aging. Therefore, new TWCs promote the water-gas shift reaction significantly more than aged ones, as has been described above.

The TWC consists of a carrier (usually ceramic or metallic substrate) and coating material. First, the washcoat (aluminum oxide) is applied to enlarge the surface of the TWC. Then the additional oxygen storage material (cerium oxide) is brought on the washcoat, and, finally, stabilizers and noble metals are placed on top. The better the distribution of the various components, the better the activity of the TWC.

Modeling of Three-Way Catalytic Converters

General Remarks

Models of TWCs usually serve to predict the conversion rates of the main species and/or the composition of the output gas. They are used to gain a better understanding of the behavior of a TWC during transients and to investigate external influences such as "catalyst poisoning" or aging.[47]

[47] The expression "catalyst poisoning" is used to describe the reduction of catalytic efficiency by unwanted species that occupy active surface sites and, thus, prevent the desired species from adsorbing on these sites. Poisoning can occur irreversibly with substances such as lead (Pb), but also reversibly, for example, with sulfur dioxide (SO_2).

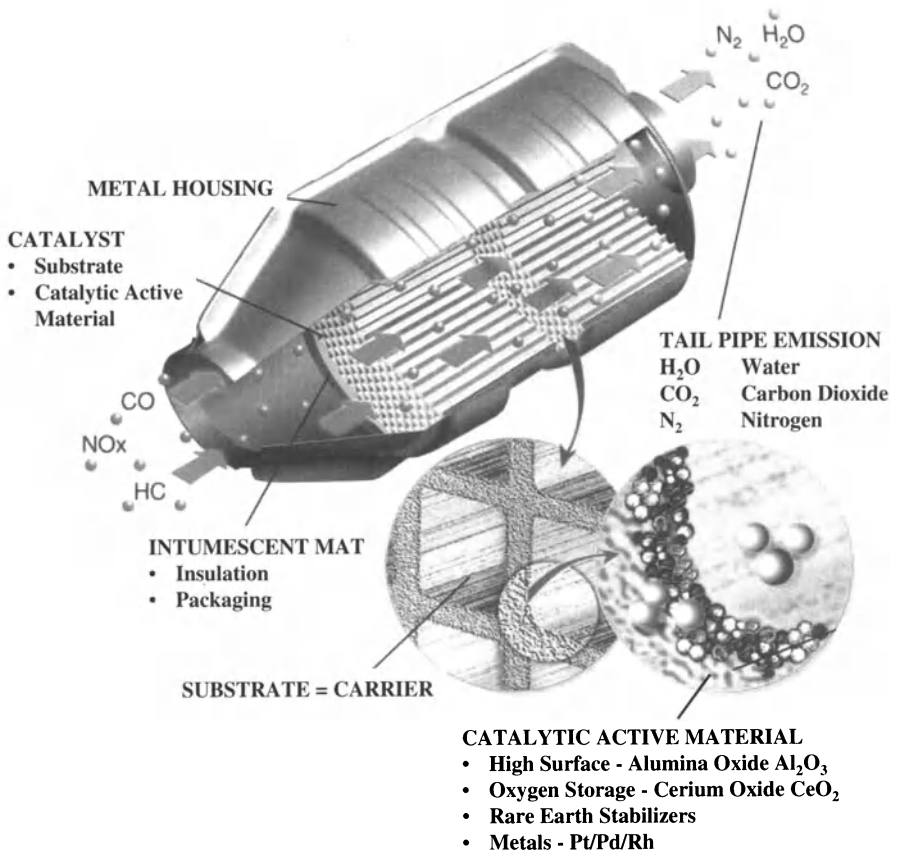


Fig. 2.61. Schematic view of a TWC's structure, reprinted with the permission of Umicore AG.

Predicting the conversion rates is very important for the on-board diagnosis (OBD) of the TWC. An OBD system must be able to warn the driver if the catalytic converter is no longer able to sufficiently remove the pollutants in the exhaust gas. In order to reach the ULEV or Euro IV standards, conversion rates of more than 96% are necessary. Accordingly, the results of the OBD algorithms must be very accurate, which is a challenging objective when considering the few sensor signals available in series production cars. Accurate catalyst models, which do not impose a large computational burden on the ECU, can help achieve these objectives.

Another very important point is the warm-up behavior of a TWC. A cold TWC is almost inactive. A TWC works efficiently only after having reached the light-off temperature.[48] The catalyst is heated by the hot exhaust gases

[48] The light-off temperature of a TWC is defined as the temperature at which a conversion efficiency for HC of 50% is reached.

and the reaction enthalpy that is produced by the exothermic chemical reactions in the catalytic converter. As soon as the light-off temperature has been reached and sufficient catalytic activity is ongoing, the chemical reactions provide enough heat to further warm the TWC, which, in turn, promotes more catalytic activity [71]. This chain reaction can be seen by a significant temperature increase after light-off. In order to reduce pollutants, the warm-up time must be kept as short as possible. Here, too, legislation will become more stringent, as cold start temperatures eventually will be lowered from +20° C to −7° C. Models of TWCs thus also serve to develop strategies aimed at minimizing the warm-up duration.

Detailed models of TWCs are complex. State-of-the-art models incorporate more than 50 reactions with approximately 10 gas species and between 10 and 15 adsorbed species [88]. These models are very stiff, since the time constants of most of the chemical reactions are much smaller than those of the heat and mass flux phenomena.

Control-oriented TWC models usually are much simpler. Often, the dynamics of the temperature are neglected, and most of the reactions are lumped into the two or three most relevant ones. This simplifies the models significantly and thus allows their use in currently available control systems. Several publications deal with this problem, for instance [146] and [18].

The objective of developing a control-oriented TWC model is mainly to predict the oxygen storage level. This quantity cannot be measured. Thus, it has to be estimated by an observer which is usually supported by a lambda sensor downstream of the TWC [25]. The objective of all TWC control systems is to keep the TWC in a state from which excursions to the lean and to the rich side can be handled equally well. It is generally believed that this can be achieved by keeping the oxygen storage level at one-half of the TWC's total storage capacity. Notice that this is a rather simplified modeling approach. In reality, not only is oxygen stored, but also a number of other components.

Since oxygen adsorbs well on the catalytically active surface and does not desorb, the oxygen storage behavior can be modeled using straightforward approaches. Using the available information on the upstream air/fuel ratio and the engine-out mass flow, this model can be used for feedforward control of the TWC. This loop is often referred to as the *buffer control system.*

For feedback control and OBD systems, more complex models are needed. In fact, a complete feedback control system for the oxygen level would be realizable only if a sensor that measures that variable was available. Unfortunately, the only sensor currently available is a (usually switch-type) lambda sensor downstream from the TWC. This sensor measures the remaining oxygen concentration in the exhaust gas if there is any. Moreover, when the engine is operated under rich conditions, this sensor is very sensitive to a number of species, mainly hydrogen, but also CO and HC [26], [15]. Therefore, in order to allow a correct interpretation of the output of this sensor, these disturbances of the sensor output must be modeled. As has been mentioned above, the H_2/CO ratio changes with the ratio of the cerium/noble metal surface,

i.e., with aging. Therefore, simplified TWC models which are supported by sensors downstream, e.g., by means of an observer, must take into account at least two storage capacities, namely the one of cerium and the one of the noble metal [16].

In the following, a model of a TWC will be described which was presented in [125] (several earlier papers discuss the COM of a TWC, see, for instance, [100] or [28]). The effects considered in this model are illustrated in Fig. 2.63. Using this approach, the TWC is assumed to consist of three elements:

- Gas phase: Only convective mass and heat transfer in the axial direction are considered.
- Pores/Washcoat: Here, no convection takes place in the direction of the gas flow, but mass transport to and from the gas phase and to and from the solid phase by desorption and adsorption.
- Solid phase: Heat transfer occurs through heat conduction in the solid phase, heat exchange with the gas phase, and heat production from the reaction heat of the adsorbed species.

Mass balances for each species can be formulated for the gas phase and the pores while heat balances are formulated for the gas phase and the solid phase.

A catalytic converter consists of many channels that are much longer than their diameter (see Fig. 2.61). Therefore, a lumped-parameter modeling approach is not possible. Instead, a formulation using partial differential equations must be used to describe the physical phenomena within the channels. Fortunately, the channels may be assumed to be perfectly mixed in the direction *perpendicular* to the flow direction such that only partial derivatives in the axial direction need be considered.

The concentrations of the components that react in the catalytic converter are much smaller than those of nitrogen, carbon dioxide and water in the exhaust gas. Therefore, a constant mass flow $\dot{m}$ through the catalytic converter may be assumed. If no additional mass flow into or from the control volume occurs, the mass balance for the species i in the control volume depicted in Fig. 2.62 can be expressed by the following equation

$$\varepsilon \cdot A \cdot \partial x \cdot \rho_g \cdot \frac{\partial \xi_{i,g}}{\partial t} = \dot{m} \cdot (\xi_{i,g}(x,t) - \xi_{i,g}(x + \partial x, t)) \tag{2.177}$$

where ε denotes the volume fraction of the catalyst which is filled with exhaust gas, A is the cross section area of the channels through which the exhaust gas flows, ρ_g is the exhaust gas density, and $\xi_{i,g}$ is the mass fraction of the species i in the exhaust gas. The mass in the control volume depicted in Fig. 2.62 is thus given by $\varepsilon \cdot A \cdot \partial x \cdot \rho_{ga}$.

Taking the limit case of an infinitesimally small cell size, (2.177) yields a partial differential equation of the form

$$\varepsilon \cdot A \cdot \rho_g \cdot \frac{\partial \xi_{i,g}}{\partial t} = -\dot{m} \cdot \frac{\partial \xi_{i,g}}{\partial x} \tag{2.178}$$

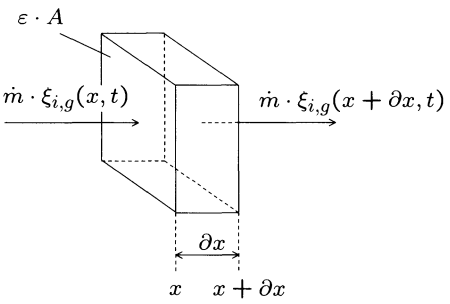


Fig. 2.62. Mass flow through the channel of a catalytic converter.

for all species i. This PDE forms the basis for all subsequent mass conservation equations. The energy conservation equations are derived in a similar way. Its numerical solution is well documented, see, for instance, [126]. However, to obtain state space models which are then suitable for subsequent control system designs, a spatial discretization is usually necessary.

Mass Balances

Figure 2.63 shows the mass and energy flows in the catalytic converter. Flow 1 is the exhaust gas mass flow through the channel. Its dynamics are obtained by combining the basic mass conservation (2.178) with the additional mass flow between pores and gas phase as indicated by Flow 2 in Fig. 2.63. This mass balance for the species i in the gas phase may be formulated as follows

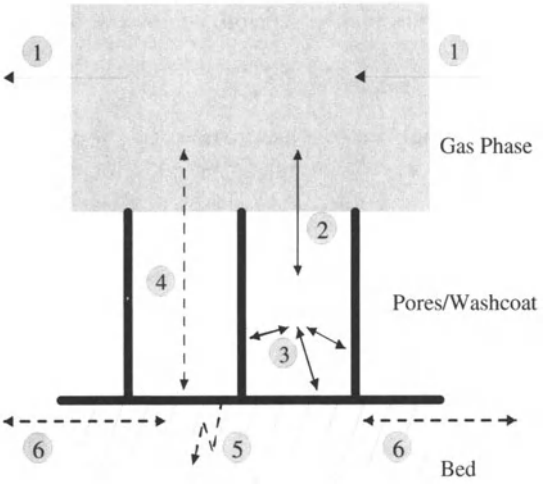
$$\varepsilon \cdot A \cdot \rho_g \cdot \frac{\partial \xi_{i,g}}{\partial t} = -\dot{m} \cdot \frac{\partial \xi_{i,g}}{\partial x} - k_{i,g} \cdot A \cdot a_V \cdot \rho_g \cdot (\xi_{i,g} - \xi_{i,pores}) \qquad (2.179)$$

where $k_{i,g}$ represents the mass transfer coefficient to the washcoat pores, a_V is the cross-sectional area, separating gas phase and washcoat, per catalytic converter volume for the mass transfer to the pores, and $\xi_{i,pores}$ is the concentration of the species i in the pores (the other parameters have been defined in (2.177)). The first term on the right-hand side considers the mass transfer due to the flow through the channel (Flow 1 in Fig. 2.63), the second term describes the mass increase due to mass transfer from the gas to the pores (Flow 2 in Fig. 2.63).[49] Notice that the diffusion in the axial direction has been neglected.

A second mass balance must be formulated for the species i in the pores

$$4\,\varepsilon_W d_W d_b \rho_g \frac{\partial \xi_{i,pores}}{\partial t} = k_{i,g} A a_V \rho_g \cdot (\xi_{i,g} - \xi_{i,pores}) - a_{cat} A R_i M_i \qquad (2.180)$$

[49] Since the partial pressures of the components are not guaranteed to equal a constant value, the Stefan-Maxwell equation would be a more accurate choice. Nevertheless, because p_{N_2} is dominant, this effect can be compensated by slightly adjusting p_{N_2}.



Mass Balance
1 Exhaust Massflow
2 Mass Transfer to/from Pores
3 Adsorption/Desorption

Energy Balance
1 Exhaust Massflow
4 Heat Exchange solid/gas phase
5 Reaction Heat
6 Heat Conduction in Solid Phase

Fig. 2.63. Mass and heat transfer term in the balance equations.

where ε_W denotes the porosity of the washcoat, a_{cat} is the catalytically active area per catalyst volume, d_W stands for the thickness of the washcoat, and d_b is the length of one side of the quadratic monolith channel. The free volume per length of the washcoat that is accessible by the gas is assumed to be $4\,\varepsilon \cdot d_W \cdot d_b$.

The mass production term R_i expresses the mass transfer between the solid phase and the gas phase in $[mol\,m^{-2}s^{-1}]$. It stands for the adsorption/desorption term, and may be calculated from the adsorption ($r_{i,ads}$) and desorption rate ($r_{i,des}$) of the species i using

$$R_i = (r_{i,ads} - r_{i,des}) \cdot L_t \tag{2.181}$$

where L_t denotes the maximum capacity of adsorbed species in $[mol/\,m^2]$. The occupancy θ_i is a relative parameter which expresses the fraction of coverage of the species i, with 0 meaning no coverage and 1 standing for one-hundred percent coverage of the species i.

The dynamics of occupancy are often calculated with the *Langmuir-Hinshelwood* (LH) kinetics or with the *Eley-Rideal* (ER) kinetics. The former allows only adsorbed reactants, whereas in the latter, only one reactant has to be adsorbed; the others may also be in the gas phase. Here, the following assumptions are made:

- All adsorption sites are equivalent, i.e., all species compete for equivalent sites.
- No interactions such as attractive forces between adsorbed species occur.

- Adsorption energies (activation energies for desorption, see below) are independent of the occupancy.
- The total number of adsorption sites is constant.

The reactions modeled depend strongly upon the purpose of the model, e.g., if the downstream switch type sensor voltage has to be used for control purposes, hydrogen has to be taken into account due to the strong sensitivity of the switch type sensor to this component.

In order to illustrate the LH kinetics, a simplified reaction scheme for the oxidation of CO by O_2 is considered as a first step. The reaction occurs in three steps: Adsorption of the two species, reaction and, finally, desorption of the product, CO_2. These three steps are described by

$$CO_g + v^* \leftrightarrow CO^* \qquad O_{2,g} + 2v^* \rightarrow 2O^* \tag{2.182}$$

$$CO^* + O^* \rightarrow CO_{2,g} \tag{2.183}$$

The symbol v^* denotes a free site on the catalyst surface, the subscript g denotes a gas species, and the asterisk an adsorbed species.

According to the LH kinetics, the rate of adsorption of the species i is only dependent upon the fraction of vacant sites, θ_V, and its partial pressure in the gas phase, p_i. For CO, the adsorption rate is thus

$$r_{CO,ads} = k_{ads,CO} \cdot p_{CO} \cdot \theta_V \tag{2.184}$$

If all the vacant sites are equivalent, their fraction may be expressed as

$$\theta_V = 1 - \theta_O - \theta_{CO} \tag{2.185}$$

The partial pressure of the component i can be calculated using the ideal gas law

$$p_i = \rho_g \cdot \xi_{i,g} \cdot R_i \cdot \vartheta_g = \rho_g \cdot \xi_{i,g} \cdot \frac{\mathcal{R}}{M_i} \cdot \vartheta_g \tag{2.186}$$

The rate constant $k_{ads,CO}$ is derived using ideas of statistical thermodynamics by computing the number of collisions of gas particles on the catalyst surface per time and catalyst area (interested readers may find more information about this in [60] and [13])

$$k_{ads,CO} = \frac{1}{L_t} \cdot \frac{s}{\sqrt{2\pi \cdot M_{CO} \cdot \mathcal{R} \cdot \vartheta_{gas}}} \tag{2.187}$$

Here, M denotes the molecular mass, $\mathcal{R}$ the universal gas constant, and s the sticking probability. The last factor expresses the fraction of collisions which actually lead to an adsorption reaction. It is dependent upon the surface material and upon the gas species. Adsorption of O_2 is calculated accordingly.

The reactivity of two or more reactants is dependent upon their availability, i.e., their surface occupancy, $\theta_{...}$, and upon the rate constant, k_{reac}

$$r_{CO,reac} = r_{O,reac} = k_{reac,CO_2} \cdot \theta_{CO} \cdot \theta_O \tag{2.188}$$

Not only is CO_2 being produced and its occupancy increased, but also CO and O are consumed and, thus, their occupancies decrease. The rate constant is usually calculated following the law of Arrhenius

$$k_{reac} = A_{reac} \cdot e^{-\frac{E_a}{\mathcal{R} \cdot \vartheta_{cat}}} \tag{2.189}$$

For a detailed explanation see [60] and [13]. The symbol E_a denotes the activation energy, which can be seen as an energy barrier that must be exceeded in order to start a reaction, whereas A_{reac} denotes the pre-exponential factor with unit s^{-1}. Both parameters are dependent upon the species and the surface material. According to this law, the reactivity increases exponentially with the catalyst temperature

The desorption rate of a species is dependent upon its occupancy and upon a rate constant that follows Arrhenius' law as well

$$r_{CO,des} = k_{des,CO} \cdot \theta_{CO} \tag{2.190}$$

The desorption rates of O_2 can be derived in the same way.

Now, the total reaction rate of adsorbed CO can be expressed as follows

$$\begin{aligned} \frac{\partial \theta_{CO}}{\partial t} &= r_{CO,ads} - r_{CO,des} - r_{CO,reac} \\ &= k_{ads,CO}\, p_{CO}\, \theta_V - k_{des,CO}\, \theta_{CO} - k_{reac,CO_2}\, \theta_{CO}\, \theta_O \end{aligned} \tag{2.191}$$

The mass production term in (2.180) only includes the adsorption and the desorption terms, for CO

$$\begin{aligned} R_{CO} &= (r_{CO,ads} - r_{CO,des}) \cdot L_t \\ &= (k_{ads,CO} \cdot p_{CO} \cdot \theta_V - k_{des,CO} \cdot \theta_{CO}) \cdot L_t \end{aligned} \tag{2.192}$$

Figure 2.64 illustrates a selection of chemical reactions on a TWC.

Buffer Model for the Oxygen Storage of a TWC

From a control engineering point of view, the concentrations of the individual exhaust gas components are of minor interest. It is only important to keep the oxygen storage of the catalytic converter at a level which allows the reduction of the nitrogen oxide and the oxidation of the hydrocarbons and carbon monoxides. As can be seen in Fig. 2.65, this is the case if the oxygen content (relative oxygen level, ROL) is between 0 (no stored oxygen) and 1 (all storage places occupied with oxygen). The relative oxygen level can be calculated according to the following equation (all delays have been omitted here for the sake of clarity; further details can be found in [133], [85] and [16])

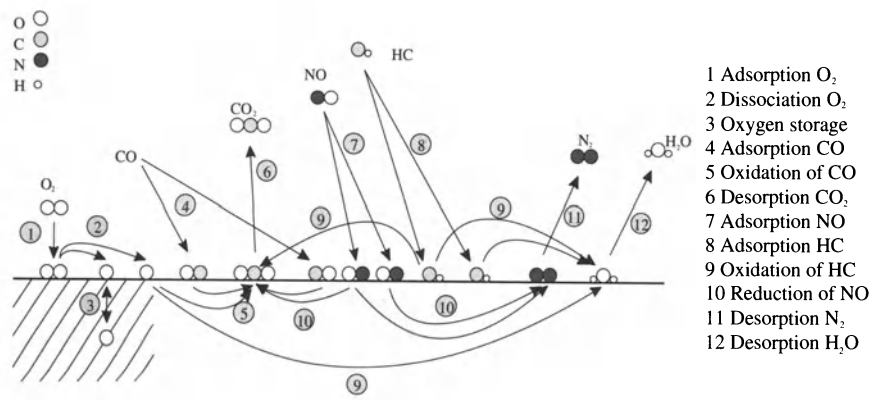


Fig. 2.64. Overall chemical reactions in a TWC.

$$ROL(t) = \lim_{0,1} \int 0.21 \cdot \frac{(\dot{m}_\beta(t) - \sigma_0 \cdot \dot{m}_\varphi(t))}{C_{cat}} \, dt$$

$$= \lim_{0,1} \int 0.21 \cdot \frac{\dot{m}_\beta(t)}{C_{cat}} \cdot (1 - \frac{1}{\lambda(t)}) \, dt \qquad (2.193)$$

Here, C_{cat} represents the maximum oxygen storage capacity of the catalytic converter. Since only oxygen has to be considered, the mass of air must be multiplied by 0.21 (assuming 21% oxygen in the air). If the buffer is full of oxygen ($ROL = 1$), no more oxygen can be stored; thus, the integrator has to be limited. For an empty buffer similar considerations hold.

In Fig. 2.65, measurement results are shown that were obtained on an SI engine equipped with a standard TWC system. In this experiment, the engine runs at a constant speed and manifold pressure, resulting in a constant air mass flow. The nominal fuel mass flow command signal is modulated with a square wave around stoichiometry with an amplitude of approximately ±10%.

The top plot of Fig. 2.65 shows the signals of wide-range lambda sensors upstream and downstream of the TWC. The fuel path dynamics can be neglected for this investigation because they are much faster than the filling and emptying of the TWC.

In the second plot of Fig. 2.65, measurement results are depicted that were obtained using fast CO and NO sensors. It is interesting to note that the NO concentration drops immediately when the upstream lambda switches from lean to rich; whereas the CO concentration remains low until the oxygen storage of the TWC is depleted and no more oxidation can take place. When the input lambda is switched from rich to lean, the CO concentration drops with a certain time constant due to desorption processes within the TWC. This is in contrast to the behavior of the NO concentration, which drops immediately with a switch from lean to rich. This is due to negligible desorption rate of oxygen compared to the desorption of CO. Since CO desorbs spontaneously

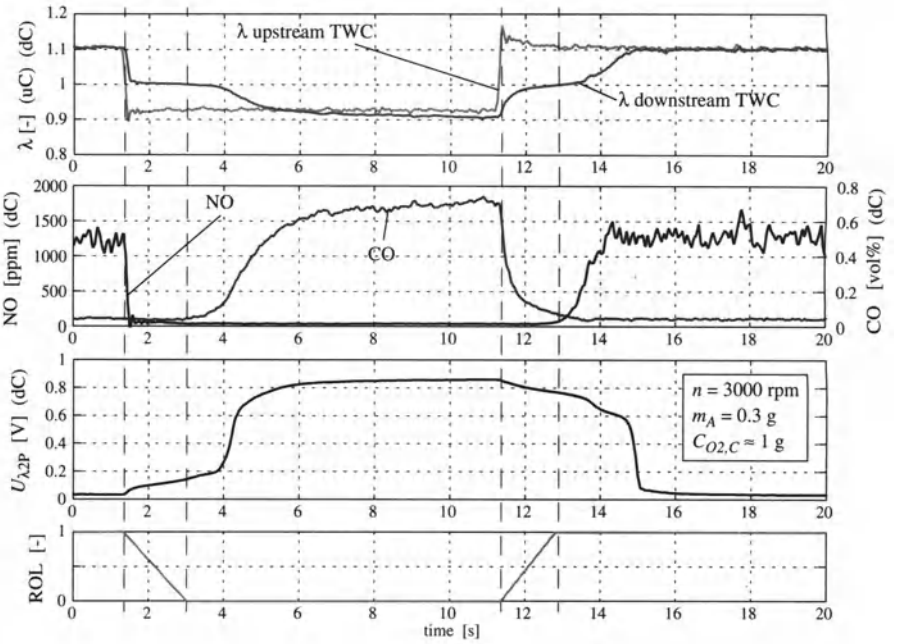


Fig. 2.65. Measurements of λ and emissions downstream of a TWC imposing step changes in the fuel mass flow as modulations.

into the gas and since the LH kinetics are dominant (which does not allow reactions in the gas phase), the desorbed CO will appear downstream of the TWC. Again, when the oxygen storage is full, the NO concentration rises again.

The third plot in Fig. 2.65 shows the switch-type lambda sensor voltage downstream of the TWC. Depending on the previous inputs, the sensor "switches" after the oxygen storage is either full or empty. From the results of the emission measurements, an oxygen storage capacity can be estimated and thus a relative oxygen level can be calculated. This variable is depicted in the bottom plot of Fig. 2.65.

Energy Balances

The energy balance for the gas phase includes the heat flows shown in Fig. 2.63. Heat is transported convectively with the exhaust gas mass flow. Additionally, heat is exchanged with the solid catalyst bed. The catalyst is assumed to be adiabatic, i.e., heat exchange with the environment is neglected. Thus, the heat balance of the gas phase can be formulated as follows

$$\varepsilon \cdot \rho_g \cdot c_{v,g} \cdot \frac{\partial \vartheta_g}{\partial t} = -\frac{\dot{m}}{A} \cdot c_{p,g} \cdot \frac{\partial \vartheta_g}{\partial x} - \alpha \cdot a_V \cdot (\vartheta_g - \vartheta_{cat}) \tag{2.194}$$

where $c_{p,g}$ denotes the specific heat capacity of the exhaust gas, and ϑ_g and ϑ_{cat} the gas and catalyst temperatures. The symbol α represents the heat transfer coefficient.

The energy balance for the solid phase also includes the heat exchange term with the gas phase. In addition, heat conduction occurs in the solid phase. Heat production from the exothermic chemical reactions on the catalyst surface also has to be included thus yielding

$$(1-\varepsilon)\cdot\rho_{cat}\cdot c_{cat}\cdot\frac{\partial\vartheta_{cat}}{\partial t}=\lambda_{cat}\cdot(1-\varepsilon)\cdot\frac{\partial^2\vartheta_{cat}}{\partial x^2}+\alpha\cdot a_V\cdot(\vartheta_g-\vartheta_{cat})$$
$$+a_{cat}\sum_k(-\Delta H_r)_k\cdot r_{reac,k}\cdot L_t \qquad (2.195)$$

Here, ρ_{cat} and c_{cat} stand for the density and the specific heat capacity of the solid phase, respectively. The symbol λ_{cat} denotes the heat conduction coefficient, whereas $(-\Delta H_r)_k$ and $r_{reac,k}$ are the reaction enthalpy and the reaction rate of the reaction k.

2.8.3 Pollution Abatement Systems for Diesel Engines

Motivation

It is anticipated that future limits imposed on the pollutant emissions of Diesel engines will no longer be met by reducing only engine-out pollutants. Most probably, an aftertreatment system will become necessary to reduce the nitrogen oxide or the particulate matter emissions to the future legislation limits shown in Fig. 2.66.

The type of aftertreatment system selected will depend on the engine application (light/medium/heavy duty) and on the chosen engine calibration. With a high-efficiency NO_x reduction system in place, for instance, the engine can be operated close to its optimal bsfc and at low PM conditions.[50]

Three types of pollution abatement systems for Diesel engines are considered at the moment:

- particulate filters;
- NO_x traps; and
- selective catalytic reduction (SCR) systems.

The first two systems have the drawback that for their regeneration, high temperatures (particulate filters) or rich exhaust gases (NO_x traps) are necessary. These regeneration phases will increase fuel consumption. The SCR approach requires an additional agent, most probably urea, in a separate storage

[50] If, however, the targeted emission limits are very low, a combination of engine tuning, in-engine measurements, and an exhaust gas aftertreatment system will have to be setup, and the engine may *not* be re-tuned fuel efficiently.

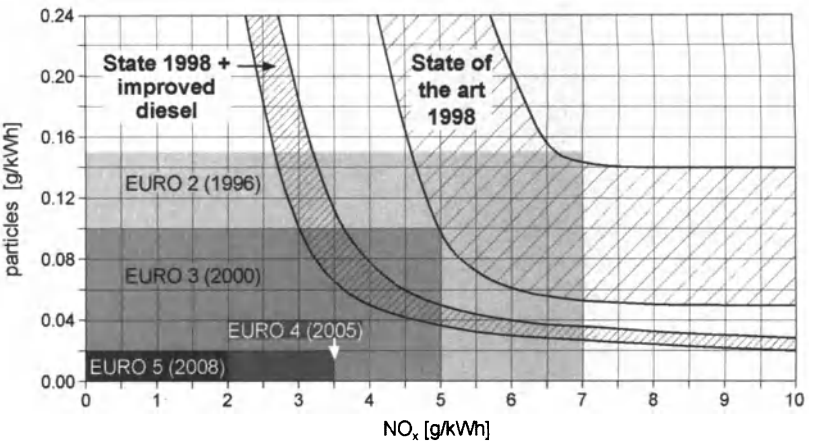


Fig. 2.66. Legislation and emission trade-off for heavy-duty Diesel engines: particulate matter (PM) versus nitrogen oxide (NO_x), adopted from [48].

system. Fortunately, this substance is innocuous and not expensive. A comprehensive preliminary assessment of various exhaust aftertreatment systems is presented in [121], and a state-of-the-art overview in [89].

An SCR exhaust gas aftertreatment system using urea solution as a reductant has a high NO_x reduction potential and is a well-known technique for stationary applications, [24]. In mobile applications, however, the volume of the catalytic converter is limited and the engine is operated under transient conditions most of the time. Due to the fast changes in the exhaust gas temperature and the high gas hourly space velocity (GHSV), ammonia slip is likely to occur unless special control actions are applied. Model-based control concepts, as presented for instance in [147], are therefore a must in this case.

Structure and Working Principle of SCR Systems

A catalytic converter for Diesel engines cannot work the same way as a TWC for SI engines because an excess of oxygen is always present in the exhaust gas. One approach to resolve this problem is to use an additional reducing agent. A convenient reducing agent showing a good selectivity toward NO_x is ammonia (NH_3). For dosage and safety reasons, ammonia is often injected as a water-urea solution that decomposes in the catalytic converter to form ammonia and CO_2.[51]

The primary reactions occurring in this catalytic system are:

- Evaporation of the urea solution and decomposition of the urea molecules (where s denotes solid)

[51] For the engine under investigation 1 liter urea solution is needed for each 20 liter of Diesel fuel.

$$((NH_2)_2CO)_s + H_2O \rightarrow 2NH_3 + CO_2 \tag{2.196}$$

- Adsorption and desorption of ammonia on the surface of the catalytic converter (where * denotes an adsorbed species)

$$NH_3 \leftrightarrow NH_3^* \tag{2.197}$$

- Selective catalytic reduction consuming adsorbed ammonia and gaseous NO_x (Eley-Rideal mechanism)

$$4NH_3^* + 2NO + 2NO_2 \rightarrow 4N_2 + 6H_2O \tag{2.198}$$

$$4NH_3^* + 4NO + O_2 \rightarrow 4N_2 + 6H_2O \tag{2.199}$$

  The first reaction is faster than the second. Up to 95% of the NO_x leaving the engine is NO. It is useful to mount an oxidation catalytic converter in front of the SCR catalyst. This device changes the ratio of NO to NO_2 (but will not affect the amount of NO_x because of excess oxygen), which is favorable for the SCR reaction.[52] Additionally, this catalytic converter transforms hydrocarbons and soot into CO_2 and simultaneously transforming NO to NO_2, and prevents the SCR catalyst from being poisoned.
- Oxidation of adsorbed ammonia

$$4NH_3^* + 3O_2 \rightarrow 2N_2 + 6H_2O \tag{2.200}$$

For reasonably sized catalytic converters, the temperature window where an SCR system works best is 250-400 °C. Below 200 °C the reactions inside the SCR are too slow, above 450 °C an increasing amount of ammonia is converted directly to nitrogen.

Modeling of an SCR System

As shown in Fig. 2.67, the complete SCR system consists of an oxidation catalyst, a device to inject urea solution and the SCR catalytic converter.

The subscripts u, m and d stand for upstream, middle and downstream of the SCR system. The plant is modeled by an oxidation cell and n identical SCR cells. Notice that the oxidation cell includes the oxidation catalytic converter *and*, immediately downstream of that converter, the urea injection system.

The models derived below will disregard the hydrocarbon and soot present in the exhaust gases. This simplification is based on the assumption that these species will not affect the main NO_x reduction pathways.

[52] In order to avoid additional very slow reactions, the ratio of NO to NO_2 should not become less than 1.

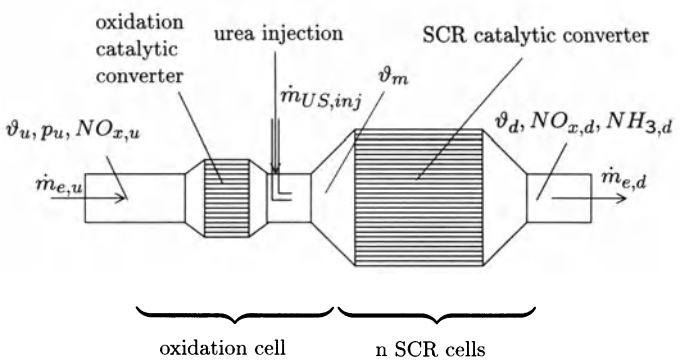


Fig. 2.67. Complete SCR system consisting of the oxidation catalytic converter, the urea injection and the SCR catalytic converter.

Oxidation Cell

The oxidation catalytic converter and the urea injection system are modeled first. This component will be referred to as the *oxidation cell*. In general, the oxidation cell has a different coating compared to the SCR cell.

The assumptions made to simplify the modeling of the oxidation cell are:

- The injected urea solution mass flow $\dot{m}_{US,inj}$ (a water-urea mixture with a urea concentration $\xi_U S$) is heated to the temperature of the exhaust gas, however, no urea decomposition takes place.
- The molar flow of NO_x is not affected.
- The oxidation catalyst is assumed to be a perfect heat exchanger and the exhaust gas leaving the catalyst is assumed to have the same temperature as the catalyst ($\vartheta_{OC} = \vartheta_m$).
- The oxidation catalyst can be modeled accurately as a perfectly stirred reactor with concentrated model parameters.

Note that only the total NO_x flow is taken as input to the system and that the primary effects modeled are the thermal behavior of the oxidation cell.

Modeling the system using two thermal energy reservoirs often yields satisfactory results. These two reservoirs model the behavior of the converter and its housing. For the first level variable, the temperature of the converter ϑ_{OC}, the following equation holds

$$\frac{d}{dt}\vartheta_{OC} \cdot c_{OC} \cdot m_{OC} = (\vartheta_u - \vartheta_{OC}) \cdot c_{p,g} \cdot \dot{m}_{e,u} - \dot{Q}_{US} - \dot{Q}_{cond} \qquad (2.201)$$

where the two heat flows $\dot{Q}_{US}$ and $\dot{Q}_{cond}$ are modeled by the following equations

$$\dot{Q}_{US} = \Big[\big((\vartheta_b - \vartheta_a) \cdot c_{p,H_2O(l)} + r_{v,H_2O} + (\vartheta_{OC} - \vartheta_b) \cdot c_{p,H_2O(g)}\big) \cdot (1 - \xi_{US}) + (\vartheta_{OC} - \vartheta_a) \cdot c_{p,U} \cdot \xi_{US}\Big] \cdot \dot{m}_{US,inj} \tag{2.202}$$

$$\dot{Q}_{cond} = \alpha_{cond} \cdot A_{cond} \cdot (\vartheta_{OC} - \vartheta_{OCh}) \tag{2.203}$$

The heat capacity of the oxidation converter is given by c_{OC} and m_{OC} is the mass of the oxidation converter. The first term on the right-hand side of (2.201) describes the enthalpy increase by the mass flow through the oxidation converter. The heat capacity of the exhaust gas is given by $c_{p,g}$ and $\dot{m}_{e,u}$ is the exhaust gas mass flow at the inlet of the oxidation cell. The variable $\dot{Q}_{US}$ denotes the energy flow necessary to heat and vaporize the injected urea solution. The urea solution is an eutectic solution ($\xi_{US} = 0.325$).

Note that the vaporization of the urea solution must be divided into three steps: heating from the ambient temperature ϑ_a to the boiling point ϑ_b, vaporization and, finally, heating from the boiling point to the gas temperature. Urea is assumed to enter the system as a solid substance.

The variable $\dot{Q}_{cond}$ describes the conductive heat losses to the converter housing. The housing of the oxidation converter is modeled according to Sect. 2.6.3. The temperature of the oxidation converter housing, ϑ_{OCh}, is the second level variable. Its dynamics are described by

$$\frac{d}{dt}\vartheta_{OCh} \cdot c_{OCh} \cdot m_{OCh} = \dot{Q}_{cond} - \dot{Q}_{out,OCh} \tag{2.204}$$

where

$$\dot{Q}_{out,OCh} = k_{out,OCh} \cdot A_{out,OCh} \cdot (\vartheta_{OCh}^4 - \vartheta_a^4) + \alpha_{out,OCh} \cdot A_{out,OCh} \cdot (\vartheta_{OCh} - \vartheta_a) \tag{2.205}$$

The heat losses $\dot{Q}_{cond}$ in (2.201) represent the heat increase in (2.204). The second term models the heat losses of the housing to the ambient by radiation and convection.

Figure 2.68 shows the mass and heat flows as well as the inlet and outlet temperatures in the oxidation converter where $\dot{n}_i$ denotes the molar flow of species i. The level variables, i.e., ϑ_{OC} and ϑ_{OCh}, are encircled.

The liquid urea solution is vaporized and a molar flow $\dot{n}_{U,m}$ of urea is produced. The NO_x mass flow remains unchanged, while the total mass flow $\dot{m}_{e,m}$ is increased by the mass of the injected urea solution ($\dot{m}_{e,m} = \dot{m}_{e,u} + \dot{m}_{US,inj}$).

As mentioned, the expressions (2.202) and (2.203) describe only the heating of the urea solution and the heat losses to the oxidation converter housing. The NO/NO_2 ratio change is not modeled and therefore not visible in these equations. The main justification for this simplification is that both NO and NO_2 "consume" one ammonia molecule when reduced to molecular nitrogen.

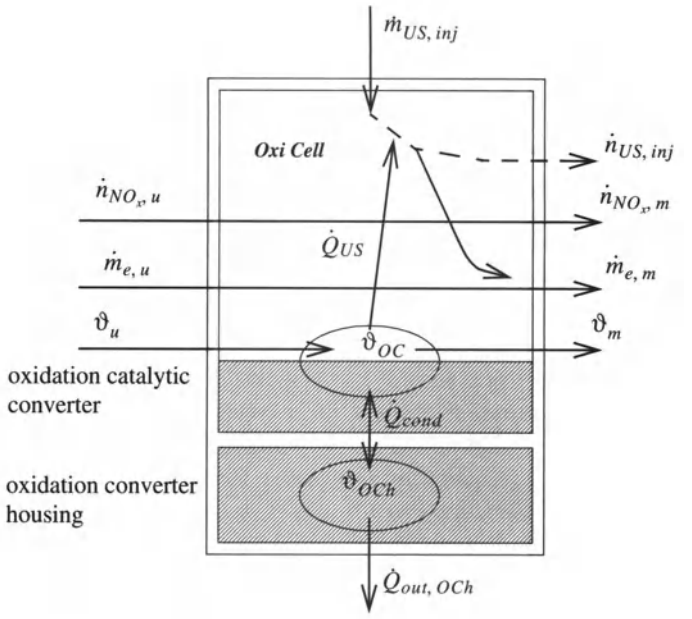


Fig. 2.68. Mass and heat flows within the oxidation cell.

Moreover, the kinetics of both reactions are faster than the ammonia adsorption such that the latter dominates the overall dynamics. Of course these global reaction parameters must be experimentally identified and are valid only for the specific configuration used in the experiments.

SCR Cell

In this part, a model of the SCR cell is introduced. A distributed-parameter formulation using partial differential equations must be used for this subsystem. Of course, this model will have to be discretized for use in a later control system. The derivation of the PDEs follows the approach used in Sect. 2.8.2.

The assumptions made when deriving the model of the SCR cell are:

- Only one solid component, urea, and only two gases, NO_x and NH_3, are considered in the model.
- The reaction for the urea decomposition is modeled using (2.196).
- Adsorption and desorption of ammonia on the surface of the catalytic converter are described using (2.197).
- The oxidation of the ammonia adsorbed on the surface of the catalytic converter is modeled using (2.200)
- An Eley-Rideal reaction kinetics is assumed for the SCR component. The gas NO_2 is considered to react the same as NO, consuming one NH_3 molecule per NO_x molecule, see (2.199) and (2.198). Depending on the activity of the oxidation cell, a certain average NO/NO_2 ratio will be

present at the input of the SCR converter. This will result in an average SCR reduction rate which has to be identified by measurements.
- No adsorption of reaction products is considered.
- No mass storage effects are assumed to take place in the *gas* phase, i.e, the exhaust mass flow is assumed to be large compared to the mass flow of the reacting species.
- The influence of any washcoat diffusion effects is neglected.
- Compared to the oxidation catalytic converter, the SCR converter has a much smaller surface to volume ratio. For that reason, only one thermal energy reservoir must be considered.
- The reaction enthalpies are neglected since they do not have a significant influence on the temperature of the system.
- Regarding the thermal effects, the SCR converter is assumed to be a perfect heat exchanger that satisfies the lumped-parameter assumptions. Accordingly, the exhaust gas temperature at the outlet, ϑ_d is assumed to be equal to the converter temperature ϑ_C. The dynamics of this variable can be approximated well using an ODE.

Therefore, the simplified thermal energy dynamics for the SCR catalytic converter can be described by

$$c_{SCR} \cdot m_{SCR} \cdot \frac{d}{dt}\vartheta_C(t) = c_{p,g} \cdot \dot{m}_{e,m} \cdot (\vartheta_m - \vartheta_C) - \dot{Q}_{out,SCR} \tag{2.206}$$

The heat losses to ambient are calculated similar to (2.205)

$$\begin{aligned}\dot{Q}_{out,SCR} = k_{out,SCR} \cdot A_{out,SCR} \cdot (\vartheta_C^4 - \vartheta_a^4) \\ +\alpha_{out,SCR} \cdot A_{out,SCR} \cdot (\vartheta_C - \vartheta_a)\end{aligned} \tag{2.207}$$

Applying mass balances, kinetic gas theory for the adsorption and Arrhenius reaction kinetics for the urea decomposition, the desorption, the SCR reaction, and the ammonia oxidation, a system of coupled PDE's is obtained. These equations describe the dynamics of the various gas concentrations and of the surface coverage. Note the similarities to the corresponding set of equations presented in Sect. 2.8.2. The urea decomposition is described by

$$\begin{aligned}\varepsilon \cdot A \cdot \rho_g \cdot \frac{\partial \xi_U}{\partial t} &= -\dot{m}_{e,m} \cdot \frac{\partial \xi_U}{\partial x} - \varepsilon \cdot A \cdot \rho_g \cdot r_{UD} \\ r_{UD} &= A_{UD} \cdot e^{\left(-\frac{E_{a,UD}}{\mathcal{R}\vartheta_C(t)}\right)} \cdot \xi_U\end{aligned} \tag{2.208}$$

The first term on the right-hand side models the increase in concentration by the flow through the converter. The second term describes the urea decomposition. The modeling is very similar to the one of the TWC.

The PDE for the ammonia concentration is given next

$$\varepsilon \cdot A \cdot \rho_g \cdot \frac{\partial \xi_{NH_3}}{\partial t} = -\dot{m}_{e,m} \frac{\partial \xi_{NH_3}}{\partial x} + \varepsilon \cdot A \cdot \rho_g \cdot 2 \cdot r_{UD} + \varepsilon \cdot A \cdot a_{cat} \cdot L_t \cdot (r_{des} - r_{ads}) \cdot M_{NH_3} \tag{2.209}$$

$$r_{ads} = \frac{1}{L_t} \cdot \frac{s}{\sqrt{2\pi \cdot M_{NH_3} \cdot \mathcal{R} \cdot \vartheta_C}} \cdot p_{NH_3} \cdot (1 - \theta_{NH_3}) = \rho_g \cdot \frac{s}{L_t \cdot M_{NH_3}} \cdot \sqrt{\frac{\mathcal{R} \cdot \vartheta_C}{2\pi \cdot M_{NH_3}}} \cdot \xi_{NH_3} \cdot (1 - \theta_{NH_3})$$

$$r_{des} = A_{des} \cdot e^{(-\frac{E_{a,des}}{\mathcal{R}\vartheta_C})} \cdot \theta_{NH_3}$$

In addition to the already known terms, the adsorption and desorption of urea on the surface of the catalytic converter must be considered. The adsorption rate is primarily dependent on the number of free sites on the surface $1-\theta_{NH_3}$ and upon the ammonia concentration in the gas. The desorption rate is assumed not to be dependent upon the gas concentration. The PDE for the NO_x concentration contains the SCR reaction

$$\varepsilon \cdot A \cdot \rho_g \cdot \frac{\partial \xi_{NO_x}}{\partial t} = -\dot{m}_{e,m} \cdot \frac{\partial \xi_{NO_x}}{\partial x} - \varepsilon \cdot A \cdot a_{cat} \cdot L_t \cdot r_{SCR} \cdot M_{NO_x} \tag{2.210}$$

$$r_{SCR} = \frac{1}{L_t} \cdot \frac{s}{\sqrt{2\pi \cdot M_{NO_x} \cdot \mathcal{R} \cdot \vartheta_C}} \cdot p_{NO_x} \cdot \theta_{NH_3} = \rho_{gas} \cdot \frac{s}{L_t \cdot M_{NO_x}} \cdot \sqrt{\frac{\mathcal{R} \cdot \vartheta_C}{2\pi \cdot M_{NO_x}}} \cdot \theta_{NH_3} \cdot \xi_{NO_x} \tag{2.211}$$

Finally the surface coverage can be calculated

$$\frac{\partial}{\partial t}\theta_{NH_3} = r_{ads} - r_{des} - r_{SCR} - r_{ox} \tag{2.212}$$

$$r_{ox} = A_{ox} \cdot e^{(-\frac{E_{a,ox}}{\mathcal{R}\vartheta_C})} \cdot \theta_{NH_3} \tag{2.213}$$

It is assumed that there is always excess oxygen in the exhaust gas and thus no dependency on the oxygen partial pressure arises. The oxidation of ammonia on the surface of the catalytic converter has been taken into account additionally. The mass and energy within the SCR cell are depicted in Fig. 2.69.

The level variables are the concentration of urea, the concentrations of ammonia and NO_x in the gas phase, the surface coverage of the catalytic converter with ammonia, and the temperature of the catalytic converter. Only the listed components are indicated, therefore the products N_2 and H_2O of the ammonia oxidation are not shown. The same is valid for the SCR reaction, which, again, produces only N_2 and H_2O.

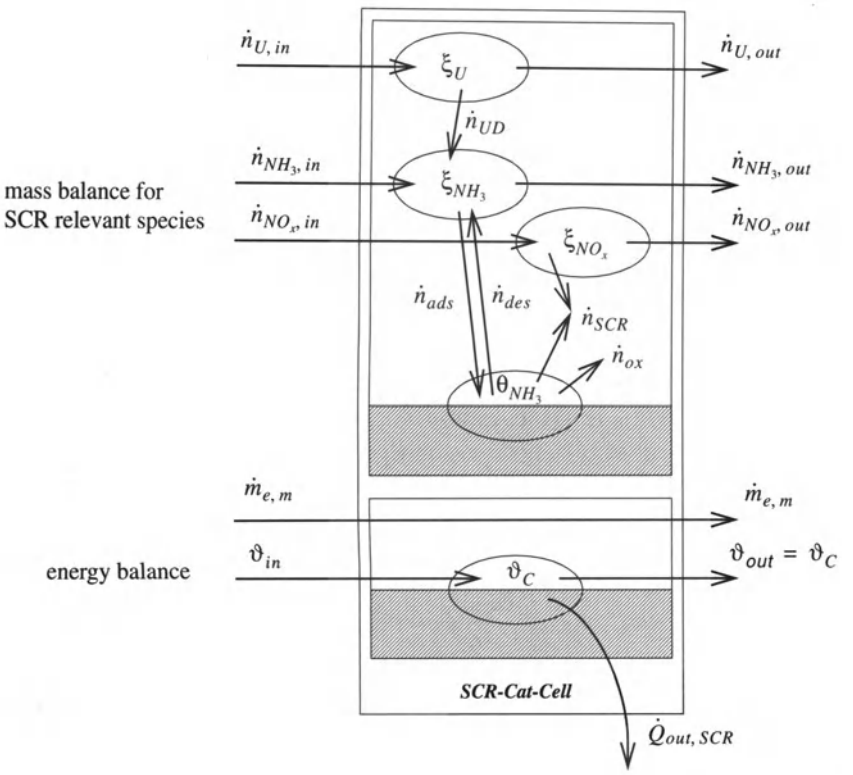


Fig. 2.69. Mass and energy flows within the SCR cell.

Usually, these equations are not solved as partial differential equations. They are approximated by ordinary differential equations by discretizing the converter into n_{cells} cells along its flow axis. Temperature distribution and gas concentrations are assumed to be homogenous in each cell (lumped-parameter assumption). The order of the discretized complete SCR system is then $2 + 5n_{cells}$, two states from the oxidation cell, and five states for each SCR cell. Even for a small number of cells, the solution of these equations requires a substantial amount of numerical computations. Therefore, such a model is not useful in real-time applications.

Simplified Model for the SCR System

In order to obtain a model that is useful for the design of model-based control algorithms, additional assumptions must be made:

- The urea decomposition is fast and therefore the ammonia flow is statically calculated from the amount of urea at the entrance of the SCR catalytic converter.

- The time constants of the storage effects in the gas phase are much smaller than that for the storage of thermal energy and of ammonia on the catalyst surface. Thus, the NO_x concentration as well as the gaseous ammonia concentration are calculated statically.
- The reaction enthalpies can be neglected compared to thermal convection and radiation.

This results in the following two ordinary differential equations for an SCR cell

$$\begin{aligned}\frac{d\theta_{NH_3,i}}{dt} = &\frac{s \cdot \rho_g}{L_t \cdot M_{NH_3}} \cdot \sqrt{\frac{\mathcal{R} \cdot \vartheta_{C,i}}{2\pi \cdot M_{NH_3}}} \cdot (1 - \theta_{NH_3,i}) \cdot \xi_{NH_3,i} \\ &-\theta_{NH_3,i} \cdot \left[A_{des} \cdot e^{\left(\frac{-E_{a,des}}{\mathcal{R} \cdot \vartheta_{C,i}}\right)} + A_{ox} \cdot e^{\left(\frac{-E_{a,ox}}{\mathcal{R} \cdot \vartheta_{C,i}}\right)} \right. \\ &\left. + \rho_g \cdot \frac{s}{L_t \cdot M_{NO_x}} \cdot \sqrt{\frac{\mathcal{R} \cdot \vartheta_{C,i}}{2\pi \cdot M_{NO_x}}} \cdot \xi_{NO_x,i} \right]\end{aligned} \tag{2.214}$$

$$\begin{aligned}\frac{d\vartheta_{C,i}}{dt} = &\frac{n_{cells} \cdot c_{p,g}}{c_{SCR} \cdot m_{SCR}} \cdot \dot{m}_{e,m} \cdot (\vartheta_{C,i-1} - \vartheta_{C,i}) \\ &- \frac{A_{out,SCR}}{c_{SCR} \cdot m_{SCR}} \cdot (k_{out,SCR} \cdot (\vartheta_{C,i}^4 - \vartheta_a^4) + \alpha_{out,SCR} \cdot (\vartheta_{C,i} - \vartheta_a))\end{aligned} \tag{2.215}$$

The gas concentration mass fractions must be calculated with the following algebraic equations

$$\xi_{NO_x,i} = \frac{\xi_{NO_x,i-1}}{1 + \frac{\rho_g \cdot V_C}{n_{cells} \cdot \dot{m}_{e,m}} \cdot a_{cat} \cdot \theta_{NH_3,i} \cdot s \cdot \sqrt{\frac{\mathcal{R} \cdot \vartheta_{C,i}}{2\pi \cdot M_{NO_x}}}} \tag{2.216}$$

and

$$\xi_{NH_3,i} = \frac{\xi_{NH_3,i-1} + \frac{V_C}{n_{cells} \cdot \dot{m}_{e,m}} \cdot a_{cat} \cdot L_t \cdot \theta_{NH_3,i} \cdot M_{NH_3} \cdot A_{des} \cdot e^{-\frac{E_{a,des}}{\mathcal{R} \cdot \vartheta_{C,i}}}}{1 + \frac{\rho_g \cdot V_C}{n_{cells} \cdot \dot{m}_{e,m}} \cdot a_{cat} \cdot (1 - \theta_{NH_3,i}) \cdot s \cdot \sqrt{\frac{\mathcal{R} \cdot \vartheta_{C,i}}{2\pi \cdot M_{NO_x}}}} \tag{2.217}$$

The algebraic equations can be derived from the differential equations (2.210) and (2.209) by setting the time derivative to zero and discretizing, by splitting the converter into n_{cells} identical slices. Note that $\frac{\rho_{gas} \cdot V_C}{n_{cells} \cdot \dot{m}_{e,m}}$ indicates the mean residence time of the exhaust gas in the SCR cell. The resulting system now has $2 \cdot n_{cells}$ states. Simulations with varying n_{cells} show that it is usually sufficient to consider two or three cells. This reduced complexity is acceptable such that this model is a basis for the later control system design.

3

Discrete-Event Models

Discrete-event models (DEM) of selected engine subsystems are described in this chapter. The input and output signals of these systems are not defined or not relevant for all time, but only at certain discrete instances. Thus, the system behavior associated with the sampling of the signals, the delays caused by the timing relations and the interaction with the ECU, which operates in a discrete-time/event way as well, become important.

Crank-angle based representations are often used for IC engine models. This change in the independent variable from time t to crank angle ϕ has advantages for certain control problems. Specifically the feedforward action of the air/fuel ratio control system needs this special representation because it must be realized with maximum bandwidth, and, thus, is very sensitive with respect to errors in modeling the delays and timing relations.

The following items are discussed in this chapter:

- *When are DEM required?*
- *Various ECU operation modes and time scales.*
- *Precise timing of injection and ignition commands.*
- *DEM of torque production in IC engines.*
- *DEM of the air flow in IC engines.*
- *DEM of the fuel flow in IC engines.*
- *DEM of back-flow dynamics of CNG engines.*
- *DEM of residual gas dynamics.*
- *DEM of exhaust pipe dynamics.*
- *DEM based on measurements of the cylinder pressure.*

Most of these models are control-oriented *models. Only the last model represents a first step toward thermodynamically more detailed descriptions.*

One fundamental assumption adopted that underlies the derivation of all of these models is that the engine speed does not significantly change during one engine cycle. This assumption has been justified in Sect. 2.5.2. Therefore, the models, although being defined in the crank-angle domain, are formulated on the basis of constant sampling times.

L. Guzzella et al., *Introduction to Modeling and Control of Internal Combustion Engine Systems*

3.1 Introduction to DEM

3.1.1 When are DEM Required?

A model of an engine system always includes continuous-time and discrete-event subsystems. Examples of the continuous-time subsystems are the intake manifold dynamics, the acceleration of the crankshaft, the turbocharger speed dynamics, etc. Subsystems that are often modeled as discrete-event systems are the torque production, the gas exchange of the individual cylinders, the injection and the ignition processes, etc.

The ECU is another part of the complete system that is best described using a discrete-time approach. In fact, modern ECUs are always microprocessor systems running various tasks with different sampling times. Section 3.1.3 provides a short overview of the software structure implemented in such ECUs.

The derivation of continuous-time mean-value models for the various engine subsystems was shown in Chap. 2. The discrete working principle of the subsystems must be considered in the following cases:

- When the system representation, and therefore the subsequent controller realization, is simpler in the crank-angle domain.
- When the control system has to achieve bandwidths that make it necessary to take into account the synchronization problems. This is often the case for feedforward control action, since this type of controller usually requires maximum bandwidth.
- When cylinder-individual effects have to be analyzed, e.g., single-cylinder air/fuel control, misfire detection with engine speed measurements, the interpretation of cylinder-pressure measurements, etc.

Note that it is usually not possible to know at the outset which approach is better suited. A robustness analysis can help to decide whether a time or a crank-angle discretization achieves the best performance. Moreover, the available CPU power does not allow the calculation of control tasks in the so-called "synchro mode" (synchronized with the engine's crank angle, i.e., using a discrete-event approach) unless it is mandatory to do so.

3.1.2 Discrete-Time Effects of the Combustion

Due to its reciprocating action, an internal combustion engine by its very nature is a discrete-event system. In this chapter, the focus will be on four-stroke spark ignited (SI) engines. Almost identical models may be derived for two-stroke gasoline engines and Diesel engines.

A cylinder-pressure trajectory and the scaled valve-lift curves are depicted in Fig. 3.1. The timing of the different strokes is indicated as well. As the diagram clearly shows, a single-cylinder engine produces torque every two engine revolutions. This torque acts on the crankshaft during approximately

one-third of a revolution. In the remaining parts of the two revolutions, the engine rotates due to the inertia of the flywheel. This leads to very strong engine speed oscillations and rough engine behavior. Having more cylinders and distributing their torque peaks evenly over the two revolutions results in a much smoother behavior. Accordingly, in a six-cylinder engine, a cylinder fires every one-third of a revolution, resulting in a rather constant engine torque over one complete engine cycle.

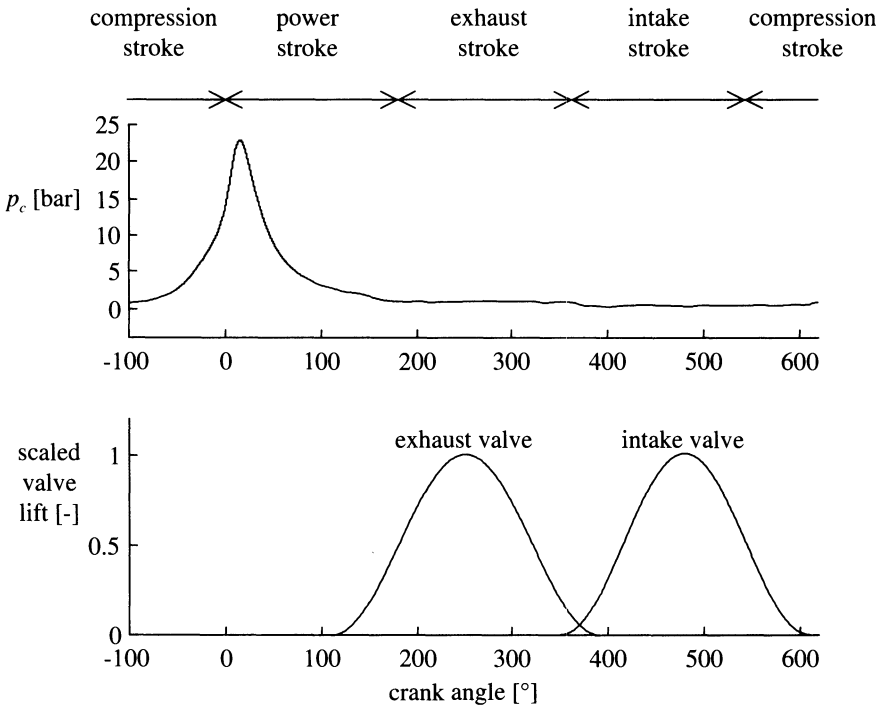


Fig. 3.1. Timing diagram of a four-stroke engine.

At a certain crank angle (for port-fuel injected Otto engines this is the crank angle where the ignition occurs) all boundary conditions for the combustion are fixed and can no longer be changed. Hence, the torque production acts as a discrete system with a sampling period of

$$\phi_{seg} = \frac{2\pi \cdot N}{n_{cyl}} \tag{3.1}$$

$$\tau_{seg} = \frac{2\pi \cdot N}{\omega_e \cdot n_{cyl}} \tag{3.2}$$

where N indicates a two-stroke ($N = 1$) or a four-stroke engine ($N = 2$), n_{cyl} is the number of cylinders of the engine, ϕ_{seg} is the sampling angle in radians, and τ_{seg} is the sampling time in seconds, usually referred to as the *segment time.*

The sampling interval is constant in the crank-angle domain and depends upon engine speed in the time domain. In discrete-time control theory, the sampling of the signals is usually assumed with constant sampling times. However, this assumption does not hold for crank-angle-based sampling. As shown in Sect. 2.5.2, changes in engine speed normally are at least a factor of ten slower than the changes in manifold pressure, which set the boundary conditions for the combustion in the cylinder. Thus, crank-angle-based sampling can be assumed to take place at evenly spaced points in time.

In four-stroke engine applications, the synchronization of any device (control system, measurement equipment or actuator) with the sampling behavior of the engine requires a crank angle *and* a camshaft position information. Usually this is accomplished by using two sensors that detect the variations in a magnetic field caused by the teeth on the outer rim of the flywheel and by the camshaft.

3.1.3 Discrete Action of the ECU

Before the advent of electronic control units, the ignition was changed continuously by a centrifugal advance mechanism (for speed adjustment) and a vacuum unit (connected to the manifold for load adaptation). Therefore, the influences of the engine speed and of the manifold pressure on the ignition angle were separate, and the characteristics of that control action were given by the shape of the mechanical cam mechanisms. The time during which the ignition current was on, building up energy in the ignition coil, was given by the adjustment of the ignition contact breaker and was constant in the crank-angle domain yielding a constant dwell angle, but a varying dwell time. No correction for low battery voltage was possible and the coils had to be short-circuit resistant, because at start-up, whenever the ignition was on, the full current was flowing in the coil for an extended period of time.

Fuel was brought into the cylinder by means of carburetors that continuously evaporated fuel into the air stream. The resulting air/fuel ratio was not constant over the entire operating region, which caused drivability problems, high pollutant emissions and fuel economy problems. Mechanical devices, i.e., accelerator pumps, various adjustable nozzles, etc., were added in an attempt to deal with these problems. However, they just made carburetors costly, difficult to adjust and unreliable.

The introduction of emission legislation required the introduction of three -way catalytic converters. As shown in Sect. 1.2.2, for these devices to work properly, a precise air/fuel ratio control system is mandatory. For that reason, carburetors with some electronically controllable elements were introduced that were able to utilize the signal of the air/fuel ratio sensor to compensate

for non-stoichiometric mixtures. However, those systems turned out to be more expensive than the new injection systems, which injected the fuel using electronically controlled solenoid valves. The first devices of that type were central fuel injection systems that injected fuel at an upstream position in the manifold from where it was distributed to several cylinders. Such a system is simple to realize but difficult to control, due to the large wall-wetting effects and the long transport time delays. The better solution proved to be a port fuel injection, i.e., a configuration in which the fuel is injected just upstream of the intake valve.

Since that approach required electronic control units with electronic power amplifiers, the next step was the integration of a solid state ignition system, which made it possible to control injection and ignition simultaneously. Around the year 1980, the first of these SI engine control units with microprocessors appeared on the market, [61]. They combined the control of ignition and injection in one control unit, and were able to implement more complex control algorithms.

The core of all ECUs is a microcontroller. This device runs in a discrete-event mode, i.e., performing some of its tasks repetitively in constant crank angle intervals. In addition, the ECU runs several tasks using fixed sampling time intervals. A real-time operating systems coordinates these various tasks and resolves computation-time conflicts on the basis of predefined priorities (see below).

Two operating modes are possible for an ECU:

- *Fully Synchronized Operating Mode*: The ECU processor continuously counts pulses from a sensor that picks up the variations in the magnetic field caused by the teeth on the outer rim of the flywheel (60 teeth is a common number) and compares the demanded values for injection and ignition crank angle with the actual values of the crank angle. This operation mode guarantees minimal additional delays. On the other hand, a great deal of precious processor time is spent on pulse counting. Therefore, modern ECUs avoid this fully synchronized operation mode.
- *Asynchronous Operation Mode*: The ECU processor has a dedicated slave processor, often referred to as a time-processing unit (TPU), which relieves the ECU from the mentioned synchronization tasks [64]. In fact, the microprocessor sends the new values for injection and ignition timing asynchronously to the TPU. The TPU obtains crank angle and cam shaft pulses from the engine and triggers the injection and ignition events at the correct crank angle values. This operation mode frees up more time for "intelligent" tasks performed by the processor. In production-type ECUs, these interface chips are directly mounted on the ECU board or, in more recent systems, are integrated on the same chip as the microcontroller. For research purposes, such TPUs with a very high degree of flexibility have been designed using field programmable gate arrays [56].

Even the most modern engine ECUs have processors with clock rates that are much slower than the processors of modern desktop computers. This is due to the harsh environmental conditions in which they must work reliably, e.g., temperatures ranging from $-40°C$ to $+100°C$, high humidity and electromagnetic fields (radio signals, GSM, ignition system, etc.). A consequence of these slow clock rates is that the computing resources must be carefully managed. This is achieved by utilizing several clock rates:

- 1 ms: very fast tasks, sensor signal post processing, etc.;
- 20 ms: normal time scale for control functions;
- 100 ms: slow (adaptive) functions, temperature effects, etc;
- 1 segment (3.1): control functions which have to be synchronized with the engine's reciprocating behavior; and
- 1/2 segment or even faster: data processing for special tasks, i.e., cylinder-individual air/fuel control, cylinder-pressure interpretation, etc.

Computational power is limited such that every task must be run in exactly the time frame that is required by the physical constraints, not faster. Every task has a priority, which enables the operating system to resolve conflicts in case the computational resources are not sufficient. At high engine speeds, due to the large amount of calculation time spent in the high-priority segment tasks, the low-priority tasks are hardly ever executed.

This aspect can be illustrated by considering the available computation intervals at different engine speeds and for various crank-angle intervals. Table 3.1 lists the corresponding values for engine speeds ranging from 600 *rpm* to 6000 *rpm*. These values show that at low engine speeds there is sufficient time for many calculations, but the time delays associated with the crank-angle domain become very large rendering the feedback control problems hard to solve. On the other hand, at high speeds, these time delays are short but the time available for computations is so small that not many tasks can be run simultaneously.

Table 3.1. Time domain values for engine events.

	crank angle	$n_e = 600$ *rpm*	$n_e = 6000$ *rpm*
engine cycle (4-stroke)	4π	200 ms	20.0 ms
engine revolution	2π	100 ms	10.0 ms
engine segment (6-cylinder)	$\frac{2\pi}{3}$	33 ms	3.3 ms
engine segment (4-cylinder)	π	50 ms	5.0 ms

3.1.4 DEM for Injection and Ignition

The primary task of the ECU is to guarantee the injection of the correct amount of fuel into the manifold and the ignition of the mixture in the cylinder at the desired crank-angle position with sufficient energy.

The fuel-injection module is analyzed in a first step. The timing of the injection is derived from Fig. 3.2. The lift curves of the intake and the exhaust valves of an engine are assumed to be constant in the crank-angle domain (with variable-valve timing this is not valid, of course). Both valves are open for approximately two-thirds of an engine crankshaft revolution. It is advantageous to inject the fuel on a closed valve[1] such that sufficient time for fuel evaporation is available. Moreover, since the intake valve is very hot compared to the intake runners, the evaporation of the injected fuel is further improved.

Following this approach, the time available for fuel injection at 6000 *rpm* amounts to approximately 13 ms. This value should not be exceeded except for very harsh transients. At idling, the other extreme, a minimal amount of fuel must be injected. Opening and closing the fuel injector takes approximately 1 ms, a quantity that strongly depends upon manufacturing tolerances, wear, bouncing and fuel quality. To prevent large variations in the amount of injected fuel, the minimum injection time should not be smaller than approximately 2 ms. These considerations determine the dimensions of the injection valve. Obviously, a trade-off has to be made between low hydrocarbons at full load and fueling precision at idle, i.e., the more fuel that needs to be injected, the more critical the manufacturing tolerances become at idle.

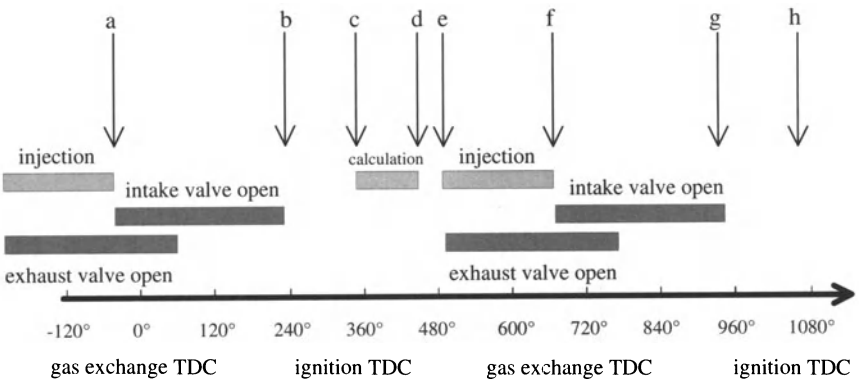


Fig. 3.2. Timing diagram for the injection.

The timing diagram for the injection depicted in Fig. 3.2 will now be discussed in detail.

[1] Injecting fuel directly into the cylinder through the open valve would increase hydrocarbon emissions.

- In normal operation, the injection is completed at point **a**. In the ECU, this event is fixed. The start of injection is set at the necessary crank angle, which is computed using the air mass flow information obtained at an earlier crank angle.
- Between the points **a** and **b**, the air mass and the fuel mass flow into the cylinder. Changes in these mass flows into the cylinder during that period can no longer be corrected with a regular injection anymore.
- In normal operation, point **b** is the earliest possible start of injection for the next cycle. For high loads, the resulting injection duration must be sufficient.
- At point **c**, the measurements for the calculation of the duration of the injection are made. This event is usually triggered at all crank-angle intervals ϕ_{seg} (3.1). This is the relevant event for the calculation of the injection-to-torque delay, because the measurement signals valid at **c** are responsible for the combustion at **h**. Between **c** and **d**, all calculations associated with the injection control algorithms are performed. This event will be referred to as the *update injection event*.
- At **d**, new values for injection duration (as well as ignition timing) are sent to the time processing unit (TPU).
- Until **e** is reached no new information for the injection timing is available. An average delay between **d** and **e** of one-half sample interval τ_{sample} can be assumed, reflecting that an even distribution of the time delays between zero (the calculations finished just in time and the injection starts immediately after the update event) and one sample interval (the calculations required too much time such that the result is transferred to the TPU only in the next update injection event) may be expected.
- At point **e**, the injection with the most recent air mass information starts.
- Depending on the ECU, between **e** and **f**, the injection pulse can be extended if necessary, but only if critical air/fuel ratios are to be expected (i.e., misfire). Most ECUs, on the other hand, allow an immediate closing of the injection valve.
- At point **f**, the regular injection pulse is finished. During very heavy transients, some ECUs offer the possibility of injecting a small amount of fuel onto the open intake valve. Of course, fuel can only be added, thus if the air mass is reduced, very rich conditions must be accepted.
- After **g**, the amounts of air and fuel in the cylinder are fixed and the combustion starts at **h**.

Summarizing these considerations, the time delay between the start of the calculation (update injection event) and the injection start event can be approximated by

$$\tau_{c)-e)} = \tau_{calculation} + \frac{1}{2} \cdot \tau_{sample} \tag{3.3}$$

For the ignition, similar considerations exist, see Fig. 3.3. Modern ignition coils must be connected to the full battery voltage for approximately 4 ms

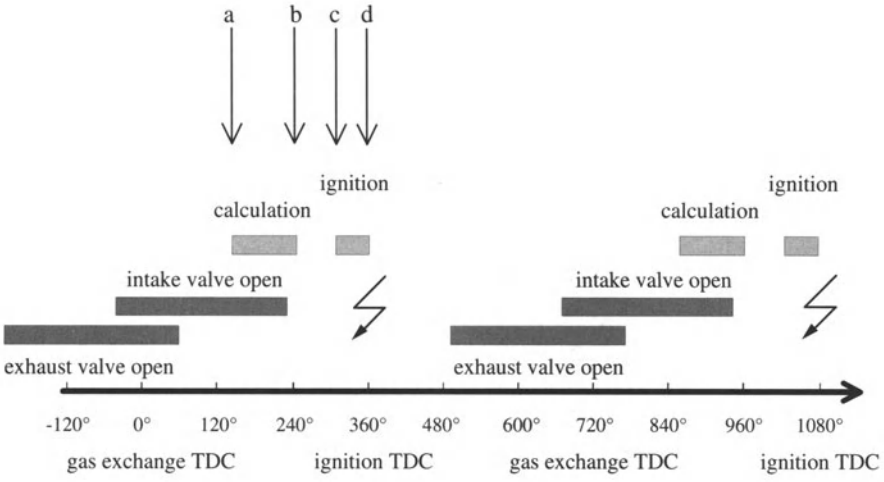


Fig. 3.3. Timing diagram for the ignition.

to build enough energy for a sufficient spark. If the coil is connected to the battery voltage for more than 10 ms it may suffer irreversible damage.

- At point **a**, the measurements of the engine variables necessary for the determination of the ignition timing are triggered by the ECU. This point will be referred to as the *update ignition event* below. In general, this will not be the same crank angle at which the update injection event takes place.
- Calculations are performed up to point **b**, at which the calculated values are sent to the TPU.
- Again, until ignition starts, an additional time delay occurs. For this delay, a value of one-half of the sample time may be assumed. At **c**, the coil is connected to the battery terminals.
- At **d**, the ignition (or spark advance) angle, the circuit breaker is opened causing the spark to start. The time between **c)** and **d)** is the dwell angle. Both variables are chosen such that the thermodynamic conditions of combustion are optimized.

Changing the spark advance between **c** and **d** is possible within small limits.[2] The case of missing a spark event must be strictly avoided because such misfires can cause severe damage to the TWC. Accordingly, increasing the spark advance will always be more critical.

The calculation of the spark advance must be done considering the available information on engine speed, air mass in cylinder, knock onset during

[2] Increasing the dwell time up to 10 *ms* for one ignition is possible, but such large values may not be sustained for longer periods without causing damage to the ignition coils.

the last combustion event, etc. Since the dwell-time is rather short and the ignition takes place during the compression of the gases in cylinder, the ignition has a smaller inherent delay than the injection. The ignition is thus the fastest actuator for a port-fuel injected SI engine.

3.2 The Most Important DEM in Engine Systems

3.2.1 DEM of the Mean Torque Production

In a mean-value setting, all the delays present in the torque production were lumped into the IPS time delay (2.1). This section presents a more detailed analysis of the reciprocating torque generation. The main idea is to identify for each relevant subsystem x a characteristic crank angle ϕ_x, at which the actual boundary conditions (manifold pressure, air mass flow, etc.) must be sampled. If this information permits to predict the torque production, a crank-angle-based DEM of the corresponding subsystem can be formulated.

Note that this torque is still assumed to be constant over one engine segment, i.e., the models presented in this section are only able to predict mean-torque values. If the variations in engine speed due to the pulsating combustion are of interest, the crank-angle-resolved cylinder pressure and its effect on the cranking mechanism must be taken into account. For this purpose, some simple models are introduced in Sect. 3.3 .

To model the mean torque taking into account the various delays present in an engine system, a characteristic crank-angle center position is defined for each stroke:

- Intake Stroke: The intake center corresponds to the crank angle at which the relevant air mass flow enters the cylinder.
- Compression Stroke: The events relevant to the ignition process, as discussed in Sect. 3.1.4, take place.
- Power Stroke: The torque center represents the crank angle of the relevant torque production.
- Exhaust Stroke: Similar to the intake center, the exhaust center is defined as that crank angle at which the relevant exhaust gas mass flow leaves the cylinder.

The crank-angle position of these events is not easy to determine. For the intake and exhaust center, a first approximation of $\pm 90°$ crank angle before or after TDC can be given (roughly where the piston moves with maximum speed).[3]

The location of the intake and exhaust center is influenced by the cylinder pressure and therefore by the operating point, by the geometry of the cranking mechanism and by the engine speed. Accordingly, the position of the intake,

[3] The experiments described below show that a value of $\pm 110°$ is more realistic.

the exhaust and the torque center in the crank-angle domain must be determined by experiments. Since only time delays can be measured, but not crank-angle delays, the experiments must be conducted on a test bench with a high-bandwidth engine-speed controller. Once the time delays are identified, they can be converted to crank angles by considering the engine speed.

Measuring Delays

To determine the torque-center crank angle, the input variables influencing the torque production must first be analyzed. In Sect. 2.5.1 it was shown that the following equation can be used to approximate the engine torque

$$T_e = \frac{H_l}{4\pi} \cdot m_\varphi \cdot \eta_{e0}(m_\varphi, \omega_e) \cdot \eta_\lambda(\lambda) \cdot \eta_\zeta(\zeta) \cdot \eta_{egr}(x_{egr}) \tag{3.4}$$

Accordingly, the following "inputs" can be used to trigger changes in the engine torque:

- the air mass in the cylinder;
- the fuel in the cylinder;
- the spark advance; and
- the exhaust gas recirculation rate (internal and external).

The signal flowchart of the corresponding system is depicted in Fig. 3.4. The dynamics of the different blocks can be determined by keeping all inputs with the exception of one constant. As discussed in Appendix A, frequency response measurements are a convenient tool to determine the time delays for linear or linearized systems.

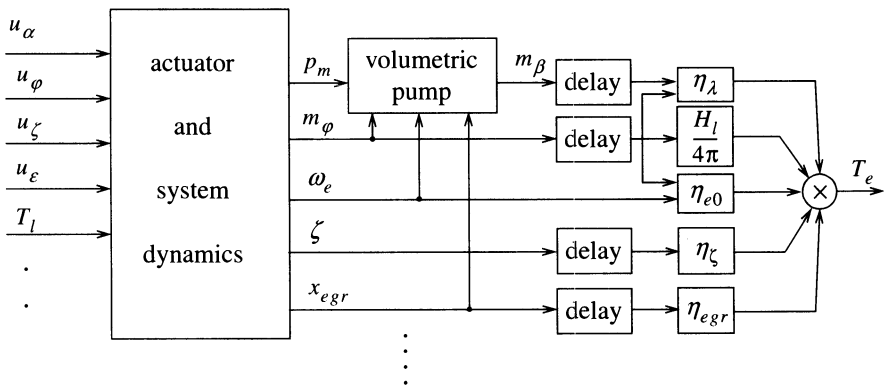


Fig. 3.4. Block diagram of the torque production considering the delays in the signal paths.

As mentioned above, the only input which can be assumed to influence the torque production almost instantaneously is the spark signal. Measured

frequency responses with the spark signal as input and the engine torque as output are therefore the best choice to analyze the time delay between the ignition update event (chosen by the ECU, as explained in Sect. 3.1.4) and the torque center. Of course, using these measurements, the engine speed, the throttle position, the EGR valve position and the injection duration must be kept constant.

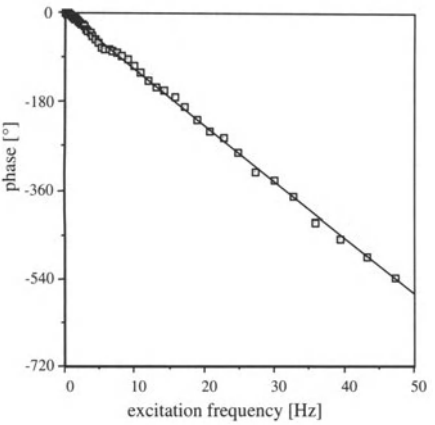


Fig. 3.5. Phase of the frequency response measurement from ignition input u_ζ to engine torque output T_e. Simulation (solid) and measurement (squares indicating the excitation frequencies at which measurements were done).

Figure 3.5 shows the phase plot of such a frequency response measurement. The spark angle is changed sinusoidally while the update ignition event is kept constant. Thus, all variations in engine torque phasing are due to the delay between update ignition event and torque center. The results shown in Fig. 3.5 confirm that the dynamics of this subsystem can be approximated well by a single delay element. In fact, only a pure time delay shows the constant phase roll-off over the excitation frequency ($\varphi = -\omega \cdot \tau_{delay}$) shown in that plot. From this data, the time delay, and, with the knowledge of the engine speed, the crank-angle delay, can be determined using the following relationship

$$\tau_{delay} = \frac{\Delta\varphi(\Delta f)}{360 \cdot \Delta f} \tag{3.5}$$

where the phase lag φ is assumed to be given in degrees and the excitation frequency f of the spark angle modulation in Hz.

For the example shown in Fig. 3.5, an update ignition event at 90° crank angle before TDC and an engine speed of 900 rpm is chosen. Using (3.5) a $\tau_{delay} \approx 31\ ms$ is identified, which, for the engine speed mentioned, is equivalent to a crank angle of 170°. Accordingly, for the torque center, a characteristic crank angle of 80° after TDC is obtained.

Once the position of the torque center is known, the intake center crank-angle position can be determined with a frequency response measurement that modulates the throttle position and utilizes the manifold pressure as input signal and the engine torque as output signal. Obviously, this approach relies on the assumption that, for constant engine speeds, the air mass in the cylinder is a static function of the manifold pressure (see Sect. 2.3). Similarly, mounting a fast exhaust gas sensor as close as possible to the exhaust valve and exciting the spark advance, the exhaust center can be determined.

Table 3.2 summarizes a set of measured values for these engine parameters. The values given in that table are valid at idle conditions. They are influenced by the cylinder pressure and valve timing and, therefore, by the operating point. Additionally, they can be altered by the algorithms implemented in the ECU by changing the ignition timing and, in variable-cam systems, the cam phasing.

Table 3.2. Timing events of a four-stroke engine at idle speed.

engine event	*position in ° crank angle after ignition TDC*
intake center (IC)	470° (110° after TDC)
torque center (TC)	80°
exhaust center (EC)	250° (110° before TDC)
update ignition (U_{ign}) update injection (U_{inj})	defined in the ECU $\phi_{update} = \phi_{offset} + \phi_{seg} \cdot k$ $k = 0, 1, 2, 3, 4, \ldots$

Using the numerical values given in Table 3.2 to localize the events discussed above, yields the final timing diagram shown in Fig. 3.6. For the chosen engine speed (idling), an induction-to-power-stroke delay of 330° and an induction-to-exhaust delay of 500° result, which corresponds well to the values introduced in Sect. 2.1, where $\tau_{ips} \approx \frac{2\pi}{\omega_e}$ and $\tau_{eg} \approx \frac{3\pi}{\omega_e}$ were given as approximations.

Time Domain Description

Taking into account all delays introduced so far, the engine torque can be approximated by

$$T_e(t) = \frac{H_l}{4\pi} \cdot m_\varphi(t - \tau_{Uinj \to TC}) \cdot \eta_{e0}(m_\varphi(t - \tau_{Uinj \to TC}), \omega_e(t))$$

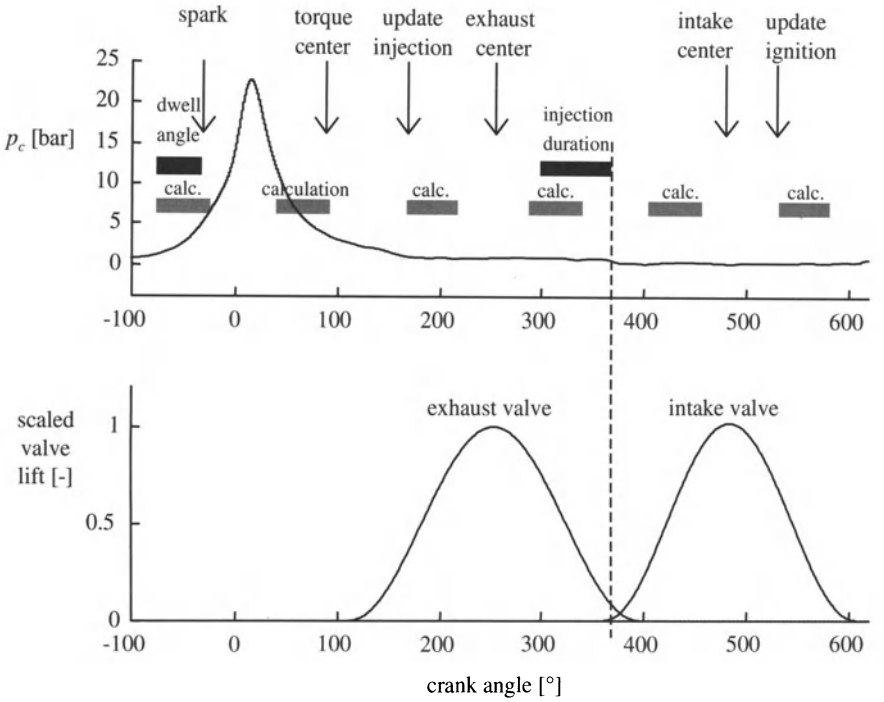


Fig. 3.6. Timing diagram of a four-stroke six-cylinder engine. Note that the calculation is performed every segment, in this case every third of a revolution. Depending on injection duration, spark advance and dwell angle the relevant calculation event may vary.

$$\cdot\eta_\lambda\left(\frac{m_\beta(t-\tau_{IC\to TC})}{m_\varphi(t-\tau_{Uinj\to TC})}\cdot\frac{1}{\sigma_0}\right)\cdot\eta_\zeta(\zeta(t-\tau_{Uign\to TC}))$$
$$\cdot\eta_{egr}(x_{egr}(t-\tau_{tra}-\tau_{IC\to TC})) \tag{3.6}$$

In this formulation, the two delays between the update-injection to intake-center event and the intake-center to torque-center event are lumped into a single delay $\tau_{Uinj\to TC}$.

Using the numerical values indicated in Table 3.2, the delays introduced above can be written as

$$\tau_{IC\to TC} \approx \frac{330\,^\circ}{360\,^\circ}\cdot\frac{2\pi}{\omega_e} \tag{3.7}$$

$$\tau_{Uinj\to TC} = \tau_{calc} + \frac{1}{2}\tau_{sample} + \tau_{inj} + \tau_{IVO\to IC} + \tau_{IC\to TC} \tag{3.8}$$

$$\tau_{Uign\to TC} = \tau_{calc} + \frac{1}{2}\tau_{sample} + \tau_{dwell} + \tau_{sp\to TC} \tag{3.9}$$

where the following abbreviations have been used:

sp	spark
IVO	intake valve opening
calc	calculation
τ_{tra}	transport delay from EGR valve to intake manifold

Crank Angle Domain Description

Changing to a crank-angle domain formulation, the corresponding delays must be divided by the segment time $\tau_{seg} = (2\pi \cdot N)/(\omega_e \cdot n_{cyl})$.

In this formulation, the advantages of a crank-angle-based representation become obvious. In fact, in this domain most delays are constant. For instance, assuming a six-cylinder engine and neglecting the calculation times, the dwell time, and the injection duration (which is the most critical assumption but holds well for idle) the following discrete event formulation can be derived for the torque production at idle

$$\begin{aligned} T_e(t_k) = \frac{H_l}{4\pi} \cdot m_\varphi(t_{k-5}) \cdot \eta_{e0}(m_\varphi(t_{k-5}), \omega_e(t_k)) \\ \cdot \eta_\lambda\left(\frac{m_\beta(t_{k-3})}{m_\varphi(t_{k-5})} \cdot \frac{1}{\sigma_0}\right) \cdot \eta_\zeta(\zeta(t_{k-2})) \\ \cdot \eta_{egr}(x_{egr}(t_{k-3} - \tau_{tra})) \end{aligned} \tag{3.10}$$

In this equation, the delays have been rounded up to the next higher integer multiple of the sampling time. Of course, for other operating points the delays will be different. Moreover, for high loads and high engine speeds it is not possible to neglect the injection duration.

3.2.2 DEM of the Air Flow Dynamics

Besides the fuel flow, the air flow of an engine is the most important subsystem to be considered when designing an open-loop air/fuel ratio control system. The best possible drivability and the lowest possible emissions are only realized if a detailed model of the air flow dynamics is used in that design problem.

As shown in Sect. 2.3, the mean value model for the manifold dynamics can be described by the following equations (no EGR is considered here)

$$\frac{d}{dt} m_m(t) = \dot{m}_\alpha(t) - \dot{m}_\beta(t) \tag{3.11}$$

An approximate description of the mass flow $\dot{m}_\beta(t)$ entering the cylinder is given by

$$\dot{m}_\beta(t) = \rho_m(t) \cdot \dot{V}(t) = \rho_m(t) \cdot \lambda_l(p_m(t), \omega_e(t)) \cdot \frac{V_d}{N} \cdot \frac{\omega_e(t)}{2\pi} \tag{3.12}$$

For convenience, the definition (2.20) of the volumetric efficiency has been used here.[4]

Using (3.12) and $m_m = \rho_m \cdot V_m$, it is possible to rewrite (3.11) as follows

$$\frac{d}{dt}\left(\frac{\dot{m}_\beta(t)}{\lambda_l(p_m(t),\omega_e(t))} \cdot \frac{V_m}{V_d} \cdot N \cdot \frac{2\pi}{\omega_e(t)}\right) = \dot{m}_\alpha(t) - \dot{m}_\beta(t) \tag{3.13}$$

This equation must be integrated over the time interval of one engine segment yielding

$$\int_{t_{k-1}}^{t_k} \frac{d}{dt}\left(\frac{\dot{m}_\beta(t)}{\lambda_l(p_m(t),\omega_e(t))} \cdot \frac{V_m}{V_d} \cdot N \cdot \frac{2\pi}{\omega_e(t)}\right)dt \tag{3.14}$$

$$= \int_{t_{k-1}}^{t_k} \dot{m}_\alpha(t)dt - \int_{t_{k-1}}^{t_k} \dot{m}_\beta(t)dt \tag{3.15}$$

$$\tag{3.16}$$

where t_k is equal to $k \cdot \tau_{seg}$ and $k = 1, 2, 3, \ldots$ is the engine segment counter. Obviously this equation can be reformulated as follows

$$m_{\alpha,k} - m_{\beta,k} = \frac{V_m}{V_d} \cdot N \cdot \frac{\dot{m}_\beta(t_k)}{\lambda_l(p_m(t_k),\omega_e(t_k))}\frac{2\pi}{\omega_e(t_k)} \tag{3.17}$$

$$-\frac{V_m}{V_d} \cdot N \cdot \frac{\dot{m}_\beta(t_{k-1})}{\lambda_l(p_m(t_{k-1}),\omega_e(t_{k-1}))}\frac{2\pi}{\omega_e(t_{k-1})}$$

The symbol $m_{\alpha,k}$ represents the mass of air flowing through the throttle during one segment. Similarly, the symbol $m_{\beta,k}$ represents the mass of air entering the cylinder during the event k. The corresponding mass flows $\dot{m}_\ldots(t)$ do not necessarily have to be constant. It is only assumed that they are periodic with period τ_{seg} such that the integration is independent of the location of the sampling event.

Since $\dot{m}_\beta$ and ω_e vary much slower than $\dot{m}_\alpha$, an Euler integration is usually acceptable for the computation of the mass entering the cylinder

$$m_k = \dot{m}_\beta(t_k) \cdot \tau_{seg} \tag{3.18}$$

The variable $m_{\alpha,k}$ is the forcing term in the manifold system description (3.17). If the throttle angle command u_α is the actual input signal, the mass flow through the throttle must be calculated using (2.10) and numerical integration. However, often the mass flow $\dot{m}_\alpha$ entering the manifold is measured with a hot film anemometer. In that case, this measurement can be used to calculate $m_{\alpha,k}$ directly using, again, a suitable numerical integration method.

[4] The error that has to be expected thereby is in the order of 5% for gasoline fuels. If the engine is fueled by natural gas, the error would be larger. In this case, the more precise, but also more cumbersome, approximation (2.19) is recommended.

Therefore, in the following text the variable $m_{\alpha,k}$ will be taken as the input signal that drives the manifold dynamics.

Combining (3.17) and (3.18) yields

$$m_{\alpha,k} - m_{\beta,k} = \frac{(m_{\beta,k} - m_{\beta,k-1})}{\lambda_l(p_m, \omega_e)} \frac{V_m}{V_d} \cdot N \cdot \frac{2\pi}{\omega_e} \cdot \frac{1}{\tau_{seg}} \tag{3.19}$$

$$= \frac{(m_{\beta,k} - m_{\beta,k-1})}{\lambda_l(p_m, \omega_e)} \cdot \frac{V_m}{V_d} \cdot n_{cyl} \tag{3.20}$$

and introducing the parameter

$$X = \lambda_l(p_m, \omega_e) \cdot \frac{V_d}{V_m} \cdot \frac{1}{n_{cyl}} \tag{3.21}$$

a discrete event formulation for the mass m_k entering the cylinder during the k-th engine cyle is obtained

$$m_{\beta,k} = m_{\beta,k-1} \cdot \frac{1}{1+X} + m_{\alpha,k} \cdot \frac{X}{1+X} \tag{3.22}$$

Remarkably, assuming a slowly varying volumetric efficiency, the manifold dynamics become invariant in the crank-angle domain.

The last equation is the zero-order hold discrete-time equivalent of a continuous-time first-order lag element with gain equal to one (see Appendix A). The time constant τ of that continuous-time element is related to the parameter X by the following well-known relationship

$$e^{-\tau_{seg}/\tau} = \frac{1}{1+X} \tag{3.23}$$

Assuming that τ_{seg} is much smaller than the time constant τ, the following approximation is valid

$$\frac{1}{1+X} \approx \frac{1}{1+\frac{\tau_{seg}}{\tau}} \quad \Rightarrow \quad \tau \approx \frac{\tau_{seg}}{X} \tag{3.24}$$

and using the definition (3.21) of the parameter X, the following expression for the equivalent continuous-time time constant τ is obtained

$$\tau \approx \tau_{seg} \cdot n_{cyl} \cdot \frac{V_m}{V_d} \cdot \frac{1}{\lambda_l(p_m(t), \omega_e(t))} \tag{3.25}$$

Assuming $\lambda_l = 1$, this time constant corresponds to the time of one segment multiplied by the size of the manifold volume relative to the cylinder displacement. Moreover, inserting the definition of the segment time (3.2) into the last equation the following expression for τ is obtained

$$\tau \approx \frac{4\pi}{\omega_e} \cdot \frac{V_m}{V_d} \tag{3.26}$$

that is equal to the manifold time constant $\widetilde{\tau}_p$ (2.133) obtained in Sect. 2.5.2 in a rather different context.

3.2.3 DEM of the Fuel-Flow Dynamics

Introduction

The fuel injection in a sequentially injected SI engine is inherently synchronized with the engine's crank angle. The correct identification of the delays present in this loop is important for the design of the feedforward and the feedback action of the corresponding control system. A simple but useful discrete-event formulation of the fuel dynamics will be shown below.

The primary question that must be answered in this section is how the single command issued by the fuel-control system is branched to the n_{cyl} different cylinders of the engine. Similarly, the same cylinders produce n_{cyl} different air/fuel ratio signals that are all recorded by only one air/fuel ratio sensor. Of course, one approach would be to treat all cylinders individually. This solution offers the greatest flexibility but imposes a large computational burden. The models of the fuel-flow dynamics presented in this section offer an optimal compromise between computational efforts and performance of the control loop designed based on these models.

Multi-Cylinder Engines as Multiplexed Systems

In a first step, a single cylinder of the engine is analyzed. The mean-value model of the injection dynamics derived in Sect. 2.4.2 assumes a continuous flow of fuel into the cylinders. A discrete-event formulation, which predicts the total amount of fuel that enters the cylinder in the k-th cycle, can be obtained by using a $\mathcal{Z}$-transformation approach. Assuming first-order fuel dynamics (2.65)

$$G(s) = (1 - \kappa) + \kappa \cdot \frac{1}{s\tau + 1} \tag{3.27}$$

and a constant engine speed, the following discrete-time system description can be found for one cylinder

$$G(z) = (1 - \kappa) + \kappa \cdot \frac{1 - e^{-\frac{n_{cyl}\tau_{seg}}{\tau}}}{z - e^{-\frac{n_{cyl}\tau_{seg}}{\tau}}} \tag{3.28}$$

Note that in this step the relevant delay is the *cycle* time defined by the expression

$$\tau_{cycl} = n_{cyl} \cdot \tau_{seg} \tag{3.29}$$

In the second step, all cylinders of the engine are analyzed simultaneously. Of course, in a port-injected engine the fuel path includes n_{cyl} separate wall-wetting dynamics. Similarly, the gas mixing dynamics of those parts of the exhaust manifold immediately after the exhaust valves evolve in parallel as well. These dynamic elements obtain their inputs in a sequential way synchronized to the crank angle, and they deliver their output signals in the same manner.

On an abstract level, such a multi-cylinder system can be modeled as a *multiplexed system*. Using this approach, a multiplexing element is assumed to act at the input and at the output of the system. Figure 3.7 shows the structure of such a system, including the necessary sample-and-hold elements.

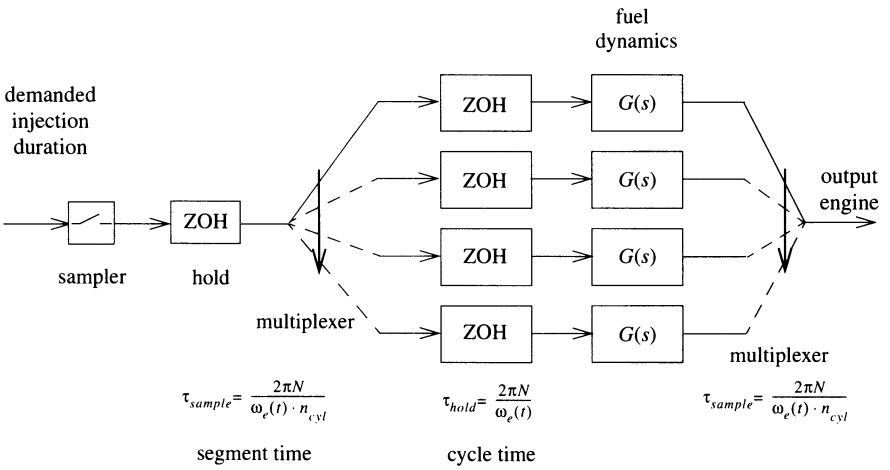


Fig. 3.7. Continuous-time fuel dynamics of a port-fuel-injected four-cylinder engine.

The input signal is sampled with the segment time τ_{seg}. Accordingly, in each segment a new value enters the multiplexer. In accordance with the chosen firing sequence, these values are distributed to the different injection valves. The individual wall-wetting dynamics are synchronized with the engine cycle time τ_{cycl} defined in (3.29). Therefore, the block diagram seen in Fig. 3.7 can be redrawn as shown in Fig. 3.8. The input and output of that system are continuous-time signals. Therefore, there must be a continuous-time transfer function $G_{total}(s)$ that describes the dynamics of that module. Again assuming a first-order fuel dynamics for each cylinder, the system depicted in Fig. 3.8 is represented by the following transfer functions

$$G_{total}(s) = G_{hold}(s) \cdot G_{fd}(z)|_{z=e^{s\tau_{seg}}} \tag{3.30}$$

$$= \frac{1 - e^{-s\tau_{seg}}}{s\tau_{seg}} \cdot \left((1-\kappa) + \kappa \cdot \frac{1 - e^{-\frac{n_{cyl}\tau_{seg}}{\tau}}}{z^{n_{cyl}} - e^{-\frac{n_{cyl}\tau_{seg}}{\tau}}} \right) \Bigg|_{z=e^{s\tau_{seg}}}$$

In this expression, the sample-and-hold block shown in Fig. 3.7 is represented by $G_{hold}(s)$. The second term represents the wall-wetting dynamics. The only difference to the single-cylinder case is that the z-variable in the single-cylinder formulation (3.28) must be substituted with $z^{n_{cyl}}$ because the sampling in the multi-cylinder case is n_{cyl} times faster than in the single-

cylinder case. Accordingly, the single cylinder, which still operates on the same cycle length as before, must be delayed by n_{cyl} sampling intervals.

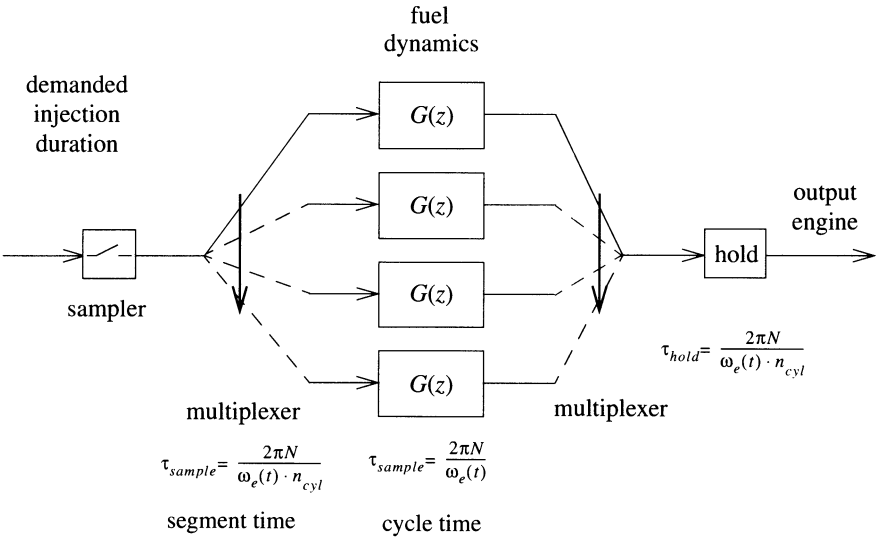


Fig. 3.8. Discrete fuel dynamics of a port-fuel-injected four-cylinder engine.

Assuming a four-cylinder engine, the state-space description of this system has the form

$$x_{k+1} = \begin{bmatrix} 0 & 0 & 0 & F \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot x_k + \begin{bmatrix} 1-F \\ 0 \\ 0 \\ 0 \end{bmatrix} \cdot u_k \qquad (3.31)$$

$$y_k = \begin{bmatrix} 0 & 0 & 0 & \kappa \end{bmatrix} \cdot x_k + (1-\kappa) \cdot u_k$$

where $F = e^{\frac{-n_{cyl} \cdot \tau_{seg}}{\tau}}$ is the only non-trivial parameter. For each new segment, a new fueling value is calculated, discretizing the associated continuous-time description with one engine cycle. Then this value is shifted $(n_{cyl} - 1)$ times until it reaches the output. This permits each cylinder to operate with its own fuel dynamics. Therefore, defining $a = 1 - F$ the block diagram of this discrete system can be drawn as shown in Fig. 3.9.

Figure 3.10 shows the response of the system (3.31) to a step input (left plot) and to a harmonic input (right plot). The first plot suggests that the system includes a sample-and-hold element with a delay equal to $n_{cyl} \cdot \tau_{seg}$, while in the second plot the same system seems to include a phase lag and a much shorter sample-and-hold delay equal to τ_{seg}. This is easy to explain with the system description (3.31):

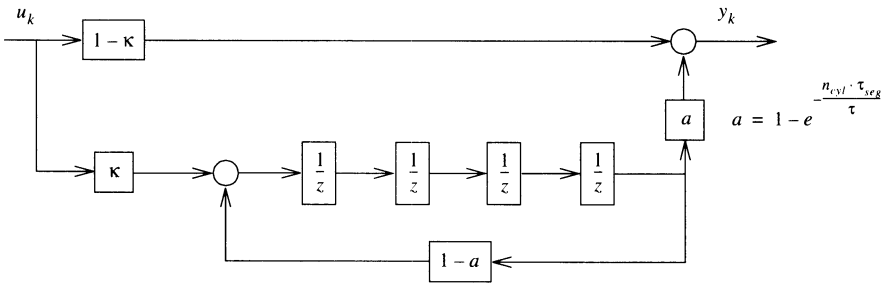


Fig. 3.9. Block diagram of the discrete-time dynamics of the fuel system.

- When a step-wise input is applied, all cylinders obtain the same new values for the injection timing such that the output is similar to the response of a single-cylinder engine.
- When a sinusoidal input is applied, the injection command changes from segment to segment. Accordingly, each of the output signals changes in the same way.

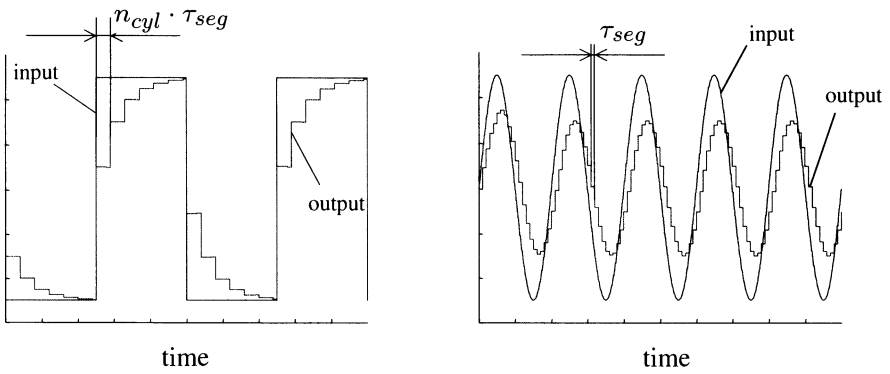


Fig. 3.10. Simulation of the multiplexed system.

Analysis of the Complete System

Many dynamic effects in the fueling system can be best understood by measuring frequency responses. The following measurements were all conducted on a 1.8-liter, four-cylinder, four-valves-per-cylinder, SI engine. Figure 3.11 shows a typical measurement setup for such an experiment.

Frequency response measurements are chosen because they permit the correct interpretation of the aliasing effects that are caused by the discrete-event behavior. Although the system is sampled in a segment mode, the individual fuel dynamics are excited only once every engine cycle. Therefore, for

the identification of the fuel dynamics, only frequencies up to $\omega_e/4$ satisfy the Shannon theorem, by which the slowest sampling defines the limits. The input modulation frequency can be chosen higher than this limit, and the resulting frequency response must be matched by the system model derived above using physical first principles, provided the latter is sufficiently close to the real system. Of course, the inverse is not true, i.e., at higher frequencies it is not possible to find a unique mathematical model using only the measured frequency response. The ambiguities introduced by the sampling process at frequencies higher than the Nyquist frequency $\omega_e/4$ preclude this.

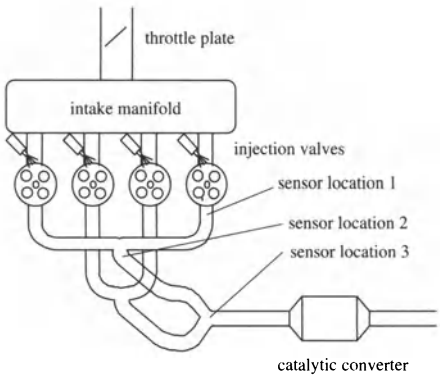


Fig. 3.11. Measurement setup for the fuel path investigations.

Depending on the output chosen the fuel dynamics incorporate different subsystems:

- *Engine Torque:* The fuel dynamics contain wall-wetting phenomena and back flow from the cylinder into the intake manifold.
- *Air/Fuel Ratio:* The fuel dynamics contain wall-wetting phenomena, back flow from the cylinder into the intake manifold, as well as gas mixing in the exhaust manifold and transport delays between the actuator and sensor locations.

The engine used in the subsequent tests is equipped with a symmetric 4-in-2-in-1 exhaust manifold, as shown in Fig. 3.11. The main advantage of this configuration is that the gas mixing dynamics may be assumed to be identical for all cylinders. Output measurements were made at three locations with a BOSCH LSU 4 wide-range air/fuel sensor, which can be modeled well as a first-order lag with a time constant of approximately 20 ms. All measurements were performed at 2000 rpm and 50% of full load. As shown in Fig. 3.12, the chosen input signal is the fuel command u_φ, while the output is the signal $y_{\lambda,i}$ of one of the three wide-range air/fuel ratio sensors.

The structure of the complete system consisting of several delays, the wall-wetting dynamics, the sensor dynamics and the gas mixing dynamics, is

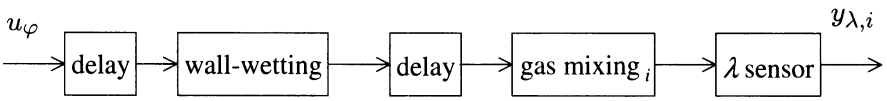


Fig. 3.12. Main dynamic subsystems in the fuel dynamics of an SI engine.

shown in Fig. 3.12. Such an input-output setting of course does not permit the identification of the contribution of any single module to the total frequency response of a system. Therefore, the individual sub-blocks of the fuel path of an SI engine must be identified and validated sequentially.

The dynamic behavior of the torque generation has been described in Sect. 3.2.1 such that in this section the focus is on the air/fuel ratio dynamics. As a first step, the air/fuel ratio immediately after the exhaust valves is chosen as the output signal. To avoid any gas mixing dynamics, the corresponding sensor must be placed as close as possible to one of the exhaust valves. Figure 3.13 shows the Bode plot of the frequency response from the injection quantity u_φ to the corresponding air/fuel ratio signal $y_{\lambda,1}$.

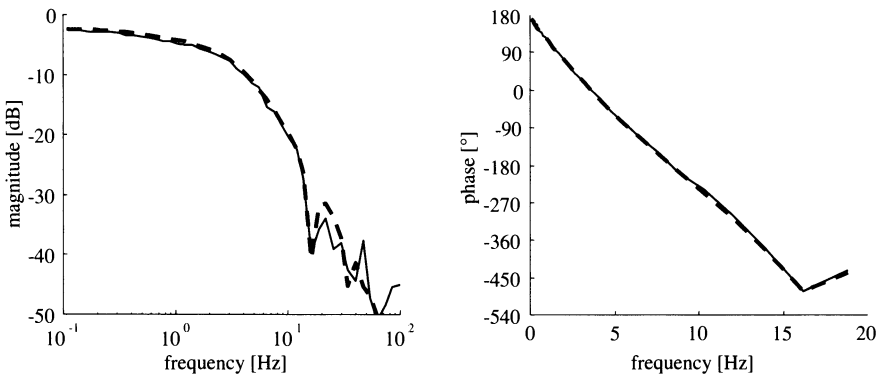


Fig. 3.13. Frequency response measurement (solid) and simulation (dashed) from fuel input u_φ to the output $y_{\lambda,1}$ of the LSU sensor at the exhaust valve.

To identify the parameters of this system, first the structure of the transfer function must be fixed. A series connection of a sample-and-hold element with sampling time $4\tau_{seg}$, an additional delay[5] τ_{delay}, the wall-wetting dynamics, and the first-order lag with time constant τ_{LSU} caused by the sensor dynamics are proposed.

The wall-wetting subsystem is discussed first. At low engine temperatures a first-order wall-wetting model does not sufficiently well reproduce all observed

[5] This delay includes all delays from the update injection to the exhaust center event.

dynamic effects. Therefore, a second parallel first-order lag is added,[6] resulting in the following continuous-time representation for the wall-wetting dynamics

$$G(s) = (1 - \kappa_1 - \kappa_2) + \frac{\kappa_1}{\tau_1 s + 1} + \frac{\kappa_2}{\tau_2 s + 1} \tag{3.32}$$

When the engine is fully warmed, typical values for the parameters are

$$1 - \kappa_1 - \kappa_2 \approx 0.5, \quad \kappa_1 \approx 0.15, \quad \tau_1 \approx 0.5\ s, \quad \tau_2 \approx 0.05\ s \tag{3.33}$$

For a cold engine, the following values are more realistic choice is

$$1 - \kappa_1 - \kappa_2 \approx 0.4, \quad \kappa_1 \approx 0.4, \quad \tau_1 \approx 1\ s, \quad \tau_2 \approx 0.1\ s. \tag{3.34}$$

Of course, these values will be different for each operating point and each engine system. The numerical values given here just show the orders of magnitude that can be expected.

The wall-wetting dynamics (3.32) of the four-cylinder engine analyzed above have the following discrete-time representation

$$G_{wwd}(z) = (1 - k_1 - k_2) + k_1 \cdot \frac{1 - e^{-\frac{4\tau_{seg}}{\tau_1}}}{z - e^{-\frac{4\tau_{seg}}{\tau_1}}} + k_2 \cdot \frac{1 - e^{-\frac{4\tau_{seg}}{\tau_2}}}{z - e^{-\frac{4\tau_{seg}}{\tau_2}}} \tag{3.35}$$

The transfer function of the complete system can be written as follows

$$G(s) = \frac{1 - e^{-4\tau_{seg}s}}{4\tau_{seg}s} \cdot e^{-\tau_{delay}s} \cdot G_{wwd}(z)|_{z=e^{4\tau_{seg}s}} \cdot \frac{1}{\tau_{LSU}s + 1} \tag{3.36}$$

Note that in the case where the sensor is placed close to the exhaust valve of one cylinder, the air/fuel ratio dynamics are essentially those of a single-cylinder engine. Accordingly, in (3.36) the sampling time is equal to the cycle time $\tau_{cycl} = 4\tau_{seg}$.

If the structure, the gains and the time constants are correct, the frequency response obtained with (3.36) should be close to that measured at even higher frequencies than the Nyquist frequency. As Fig. 3.13 shows, this is the case indeed. A crucial element in the loop is the sample-and-hold element introduced to model the reciprocating engine behavior. This element causes the strong drop in amplitude at approximately 16 Hz, which corresponds to the sampling frequency of a single cylinder at 2000 *rpm*. Modulating the fuel input with this frequency results in a constant injection duration for each cylinder, i.e., every cylinder operates under constant but different conditions. This also holds for integer multiples of that frequency.

The approaches developed so far are now applied to the multi-cylinder case. The structure of that system is depicted in Fig. 3.14. Combining the

[6] This additional element also models back flow of combustion gases into the manifold after BDC when the intake valve is not yet completely closed.

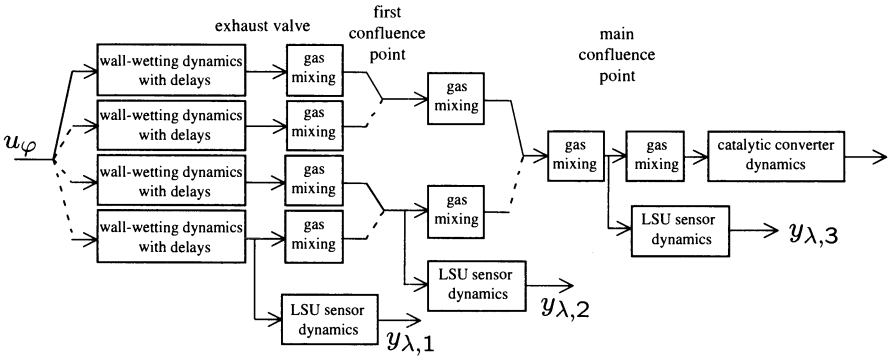


Fig. 3.14. Flowchart of the complete system investigated.

multiplexing modeling approach of the wall-wetting dynamics with the gas mixing dynamics, and using again the ideas of multiplexing, the structure of the complete system with input u_φ and output $y_{\lambda,3}$ is found to be

$$G(s) = \frac{1 - e^{-\tau_{seg} s}}{\tau_{seg} s} \cdot \widetilde{G}_{wwd}(z)\Big|_{z=e^{\tau_{seg} s}} \cdot \left(\frac{z^4 (b_1 z^4 + b_2)}{z^8 - a_1 z^4 - a_2} \right)\Bigg|_{z=e^{\tau_{seg} s}} \tag{3.37}$$

$$\cdot \left(z^2 \frac{1 - e^{-\frac{2\tau_{seg}}{\tau_3}}}{z^2 - e^{-\frac{2\tau_{seg}}{\tau_3}}} \right)\Bigg|_{z=e^{\tau_{seg} s}} \cdot \left(\frac{\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2} \right) \cdot \frac{e^{-\tau_{delay} s}}{1 + \tau_{LSU} s}$$

where

$$\widetilde{G}_{wwd}(z) = \left((1 - k_1 - k_2) + k_1 \frac{1 - e^{-\frac{4\tau_{seg}}{\tau_1}}}{z^4 - e^{-\frac{4\tau_{seg}}{\tau_1}}} + k_2 \frac{1 - e^{-\frac{4\tau_{seg}}{\tau_2}}}{z^4 - e^{-\frac{4\tau_{seg}}{\tau_2}}} \right) \tag{3.38}$$

The various parts in this expression model the following effects:

1. The sample-and-hold element is represented in its usual form.
2. The wall-wetting dynamics (in fact all dynamic effects taking place between injection location and exhaust valve) are approximated with a second-order element.
3. The gas mixing phenomena in the exhaust tubes from the exhaust valve to the first confluence point are described with a second-order element that is multiplexed. The structure of this element is chosen such that it reflects the underlying phenomena caused by the multiplexed operation of four exhaust parts. This choice reduces the number of parameters that must be identified experimentally to the bare minimum.
4. Similarly, the exhaust pipes between first confluence point and main confluence point can be approximated with a first-order element, multiplexed with one-half of the segment frequency.

5. The gas mixing in the main confluence point is then again a continuous element, which can be described by two additional states. Interestingly, these dynamics are best described by a second-order system. This approach permits the combination of the two extreme situations of *plug flow* and *perfectly stirred reactor* conditions, as described in [107] and [184].
6. The overall delay and the LSU sensor dynamics follow as the last elements.

The corresponding parameters must be identified using a sequential approach. Starting with the model description (3.36), obtained for the case where the sensor is mounted close to one of the exhaust valves, the subsequent dynamic elements can be identified by using intermediate measurement signals. Figure 3.15 shows the measured and the predicted frequency response of the full system with the air/fuel ratio sensor at position 3.

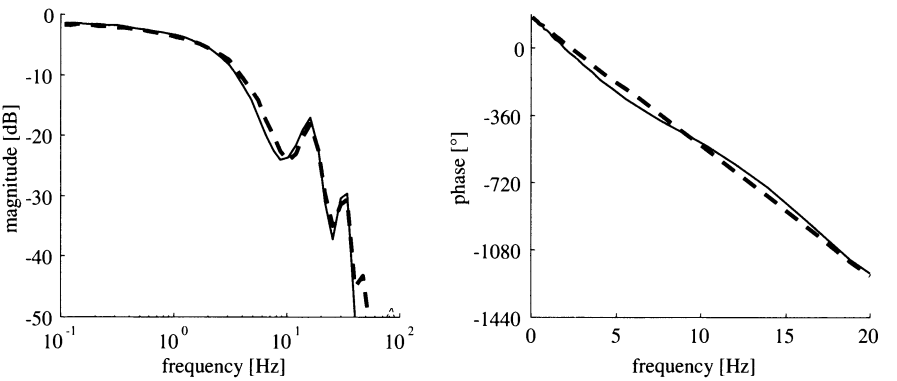


Fig. 3.15. Frequency response measurement (solid) and simulation (dashed) from fuel excitation to the output $y_{\lambda,3}$ of the LSU sensor at the main confluence point.

As Fig. 3.15 shows, the model represented by (3.37) can well predict the measured frequency response. The system is a low-pass element with a first corner frequency at approximately 2 Hz. At 16.66 Hz a local maximum is attained because at that frequency the fuel path of all cylinders operates at steady-state conditions,[7] which eliminates the influence of the wall-wetting dynamics. The same observation is true at $2 \cdot 16.66$ Hz and $3 \cdot 16.66$ Hz. At $4 \cdot 16.66$ Hz (not shown in Fig. 3.15) the sample-and-hold element introduces a zero such that at that frequency no local maximum is attained. The continuous-time elements in (3.37) cause a steady drop in magnitude and phase. The latter is dominated by the delays, as the left part of Fig. 3.15 shows (notice that the frequency in this diagram is linear).

[7] The fuel modulation and the multiplexing frequency are identical. Hence, all cylinders obtain different but constant fuel quantities.

3.2.4 DEM of the Back-Flow Dynamics of CNG Engines

For several reasons, compressed natural gas (CNG) is an interesting fuel. For instance, it permits a clean and quiet combustion, produces low non-methane emissions even at cold start, features low CO_2 emission, and its high octane numbers are well suited to supercharged engine applications. Moreover, it is readily available in many places at relatively low costs. Compared to energy carriers based on liquid hydrocarbons, its main disadvantages are smaller energy density, additional compression energy demand, leakage losses in production and transportation and the currently small number of operational fueling stations.

The majority of all CNG engines are spark-ignited, port-injected engines that operate in an Otto cycle. In their dynamic behavior, such CNG engines are similar to SI engines. The cause and effect diagram shown in Fig. 2.5 remains valid in principle for such CNG engines as well.

The only three major differences that must be considered are:

- Compared to gasoline, natural gas occupies a substantial part of the aspirated volume. This requires that the detailed model of the engine mass flow dynamics, as expressed in equation (2.27), is used in simulations.
- Although with a gaseous fuel no wall-wetting phenomena occur, the fuel path exhibits a similar dynamic effect due to a substantial back flow of aspirated mixture.
- Since the fuel is gaseous, it can be injected while the intake valves are open without having to accept an associated penalty in hydrocarbon emissions. This permits a certain amount of charge stratification. Therefore, this effect must be included in the model of the back-flow dynamics.

Few authors have discussed the control-oriented modeling of CNG engines, e.g., [138]. In the following, a model of the back-flow dynamics is presented which was first introduced in [45]. The discrete-event formulation of this model, as shown below, requires the following variables to be defined first:

m_e Mass of NG gas entering the cylinder.
m_b Mass of NG flowing from the cylinder back into the intake manifold.
m_c Mass of NG remaining in the cylinder after the intake valve is closed.
m_l Mass of NG injected after back flow has started.
m_t Total mass of NG injected into the intake manifold.
V_a Cylinder volume filled with air.[8]
V_m Cylinder volume filled with air and NG mixture.
V_b Cylinder volume that flows back to the intake manifold.

Figure 3.16 shows the status of the cylinder at its bottom dead center (BDC). Simplifying the complex interactions that take place in a real gas-exchange process, it is assumed that the cylinder charge consists of a first

[8] The burnt gases inside the cylinder are assumed to be inert gases, i.e., the mixtures are assumed to be close to stoichiometric.

volume, V_a, filled with air and a second volume, V_m, filled with an NG/air mixture. It is assumed that the maximum amount of NG entering the cylinder m_e is achieved at BDC, that the back flow takes place after that point and that no mixing between V_m and V_a must be considered.

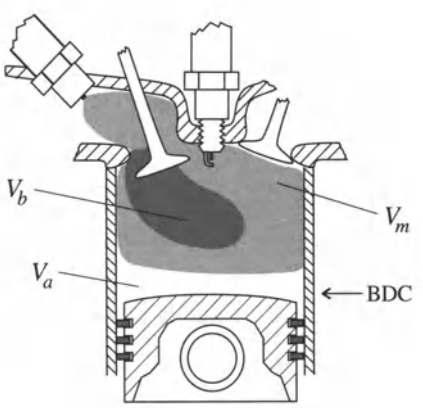


Fig. 3.16. Simplified engine stratification model used in the derivation of the back-flow model.

The following three dimensionless variables will be used to compactly formulate the model. The *back-flow* parameter α defines the relative amount of NG flowing back into the intake manifold with respect to the maximum aspirated amount of NG. Assuming a homogeneous mixture yields

$$\alpha = \frac{V_b}{V_m + V_a} \tag{3.39}$$

Obviously, a value of $\alpha = 0$ indicates no back flow.

The *stratification* parameter β describes the ratio between the total charge and the mixture

$$\beta = \frac{V_m + V_a}{V_m} \tag{3.40}$$

In the case of an unrealistically large stratification factor $\beta = 1/\alpha$, all of the aspirated NG would flow back again.

Finally, the *overlap* factor ξ indicates the relative amount of fuel that is injected after back flow has started, i.e., that part of the fuel that will not enter the cylinder in the actual cycle. Obviously, a $\xi = 0$ indicates that all injected fuel enters the cylinder (but does not necessarily stay there due to back flow), while a $\xi = 1$ indicates that all of the fuel injected in the current cycle will be aspirated into the cylinder in the next cycle.

Using the definitions introduced above, the dynamic behavior of one cylinder of an n-cylinder engine is described in the k-th segment by the following set of equations

$$m_b(k) = \alpha(k) \cdot \beta(k) \cdot m_e(k) \tag{3.41}$$

$$m_l(k) = \xi(k) \cdot m_t(k) \tag{3.42}$$

$$m_e(k) = m_b(k-n) + m_l(k-n) + m_t(k) - m_l(k) \tag{3.43}$$

$$m_c(k) = m_e(k) - m_b(k) \tag{3.44}$$

These equations essentially represent a mass balance with one reservoir (the mass $m_b(k-n) + m_l(k-n)$ is stored in the intake manifold between two cycles), one input (the total mass $m_t(k)$ of NG injected) and one output (the mass $m_c(k)$ of NG taking part in the combustion in the k-th cycle).

The variables $\{\alpha(k), \beta(k), \xi(k)\}$ completely describe the intrinsic dynamic properties of the back flow dynamics. As is the case in the wall-wetting dynamics in port-injected gasoline engines, these parameters are, in general, nonlinear functions of the engine's load and speed and of many other influencing variables.

For small changes around a nominal operating point, the parameters $\{\alpha(k), \beta(k), \xi(k)\}$ may be assumed to be constant such that the following DEM transfer function formulation may be derived

$$m_c(z) = P(z) \cdot m_t(z) \tag{3.45}$$

with

$$P(z) = (1 - \alpha \cdot \beta) \cdot (1 - \xi) + \frac{(1 - \alpha \cdot \beta) \cdot (\xi + \alpha \cdot \beta \cdot (1 - \xi))}{z^n - \alpha \cdot \beta} \tag{3.46}$$

As in the wall-wetting dynamics, a feed-through element and a "memory" element can be seen in this expression. As expected, the static gain $P(1)$ of the transfer function (3.46) is equal to 1.

The parameters $\{\alpha, \beta, \xi\}$ have to be identified using experiments that are essentially the same as those used in Sect. 3.2.3 in the case of gasoline engines. The well-known correlation between air/fuel ratio and NO_x emissions and the availability of very fast NO_x sensors permit the identification of the unknown parameters using step responses.

Figure 3.17, taken from [45], shows the results for one specific case in which the values identified were $\{\alpha = 0.16, \beta = 1.6, \xi = 0\}$.

3.2.5 DEM of the Residual Gas Dynamics

The residual gas is the burnt gas remaining in the cylinder after the exhaust valve has closed. Especially for varying air/fuel ratios, this mass and its corresponding air/fuel ratio may not be neglected. In this section, a DEM of the residual gas dynamics is introduced that is useful for an understanding of the main effects.

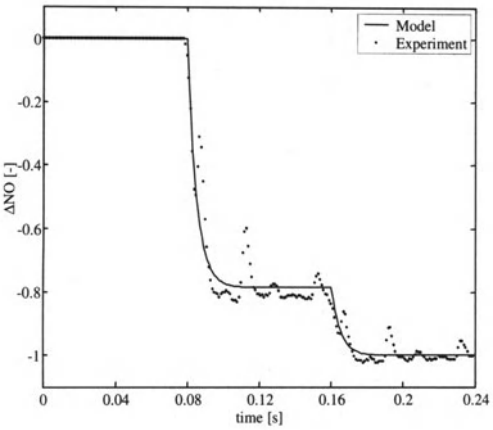


Fig. 3.17. Model/experiment validation of the back flow dynamics for a 1.8-liter, four-cylinder SI engine. Normalized step response of the transfer function (3.46). Engine operating point $\omega_e = 150\ rad/s$, $p_{me} = 3.8\ bar$. The simulation combines the discrete-event model with a continuous-time model of the sensor dynamics.

The variable $\tilde{\lambda}(k)$ is used to denote the air/fuel ratio of the mixture that enters the cylinder in the k-th cycle. According to Sect. 2.4.3, this variable is defined as

$$\tilde{\lambda}(k) = \frac{m_\beta(k)}{m_\varphi(k) \cdot \sigma_0} \tag{3.47}$$

Using this definition, the total mass of fresh mixture in the cylinder in the k-th cycle can be expressed as follows

$$m_{fc}(k) = m_\beta(k) + m_\varphi(k) \tag{3.48}$$

$$= m_\beta(k) \left(1 + \frac{1}{\tilde{\lambda}(k) \cdot \sigma_0}\right) \tag{3.49}$$

$$= m_\varphi(k) \cdot (\tilde{\lambda}(k) \cdot \sigma_0 + 1) \tag{3.50}$$

Similar equations describe the dynamics of the mass of residual gas $m_{rg}(k)$. The air/fuel ratio of the residual gas is equal to the air/fuel ratio of the total mass[9] at event t_{k-1}, assuming a sampling time of one engine cycle. Therefore, formulating a mass balance of the fresh charge *and* the burnt gases present in the cylinder after the intake valve closed, yields the following equation for the air/fuel ratio $\lambda(k)$ of the cylinder contents

[9] This definition is based on a carbon-to-hydrogen balance, which is valid even after the combustion. This approach is used in [81], for instance, to derive the air/fuel ratio by analyzing the exhaust gases.

$$\lambda(k) = \frac{m_{fc}(k) \cdot \frac{\tilde{\lambda}(k)\cdot\sigma_0}{1+\tilde{\lambda}(k)\cdot\sigma_0} + m_{rg}(k) \cdot \frac{\lambda(k-1)\cdot\sigma_0}{1+\lambda(k-1)\cdot\sigma_0}}{m_{fc}(k) \cdot \frac{1}{1+\tilde{\lambda}(k)\cdot\sigma_0} + m_{rg}(k) \cdot \frac{1}{1+\lambda(k-1)\cdot\sigma_0}} \cdot \frac{1}{\sigma_0} \tag{3.51}$$

This equation can be transformed into the following (nonlinear) equation

$$\lambda(k) = \frac{m_{fc}(k)\tilde{\lambda}(k)(1+\lambda(k-1)\sigma_0) + m_{rg}(k)\lambda(k-1)(1+\tilde{\lambda}(k)\sigma_0)}{m_{fc}(k)(1+\lambda(k-1)\sigma_0) + m_{rg}(k)(1+\tilde{\lambda}(k)\sigma_0)} \tag{3.52}$$

As expected, for steady-state conditions, i.e., for $\lambda(k) = \lambda(k-1)$, the identity $\lambda = \tilde{\lambda}$ results. Moreover, since in most situations $\lambda(k) \cdot \sigma_0 >> 1$, the last equation may be simplified

$$\lambda(k) \approx \frac{m_{fc}(k) \cdot \tilde{\lambda}(k) \cdot \lambda(k-1) + m_{rg}(k) \cdot \lambda(k-1) \cdot \tilde{\lambda}(k)}{m_{fc}(k) \cdot \lambda(k-1) + m_{rg}(k) \cdot \tilde{\lambda}(k)} \tag{3.53}$$

$$\approx [m_{fc}(k) + m_{rg}(k)] \cdot \left[\frac{m_{fc}(k)}{\tilde{\lambda}(k)} + \frac{m_{rg}(k)}{\lambda(k-1)}\right]^{-1} \tag{3.54}$$

This equation can be reformulated as

$$\frac{m_{fc}(k) + m_{rg}(k)}{\lambda(k)} \approx \frac{m_{fc}(k)}{\tilde{\lambda}(k)} + \frac{m_{rg}(k)}{\lambda(k-1)} \tag{3.55}$$

which can be interpreted as a weighted sum of the corresponding amounts of gases. If, in addition, it can be assumed that the air/fuel ratio does not deviate too much from its stoichiometric value equal to 1, then the following simplification can be made

$$\frac{1}{\lambda} \approx 1 - (\lambda - 1) = 2 - \lambda \tag{3.56}$$

This leads to the formulation that has already been used in Sect. 2.4.3

$$\lambda(k) \approx \frac{m_{fc}(k) \cdot \tilde{\lambda}(k) + m_{rg}(k) \cdot \lambda(k-1)}{m_{fc}(k) + m_{rg}(k)} \tag{3.57}$$

Clearly, (3.57) is still a nonlinear equation because m_{fc} and m_{rg} are varying quantities. However, if the fuel quantity is varied only, m_{fc} and m_{rg} can be approximated well by constant values. In this case, the air/fuel ratio dynamics are well approximated by a discrete first-order lag element.

It is usually not easy to obtain information on the precise amount of burnt gas that remains inside the cylinder. A very rough estimation would be a value for m_{rg} of 5% of the full-load total mass in the cylinder. A better approximation is obtained if the mass of residual gas is estimated using an affine relationship of the form

$$m_{rg} = a_{rg} \cdot \frac{1}{\omega_e} + b_{rg} \tag{3.58}$$

Figure 3.18 shows as an example the mass of burnt gas that remains in the cylinder for a six-cylinder 3.2-liter SI engine, which at full load has a total cylinder charge mass of approximately 450 mg.

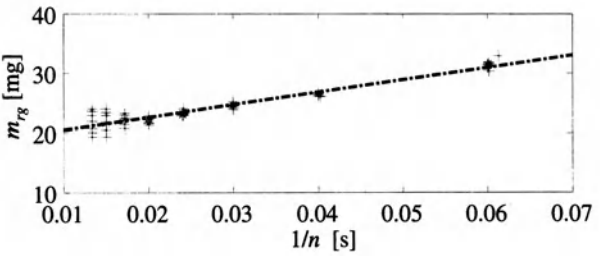


Fig. 3.18. Residual burnt-gas mass as a function of the time per revolution of a six-cylinder 3.2-liter SI engine. Stoichiometric combustion, engine speed $1000 - 4500$ rpm, engine load $p_{me} = 0 - 6$ bar.

3.2.6 DEM of the Exhaust System

In this section, a DEM of the effects taking place in the exhaust system is presented. One motivation for doing this is the objective to estimate the air/fuel ratios of each cylinder[10] using one air/fuel ratio sensor only. The main simplifications assumed in this section are steady-state operating conditions, i.e., each cylinder operating at constant but different air/fuel ratios, and a symmetric layout of the exhaust manifold. Under these conditions a compact model formulation can be derived as first proposed in [72].

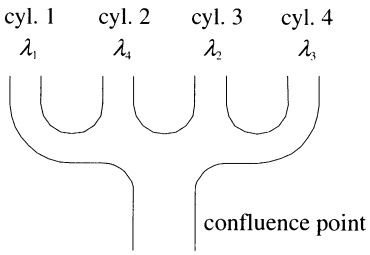


Fig. 3.19. Idealized exhaust manifold model.

[10] Manufacturing tolerances for the injection valves and intake ports, as well as different flow characteristics in the intake runners, result in different air/fuel ratios for every cylinder. Variations in the air/fuel ratios can cause drivability and emission problems, especially when the engine is cold.

The symmetric exhaust layout assumed in this section is shown in Fig. 3.19. Since all cylinders are operated in constant conditions, such a symmetric layout will result in periodic air/fuel ratio signals at the confluence point where the sensor is assumed to be placed. Accordingly, the air/fuel ratio λ_{cfp} of the exhaust gas that the sensor observes can be approximated by the following expression[11]

$$\lambda_{cfp}(k) = c_1(k) \cdot \lambda_1 + c_2(k) \cdot \lambda_2 + c_3(k) \cdot \lambda_3 + c_4(k) \cdot \lambda_4 \tag{3.59}$$

where k is the segment counter and λ_i the air/fuel ratio of the i-th cylinder. The coefficients $c_i(k)$ are functions of the segment counter. They are periodic and their sum is equal to 1 for all k. Figure 3.20 shows an example of these variables. The specific form of the curves will depend on the engine operating condition and on the manifold layout.

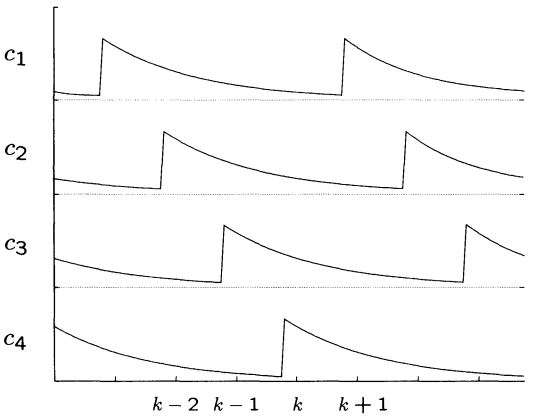


Fig. 3.20. Trajectories of the weighting functions $c_i(k)$.

Instead of using periodic coefficients $c_i(k)$, a cyclic permutation of the four λ_i signals can be used to describe the same behavior. In this case, the following permutation law must be adopted

$$\begin{bmatrix} \widetilde{\lambda}_1(k+1) \\ \widetilde{\lambda}_2(k+1) \\ \widetilde{\lambda}_3(k+1) \\ \widetilde{\lambda}_4(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \widetilde{\lambda}_1(k) \\ \widetilde{\lambda}_2(k) \\ \widetilde{\lambda}_3(k) \\ \widetilde{\lambda}_4(k) \end{bmatrix} \qquad \begin{bmatrix} \widetilde{\lambda}_1(0) \\ \widetilde{\lambda}_2(0) \\ \widetilde{\lambda}_3(0) \\ \widetilde{\lambda}_4(0) \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \tag{3.60}$$

where λ_i is the air/fuel ratio of the cylinder i and $\widetilde{\lambda}_i(k)$ the i-th fictitious air/fuel ratio signal that is used to define the output signal

[11] This formulation is valid for a four-cylinder engine. The model can be easily adapted to an arbitrary number of cylinders.

$$\lambda_{cfp}(k) = \begin{bmatrix} \widetilde{c}_1 & \widetilde{c}_2 & \widetilde{c}_3 & \widetilde{c}_4 \end{bmatrix} \cdot \begin{bmatrix} \widetilde{\lambda}_1(k) \\ \widetilde{\lambda}_2(k) \\ \widetilde{\lambda}_3(k) \\ \widetilde{\lambda}_4(k) \end{bmatrix} \tag{3.61}$$

where the *constants* $\widetilde{c}_i$ are obtained by taking the explicit numerical values of $c_i(k)$ for any k. The variable $\lambda_{cfp}(k)$ is the air/fuel ratio at the confluence point at event k. Of course, the sensor dynamics must be taken into account as well. Assuming the usual first-order time lag model with linearized static behavior, the following result is obtained

$$\lambda_{meas}(k+1) = a \cdot \lambda_{meas}(k) + (1-a) \cdot \lambda_{cfp}(k) \tag{3.62}$$

where

$$a = e^{-\frac{\tau_{seg}}{\tau_{sensor}}} \tag{3.63}$$

Since the sensor time constant is fixed in the time domain, the discrete representation of the sensor dynamics must be recalculated every time a sample is taken.

3.3 DEM Based on Cylinder Pressure Information

3.3.1 General Remarks

The pressure of the gases in the cylinder provides the most direct signal available for engine control purposes. Since the interpretation of this variable is a well known tool in ICE research and development, numerous control algorithms based on that information have been proposed (see, e.g., [123], [124], [152] [180]). However, the equipment used in those experiments is far too expensive to be included in series production engines. Recently, new in-cylinder pressure sensors have been developed that are sufficiently robust and reasonably priced such that their inclusion in series production vehicles appears possible (see e.g., [79] and [174]).

Since full process calculations are not yet possible in the ECU, algorithms requiring a reduced amount of computational burden are of high interest. In [116], such simplified correlations useful for a control-oriented cylinder pressure interpretation have been derived. The primary points of that approach are reviewed in this section.

The pressure in the cylinder is determined by the movement of the piston and the corresponding compression of the air/fuel mixture inside the cylinder, the heat release caused by the combustion of this mixture and the heat transfer from the gas to the cylinder walls. The position of the piston can be described by

$$s(\phi) = r_{cs} \cdot \left(\frac{l + r_{cs}}{r_{cs}} - \sqrt{\frac{l^2}{r_{cs}^2} - \sin^2\phi} - \cos\phi \right) \tag{3.64}$$

where $s(\phi)$ is the distance of the piston from its TDC position, ϕ is the crank angle ($\phi = 0$ at TDC), r_{cs} is the radius of the crankshaft, i.e., one-half of the stroke, and l is the length of the connecting rod. A good approximation of (3.64) is given by

$$s(\phi) = r_{cs} \cdot (1 - \cos\phi + \frac{r_{cs}}{2 \cdot l} \sin^2 \phi) \tag{3.65}$$

In particular the equations describing the piston speed and the piston acceleration are easier to formulate using (3.65). Denoting the compression ratio by ε and the displaced volume by V_d, the following equation holds for the cylinder volume V_c

$$V_c(\phi) = V_d \cdot \left(\frac{1}{\varepsilon - 1} + \frac{s(\phi)}{2 \cdot r_{cs}} \right) \tag{3.66}$$

3.3.2 Estimation of Burned-Mass Fraction

In "forward" thermodynamic process calculations, the development of combustion is often described with the burnt-mass fraction $x_B(\phi)$ defined as

$$x_B(\phi) = \frac{m_{\varphi b}(\phi)}{m_{\varphi t}} \tag{3.67}$$

where $m_{\varphi b}(\phi)$ is the mass of burnt fuel at crank angle ϕ, and $m_{\varphi t}$ is the total burnt fuel at the end of combustion. For rich conditions, the total mass of burnt fuel must be calculated with the available mass air, thus $m_{\varphi t} = \lambda \cdot m_{\varphi}$ where m_{φ} is the fuel mass in the cylinder at IVC. Under lean conditions, $m_{\varphi t}$ is simply m_{φ}.

The burnt-mass fraction (3.67) is usually parameterized using the well-known Vibe function, [177]

$$x_B(\phi) = 1 - e^{c \cdot (\frac{\phi - \phi_s - \Delta\phi_d}{\Delta\phi_{bd}})^{m_v + 1}} \tag{3.68}$$

The parameter $\Delta\phi_d$ indicates the spark delay, $\Delta\phi_{bd}$ represents the duration of the combustion, $c \approx \log(0.001)$ indicates that the end of combustion is assumed at $x_B = 0.999$, and m_v is used to shape the resulting curve for the burnt-mass fraction.

Note that if a symmetric burnt-mass fraction is acceptable, an even simpler approximation can be used as well

$$x_B(\phi) = \sin^2(\frac{\phi - (\phi_s + \Delta\phi_d)}{\Delta\phi_{bd}} \cdot \frac{\pi}{2}) \tag{3.69}$$

Using a heat release function, combined with a heat-transfer approximation and the thermodynamic conservation laws, the pressure inside the cylinder can be estimated [81]. This information can then be used to compute

indicated efficiencies or, combined with a friction model similar to (2.111), to compute brake efficiencies.

In [41], functions were derived that describe the variations of the three Vibe parameters of equation (3.68) for different operating conditions. This approach allows a control-oriented description of the combustion processes and, thus, for instance, the determination of the engine inputs for optimal fuel consumption [129].

If the cylinder pressure is measured, the burnt-mass fraction can be determined by an "inverse" thermodynamic process calculation. However, this is computationally quite demanding. Therefore, an approximation first described in [141] is used in this text. The primary points of that approach are visualized in Fig. 3.21.

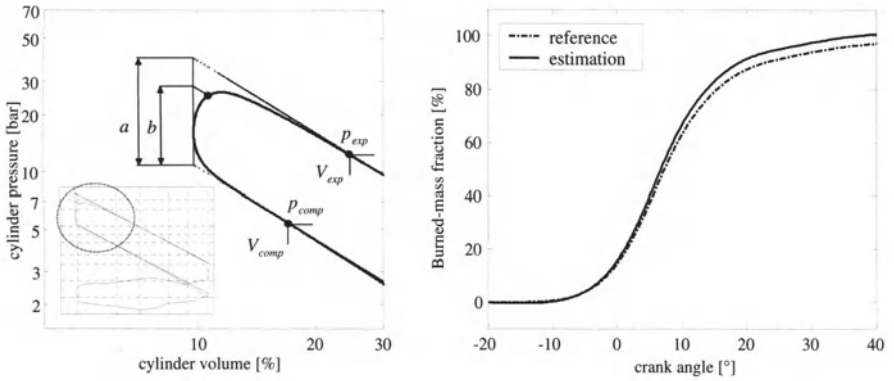


Fig. 3.21. Estimation of the burn rate.

The procedure starts with the selection of a pressure/volume pair, p_{comp} and V_{comp}, during the compression phase (before the combustion starts) and a pressure/volume pair, p_{exp} and V_{exp}, during expansion phase (after the combustion is completed). Extrapolating these two pairs to the minimum volume value using a polytropic approximation yields the pressure difference indicated by **a** in Fig. 3.21. The actual pressure/volume data during the combustion is extrapolated in the same way yielding the pressure difference **b**. The mass of burnt fuel is then estimated by the ratio of these two pressure differences

$$x_B(\phi) \approx \frac{b}{a} \approx \frac{p_c(\phi) \cdot \left(\frac{V_c(\phi)}{V_{TDC}}\right)^{n(\phi)} - p_{comp} \cdot \left(\frac{V_{comp}}{V_{TDC}}\right)^{n_{comp}}}{p_{exp} \cdot \left(\frac{V_{exp}}{V_{TDC}}\right)^{n_{exp}} - p_{comp} \cdot \left(\frac{V_{comp}}{V_{TDC}}\right)^{n_{comp}}} \tag{3.70}$$

$$n(\phi) = \begin{cases} n_{comp} & \text{for} \quad \frac{\partial V_c(\phi)}{\partial \phi} \leq 0 \\ n_{exp} & \text{for} \quad \frac{\partial V_c(\phi)}{\partial \phi} > 0 \end{cases}$$

Choosing the reference points to be symmetric with respect to TDC (say at 80 ° before and after[12] TDC) simplifies the computations. In fact, as illustrated in Fig. 3.22, in this case the burn rate can be estimated using the expression

$$x_B(\phi) \approx \frac{p_c(\phi) \cdot \left(\frac{V_c(\phi)}{V_{TDC}}\right)^{n_{pol}} - p_{c1}}{\Delta p_c} \tag{3.71}$$

To further simplify the computations, a constant value of $n_{pol} = 1.31$ can be used for the polytropic exponent.

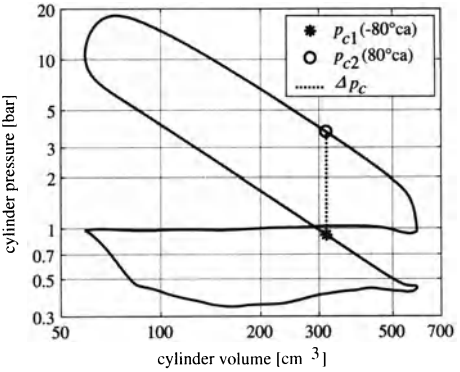


Fig. 3.22. Pressure/Volume (pV) diagram using a double-logarithmic representation including the reference points chosen for the symmetric case.

3.3.3 Cylinder Charge Estimation

Introduction

A typical application for the method introduced in the last section is the estimation of the 50% burnt-mass position. Of course, the cylinder pressure trace can be used for various other purposes. Two straightforward possibilities are the detection of knock and misfiring or the identification of the actual engine cycle.

A computationally more demanding use of the pressure trace, which is based on physical first principles, is the estimation of the fresh air, fuel, and burnt-gas mass in the cylinder. This application is discussed in this section. Assuming each cylinder to be equipped with a pressure sensor, this approach offers the possibility of cylinder-individual control, resulting in a more uniform load and air/fuel ratio distribution. Moreover, if the estimation results are sufficiently accurate, the air mass sensor can be left out.

[12] Note that the combustion must be completed at that point. This is the case for most regular combustion systems, but not necessarily for new "exotic" engines.

The total mass in the cylinder, defined by $m_{tot} = m_\beta + m_\varphi + m_{bg}$, is fixed as soon as the intake valve has closed. The variable $m_{bg} = m_{rg} + m_{egr}$ is the total mass of burnt gas, i.e., the sum of the residual gas and the externally recirculated exhaust gas. Choosing a reference position during the compression stroke (indicated by an asterisk in Fig. 3.22), and using the measured cylinder pressure p_{c1} and the known cylinder volume (3.66), the ideal gas law

$$m_{tot} = \frac{p_{c1} \cdot V_{c1}}{R \cdot \vartheta_{c1}} \tag{3.72}$$

yields an estimation of m_{tot} by *assuming* the temperature ϑ_{c1} at the same reference point to have a value of approximately $470°\ K$. Note that (3.72) is valid only because the gas constant R of the air/fuel mixture and burnt gases is quite similar (a value of $287\ J/kgK$ is recommended). Once the total mass is known, the amount of burnt gas can be estimated if the mass of aspirated mixture is estimated with any of the methods presented in either Sect. 2.3 or Sect. 3.2.2, or with the approaches shown below.

The estimation (3.72) is straightforward but does not yield sufficiently precise results that would allow for the omission of the air mass flow sensor. To achieve this goal, additional information must be obtained by analyzing the complete pressure trajectory $p_c(\phi)$. However, in this case some information on the heat release during the combustion must be considered.

Basic Model

Using the lower heating value H_l of the fuel and introducing a parameter C_c ("completeness of combustion") that indicates how much of the fuel is actually burnt until the exhaust valve starts to open, the following equation describes the total heat release

$$Q_c = m_\varphi \cdot H_l \cdot C_c = m_{tot} \cdot \frac{H_l}{1 + \lambda \cdot \sigma_0} \cdot (1 - x_{bg}) \cdot C_c \tag{3.73}$$

At ideal conditions, i.e., for $m_{bg} = 0$, $\lambda = 1$, and $C_c = 1$, the heat released is

$$Q_{c,ideal} = m_{tot} \cdot \frac{H_l}{1 + \sigma_0} = \frac{p_{c1} \cdot V_{c1}}{R \cdot \vartheta_{c1}} \cdot \frac{H_l}{1 + \sigma_0} \tag{3.74}$$

Thus, a new variable η_c, referred to as the *cylinder charge efficiency*, can be defined as follows

$$\eta_c = \frac{Q_c}{Q_{c,ideal}} = (1 - x_{bg}) \cdot C_c \cdot \frac{1 + \sigma_0}{1 + \lambda\sigma_0} \tag{3.75}$$

For a given m_{tot}, this parameter quantifies the ratio between the actual and the maximum theoretically possible heat released in the combustion.

The relationship between heat release and pressure in the cylinder is analyzed next. For the two reference positions indicated above, the pressure difference

$$\Delta p_c = p_{c2} - p_{c1} \tag{3.76}$$

can be calculated. For the relevant operating range of a six-cylinder 3.2-liter SI engine, the pressure difference as a function of the combustion energy is depicted in Fig. 3.23. The air/fuel ratio is set to stoichiometry and the 50% burnt-mass position, ϕ_{50}, is always kept near its optimum value of approximately 8 ° crank angle after TDC.

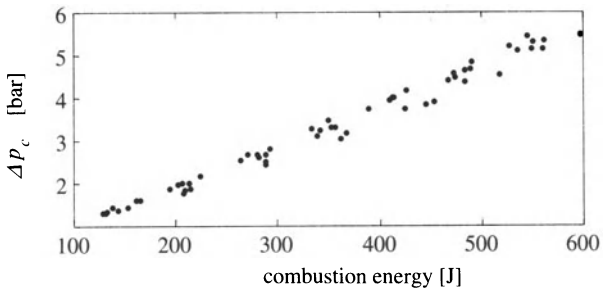


Fig. 3.23. Pressure difference as a function of the combustion energy. Engine operating range $n = 1000 - 4500\ rpm$ and $p_{me} = 0 - 6\ bar$. SI engine, $\lambda = 1$, $\phi_{50} \approx 8°$ crank angle after TDC.

As expected, an almost linear dependency can be observed. Note that this is just another confirmation of the remarks made in Sect. 2.5.1. In fact, the parameter e in the Willans approximation shown in Fig. 2.29 can be interpreted as the "indicated" (internal) efficiency that describes the ratio between combustion energy and pressure difference that eventually results in mechanical work on the piston.

Improved Model

The deviations of the single pressure difference point from a purely linear dependency occur mainly for two reasons:

- Variations in the engine speed influence the time available for the combustion and heat exchange with the cylinder walls.
- Variations in the combustion center, represented by the 50% burnt-mass position, influence the conversion of heat to mechanical work and, hence, the heat balance.

These observations lead to the following parametrization of the pressure difference as a function of engine speed and heat-release center

$$\begin{aligned} \Delta p_c &= k(\omega_e, \phi_{50}) \cdot Q_c \\ &= k(\omega_e, \phi_{50}) \cdot \eta_c \cdot Q_{c,ideal} \end{aligned} \tag{3.77}$$

$$= k(\omega_e, \phi_{50}) \cdot \eta_c \cdot \frac{p_{c1} \cdot V_{c1}}{R \cdot \vartheta_{c1}} \cdot \frac{H_l}{1 + \sigma_0}$$

To investigate these influences, the effects of varying the 50% burnt-mass position and the engine speed can be estimated using thermodynamic process simulations. Assuming an ideal combustion with $\eta_c = 1$, the factor $k(\omega_e, \phi_{50})$ can be estimated by using (3.78) because all other terms on the right-hand side of this equation can be assumed to be constant

$$\frac{\Delta p_c \cdot \vartheta_{c1}}{p_{c1}} = k(\omega_e, \phi_{50}) \cdot \frac{V_{c1}}{R} \cdot \frac{H_l}{1 + \sigma_0} \tag{3.78}$$

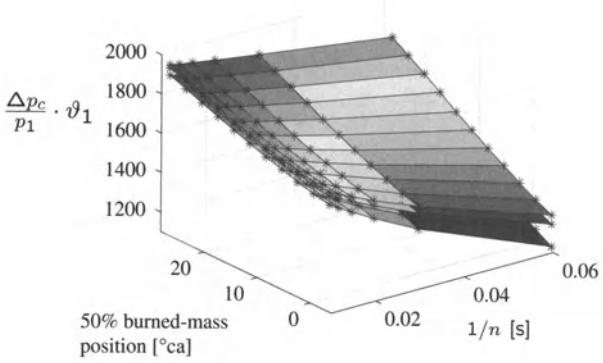


Fig. 3.24. Numerical value of $\frac{\Delta p_c}{p_{c1}} \vartheta_{c1}$ for simulated variations in engine speed n and 50% burned-mass position ϕ_{50}.

Figure 3.24 shows the calculated variation for three different total mass values (representing idle, half-load and full-load conditions). Obviously, the time available for the combustion is an important parameter. Accordingly, in this figure the inverse engine speed is used as an independent variable. A bilinear parametrization is proposed here

$$k(\phi_{50}, \omega_e) \cdot \frac{V_{c1}}{R} \cdot \frac{H_l}{1 + \sigma_0} = \sum_{j \in corners} \left(\frac{\vartheta_{c1} \cdot \Delta p_c}{p_{c1}}\right)_j \cdot \mu_j \tag{3.79}$$

where

$$\mu_j = \left(1 - \frac{\mid \phi_{50,OP} - \phi_{50,j} \mid}{\phi_{50,max} - \phi_{50,min}}\right) \cdot \left(1 - \frac{\mid \omega_{OP}^{-1} - \omega_j^{-1} \mid}{\omega_{min}^{-1} - \omega_{max}^{-1}}\right) \tag{3.80}$$

where μ_j are the weighting parameters that allow for the calculation of the value of an arbitrary operating point OP within a rectangle defined by the corner values.

With these preparations, the cylinder charge efficiency can be calculated based on the knowledge of p_{c1} and ϕ_{50}, and an estimation for ϑ_{c1}

$$\eta_c = \frac{\Delta p_c}{k(\omega_e, \phi_{50})} \cdot \frac{R \cdot \vartheta_{c1}}{p_{c1} \cdot V_{c1}} \cdot \frac{1 + \sigma_0}{H_l} \tag{3.81}$$

Choosing $\vartheta_{c1} = 470°\ K$ as a first approximation of the temperature of the cylinder charge at the IVC position, an estimation for the charge efficiency with an accuracy of typically $\pm 5\%$ is obtained. If a better estimation of ϑ_{c1} can be obtained from a detailed thermodynamic process calculation, the approximation of the charge efficiency can be improved. As Fig. 3.25 shows, an estimation of (3.81) with an accuracy of $\pm 3\%$ is feasible (see [115] for details).

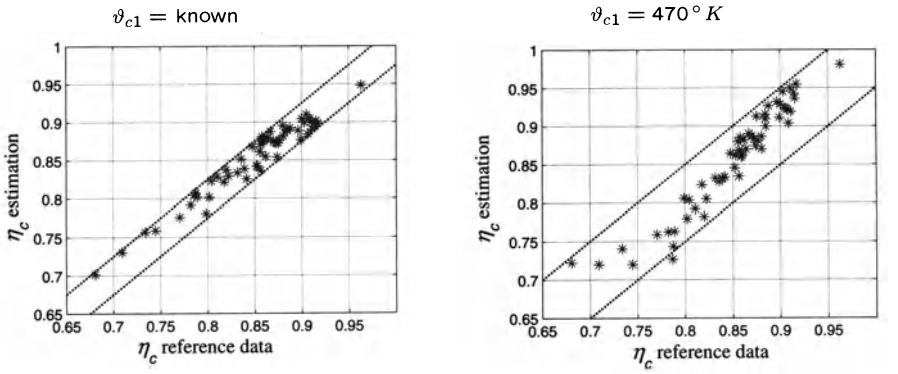


Fig. 3.25. Estimation of the charge efficiency.

Finally, assuming a minimum and a maximum completeness of combustion of 0.93 and 0.98, respectively, the second-order effects of the amount of injected fuel, of the 50% burned-mass position, of the combustion duration, etc., may all be collected in a linear function depending on p_{c2}

$$C_c = 0.93 + \frac{0.05(p_{c2} - p_{c2,min})}{p_{c2,max} - p_{c2,min}} \tag{3.82}$$

Reasonable values for a naturally aspirated engine for $p_{c2,max}$ and $p_{c2,min}$ are

$$p_{c2,min} = 1.5 \text{ bar} \quad \text{and} \quad p_{c2,max} = 15 \text{ bar}$$

Now, using (3.75), if λ is known from sensor information the burned-gas fraction at IVC can be estimated. The estimated values plotted against the reference values (obtained with a two-zone process calculation) are shown in Fig. 3.26.

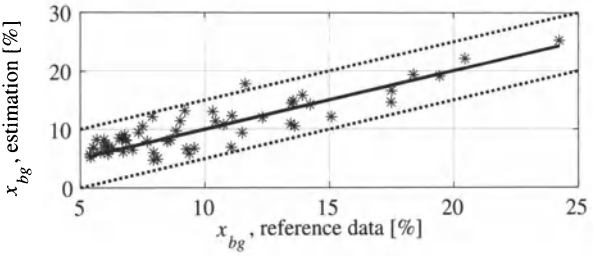


Fig. 3.26. Estimation of the burned-gas fraction at IVC, comparison between the detailed process calculations and the estimations using the cylinder charge efficiency.

3.3.4 Torque Variations Due to Pressure Pulsations

The varying pressure in the cylinder produces a non-uniform torque on the crankshaft. If these variations must be modeled, the following approach can be used. The starting point for this analysis is the mass of fuel from which the combustion energy can be calculated with (3.73). Using (3.77), the pressure difference Δp_c can then be calculated.

Assuming reasonable values for the Vibe parameters and, thus, for the burned-mass fraction and inverting the relationship (3.71), an approximation of the cylinder pressure is obtained.

The pressure at reference position 1, p_{c1}, can be obtained according to (3.70) by assuming a polytropic compression starting with the manifold pressure p_m at the IVC crank angle to the chosen reference point at $\phi_1 = -80\,°$ crank angle.

The pressure acts on the piston surface producing a force

$$F_p(\phi) = p_c(\phi) \cdot A_p = p_c(\phi) \cdot \frac{V_d}{2 \cdot r_{cs}} \tag{3.83}$$

The d'Alembert force F_m, which takes into account the inertia[13] of the piston and the crank-slider parts, must be considered as well

$$F_m = -m_p \cdot \ddot{s}(\phi) \tag{3.84}$$

Neglecting the friction forces and formulating a power balance for the crankshaft, the torque T_e on the crankshaft can be calculated

$$T_e \cdot \omega_e = T_e \cdot \dot{\phi} = (F_p + F_m) \cdot \dot{s}(\phi) \tag{3.85}$$

Using the approximation (3.65), the following expression for the engine torque can be derived

[13] One-third of the connecting-rod mass is usually added to the piston mass. The other two-thirds are added to the crankshaft inertia.

$$T_e = \frac{(F_p + F_m) \cdot \dot{s}(\phi)}{\dot{\phi}} = (F_p + F_m) \cdot r_{cs} \cdot (\sin\phi + \frac{r_{cs}}{2 \cdot l} \cdot \sin(2\phi)) \qquad (3.86)$$

The final ODE describing the crank-shaft dynamics is obtained using a torque balance

$$T_e - T_l = \Theta_e \cdot \ddot{\phi} \qquad (3.87)$$

The rotational inertia of the system, Θ_e, includes the inertia of the crankshaft, the flywheel, and two thirds of the mass of the piston rods.

Finally, assuming each cylinder to produce the same torque pattern, the total torque can be derived by shifting the single cylinder torque appropriately. In Fig. 3.27, this is shown for a four-, a six- and an eight-cylinder engine. In these simulations, the rotational inertia has been chosen to be very large such that a constant engine speed results.

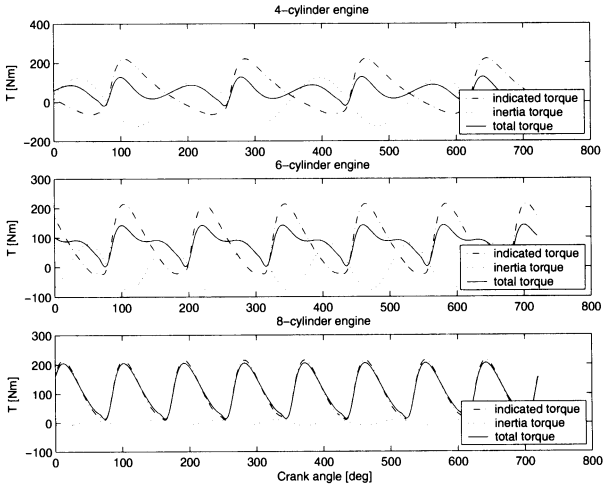


Fig. 3.27. Inertia torque, indicated torque and total torque for a four-, a six- and an eight-cylinder engine (note the different torque scales).

4

Control of Engine Systems

As discussed in the preceding chapters, engine systems incorporate a large number of control loops. For the design of these feedforward and feedback control systems, the main objectives are:

- *The driver's demands for immediate torque response, good drivability, and low fuel consumption must be met.*
- *The engine must be kept in a safe operating region where material damage or fatigue is avoided. Knocking must be prevented, the catalytic converter must not overheat, etc.*
- *The emission limits must be met. In the case of SI engines, this requires a rapid light off of the catalytic converter, precise stationary air/fuel ratio control, and good compensations of the transient phenomena.*

All these requirements are partially contradictory therefore, they must be fulfilled according to the priorities set by legislation and customer demands. Control systems play a major role as enabling technology *in this optimization.*

Complete treatment of all control problems arising in engine systems is beyond the scope of this text. The following four important case studies are presented with the intention to show how model-based methods can be used to streamline the design process:

- *The air/fuel ratio control system as a typical example of a combined feedforward and feedback control problem.*
- *The combined engine speed and air/fuel ratio control system, as an example of a MIMO control problem.*
- *The SCR control system, as an example how sensor models can help to improve the system performance.*
- *The engine temperature control system or engine "thermomanagement," as an example of a control system that must cope with a changing system structure.*

L. Guzzella et al., *Introduction to Modeling and Control of Internal Combustion Engine Systems*

4.1 Introduction

4.1.1 General Remarks

The general structure of automatic control systems is discussed in Appendix A. Readers not familiar with the main ideas of general control system theory may want to consult that section before starting this.

Figure 4.1 shows the various actuators and sensors of a modern SI engine management system. The ECU must set all actuator commands according to driver demands and available sensor signals, taking into account constraints imposed by engine safety and pollutant emission limitations.

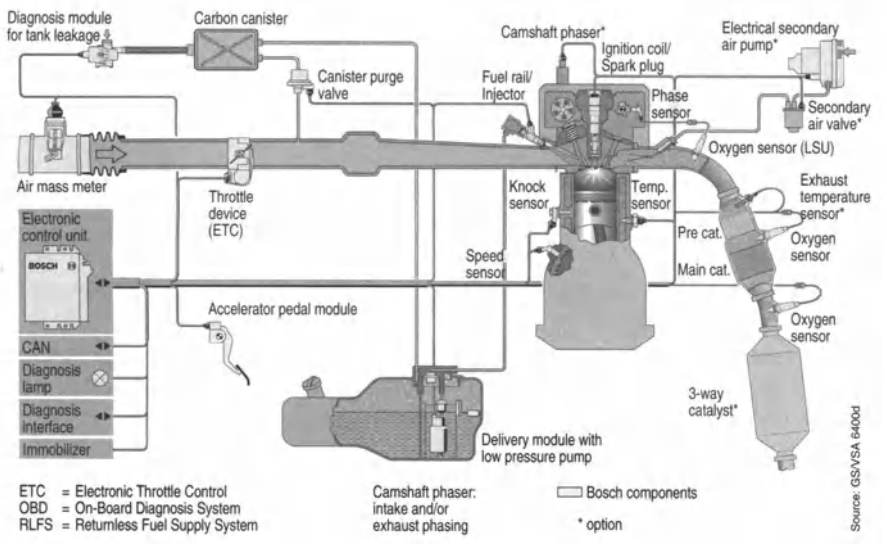


Fig. 4.1. Layout of a modern engine management system, reprinted with the permission of Robert Bosch GmbH.

In the following, it is assumed that the engine system is equipped with an electronic throttle device, i.e., the driver does not directly control the throttle plate angle. Instead, an accelerator pedal module captures the driver's input and produces an electronic signal sent to the ECU. This signal is usually interpreted as a desired engine torque. Such a decoupling permits a coordinated actuation of all involved actuators.

As discussed in Chapter 3, the engine speed sensor and the camshaft phase sensor allow synchronization of the ECU with the engine's reciprocating behavior. The throttle plate angle and the measured mass flow and temperature of the air entering the intake manifold provide the necessary information to calculate the required amount of fuel. Since the sensors are placed upstream

of the intake manifold, the mass of air entering the cylinder must be calculated using a physical model of the intake runner. This takes into account the temperature of the engine coolant as well.

The engine management system shown in Fig. 4.1 uses three catalytic converters: the close-coupled TWC optimized for rapid light off at cold start; the main TWC shortly downstream of the close-coupled TWC; and an optional underfloor catalytic converter. To control this complex exhaust aftertreatment system, three exhaust gas oxygen sensors are used: one as close as possible to the exhaust valves; one between the close-coupled and the main TWC; and one downstream of the main TWC. An optional exhaust gas temperature sensor can help to prevent damage to the catalytic converters due to overheating, and improves the overall control system performance. This configuration is adapted to the specific pollutant emission levels that are imposed by legislation. Obviously, a minimum number of TWC and sensors with which these limitations can be satisfied will be chosen in order to realize the configuration.

Two additional subsystems can be identified in Fig. 4.1:

- At cold start, the secondary air system pumps air into the exhaust manifold. Running the engine under slightly rich conditions, this air improves heating the catalytic converter system and shortens the light-off time.
- When the engine is not running, the fuel that evaporates in the tanks cannot be directed to the intake manifold for later combustion. Nevertheless, in order to prevent excessive pressures in the tank system, these fuel vapors must be vented. Since venting fuel vapors directly into the environment is not permitted, a carbon canister is used to collect them. This canister, having a limited capacity, must be purged when the engine is running. Control of that process and detecting unwanted leakages are some of the algorithms that are implemented in the ECU.

Notice that the structure shown in Fig. 4.1 does not represent the most general case. For instance, it does not contain an external EGR device or a turbocharging system. Additional sensors and actuators will be included in these situations.

4.1.2 Software Structure

The most important task of an engine is to deliver the desired amount of torque. As mentioned, the demanded torque is in general defined by the driver. However, several other subsystems influence that request, e.g., the cruise-control module, the catalyst heating system, the drive line ringing suppression control system, etc.

Figure 4.2 shows the often contradictory signal paths used in conventional engine management systems. For such systems, the design of the control algorithms is very difficult. Many control loops must be run in parallel and, due to permanent changes in the system structure, a correct multivariable control

system design is not possible. Moreover, software structures for control systems like the one shown in Fig. 4.2 are difficult to parameterize, and each new safety or monitoring task included at a later stage adds new and potentially disrupting signal couplings to the control system.

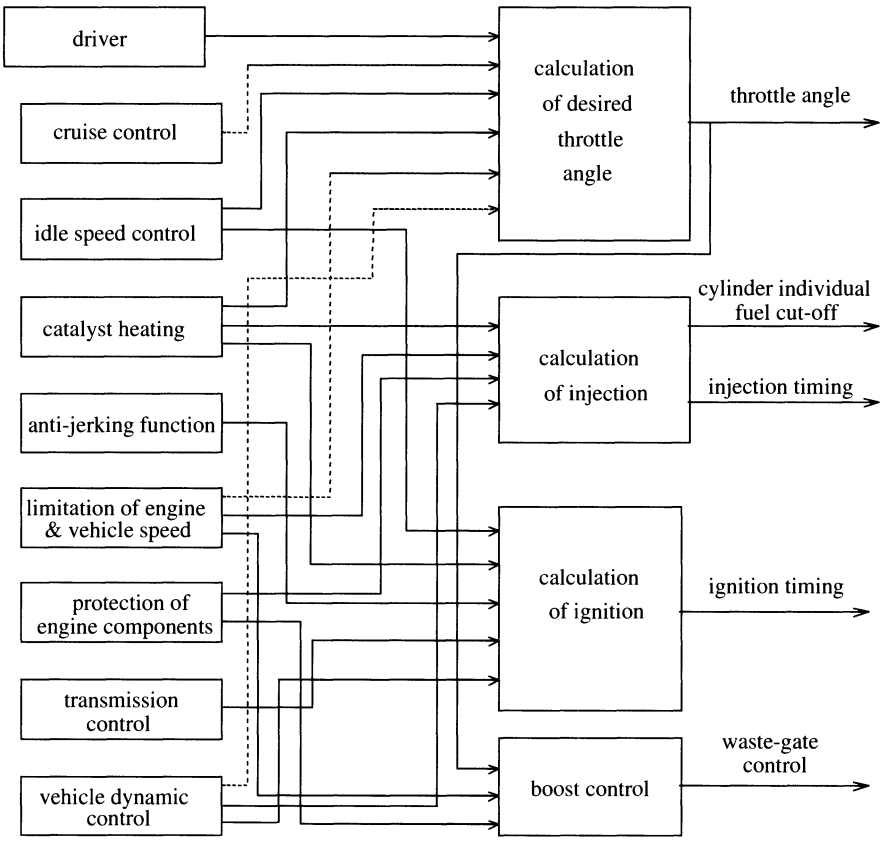


Fig. 4.2. Engine torque control structure in a conventional engine control unit (ECU).

If the *torque-based* structure shown in Fig. 4.3 is chosen as a general guideline for software development, the control system becomes structurally much simpler. This structure decouples the strategic decisions from the low-level control tasks. In fact, the *torque demand manager* collects and evaluates all demands for engine torque. Then, only one signal is transferred to the *torque conversion manager* that issues the appropriate commands to the actuators such that the demanded torque can be realized as best possible.[1]

[1] Of course, such an approach requires an electronic throttle valve.

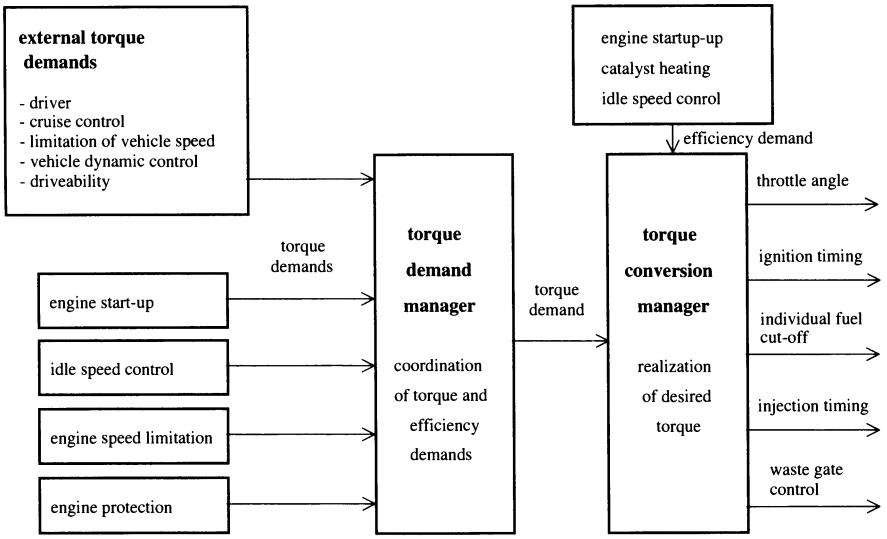


Fig. 4.3. Engine torque control structure in a torque-based engine control unit.

Since the torque-based structure reflects the underlying physics of the engine, the control algorithms implemented in the ECU are much closer to the actual processes. This makes the control system structure easier to understand and, thus, the corresponding software easier to maintain. For the design of the control system, only the blocks following the torque demand signal are important. The other blocks implement the algorithms for strategic decisions which are very important for fuel economy, pollutant emission, drivability and comfort. Thus, the set points for the various actuator input signals must be chosen mainly as a function of the desired torque at a defined engine speed. The corresponding block diagram is shown in Fig. 4.4. There, the block torque conversion contains the inversion of the torque model that is introduced in Sec. 2.5.1.

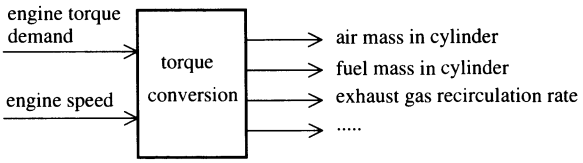


Fig. 4.4. Relevant subsystem for the control system design.

4.1.3 Engine Operating Point

Engine speed and engine torque are the most important input variables to the engine control systems. These two variables define the engine's *operating point*. The operating point characterizes the engine's main variables (air mass flows, pressures, pollutant formation, etc.). If an engine is fully warmed, most actuator inputs remain constant if the operating point is not changed.

To calculate actuator set points, instead of the actual torque, the *relative load* is often chosen as an independent variable. For each fixed engine speed, this variable indicates the actual air charge in the cylinder with respect to its maximum value. For a known full-load torque curve, the relative load can be used to derive the actual torque at a later stage.

In fully warmed conditions, the engine's actuator inputs are determined using steady-state optimizations (more will be discussed in Sect. 4.1.4). The results of these optimizations are stored in *maps* that parameterize the dependency of each actuator input on the engine operating point. It is not convenient to use maps with dimensions greater than two. Therefore, other influencing engine variables (e.g., coolant temperature, battery voltage, etc.) must be taken into account by adding one- or two-dimensional correction maps.

Using the ignition control loop as an example, Fig. 4.5 shows the most basic structure of such a map-based algorithm. Based on the measured engine speed and the relative load, a nominal spark advance is chosen (block "map1" in Fig. 4.5). Then the corrections are applied; in this simplified case a correction for engine temperature only is considered.[2] For such a separation of effects, it is important to understand the underlying thermodynamic principles. In the case of the spark advance, for instance, the effects explained in Fig. 2.33 and [57] have to be considered.

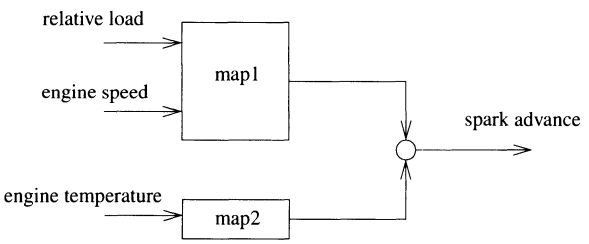


Fig. 4.5. Calculation of spark advance; example of a simple correction map structure.

Figure 4.6 shows an example of a spark advance map as a function of engine speed and load (torque) as implemented in a modern ECU. The map

[2] In a real ignition control system, additional corrections would be applied for varying battery voltages, to avoid engine overheating and knock, etc. In addition, the idle-speed control system is also influencing the ignition control loop.

is mainly defined by the best fuel-efficiency spark advance, but factors such as emission limits, knock avoidance, etc., are included as well.

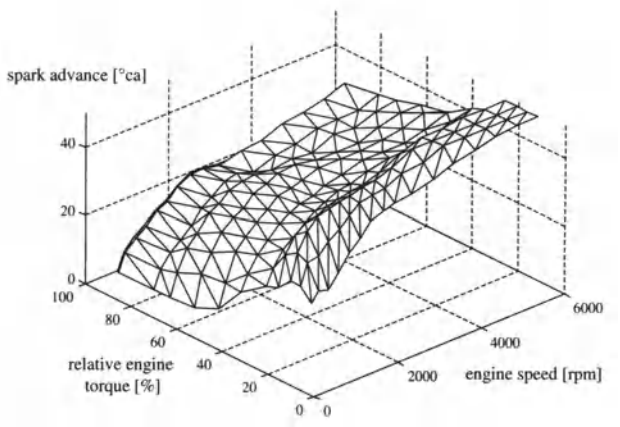


Fig. 4.6. Example of a "map1" as used in the example shown in Fig. 4.5.

Notice that, despite the fact that these maps are often referred to as feedforward control maps, they introduce feedback action in the control system. In fact, by using measured values for the actual engine speed and load to compute the actuator signals, the inherent dynamic properties of the controlled system are changed.

4.1.4 Engine Calibration

Although significant progress has been made to reduce engine-out emissions, these reductions are, in most cases, not sufficient for the engine to comply with actual and future emission regulations. From a control engineering perspective, these regulations reduce the degrees of freedom for the optimization of the engine control system by introducing additional boundary conditions. These constraints lead to increased fuel consumption since the engine control system cannot be further optimized other than with respect to minimum fuel economy.

The best-possible efficiency must be guaranteed in each operating condition and, simultaneously, the emission limitations, defined for a specific test cycle, must be met. This optimization can be performed using a Lagrange multiplier approach.

The cost function of the optimization is the fuel consumption over a driving cycle.

$$J = \int_0^{t_{dc}} \dot{m}_\varphi \cdot dt \tag{4.1}$$

where t_{dc} is the duration of the driving cycle. The boundary conditions are defined by the pollutant emission constraints that must be met:

$$m_{HC} = \int_0^{t_{dc}} \dot{m}_{HC} \cdot dt \leq \overline{m}_{HC} \tag{4.2}$$

$$m_{CO} = \int_0^{t_{dc}} \dot{m}_{CO} \cdot dt \leq \overline{m}_{CO} \tag{4.3}$$

$$m_{NO_x} = \int_0^{t_{dc}} \dot{m}_{NO_x} \cdot dt \leq \overline{m}_{NO_x} \tag{4.4}$$

Of course, the pollutant mass flows $\dot{m}_{...}$ are tail-pipe emissions, i.e., the dynamics of the aftertreatment systems must be included in the system analysis.

Equations (4.1) and (4.4) describe the system using a continuous-time approach. However, especially in the European test cycle, the engine system can often be assumed to operate in a series of quasi-stationary points. Therefore, by splitting up the driving cycle into a number of *constant* operating points with associated retention periods $t_{op,i}$, the optimization problem can be reformulated as follows

$$\begin{aligned} J = \sum_{op.points} \dot{m}_{f,i} \cdot t_{op,i} &- \lambda_{HC} \cdot (\overline{m}_{HC} - \sum_{op.points} \dot{m}_{HC,i} \cdot t_{op,i}) \\ &- \lambda_{CO} \cdot (\overline{m}_{CO} - \sum_{op.points} \dot{m}_{CO,i} \cdot t_{op,i}) \\ &- \lambda_{NO_x} \cdot (\overline{m}_{NO_x} - \sum_{op.points} \dot{m}_{NO_x,i} \cdot t_{op,i}) \end{aligned} \tag{4.5}$$

or in the form

$$\begin{aligned} J = \sum_{op.points} &(\dot{m}_{f,i} + \lambda_{HC} \cdot \dot{m}_{HC,i} + \lambda_{CO} \cdot \dot{m}_{CO,i} + \lambda_{NO_x} \cdot \dot{m}_{NO_{x,i}}) \cdot t_{op,i} \\ &- \lambda_{HC} \cdot \overline{m}_{HC} - \lambda_{CO} \cdot \overline{m}_{CO} - \lambda_{NO_x} \cdot \overline{m}_{NO_x} \end{aligned} \tag{4.6}$$

For a given set of Lagrange multipliers, the sum over all operating points is then optimal, when each summand in each operating point is optimal. Of course, this requires the single operating points to be mutually independent. This assumption is well satisfied, especially if cold-start situations are excluded, for the engine-out emissions and for fuel consumption. Engine optimization methods based on this assumptions have been presented in [30], [169], [167], and [129]. The dynamic phenomena are later considered by the feedforward and feedback loops discussed in subsequent sections.

Thus, in each operating point

$$J = \dot{m}_{f,i} + \lambda_{HC} \cdot \dot{m}_{HC,i} + \lambda_{CO} \cdot \dot{m}_{CO,i} + \lambda_{NO_x} \cdot \dot{m}_{NO_{x,i}} \tag{4.7}$$

must be minimized using the following procedure:

- Set all Lagrange multipliers to zero. The resulting optimum is the fuel optimal solution.
- Determine the sum of the masses of the limited components over the complete test cycle.
- If all limits are met, the optimization is finished. If one or more limits are exceeded, the corresponding Lagrange multipliers must be increased.
- Repeat the optimization process with this new cost function.

The optimization process consists of changing the parameters of the engine control system.[3] This can be accomplished using a purely experimental, or a partially model-based, approach [127], [165]. In both cases, the iterations can be performed using heuristics based on experience with previous engine calibration tasks or using systematic approaches, e.g., forming the gradients of the cost function (4.6) and using these vectors in a steepest-descent algorithm as discussed in [140], [91], [148] and [149].

4.2 Air/Fuel Ratio Control System

The air/fuel ratio control system is one of the most important control loops. As discussed in detail in Sect. 2.8.2, the TWC achieves its best efficiency only if the engine is operated with a stoichiometric air/fuel ratio. Due to the TWC's ability to store oxygen and carbon monoxide on its surface, short excursions of the air/fuel ratio can be tolerated as long as they do not exceed the remaining storage capacity. These two observations are the starting point for the design of an air/fuel ratio control system that comprises a slow but precise feedback control loop, and an approximate but fast feedforward control loop. These two parts will be discussed separately below.

4.2.1 Feedforward Control System

Without an appropriate feedforward component the performance of the air/fuel ratio control system would be much too slow. This is because of the many delays and lags between the input (fuel injection) and the output (air/fuel ratio sensor) of this dynamic system.

Under transient operating conditions, the correct amount of fuel to be injected must be determined as soon as possible. This is only possible if the intake manifold dynamics are considered. The available information on either the throttle plate angle, the manifold pressure, or the air mass flow into the intake manifold must be used to predict the air mass flow into the cylinder. This information is then used to estimate the amount of fuel to be injected in the next cycle. Since the fuel path has its own "wall-wetting" dynamics, that effect has to be compensated as well. These two parts will be discussed separately below.

[3] This process is often referred to as the *engine calibration*.

Air Mass Estimation

The mean value model of the manifold dynamics was introduced in Sect. 2.3 and Sect. 3.2.2. The main continuous-time domain equations describing the air flows in an SI engine are that which models the air mass flow entering the intake manifold

$$\dot{m}_\alpha(t) = c_d \cdot A(t) \cdot \frac{p_a}{\sqrt{R \cdot \vartheta_a}} \cdot \Psi(p_a, p_{in}(t)) \tag{4.8}$$

and that which models the air mass flow leaving the intake manifold and entering the cylinders

$$\dot{m}(t) = \rho_m(t) \cdot \dot{V}(t) = \frac{p_m(t)}{R \cdot \vartheta_m} \cdot \lambda_l(p_m(t), \omega_e(t)) \cdot \frac{V_d}{N} \cdot \frac{\omega_e(t)}{2\pi} \tag{4.9}$$

The amount of fuel must be calculated according to the amount of air entering the cylinder. Since a mass flow measurement at the intake of the cylinder is usually not possible, a model-based estimator must be used to obtain that information.

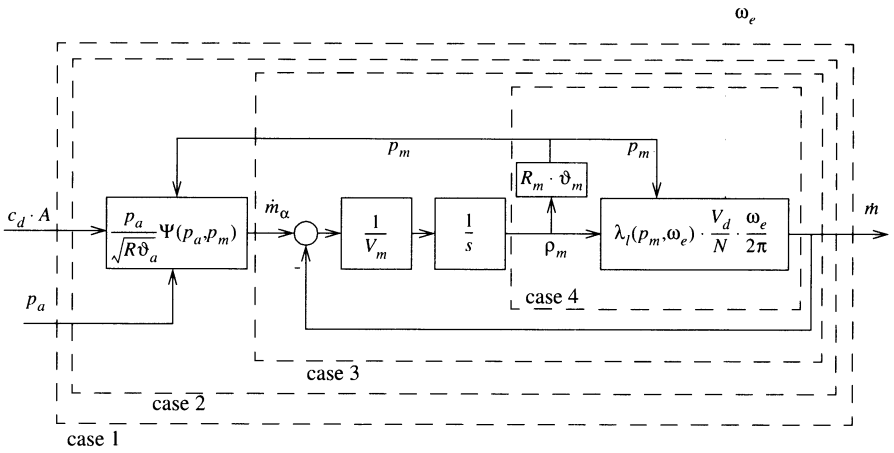


Fig. 4.7. Possible approaches to estimate the air mass in the cylinder, continuous-time formulation.

Depending upon the available sensors, four possible structures are commonly used for such an air mass flow estimator (see also Fig. 4.7):

1. The throttle angle is measured. The corresponding throttle area $A(t)$ multiplied with the discharge coefficient c_d can be calculated with a-priori measurements. The rest of the manifold dynamics must be simulated with an open-loop observer. Correct steady state operation cannot be guaranteed, since ambient pressure p_a can vary and manufacturing tolerances

result in wrong throttle area estimates. This is the cheapest solution, but obviously, also the least precise.

2. Same as case 1, except that ambient pressure p_a also is measured. Therefore, a correct altitude compensation is possible. Ambient pressure sensors are not expensive. Because they can be mounted directly on the ECU board, this improvement is usually adopted.
3. The air mass flow $\dot{m}_\alpha$ entering the intake manifold is measured with an air mass flow sensor (typically a hot-film anemometer). This guarantees a stationary correct estimation of the air mass flow into the cylinder. The compensation of the transient effects is realized using an open-loop observer. This is the preferred sensor configuration since only the precision of the air mass flow measurement is relevant under steady state conditions.
4. The intake manifold pressure p_m is measured. At first, this configuration seems to be optimal since the dynamically correct signal is used. Nevertheless, a relevant intake manifold temperature has to be estimated and the volumetric efficiency must be known with high precision. An additional disadvantage is that the information about transient changes in the throttle area $A(t)$ are perceived with substantial delay. Therefore, predicting the desired amount of fuel (which is necessary due to the timing delays between fuel and air) is rather difficult.

So far, continuous-time models have been discussed. However, to realize high-precision feedforward control systems a discrete-event formulation will be necessary. Using the discrete-event representations introduced in Sect. 3.2.2 the discrete-event signal flowchart of the cylinder air mass estimation shown in Fig. 4.8 can be derived.

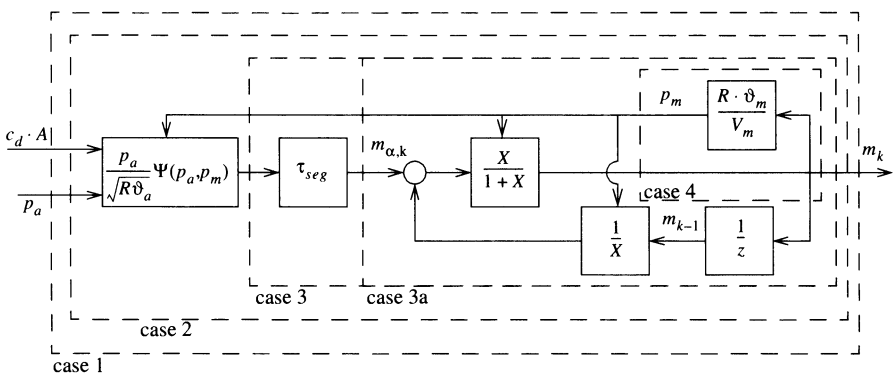


Fig. 4.8. Possible approaches for the estimation of the air mass in the cylinder, discrete-time formulation.

In the discrete-event formulation, the same three cases discussed in the continuous-time formulation are recognized. The only difference is that case

3a was not present before. In fact, in the discrete-event formulation of case 3, either the mass flow $\dot{m}_{\alpha,k}$ or the total mass $m_{\alpha,k}$ flowing into the manifold, is chosen as the input signal to the estimator. In the first case, the integration over one segment is accomplished in the ECU using a Euler forward integration (multiplication of the mass flow by τ_{seg}). In the second case, the output signal of the hot-film anemometer is integrated using a dedicated hardware component. Usually the second approach yields the more precise result.

Compensation of the Fuel Dynamics

General Approach

The discrete-event structure of the air mass dynamics has the advantage of being easily combined with the discrete-event representation of the fuel dynamics. Assuming a first-order wall-wetting dynamics and an additional timing delay of two segments, the block diagram shown in Fig. 4.9 represents the model used to determine the duration of the fuel injection.

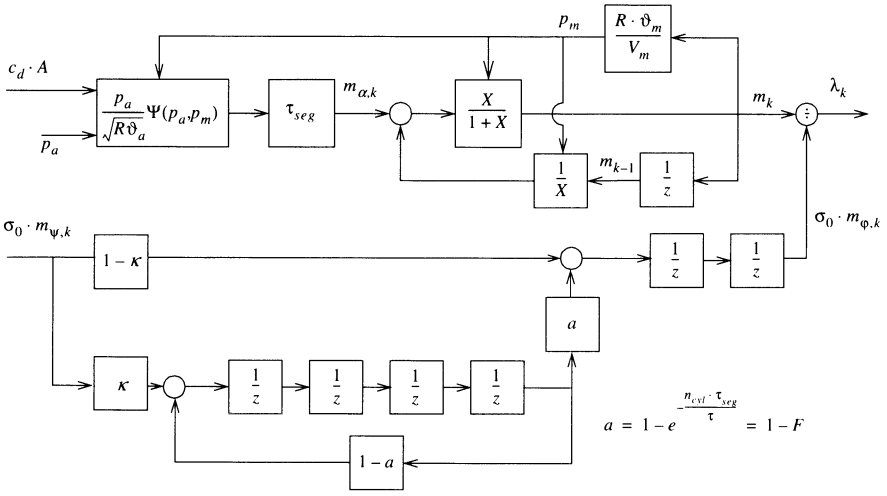


Fig. 4.9. Discrete-time air/fuel ratio model.

The mass of air in the cylinder can be estimated using one of the approaches introduced above. The wall-wetting dynamics can be compensated for as explained in Sect. 2.4.2, although the additional two-segment time delay cannot be eliminated. One way to cope with this problem is to artificially delay the input to the throttle actuator by two segments and to predict the corresponding air mass in the cylinder using a case 1 or case 2 estimator. Using this air mass information, the amount of fuel to be injected is calculated (inverting the wall-wetting dynamics) and the corresponding signal is sent to

the power amplifiers of the injectors. If the air mass flow is measured, the air mass estimator can be corrected by using the measured value and taking into account the time delays.

The drawback of this solution is that at low engine speeds, the driver has to accept a significant time delay when issuing an acceleration request (for 900 rpm and a four-cylinder engine this delay amounts to 60 ms). If this is not acceptable, the air mass must be predicted using a suitable driver command prediction algorithm.

Combining these methods (delaying the driver pedal, predicting the actuation) a good drivability and a close-to-stoichiometric air/fuel ratio mixture can be guaranteed in most transients. An additional late injection into the open intake valve may become necessary in some extreme tip-in situations .

Limitations

It is interesting to note that the inversion of the wall-wetting dynamics is only possible for physically reasonable parameters. The discrete-time transfer function $G(z)$ of the wall-wetting dynamics for one cylinder can be expressed according to (3.28) as follows

$$G(z) = (1 - \kappa) + \kappa \cdot \frac{1 - F}{z - F} \tag{4.10}$$

The inverse wall-wetting dynamics can thus be expressed with the following discrete transfer function

$$G^{-1}(z) = \frac{1}{1 - \kappa} - \frac{\kappa}{1 - \kappa} \cdot \frac{1 - \frac{F-\kappa}{1-\kappa}}{z - \frac{F-\kappa}{1-\kappa}} \tag{4.11}$$

The variable κ describes how much of the injected fuel adheres on the wall. For $\kappa = 1$ all of the injected fuel accumulates on the manifold wall, for $\kappa = 0$ all of the injected fuel enters the cylinder in the same cycle. The wall film evaporation rate is expressed by F, where a small F indicates a quick evaporation. Accordingly, these parameters satisfy the inequalities

$$0 \leq F \leq 1 \qquad \text{and} \qquad 0 \leq \kappa \leq 1 \tag{4.12}$$

Obviously, the only pole p of the discrete time representation of the inverted wall wetting dynamics (4.11) lies between -1 and 1 to guarantee that the inverted system is stable. If mere stability is not enough, but non-oscillatory behavior is requested as well, then the pole must be confined to the restricted interval $0 \leq p \leq 1$. These two requirements yield the following conditions for the parameters κ and a:

condition for stability

$$2 \cdot \kappa - 1 \leq F \leq 1 \tag{4.13}$$

and condition for non-oscillating behavior

$$\kappa \leq F \leq 1 \tag{4.14}$$

These limits are illustrated in Fig. 4.10. Obviously, for a gain $\kappa > 0.5$ the evaporation rate may not exceed certain values (F may not be too small) if unstable transients are to be avoided. If oscillations must be avoided as well, the evaporation rate may not be higher (F not be smaller) than the deposition rate κ. This fact can be explained as follows. For a gain κ close to one (almost all fuel adheres on the wall), the term $\frac{1}{1-\kappa}$ is relatively large, i.e., in this situation a full compensation of the wall-wetting dynamics forces the feedforward control system to inject a relatively large amount of fuel. A large part of this fuel is deposited on the manifold walls and the size of the wall film increases substantially. Therefore, if the evaporation process is fast (F is small), after one engine cycle a large amount of evaporated fuel will be ready for induction into the cylinder. This amount can exceed the amount of fuel needed to achieve a stoichiometric air/fuel ratio, resulting in an overshooting behavior.

With a well-designed injection system (injector position, intake runner geometry, etc.), this phenomenon can be avoided. Also notice that large gains κ are usually encountered at cold start. Fortunately, in this situation the evaporation rates are slow, i.e., large parameters F are to be expected.

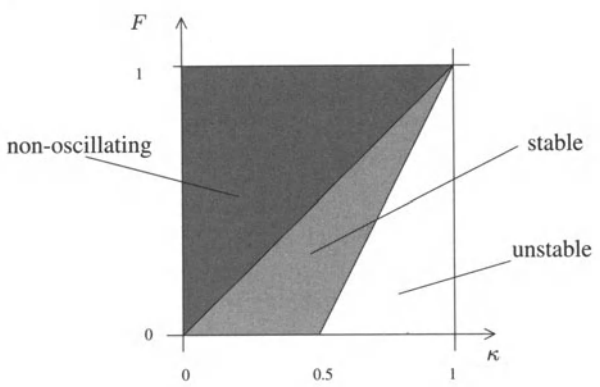


Fig. 4.10. Conditions for κ and F for stability and non-oscillating behavior of the wall-wetting compensation.

4.2.2 Feedback Control System

Although the TWC is able to cope with most dynamic air/fuel ratio deviations, the steady state error must be very small (less than 0.1%) such that the feedforward control system discussed above must be complemented with a feedback control loop. In this section, three possible air/fuel ratio feedback

control systems are introduced. The first corresponds to the classical solution adopted in many of the early systems. The second belongs to the class of model-based control systems and is a representative of the modern solutions. The third introduces a possible extension in which both the air/fuel ratio and engine speed are controlled with one control system. This approach belongs to the class of MIMO control systems.

Conventional Air/Fuel Feedback Control Systems Using Switch-Type Sensors

One straightforward approach to realize an air/fuel ratio feedback control system is to use a single switch-type air/fuel ratio sensor upstream of the TWC. The voltage function of a typical switch-type sensor is depicted in Fig. 4.11.

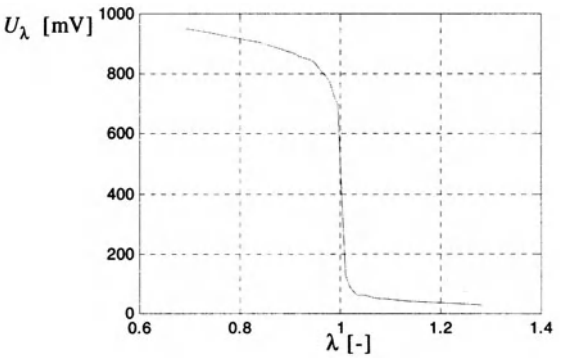


Fig. 4.11. Steady state voltage characteristic of the Bosch LSF switch-type air/fuel ratio sensor.

Typically, the sensor output is converted to a binary signal, lean for sensor voltages U_λ smaller than 450 mV, rich for U_λ greater than 450 mV. Using this binary signal, a control structure similar to a PI system is chosen. Figure 4.12 shows the block diagram of such a control system.

Instead of the usual additive approach, the correction of the nominal injection duration τ_{inj} is achieved by *multiplying* this value with the control system output F_λ. This approach is advantageous because of the multiplicative definition of the air/fuel ratio signal and because most errors in the air/fuel path have a multiplicative nature.[4]

[4] For instance, slightly blocking an injection valve with soot can produce a 10% smaller aperture area and, therefore, a 10% smaller amount of fuel injected. The correction of that error requires the nominal injection duration to be increased by 11%. If no changes in the valve aperture occur, all subsequent injection durations must be multiplied by the *same* correcting value.

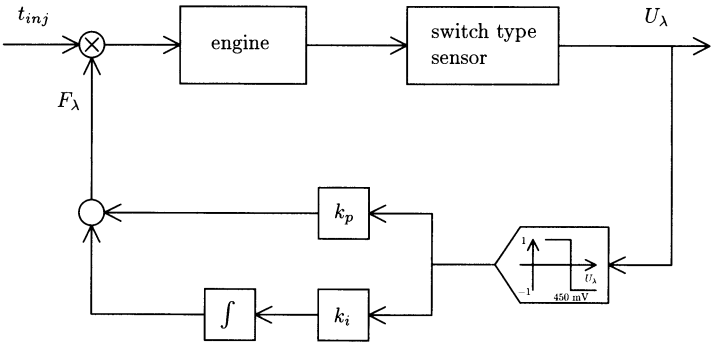


Fig. 4.12. Block diagram of the air/fuel ratio control loop with a switch-type sensor.

Due to the asymmetric behavior of the switch-type sensor with respect to steps from rich-to-lean compared to those from lean-to-rich, the control system shown in Fig. 4.12 produces a bias of the mean value of the air/fuel ratio to lean conditions. To compensate this bias and to achieve a shift of the mean air/fuel ratio to slightly richer conditions,[5] the switch from rich-to-lean is delayed by a time interval t_v.

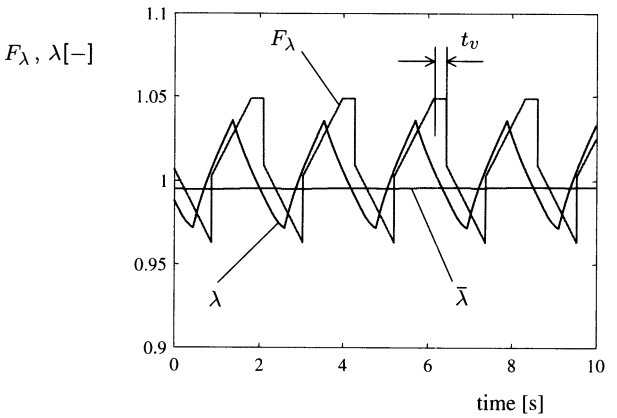


Fig. 4.13. Simulated steady-state behavior of a switch-type control system.

Figure 4.13 shows a simulation of the resulting trajectories for λ and for the control input F_λ. The fuel path of the engine is modeled as a first-order lag with a time delay. The dynamics of the switch-type sensor is neglected in that simulation. This figure also shows the mean value $\bar{\lambda}$ of λ that is obtained by

[5] As discussed in Sect. 2.8.2, the optimal catalytic converter efficiency is achieved for $\lambda = 0.99 \ldots 1.00$.

averaging λ using a moving window whose size is compatible with the oxygen storage capacity of the TWC.

In summary, this type of control has three parameters $\{k_p, k_i, t_v\}$ that must be tuned for every operating point. Since these parameters must be experimentally identified, the parametrization becomes rather time consuming.

The above approach works well as long as the emission levels that have to be attained are not too demanding. Note that with this control structure, the TWC is operated in an open-loop approach, i.e., its dynamics are not considered.

Model-Based Air/Fuel Feedback Control Systems Using Wide-Range Sensors

Control System Structure and Model Simplifications

The rather complex dynamic behavior of a TWC has been modeled in Sect. 2.8.2. It is advantageous to include these storage effects into the development of the fueling control system. In particular, the excursions of the air/fuel ratio in one direction must be compensated with equivalent corrections in the other.

The block diagram of the plant model used for to develop the feedback air/fuel ratio control system is depicted in Fig. 4.14. Only the fuel path from the injector valves to the air/fuel ratio sensor is considered. The air path is not included, and all influences of that component are treated as disturbances.

The sensor, which is placed immediately upstream of the TWC, is assumed to be a wide-range sensor, giving not only lean-rich binary information on the air/fuel ratio, but also quantitative values. This sensor is used to form a first fast air/fuel ratio control loop. In a second stage, an additional air/fuel ratio sensor downstream of the TWC is used to close an outer, slower control loop. This loop takes the oxygen storage capacity of the TWC into account and compensates for sensor drifts in the inner loop.

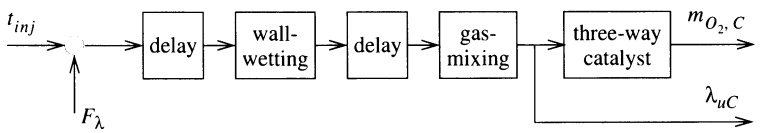


Fig. 4.14. Block diagram of the fuel path model.

As in the classical approach discussed in the last section, the output of the control system is a multiplicative factor F_λ that corrects the nominal injection duration. The error signal used as the input to the control system is defined in a multiplicative way as well.

The outputs of the plant are the air/fuel ratio value at the main confluence point λ_{uC} (upstream of the TWC) and the mass of oxygen stored in the TWC,

$m_{O_2,C}$. The system up to the main confluence point has already been analyzed in detail in Chapter 3. However, that model is too complicated to be used in the design of a feedback control system. Therefore, the complete air/fuel ratio path, from the fuel injectors to the air/fuel ratio sensor at the main confluence point, is approximated by a series connection of a first-order lag with time constant τ and a time delay δ

$$P(s) = \frac{\Delta\lambda_{uC}}{\Delta F_\lambda} = \frac{-1}{s \cdot \tau + 1} \cdot e^{-s \cdot \delta} \tag{4.15}$$

In this expression, the error is defined by $\Delta\lambda_{uC} = \lambda_{uC} - \lambda_{uCd}$ where λ_{uCd} is the demanded value for the upstream catalytic converter air/fuel ratio, and linearized input is calculated according to $\Delta F_\lambda = F_\lambda - 1$.

A typical measured and simulated step response is shown in Fig. 4.15. The engine used in this experiment is a 1.8-liter four-cylinder SI engine. The approach presented below is applicable to any port-injected SI engine. In this section, the numerical and experimental data will better illustrate that the proposed design methodology is always valid for that engine.

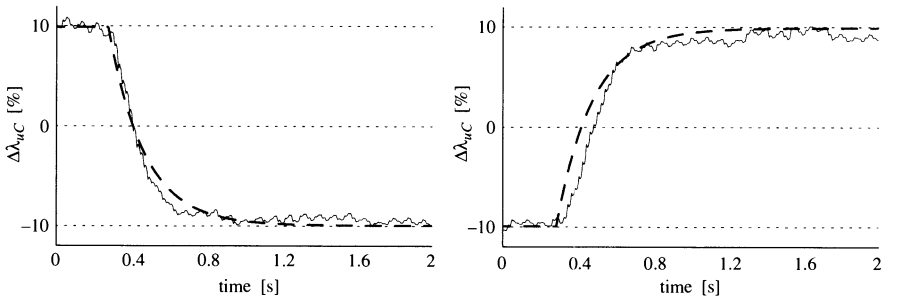


Fig. 4.15. Simulated (dashed) and measured step response of the engine mentioned in the text; operating point 1500 *rpm* and 0.1 *g* air per cylinder per cycle (approximately 25% of full-load).

The parameters of this model, i.e., the time constant τ of the lag element and the time delay δ, depend on the operating point of the engine. An on-line identification for these two parameters is possible using the approach presented in [155]. A complete map of these two parameters can easily be obtained off-line as well. Figure 4.16 shows an example of such maps for the full load range and the relevant speed range.[6]

The actual design of the feedback controller can be accomplished in many ways. Here, an approach is shown that is based on the H_∞ frequency-domain

[6] For higher engine speeds, the parameters do not change substantially because the non-engine speed dependent factors such as air/fuel ratio sensor time constant, etc. become dominant.

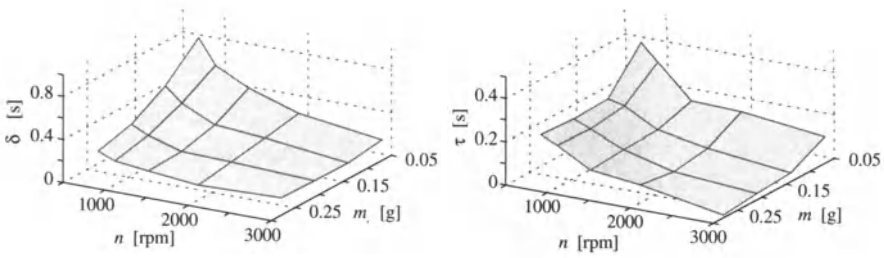


Fig. 4.16. Time delay δ and time constant τ of the fuel path of a 1.8-liter four-cylinder engine for different engine speeds n and air masses in the cylinder m.

optimization technique.[7] For this control design method, a linear and finite dimensional system description is required. Therefore, the time delay δ must be approximated using a Padé allpass element

$$e^{-s\cdot\delta} \approx \frac{\sum_{j=0}^{n_{apx}} a_j \cdot (-s\cdot\delta)^j}{\sum_{j=0}^{n_{apx}} a_j \cdot (s\cdot\delta)^j} \qquad a_j = \frac{(2\cdot n_{apx}-j)!\cdot n_{apx}!}{j!\cdot(n_{apx}-j)!\cdot(2\cdot n_{apx})!} \tag{4.16}$$

This approach has been found to be the best approximation method in this context. The order n_{apx} of the approximation is chosen as a function of the time delay. For delays up to 40 ms, a first order approximation is recommended, and for every additional 40 ms, one order should be added. This results in a rather precise approximation of the delay, but also in a rather high system order. Accordingly, in order to obtain a small-order control system, an order reduction step will be necessary.

Inclusion of the Oxygen Storage Capacity

Since the three-way catalytic converter is capable of storing oxygen as well as carbon monoxide, it is not the actual value of the air/fuel ratio that is relevant for tailpipe emissions. As long as an excess or lack[8] of oxygen can be compensated for in the TWC, no substantial emissions are expected at the tailpipe. Only if the storage capacity of the TWC is reached, is a rise in NO_x or CO and HC concentration, respectively, expected. Therefore, as suggested in (2.193) in Sect. 2.8.2, an integrator is added to the system dynamics described above to model the storage behavior of the TWC

$$\Delta m_{O_2,C} = \frac{1}{s}\cdot \dot{m}_{O_2,uC}\cdot \Delta\lambda_{uC} \tag{4.17}$$

The variable $\dot{m}_{O_2,uC}$ defines the oxygen mass flow at the same location where the upstream wide-range sensor is placed. For the linear control design, a value

[7] For a general discussion of this control system design method, the reader is referred to [44] for more information about the engineering aspects, [34] and [145] are good sources.

[8] Missing oxygen describes rich gas mixtures which result in high levels of CO.

for $\dot{m}_{O_2,uC}$ equal to 1 is assumed, i.e., a normalized system representation is used. In the controller realization, the correct mass flow will be considered.

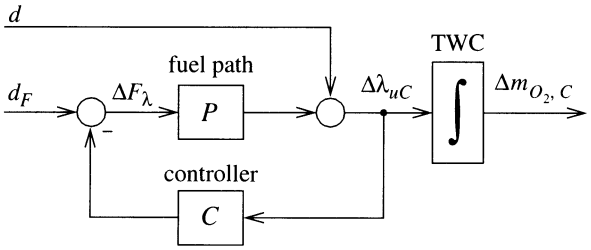


Fig. 4.17. Block diagram of the closed-loop system model used for the control system synthesis.

Inner Control Loop

The goal of the first controller's design is to minimize the impact of an air/fuel ratio deviation on the oxygen storage level in the TWC. Using a wide-range sensor, this must be achieved by measuring only the upstream air/fuel ratio (see Fig. 4.17).

The transfer function from the disturbance d to the oxygen mass $\Delta m_{O_2,C}$ stored in the TWC is given by 4.18

$$G_{Cd}(s) = \frac{\Delta m_{O_2,C}}{d} = \frac{1}{s} \cdot \frac{1}{1 + P(s) \cdot C(s)} \tag{4.18}$$

where $P(s)$ has been defined in (4.15) and (4.16) and $C(s)$ is the controller designed below.

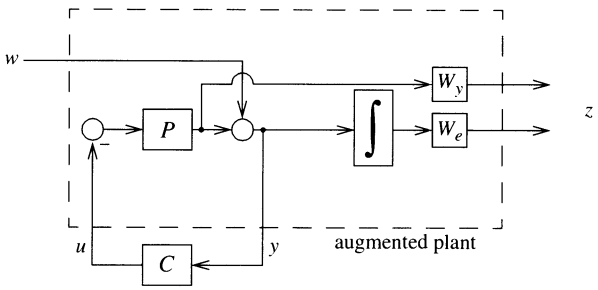


Fig. 4.18. Block diagram of the augmented control system used in the H_∞ design; w generalized disturbance input, z generalized output whose magnitude has to be minimized.

For the design of an H_∞ control system, the plant must be augmented by appropriate weighting functions. The block diagram of the resulting closed-loop system is shown in Fig. 4.18. The associated closed-loop transfer function T_{zw} has the form

$$T_{zw}(s) = \begin{bmatrix} W_e(s) \cdot G_{Cd}(s) \\ W_y(s) \cdot T_e(s) \end{bmatrix} \qquad T_e(s) = P(s) \cdot C(s) \cdot (1 + P(s) \cdot C(s))^{-1} \quad (4.19)$$

If a stabilizing controller exists that satisfies the conditions $||T_{zw}||_\infty < 1$,[9] the following two inequalities for the disturbance transfer function G_{Cd} and for the complementary sensitivity T_e will be satisfied

$$\begin{aligned} |G_{Cd}(j\omega)| &< |W_e^{-1}(j\omega)| \\ |T_e(j\omega)| &< |W_y^{-1}(j\omega)| \end{aligned} \qquad \forall\ \omega \quad (4.20)$$

The dynamic weight $W_e(s)$ thus serves as an upper bound for the magnitude of the disturbance transfer function G_{Cd}. A possible choice for $W_e(s)$ is

$$W_e(s) = \frac{1}{d_{max}} \cdot \frac{s + \omega_c}{s} \quad (4.21)$$

The corresponding magnitude plot is shown in Fig. 4.19. This diagram also illustrates the influence of the two design parameters d_{max} and ω_c on the disturbance transfer function $G_{Cd}(s)$. The parameters d_{max} limits the maximum value of $|G_{Cd}(j\omega)|$. At low frequencies, the disturbance transfer function $G_{Cd}(s)$ is forced to zero. This is achieved by the integrator element included in $W_e(s)$. The frequency at which this reduction starts can be specified by the parameter ω_c.

To avoid an unnecessarily high bandwidth of the control system, the complementary sensitivity $T_e(s)$ is limited with an appropriate choice of $W_y(s)$

$$W_y(s) = \frac{s}{k_c \cdot \omega_c} \quad (4.22)$$

where $k_c \approx 10$ is recommended. An example of the magnitude of the weighting function $W_y(s)$ is shown in Fig. 4.20.

The plant parameters τ and δ vary over quite a large range. For the engine example used in this section, the following range must be considered

$$\delta \in [0.02\ ,\ 1.0]\ s \qquad \tau \in [0.01\ ,\ 0.5]\ s$$

The first-order lag in the plant (4.15) can always be compensated and has, therefore, no influence on the weighting functions. The time delay cannot be compensated and the performance specification in the weighting matrices must be adjusted to the varying operating points and, hence, varying

[9] See (A.47) in the Appendix A for a definition of the H_∞ norm.

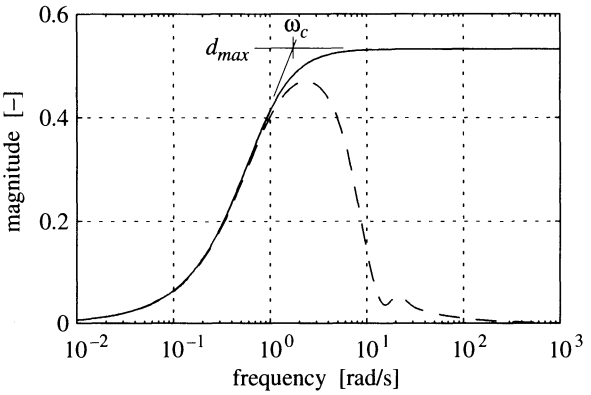


Fig. 4.19. Magnitude plot of $W_e^{-1}(j\omega)$ (solid) and $G_{Cd}(j\omega)$ (dashed).

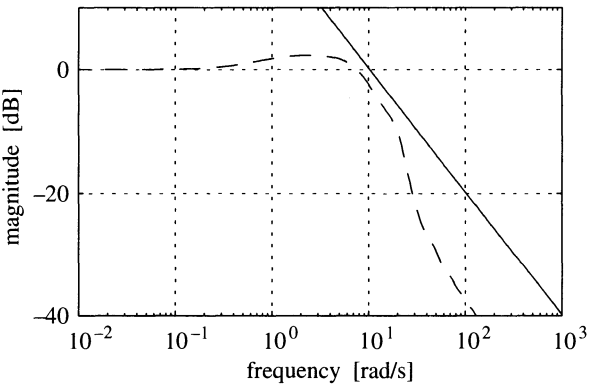


Fig. 4.20. Magnitude of $W_y^{-1}(j\omega)$ (solid) and $T_e(j\omega)$ (dashed); parameter $k_c = 13$.

delays. In other words, for larger time delays it is mandatory that the demanded bandwidth of the control system is reduced. This leads to the following parametrization of the design parameters

$$d_{max} = f_{d_{max}} \cdot \delta, \qquad \omega_c = \frac{f_{\omega_c}}{\delta} \tag{4.23}$$

This choice of the design parameters, which has been found empirically, results in an almost identical Nyquist diagram for G_{Cd} and, thus, in the same robustness properties over the whole operating range of the engine. The parameters $f_{d_{max}}$ and f_{ω_c}, which are the same for all operating points, allow a shaping of the Nyquist curve. The following values have been found to be a good choice for the engine discussed in this section

$$f_{d_{max}} = 1.9, \qquad f_{\omega_c} = 0.22 \tag{4.24}$$

Since the phase shift induced by the time delay cannot be avoided, a robust stabilizing controller must place significant actuator energy (high loop gain) at frequencies at the opposite of the Nyquist point -1. The solution $C(s)$ to the H_∞ problem formulated above produces one or several characteristic "bubbles" in the Nyquist diagram of the loop transfer function

$$L(s) = \frac{-1}{s \cdot \tau + 1} \cdot e^{-s \cdot \delta} \cdot C(s) \tag{4.25}$$

in the right half plane as shown in the Fig. 4.21.

Assuming that the error in the estimation of the time delay δ is small, this relatively high gain at frequencies higher than the crossover frequency causes a differential behavior of the loop that adds beneficial damping to the control system. Notice though that an error in the identified time delay results in an additional, usually detrimental phase shift of the Nyquist curve. Accordingly, using more than one "bubble" is only possible with a very accurate system model at hand.

The "one-bubble" form can always be realized with a fourth-order controller (the "bubble" requires a complex conjugate pair of poles). If a "no-bubble" form is chosen, then the minimal order of the controller is two, with both poles at zero in order avoid a steady state error in $\Delta m_{O_2,C}$.

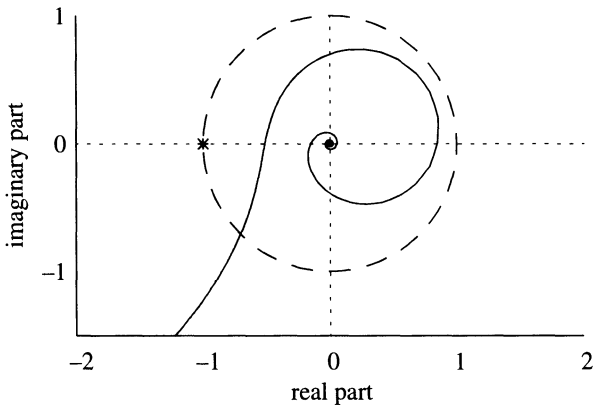


Fig. 4.21. Nyquist plot of the open-loop gain $L(j\omega)$.

For a 10% disturbance step, the control system designed in this section yields the response depicted in Fig. 4.22. The effect of varying d_{max} and ω_c on the resulting trajectories is indicated in the same figure. Decreasing d_{max} results primarily in a faster descent of the air/fuel ratio back to the stoichiometric value, and thus in a smaller peak of the oxygen to be stored in the TWC. Increasing ω_c results in a faster descent of the total oxygen stored in the TWC to its reference value. Assuming a constant air mass flow, the return

of the stored oxygen mass to its nominal value can obviously only be achieved if $\Delta\lambda_{uC}$ does not return directly to zero but overshoots or undershoots to realize the necessary compensation effects.

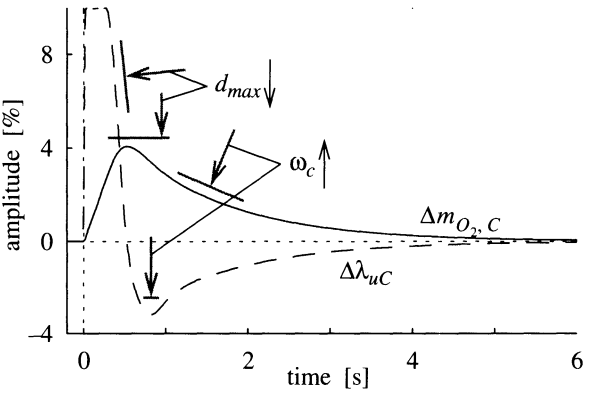


Fig. 4.22. Response of the closed-loop system to a 10% step disturbance $d(t)$.

Realization Aspects

As mentioned above, with increasing delay times, the Padé approximation of the delay element causes an increase of the plant model order and, hence, of the control system. However, a fourth-order controller is sufficient to realize the specific form proposed in the last section. Therefore, an order-reduction step is usually added after the full-order control system has been synthesized. Any of the well-known order-reduction techniques can be applied. More details of that part can be found in [145].

Several state-space forms can be chosen to realize the fourth-order controller. Because of the order reduction, the states of the control system are different in each operating point and allow no direct physical interpretation. Since the control system designed in the last section must be gain-scheduled, a fixed, and preferably physically motivated, structure must be found. This structure must incorporate the minimum number of parameters.[10] Moreover, the chosen approach must minimize the sensitivity of the control system to parameter errors which may arise due to the gain scheduling. This can be achieved by choosing a structure that utilizes as many parallel elements as possible. For the "one-bubble" solution, the following form is recommended

$$C(s) = c_1 + c_2 \cdot \frac{1}{s} + c_3 \cdot \frac{1}{s^2} + c_4 \cdot \frac{s \cdot (c_7 \cdot s + 1)}{c_5^2 \cdot s^2 + 2 \cdot c_6 \cdot c_5 \cdot s + 1} \tag{4.26}$$

[10] For a "one-bubble" solution, the number will be seven. A general fourth-order controller has nine parameters, but since in this case two poles are forced to zero, two parameters are structurally fixed.

The coefficients c_1 to c_7 are functions of the time delay, of the time constant, or of the operating point defined by engine speed and load. For each operating point, these coefficients are obtained from the result $C(s)$ of the original control system design by comparing powers of the independent variable s.

Outer Control Loop

The above control system design does not consider the drift of the sensor in front of the catalyst and the dynamics of the TWC. In fact, this device has its own dynamic effects, the most important being oxygen storage. In this section the previously designed control system is extended to compensate for sensor drifts and to consider these dynamic effects as well. This will be achieved by using the information collected by an additional air/fuel ratio sensor downstream of the TWC.

For a realistic physical model of the storage behavior, the varying mass flow through the TWC must be taken into account. This is done by multiplying the air/fuel ratio deviation $\Delta\lambda_{uC}$ with the air mass flow $\dot{m}_{O_2,uC}$ before it is integrated. The output of the integrator is again divided by the same mass flow, yielding the structure shown in Fig. 4.23. Using this approach, a correct oxygen balance is obtained without changing the input-output behavior at a fixed operating point. An alternative but equivalent formulation of this component is possible using the concept of *relative oxygen level* introduced in (2.193).

The limits 0 and C_{O_2C} of the oxygen storage capacity of the TWC must be modeled as well. To estimate C_{O_2C}, dedicated experiments similar to those whose results are shown in Fig. 2.65 must be carried out. Since the storage capacity decreases with time, an on-line identification, as proposed in [154], is necessary.

With ageing of the upstream wide-range air/fuel ratio sensor, a slowly increasing offset of its output signal is observed. This offset and the errors in the observed amount of oxygen must be corrected with an external control loop that utilizes the downstream air/fuel ratio sensor as the main input signal. Both wide-range and switch-type sensors are used for that purpose. Compared to the wide-range sensor, the switch-type sensor, in addition to being relatively cheap, has the advantage of higher precision in detecting stoichiometric air/fuel ratios.

In both approaches, the outer control loop is chosen to be a relatively "slow" PI controller. The signal from the downstream air/fuel ratio sensor is used to infer the actual oxygen storage level $m_{O_2,C}$. This estimation problem is not easy and its solution is the topic of ongoing research. The main difficulty is that as long as the TWC does not reach its storage capacity limits, the downstream air/fuel ratio sensor will not observe any substantial changes in the air/fuel ratio in the exhaust gases. Using a switch-type sensor, whose high sensitivity around stoichiometric air/fuel ratios is welcome in this context, an *approximation* of the oxygen content is obtained as follows

$$m_{O_2,C} \approx \hat{m}_{O_2,C} \cdot \begin{cases} 1 & \text{if } U_\lambda \leq 0.2\ V \\ 1 - (U_\lambda - 0.2)/0.5 & \text{if } 0.2\ V < U_\lambda \leq 0.7\ V \\ 0 & \text{if } U_\lambda > 0.7\ V \end{cases} \tag{4.27}$$

The difference between this measured oxygen level m_{O_2C} and the observed level $\hat{m}_{O_2,C}$ is the input to the a PI controller that produces a correction signal for the oxygen storage observer. This somewhat indirect approach is chosen because of, as mentioned, uncertainty with respect to the true oxygen storage dynamics.

The estimated value $\hat{m}_{O_2,C}$ of oxygen stored in the TWC is compared to the corresponding desired[11] value $\bar{m}_{O_2,C}$, and the resulting difference is used as input signal to the "buffer control" component. The resulting control system structure is shown in Fig. 4.23. Note that the gain c_3, which depends on the engine's operating point, must o be placed in front of the second (rightmost) integrator to avoid an excitation of the controller because of varying operating points. If no controller action is required, the input to the integrator will be zero. Therefore, varying c_3 will not affect the output of the controller if it is placed before the integrator.

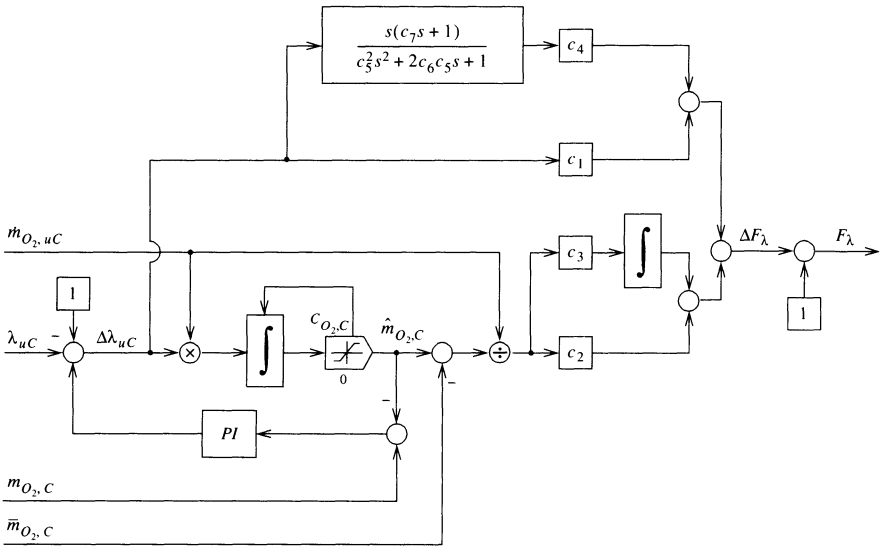


Fig. 4.23. Block diagram of the complete air/fuel ratio control system.

[11] A value of $\bar{m}_{O_2,C} = 0.6\ V$ is recommended.

Experimental Results

To demonstrate the benefits of the proposed control system, especially of its capability to automatically adapt the bandwidth of the closed loop to changing operating points, simulations and measurements at two operating points are shown below. The precise definition of the two analyzed cases is given in Table 4.1). The first operating point (OP 1) was also used for the simulations shown above. Note that in OP 1, the time delay δ corresponds to almost 14 segments. This large delay limits the achievable bandwidth of the feedback control system substantially.

Table 4.1. Operating points chosen in the experiments.

Operating point	OP 1	OP 2
Engine speed n [rpm]	1500	1500
Mass air in cylinder $m_{air,cyl.}$ [g]	0.1	0.3
Engine torque T_e [Nm]	18	80
Time delay δ [s]	0.27	0.116
Time constant τ [s]	0.192	0.146
Segment time τ_{seg} [s]	0.02	0.02
Cycle time τ_{cycle} [s]	0.08	0.08

Figure 4.24 shows the response of the upstream lambda buffer control system for a disturbance[12] step at d_F as defined in Fig. 4.17.

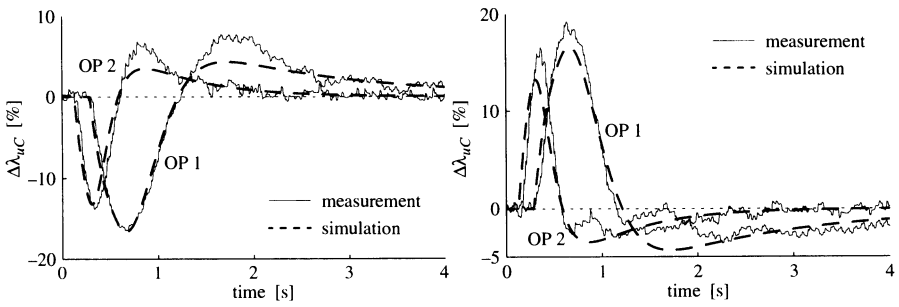


Fig. 4.24. Response of the upstream lambda buffer control to a -20% step (left) and a +20% step (right) in the disturbance d_F.

Because of the parametrization of the control system parameters proposed above, the two system responses have similar properties. However, because the

[12] The signal d_F is the more realistic choice as a disturbance source than the signal d. The latter represents the worst case situation because it acts immediately on the controller and is preferable for the design of the control system.

time delays in OP 1 and OP 2 are different, the bandwidths of the control systems vary accordingly.[13] The maximum deviation of the air/fuel ratio in OP 1 is slightly larger than in OP 2 because of the larger time delay. Also, the maximum value of the *integrated* λ error is larger in OP 1 than in OP 2. Nevertheless, because of the substantially smaller air mass flow, the error in stored oxygen is not as large.

The variations of the oxygen storage level in the TWC are used to assess the performance of the "buffer control system" to adapt to changing operating conditions. Since the oxygen storage level cannot be directly measured, the output of the switch-type air/fuel ratio sensor downstream of the TWC is taken as a measure of that quantity. It is reasonable to assume that if the oxygen storage level of the TWC remains constant, the composition of the exhaust gas remains near constant as well. Accordingly, the output voltage of the downstream sensor remains constant, too.

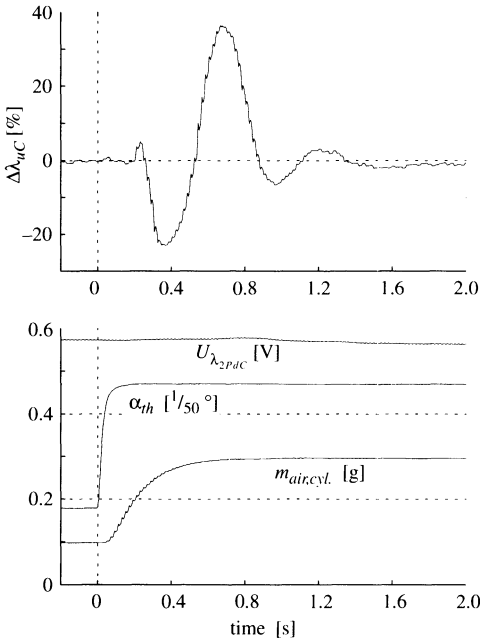


Fig. 4.25. Response of the complete control system to a step in the throttle command corresponding to a change from OP 1 to OP 2. The variable $U_{\lambda_{2PdC}}$ denotes the voltage of the switch-type air/fuel ratio sensor downstream of the catalytic converter.

[13] Notice that a scaling of the time t, according to the ratio of the delays, produces almost congruent curves.

The transient test used to assess the performance of the proposed control approach consists in rapidly (approximately 100 ms) opening the throttle plate from 9° to 24°. In steady state, the mass of air entering the cylinder in each cycle is increased by that change by a factor of three. The delayed increase in air mass visible in Fig. 4.25 is a consequence of the manifold dynamics. Moreover, to amplify the errors in the air/fuel ratio, the feedforward control action for the fuel injection has been reduced in these experiments.

The results of this test are shown in Fig. 4.25. In spite of the large increase in air mass and in spite of the reduced action of the feedforward control system, the downstream sensor output remains almost unchanged. This can be taken as an indication for a constant oxygen storage in the TWC in the steady-state conditions.

Multivariable Control of Air/Fuel Ratio and Engine Speed

The time delays in the various signal paths are largest when the engine is idling. Therefore, instead of having two parallel control loops that can disturb each other, in this situation it is advantageous to combine the air/fuel ratio and the idle-speed control problems into one single MIMO control system. The classical SISO idle speed control problem is analyzed in more detail in Appendix B. In the case studied here, [145], the correct handling of the multivariable nature of the combined control problem is emphasized.

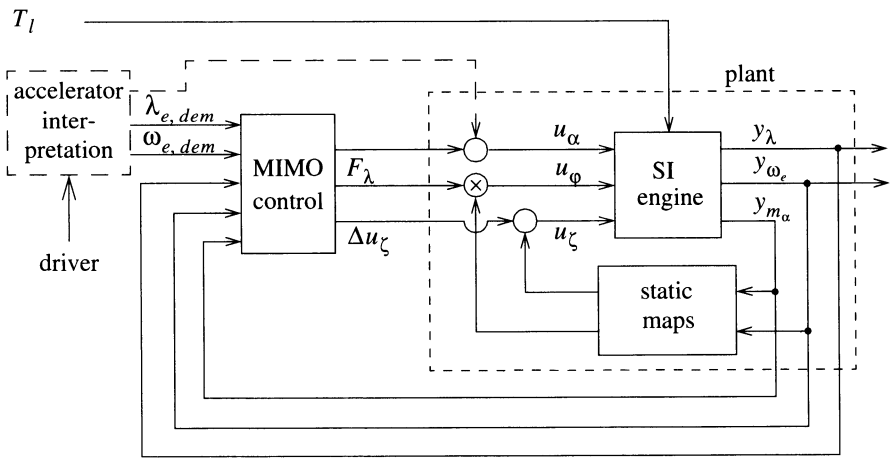


Fig. 4.26. Block diagram of the combined air/fuel ratio and idle speed control system.

The block diagram of the control system that is analyzed in this section is shown in Fig. 4.26. The reference values for the air/fuel ratio, as well as for the engine speed, are obtained from the accelerator pedal signal by an

appropriate signal interpretation module. The sensor signals that are used by the control system are the measured air/fuel ratio, the engine speed and the air mass flow into the intake manifold. Although there is no reference value for the air mass flow signal, including it into the loop improves the quality of the observer that is a part of the MIMO control system. Because of large time delays, the spark advance u_ζ must be used in addition to the throttle and the fuel command, as a third control channel. A basic control system that takes care of the steady-state control actions for ignition and fuel is implemented in a first step (block "static maps" in Fig. 4.26). This step, which optimizes the fuel consumption and minimizes the engine-out pollutant emissions, is not discussed here.

A naturally-aspirated and port-injected SI engine without external EGR is considered. The engine, from which the experimental data shown below has been obtained, has four cylinders and a displaced volume of 1.8-liter. It is mounted on a dynamic test bench that can apply load steps on the engine with a bandwidth of approximately 5 Hz. The focus of the following analysis is on the dynamic behavior of the engine system. Changes in the load torque T_l are assumed to be the main disturbances. A simple model of the corresponding engine dynamics, which will be used for the subsequent design of the control system, is shown in Fig. 4.27. Compared to the block diagram shown in Fig. 2.5, it explicitly includes the four dominant time delays and the static control maps mentioned above.

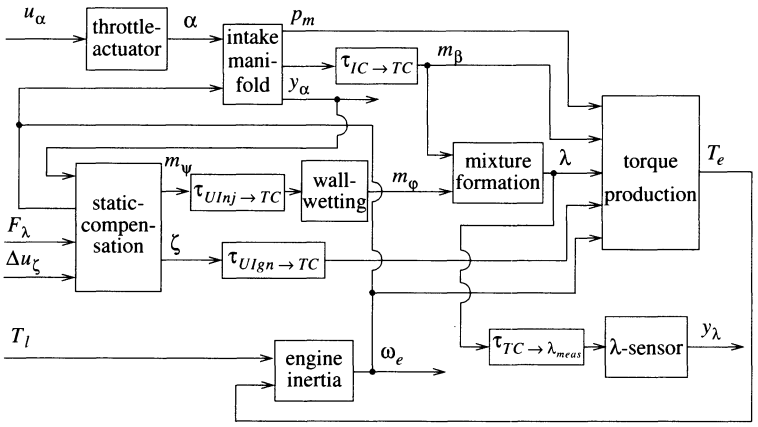


Fig. 4.27. Block diagram of the "plant" component of Fig. 4.26.

The static engine control algorithms run on a separate process controller. Only static maps for the fueling command and the spark advance are applied. The fueling command relies on the measurements of the air mass flow into the intake manifold. Thus, the controller to be designed also must take into account errors made by this feedforward component of the fuel control system.

In steady-state operation, the throttle valve correction signal must compensate all load disturbances; whereas the fueling correction signal eliminates errors in the air/fuel ratio. The spark advance correction may only be used in transients and must go back to zero under steady-state operating conditions.

For small variations of the input signal, the dynamics of the electronic throttle valve are assumed to be well approximated by a first-order lag element. For large steps, the current saturation of the DC motor becomes active, making the air flow actuator strongly nonlinear. The wall-wetting dynamics are assumed to be of first order as described in Sect. 2.4.2. The air/fuel ratio is measured with a wide range sensor. The sensor voltage is linearized and the sensor dynamics are approximated by a first order lag element with a time constant of 20 ms. The remaining two integrators describe the emptying and filling dynamic of the intake manifold and the rotational inertia of the engine. Figure 4.28 shows the block diagrm of the resulting linear engine model.

The control system to be designed must fulfill the following requirements:

- No steady-state error in engine speed and air/fuel ratio are permitted for all load disturbances in the range $[0, 20]$ Nm.
- The controller must be robust to modeling errors, i.e., the minimum return difference of the open loop may not be smaller than 0.5.
- The spark advance control channel may only be used during the transients.
- The controller must be operated in the crank-angle domain and discretized on a segment basis.

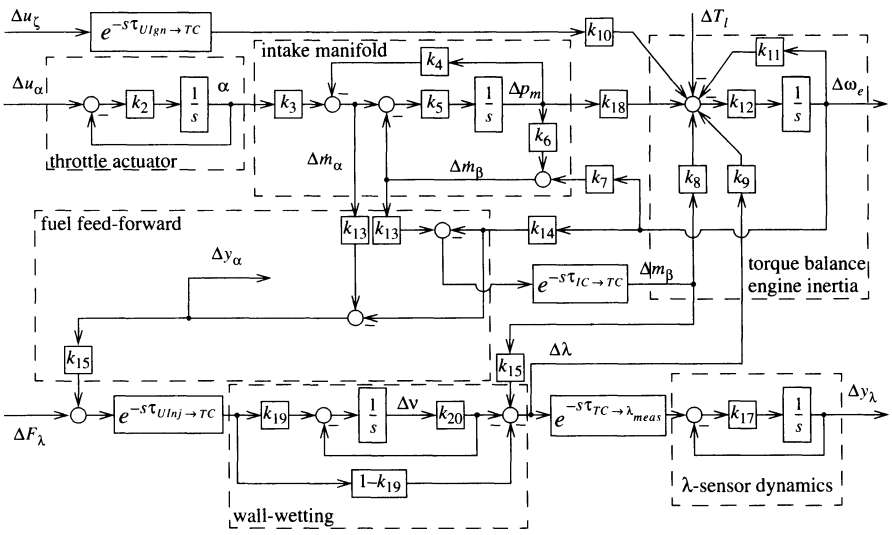


Fig. 4.28. Detailed block diagram of the linear engine model (including the delays).

Table 4.2 gives the explicit numerical values of all constants used in Fig. 4.28. These parameters are valid for a four-cylinder 1.8-liter engine at idle conditions.

Table 4.2. Parameters of the linearized engine model of Fig. 4.28.

Parameter	value	unit
k_2	40	rad/s
k_3	43.6	g/min/ °
k_4	77.5	g/min/bar
k_5	0.00386	bar·min/s/g
k_6	686	g/min/bar
k_7	0.422	g/rev
k_8	114	Nm·rev/g
k_9	-16.3	Nm
k_{10}	0.0984	Nm/ ° ca
k_{11}	0.00358	Nm/rpm
k_{12}	63.7	rpm/Nm/s
k_{13}	0.00116	1/rpm
k_{14}	0.000402	g/min/rpm/rpm
k_{15}	2.90	rev/g
k_{17}	5.7	rad/s
k_{18}	7.76	Nm/bar
k_{19}	0.3	-
k_{20}	15	rad/s
$\tau_{Uinj \to TC}$	0.14	s
$\tau_{Uign \to TC}$	0.07	s
$\tau_{IC \to TC}$	0.07	s
$\tau_{TC \to \lambda_{meas}}$	0.385	s

The control system is designed using a model-based approach by applying the H_∞ frequency-domain optimization method. The block diagram of the linearized engine model used in that synthesis procedure is shown in Fig. 4.28. The resulting observer-based controller has 13 state variables. Five state variables are necessary to observe the relevant level variables in the engine system. Two additional state variables are used to force the steady-state errors of the air/fuel ratio λ and of the engine speed ω_e to zero. The remaining six state variables are used to approximate several delays present in the system. Using this rather large number of state variables allows to establish the achievable performance limits. Of course, in a real application an order reduction step would have to follow in order to obtain a realizable control system.

A comparison between simulation and measurement is shown in Fig. 4.29 and Fig. 4.30. Figure 4.29 shows a load step. The offset in spark advance is limited to +5 crank angle. An appropriate anti-windup scheme is used.

The tracking behavior is depicted in Fig. 4.30. For steps in the demanded engine speed, the throttle must be actuated; therefore, the air/fuel path is

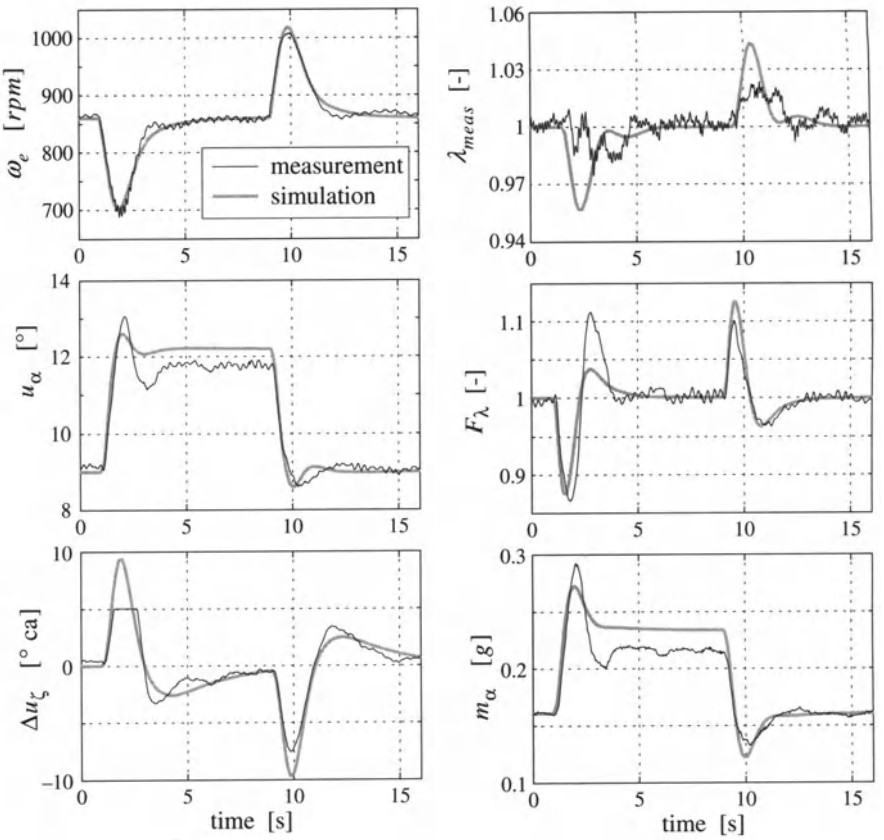


Fig. 4.29. Control system behavior, load step ± 20Nm at 1 s resp. 9s.

excited as well. Deviations of up to 7% are observed, nevertheless they are reduced within one second to values below 3%. On the other hand, excitation of the air/fuel ratio demanded result in engine speed deviations of only 20rpm, which are hardly recognizable.

Finally, in Fig. 4.31, the performance of the MIMO controller designed above is compared to the performance of the series-production type controller. This controller uses two parallel SISO control loops for the air/fuel ratio and the engine speed, respectively. Since the speed set point of the series-production type controller is 770 rpm, this value was chosen in all experiments. Moreover, since the series-production type controller was not able to handle a 20 Nm disturbance, the load step was reduced to 10 Nm in these particular experiments.

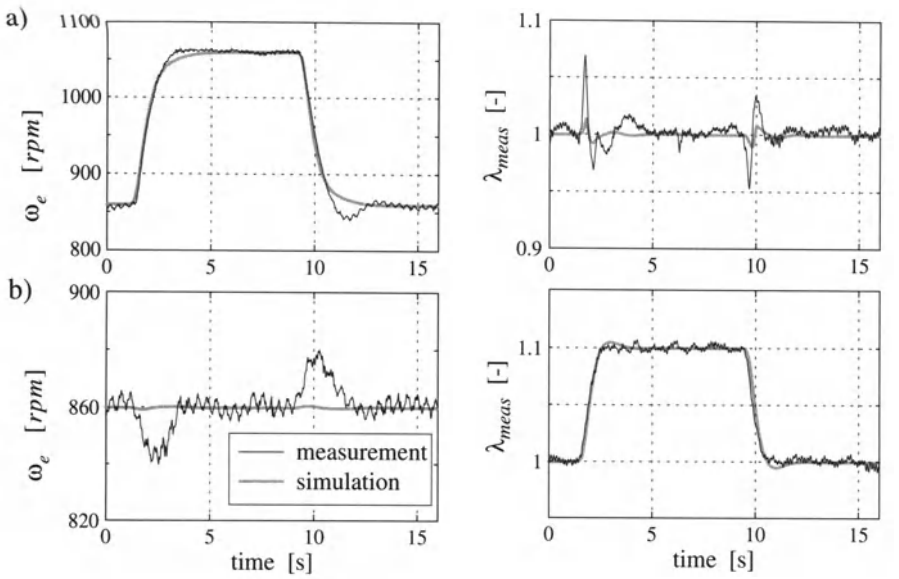


Fig. 4.30. Control system behavior, steps of the demanded values at 1 s resp. 9 s: a) engine speed steps (±200rpm), b) λ-steps (±10%).

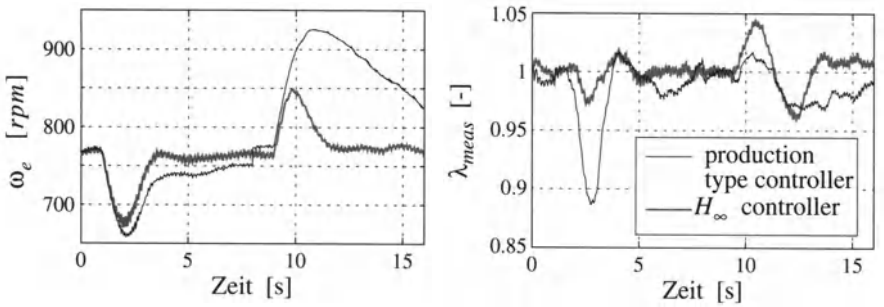


Fig. 4.31. Measured control system performance, load steps at idle speed: $\pm 10 Nm$ at 1s resp. 9s.

4.3 Control of an SCR System

The purpose of a selective catalytic reduction system was discussed in Sect. 2.8.3. The model developed in that section forms the basis for the control system synthesis discussed below. In the case studied here, [147], the tailpipe NO_x emissions of a heavy-duty Diesel engine (that was certified for the Euro II limits) have been reduced to meet the Euro V emission limits. To achieve this, a minimal NO_x reduction rate of 50% is necessary. Since NH_3 tailpipe emissions are highly undesired, a maximum mean slip of only 10 ppm of ammonia is tolerated.

The engine that is considered in this section is a 10-liter supercharged heavy-duty, direct-injection, Diesel engine with intercooler. Its rated power is 275kW. The maximum engine speed is 2150 rpm.

The amount of NO_x in the exhaust gas and the amount of injected ammonia vary strongly as functions of the engine operating point. Changing environmental conditions and the ageing components of the engine system play an important role as well. Dependency upon the engine's operating point can be considered by implementing corresponding maps for the engine-out emissions and adding a low-pass filter to account for the thermal dynamics. All other effects (plugging of the injection system, change in ambient conditions, aging) and the unavoidable modeling errors render a feedback control system necessary. Such an approach is only possible if information is available regarding the NO_x concentration in the exhaust upstream and downstream of the catalytic converter. The available solid-state NO_x sensors (Siemens-NGK "Smart NOx Sensor," SNS) show significant cross sensitivities to ammonia. Therefore, close to the optimal dosage, where it is not clear weather excess NO_x or NH_3 is present, the sensor signal cannot be interpreted in a straightforward way.

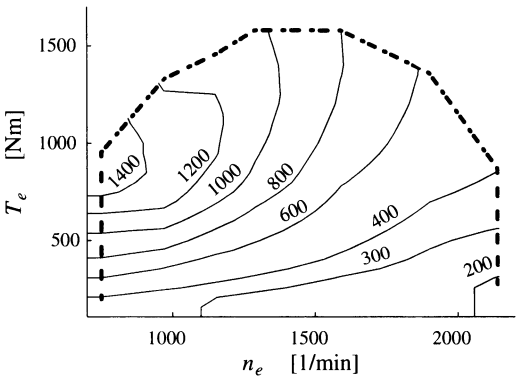


Fig. 4.32. Engine out NO_x emissions in ppm of a 10 liter, heavy duty, Diesel engine.

Figure 4.32 shows the map that is used for the calculation of the steady state engine out NO_x emissions. The output of the steady-state map is filtered with a lag-lead element with a time constant of $20s$ and a feed-through of 0.75. Figure 4.33 shows the measured and estimated engine out NO_x emissions for part of an ESC test.[14]

Figure 4.34 shows the relevant variables for a sweep in ammonia dosage. The cross-sensitivity to the ammonia concentration for higher-than-stoichiometric ammonia dosages is obvious. The "true" NO concentration is measured with

[14] The ESC test cycle is the equivalent of an MVEG-95 cycle for heavy-duty, Diesel engines.

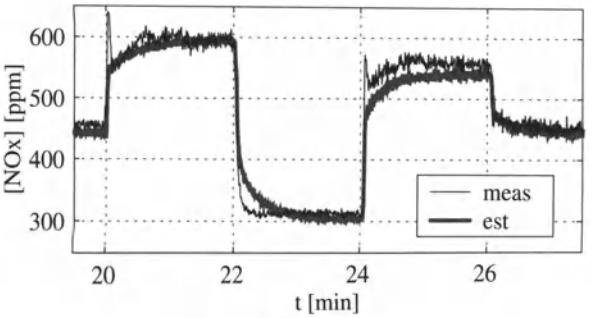


Fig. 4.33. Measured and estimated engine out NO_x emissions for part of an ESC test. Starting at 400 Nm and 1590 rpm, a step to 1365 Nm and 1900 rpm is performed. Then a step to 350 Nm, followed by a step to 1020 Nm, and finally a step to 695 Nm, all at 1900 rpm, are shown.

a standard emission testing device, whereas the ammonia concentration is measured with a dedicated instrument.

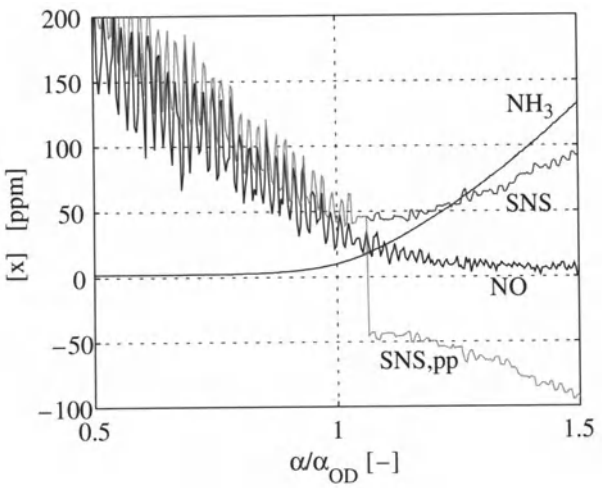


Fig. 4.34. Output signals of a solid state NO_x sensor (SNS), true NO signal (NO), and NH_3 concentration in the exhaust gas for a sweep in ammonia dosage. The post-processed sensor signal (SNS,pp) is also shown.

Due to ammonia storage effects in the catalytic converter, the SNS sensor signal is substantially less noisy when α/α_{OD}[15] is larger than one. Using this

[15] α_{OD} is the optimal dosage ammonia feed ratio, defined as the ratio $\frac{\dot{n}_{NH_3,u}}{\dot{n}_{NO_x,u}}$ where a ammonia slip of 10ppm is observed.

property, an additional periodic modulation of the dosage signal can be used to to separate the $\alpha/\alpha_{OD} < 1$ portion from the $\alpha/\alpha_{OD} > 1$ part. In fact, the amplitude reduction caused by the catalytic converter is used to decide in which regime the system is working and, thus, to calculate the output indicated by SNS,pp that represents an estimation of the NO_x concentration. Notice that this value can be negative in the case of excess ammonia.

This interpretation of the NO_x sensor signal allows for the implementation of a true SCR feedback control system. Of course, this component will have to be complemented by an appropriate feedforward control scheme in order to achieve the best possible bandwidths. Figure 4.35 shows the block diagram of the combined control scheme.

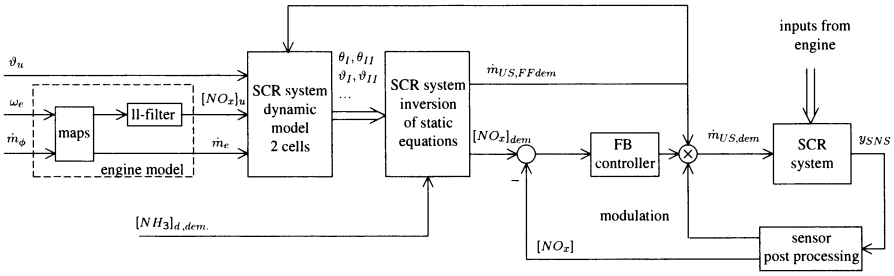


Fig. 4.35. Block diagram of the combined feedforward/feedback SCR control system.

The main elements of the proposed SCR control system are:

- Engine model: The signals available in the standard ECU (engine speed and load, temperature and pressure in the intercooler, temperature upstream and downstream of the catalytic converter, etc.) are used to calculate an estimate of the engine-out NO_x concentration and the engine out mass flow. This can be done, as indicated, by using appropriate maps. The steady-state NO_x engine out emissions are then filtered with a first order lag element with a time constant of 20 s, in order to account for the thermal dynamics of the system.
- SCR system dynamic model: The output signals of the engine model and the measured exhaust gas temperature upstream of the converter system are the inputs to an observer for the SCR system. The observer uses the simplified two cell model derived in Sect. 2.8.3. All states of the model are calculated, taking into account the actual injected amount of urea solution.
- Feedforward control: By inverting (2.216) and (2.217), the demanded value for the amount of injected urea solution, and the SCR system-out (tailpipe), NO_x emissions can be calculated. In this step, the demanded value for the tailpipe ammonia slip is chosen to be 10 ppm.
- Feedback control: The demanded amount of injected urea solution is multiplied by the output of the feedback controller. The control error is the

difference between the calculated amount of NO_x tailpipe emissions and the interpreted sensor output. The feedback element is a simple PI controller that is designed using classical frequency-domain methods, as explained in Appendix A. This PI control system may be chosen to be rather "slow" because it has only to cope with the disturbances originating from the environment and from ageing components.

- Signal processing: In order to separate the NO_x information from the NH_3 information in the sensor output signal, the product of the outputs of the feedback and feedforward control systems is modulated with a periodic signal with an amplitude of approximately 10%. This filtered output forms the reference signal for the urea dosage unit. A local dosage control system guarantees a good agreement of the injected mass flow of urea with this reference value.

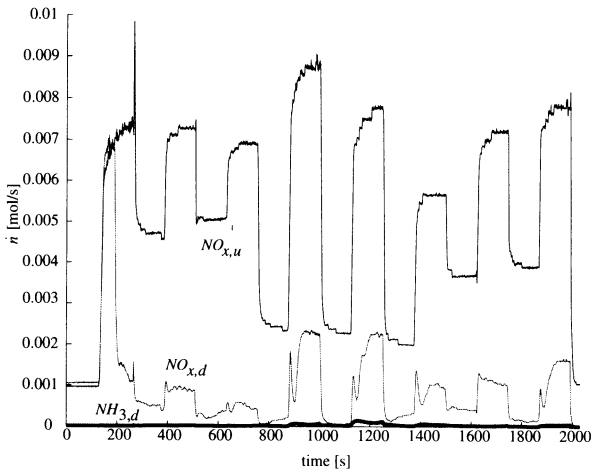


Fig. 4.36. Performance of the SCR system for an ESC test, 10-liter, supercharged, heavy-duty engine. $NO_{x,u}$ is the upstream NO_x molar flow, $NO_{x,d}$ is the downstream NO_x molar flow, whereas $NH_{3,d}$ represents the downstream ammonia molar flow.

Figure 4.36 shows the performance of the SCR control system for an ESC test cycle. The experimental verification of this control strategy showed an average NO_x reduction of almost 90%, while the ammonia slip could be kept at an average value of 10 ppm. Note that some ammonia slip is present even if the exact stoichiometric urea dosage is applied all of the time. Therefore, lower ammonia slips can be achieved only by accepting a deterioration of the NO_x reduction efficiency. Moreover, the measured value of 10 ppm lies below the limit which is considered to be harmful and is below the human perception limit as well.

A further improvement of the performance of the control system can be achieved by accepting ammonia slips that are proportional to the tailpipe NO_x emissions. As before, equations (2.216) and (2.217) can be used for the calculation of the demanded amount of urea solution and the estimated tailpipe NO_x emissions. The resulting control system structure remains essentially the same, but when necessery, higher ammonia slips are accepted for the sake of an improved NO_x reduction.

4.4 Engine Thermomanagement

4.4.1 Introduction

In automotive applications, approximately one third of the fuel energy delivered to the engine must be removed by the cooling system [81]. In high-speed and high-load operating points this amounts to significant heat fluxes. Specifically, at high ambient temperatures and low vehicle velocities, large radiators, additional fans and large thermostats must be used to guarantee that the cooling system works properly and that the engine does not overheat. Accordingly, when the engine runs in part-load conditions, which is the case most of its lifetime, the cooling effect is too large. This leads to lower fuel economy due to lower engine temperatures and therefore higher friction losses. In addition, considerable losses are caused by the mechanically driven water pump and by the inefficient thermostat used to control the coolant temperature.

While the drawbacks of a mechanically coupled water pump are obvious, the drawbacks of the conventional thermostat must be explained in more detail. As shown in Fig. 2.40, the thermostat is a valve whose position determines the flow distribution between the radiator and its bypass duct. Under stationary conditions, the position of the thermostat is defined by the coolant temperature. The active element of the thermostat (a wax element) has a dynamics that cannot be neglected. The time constant of this element amounts to approximately 30 s. For cold starts at high loads, the engine temperature must be kept from overshooting to avoid damage to the cylinder head gaskets and other components. Therefore, the thermostat's static characteristic must allow a certain amount of coolant to flow through the radiator, already at relatively low temperatures (70°C). On the other hand, at 90°C, the thermostat must be fully open to enable maximum cooling. These requirements lead to a nonlinear static gain function as shown in Fig. 4.37. The thermostat acts as a proportional controller. Since the plant itself does not have any integrating behavior, this simple control system cannot avoid errors in steady-state conditions. In particular, at higher engine speeds and at lower loads, this temperature deviation from the demanded value leads to higher friction losses in the engine and, thus, higher fuel consumption.

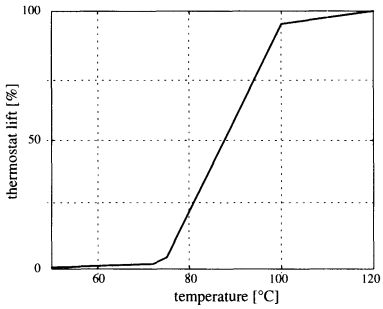


Fig. 4.37. Lift characteristic of a conventional thermostatic valve.

4.4.2 Control Problem Formulation

Modern engine cooling systems include electrically driven water pumps and electronically controllable bypass valves. The higher cost of these components must be compensated for by improvements in the overall fuel efficiency of the system. This requires improved control strategies which are able to operate the engine at part load at higher temperatures, yet still guarantee proper system cooling at high loads. A detailed mathematical model of such a cooling system was developed in Sect. 2.6.5. Here, the design of the corresponding control system is discussed.

The new cooling system offers two independent control inputs. The additional degree of freedom can be used to simultaneously control the temperatures of the coolant entering and leaving the engine. Similar to the conventional approach, the temperature ϑ_{eo} of the coolant leaving the engine will be controlled to a demanded value of $\vartheta_{eod} \approx 90\,^{\circ}\mathrm{C}$. The temperature ϑ_{ei} of the coolant entering the engine will be controlled using the actual value of ϑ_{eo}. In order to avoid detrimental mechanical stresses, the temperature difference across the engine may not be larger than approximately $15\,^{\circ}\mathrm{C}$.

Notice that in this situation the system to be controlled has a multiple-input, multiple-output (MIMO) structure. Compared to a SISO problem, the control of such MIMO systems is, in general, substantially more difficult. Therefore, whenever possible, the problem must be decoupled into several SISO loops using the knowledge of the physical properties of the plant. In the case discussed here, such a simplification is possible by identifying four different cases in which four different control strategies can be applied. Table 4.3 summarizes these four cases that are treated separately below.

Case 1 applies in the situation in which the engine is cold. The temperatures ϑ_{ei} of the coolant entering and ϑ_{eo} of the coolant leaving are lower than their demanded values $\vartheta_{...d}$.[16] Therefore, the bypass valve remains closed, and

[16] Notice that the reference values $\vartheta_{...d}$ can be functions of the engine operating point. For instance, higher reference temperatures may chosen at low loads such

Table 4.3. Summary of the decoupling strategy used in the engine-cooling control system.

	$y_2 = \vartheta_{ei}$	$y_1 = \vartheta_{eo}$	$u_1 = \dot{m}_{by}/\dot{m}_c$	$u_2 = \dot{m}_c$
case 1	$< \vartheta_{eod}$	$< \vartheta_{eod}$	100%	minimal
case 2	$> \vartheta_{eid}$	$> \vartheta_{eod}$	$\Delta\vartheta_{eo} \cdot G_{c1}(s)$	minimal
case 3	$< \vartheta_{eid}$	$> \vartheta_{eod}$	$\Delta\vartheta_{eo} \cdot G_{c1}(s)$	$\Delta\vartheta_{ei} \cdot G_{c2}(s)$
case 4	$> \vartheta_{eid}$	$> \vartheta_{eod}$	0%	$\Delta\vartheta_{eo} \cdot G_{c3}(s)$

the coolant flow through the water pump is at its minimum value.[17] In this situation, the temperatures entering and leaving the engine will rise almost simultaneously.

As soon as ϑ_{eo} reaches its demanded value (case 2), the bypass valve must start opening, otherwise the engine might overheat. The temperature difference over the engine starts to rise, until, with the coolant flow through the engine still at its minimum, the heat cannot be removed any more without violating the maximum temperature difference condition.

Therefore, the coolant mass flow must now be increased (case 3). Only in this situation is a true MIMO control problem encountered. Fortunately, the cross coupling between the two control channels is not strong, so the two SISO control loops yield satisfactory results. In the first loop, the coolant mass flow defined by the water pump speed is used to control the temperature difference across the engine. In the second loop, the bypass valve position is used to control the engine out temperature.

As soon as the bypass valve reaches its fully open position (case 4), the engine out temperature is controlled by changing only the coolant mass flow.

The design of the three SISO controllers $G_{c...}(s)$ can be accomplished using any of the classical methods presented in Appendix A. A PI structure offers sufficient degrees of freedom in most cases. The plant must be linearized around typical operating points (cold start, engine fully warmed up, full load, etc.) and the resulting PI controllers are parameterized using several possible operating modes. Figure 4.38 shows, as an example, the resulting open-loop frequency response for case 2. Because of the strongly varying plant parameters, it is important to achieve sufficient robustness. The loop gain depicted in Fig. 4.38 shows that, in this example, this objective has been achieved.

---

that fuel consumption is reduced while at full load. In order to prevent knocking, lower temperature set points must be used.

[17] The coolant flow may not be completely shut off. This could cause damage at local "hot spots." Moreover, a minimum coolant flow is necessary to guarantee the system to be observable.

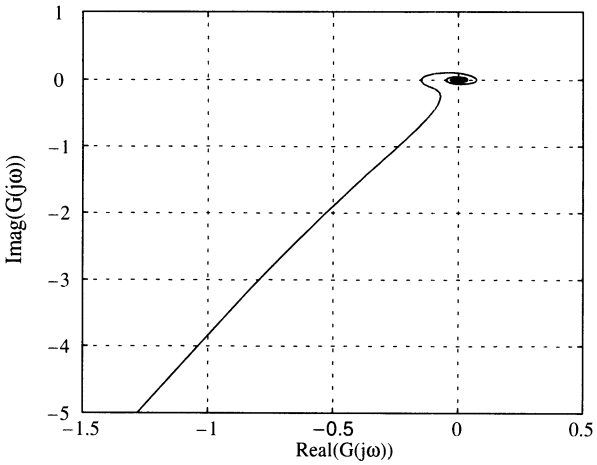


Fig. 4.38. Nyquist diagram of the open-loop transfer function with the plant channel from the input u_1 to the output ϑ_{eo} and a PI controller.

4.4.3 Feedforward Control System

Due to the large delays present in this system, the compromise between short system response times and robustness of the feedback loop cannot be satisfied by a feedback-only control system. In fact, large overshoots of the engine temperature in cold start transients and oscillations at steady-state can be observed only if a feedback control system is used.

A feedforward control system can help to improve the system response in both problem areas. The model of the external cooling circuit that has been derived in Sect. 2.6.5 forms the basis of the synthesis of this feedforward control system. The starting point is the equation

$$\vartheta_{ei}(t) = \vartheta_{mix}(t - \tau_{4\to 6}) \tag{4.28}$$

$$\vartheta_{mix}(t) = u_1(t) \cdot \vartheta_{eo}(t - \tau_{1\to 2} - \tau_{3\to 4}) + (1 - u_1(t)) \cdot \vartheta_{ro}(t - \tau_{5\to 4}) \tag{4.29}$$

Using (4.29) the control input $u_1(t)$ can be calculated

$$u_1(t) = \frac{\dot{m}_{by}(t)}{\dot{m}_c(t)} = \frac{\vartheta_{mix}(t) - \vartheta_{ro}(t - \tau_{5\to 4})}{\vartheta_{eo}(t - \tau_{1\to 2} - \tau_{3\to 4}) - \vartheta_{ro}(t - \tau_{5\to 4})} \tag{4.30}$$

Inserting the demanded values for the mixing temperature, the engine out temperature and measured values for the radiator temperature, this equation can be used for the design of the feedforward controller. Since the mixing point and the engine-in point of the coolant are usually quite close to each other, the demanded value for the engine-in temperature can be used for the mixing temperature as well. The demanded temperature for the coolant

entering the engine may be calculated from the differential equation of the engine-out temperature (ϑ_{eo}) which has been derived in Sect. 2.6.5

$$\frac{d}{dt}\vartheta_{eo}(t) = \frac{\dot{Q}_{w,c}(t) - \dot{Q}_{c,eb}(t) + c_c \cdot \dot{m}_c \cdot [\vartheta_{ei}(t) - \vartheta_{eo}(t)]}{c_c \cdot m_c} \tag{4.31}$$

and

$$\dot{Q}_{w,c}(t) = \alpha_c \cdot A_c \cdot \left[\vartheta_w(t) - \frac{\vartheta_{eo}(t) + \vartheta_{ei}(t)}{2}\right] \tag{4.32}$$

$$\dot{Q}_{c,eb}(t) = \alpha_{eb} \cdot A_{eb} \cdot \left[\frac{\vartheta_{eo}(t) + \vartheta_{ei}(t)}{2} - \vartheta_{eb}(t)\right] \tag{4.33}$$

Using these equations, the engine-in temperature can be calculated as a function of engine-out temperature, its derivative, engine-block temperature and engine cylinder-wall temperature

$$\vartheta_{ei} = f\left(\vartheta_{eo}, \frac{d\vartheta_{eo}}{dt}, \vartheta_w, \vartheta_{eb}\right) \tag{4.34}$$

The problems associated with the feedback control system arise mainly at cold start, especially in the transition phase from case 1 to case 2. For this situation the temperature of the radiator can be assumed to be constant and equal to the ambient temperature. Therefore, the expression (4.30) can be used to calculate the bypass valve position as a function of engine-in and engine-out temperatures. To calculate the demanded value for the engine-in temperature, ϑ_{eo} in (4.34) is replaced by a *constant* reference value for the engine-out temperature, ϑ_{eod}, with its derivative being equal to zero.

$$\vartheta_{eo} = \vartheta_{eod} \tag{4.35}$$

$$\frac{d}{dt}\vartheta_{eo}(t) = 0 \tag{4.36}$$

Since the temperature ϑ_w of the cylinder wall and ϑ_{eb} of the engine block cannot be measured, they must be observed. A model-based open-loop estimator, whose structure is shown in Fig. 4.39, is used for the estimation of the corresponding values $\hat{\vartheta}_w$ and $\hat{\vartheta}_{eb}$, respectively. Knowing the desired temperature ϑ_{eod} of the coolant leaving the engine, the corresponding coolant temperature at the engine input can be expressed as a function of the known variables

$$\vartheta_{eid} = f\left(\vartheta_{eod}, 0, \hat{\vartheta}_w, \hat{\vartheta}_{eb}\right) \tag{4.37}$$

The disturbance $z_1 = \dot{Q}_{g,w}$ can be calculated, according to Fig. 2.43 and Fig. 2.44, as a function of the operating point (described by ω_e and p_{me}) and of the cylinder wall temperature ϑ_w. The disturbance $z_2 = \dot{Q}_{if} - \dot{Q}_{eb.a}$ is modeled according to Sect. 2.5.1 using the following equation

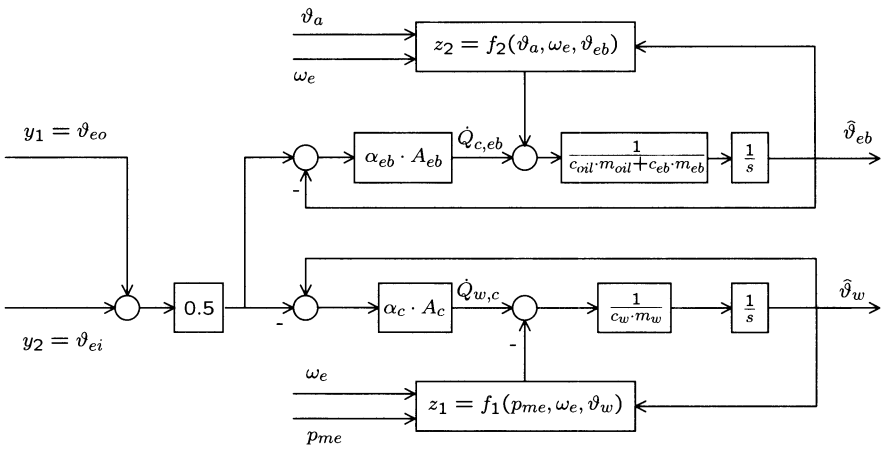


Fig. 4.39. Block diagram of the open-loop observer for engine block and engine wall temperatures.

$$z_2 = \dot{Q}_{if} - \dot{Q}_{eb.a}$$

$$= p_{me0f} \cdot V_d \cdot \frac{\omega_e}{4\pi} - A_{eb} \cdot \alpha_{eb,a} \cdot (\vartheta_{eb} - \vartheta_a) \tag{4.38}$$

$$\tag{4.39}$$

where p_{me0f} is defined in (2.111).

The block diagram of the resulting control system is depicted in Fig. 4.40.

4.4.4 Experimental Results

Figure 4.40 shows the block diagram of the control system for the case that only feedforward action is applied. Note that this system is not a true feedforward controller because it uses measured values of the engine-in and engine-out temperatures to drive the open-loop observer and that this feedforward control system is primarily designed for the warming phase and, therefore, a minimum coolant flow is assumed. This control system was implemented in a rapid-prototyping hardware, and experiments were run on the engine system described in [38]. The simulated and the measured behavior of the closed-loop control system are shown in Fig. 4.41. In the simulation, ϑ_{eo} remains exactly at the demanded value of $\vartheta_{eod} = 90\,^{\circ}\mathrm{C}$. The temperature ϑ_{ei} of the coolant entering the engine rises simultaneously with ϑ_{eod} until $\vartheta_{eod} = 90\,^{\circ}\mathrm{C}$ is reached. Then the temperature ϑ_{ei} of the coolant drops again in order to remove the appropriate amount of heat from the engine and to keep ϑ_{eo} constant.

As shown in the lower plots of Fig. 4.41, neither the engine-in nor the engine-out temperature remain at their desired values in these experiments. This had to be expected because the engine model cannot describe precisely

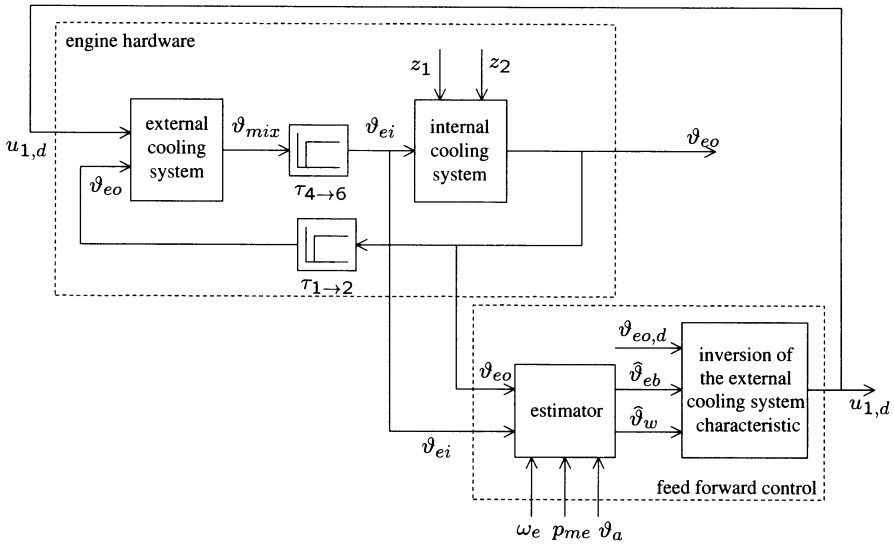


Fig. 4.40. Feedforward control system for the bypass valve channel.

the real system behavior, and because the feedforward control action does not include any error-correcting mechanisms. In addition to the temperature of the coolant entering and leaving the engine, the opening of the bypass valve is shown as well in Fig. 4.41. In the warming phase, the bypass valve is fully open and the coolant flow completely bypasses the radiator. As soon as the engine-out temperature reaches its demanded value, the bypass valve is partially closed so that the excess heat is discarded in the radiator.

The feedforward controller must be complemented with a feedback component in order to account for modeling inaccuracies and unmeasurable disturbances. The block diagram of the combined feedforward and feedback controller is shown in Fig. 4.42. As discussed above, the feedback control systems are realized as PI controllers. The specific input and output structures are defined in Table 4.3. Compared to the feedback-only loop, whose frequency response is shown in Fig. 4.38, the gain of the feedback component in this control system can be substantially reduced because of the beneficial action of the feedforward controller.

The experimental comparison between the feedback-only and the combined feedforward-feedback strategy is shown in Fig. 4.43. The feedback-only control system is able to bring the engine-out temperature to its desired value. However, this approach produces a noticeable overshoot when first reaching the desired temperature and, due to the large time delays present in the plant, a non-vanishing limit cycle behavior in steady state conditions. The combined feedforward and feedback control system avoids both unwanted effects and, moreover, reduces the settling time.

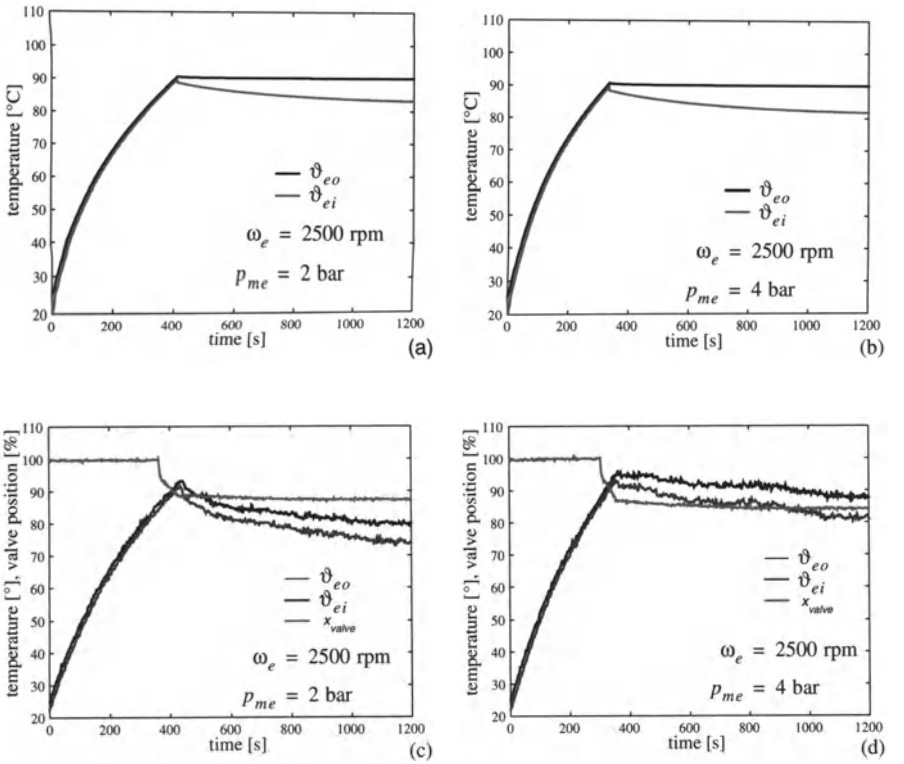


Fig. 4.41. Coolant temperatures at cold start. Curves (a) and (b): results from simulations, $p_{me} = 2bar$ and $p_{me} = 4bar$, respectively. Curves (c) and (d): results from measurements at the same operating points.

Of course the constant operating point experiments are not very informative and, no real benefit in terms of reduced fuel consumption has yet been shown. To assess the benefits of the proposed control system, experiments were conducted that compared the regular cooling system performance with the performance of the proposed novel control system. The fuel consumption benefits achievable for several constant operating points are shown in Fig. 4.44.

Because of the higher engine temperatures, a reduction in fuel consumption of up to 4.5% could be realized in that specific case. The highest relative fuel consumption reduction is achieved for the lowest load because of the largest increase in coolant temperature. Plot b) shows the larger temperature difference over the engine with the new control system. This is a consequence of the small coolant flow used. For higher engine loads, the temperature difference increases. Also note that, compared to the standard case in the advanced engine-temperature control approach, the oil temperature is significantly higher.

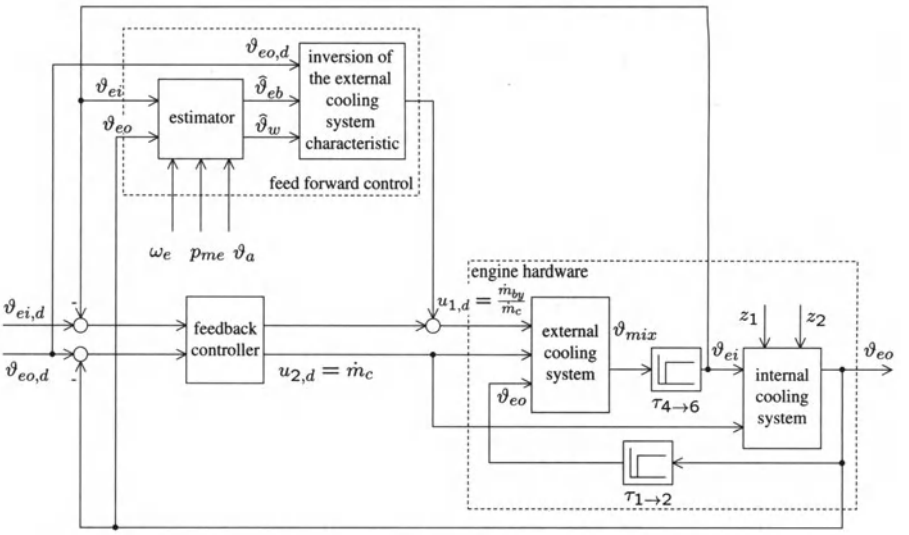


Fig. 4.42. Combined feedforward and feedback control system.

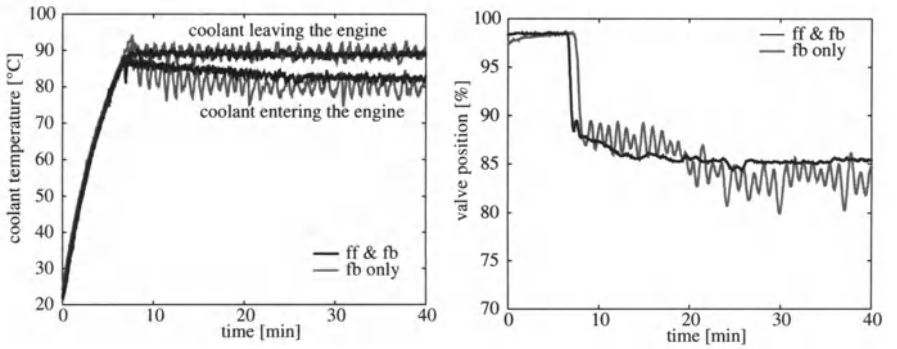


Fig. 4.43. Coolant temperature (left) and valve position (right) for the feedback-only and combined feedforward/feedback control structures.

The fuel-economy benefits realizable in transient conditions were assessed using the European test cycle. The results of these experiments are shown in Fig. 4.45. With respect to the control scheme defined in Table 4.3, it was found that case 3, where the power of the coolant pump must be increased in order to reduce the temperature over the engine, was only used for a short time at the highest acceleration in the last phase of the MVEG-95 test cycle.[18]

Two final remarks remain:

[18] Because it is calculated as the difference of two relatively large signals, the temperature difference over the engine is a very noisy signal. Therefore, an appropriate signal filter is necessary.

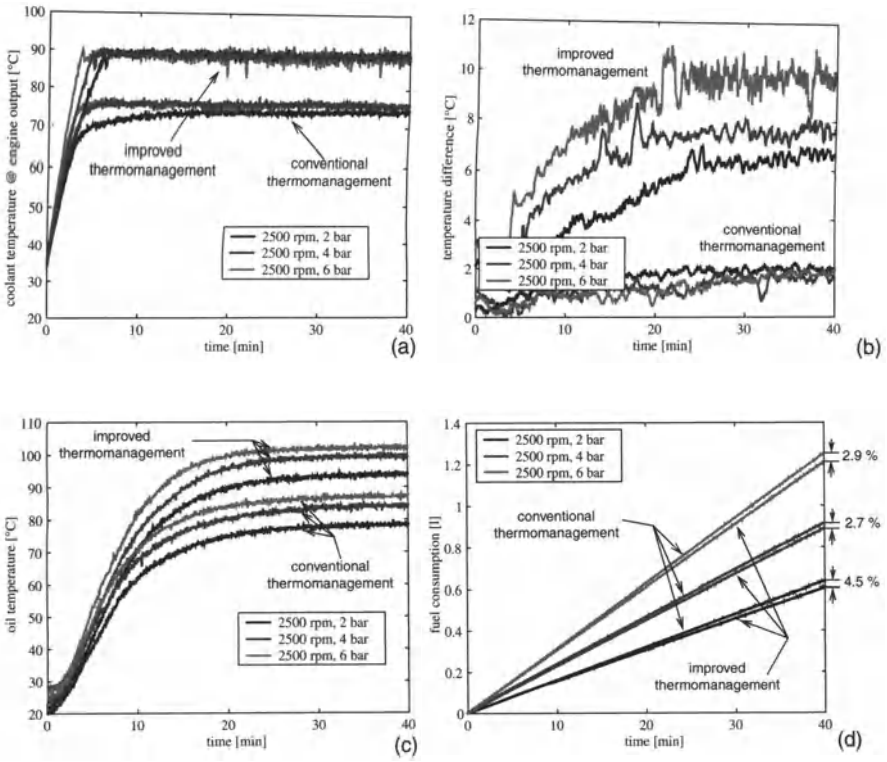


Fig. 4.44. Cold-start measurements for various steady-state engine operating points.

- Because the mechanically driven coolant pump is replaced by one that is electrically driven raises the question, of whether this results in higher energy consumption. Since the electrically driven pump works at approximately one-sixth of the nominal mass flow of the mechanical one, this is not the case.
- The proposed approach not only increases fuel efficiency, but at the same time improves comfort and safety. In fact, quicker engine heating and a higher engine temperature result in a more comfortable cabin climate and faster windshield defrosting in winter.

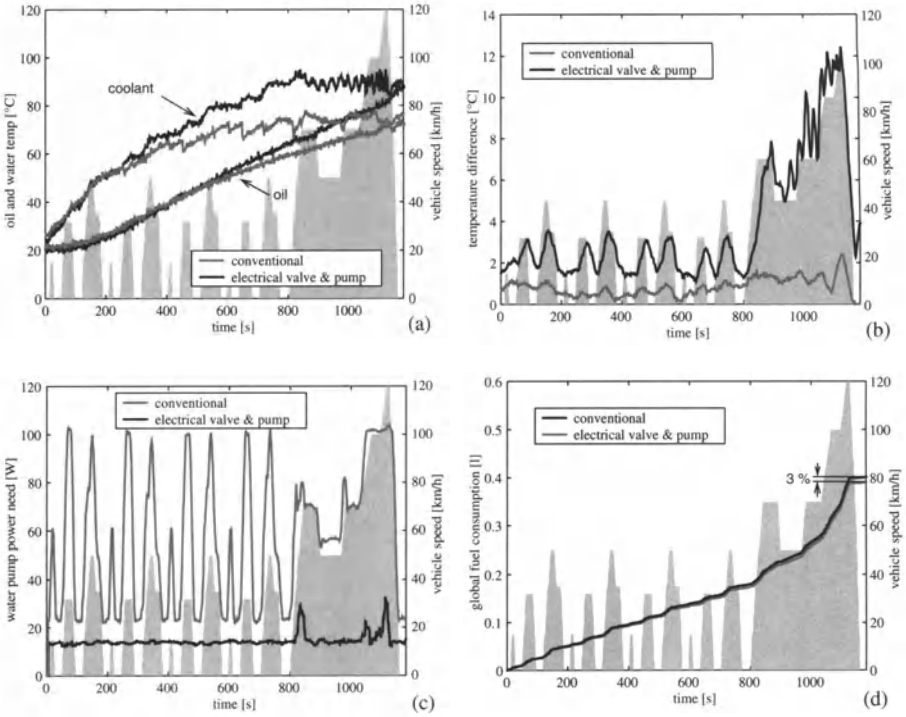


Fig. 4.45. Cold-start measurements during an MVEG-95 test cycle.

A

Appendix A

This appendix collects some of the most important definitions and results of the theory of system modeling and control systems analysis and design. To avoid complex mathematical formulations, all systems are assumed to have only one input and one output.

This appendix is not *intended as a self-contained course on control systems analysis and design (this would require much more space). This appendix is nothing but a convenient way to synchronize notation and to recall some points that are used repeatedly throughout the main text. Readers who have never had the opportunity to study these subjects are referred to the many standard text books available, some of which are listed at the end of this chapter.*

A.1 Modeling of Dynamic Systems

General Remarks

This section recalls briefly the most important ideas in system modeling for control. In general, two classes of models can be identified:

1. models based purely on (many) experiments; and
2. models based on physical first principles and (a few) parameter identification experiments.

In this text, only the second class of models is discussed. General guidelines to formulate such a control-oriented model encompass, at least, the following six steps:

1. Identify the system boundaries (what inputs, what outputs, etc.).
2. Identify the relevant *reservoirs* (mass, energy, information, ...) and the corresponding *level variables*.
3. Formulate the differential equations for all relevant reservoirs as shown in eq. (A.1).

$$\frac{d}{dt}\,(\text{reservoir content}) = \sum \text{inflows} - \sum \text{outflows} \tag{A.1}$$

4. Formulate the (usually nonlinear) algebraic relations that express the *flows* between the reservoirs as functions of the level variables.
5. Identify the unknown system parameters using experiments.
6. Validate the model with experiments other than those used to identify the system parameters.

In step two, it is important to distinguish between slow (curve c) in Fig. A.1), relevant (curve b) and fast (curve a) dynamic effects (very large, average and very small reservoirs). The relevant time scales are defined by the main variable b) that is to be controlled. Type a) variables are simplified as being algebraic functions of the other variables and type c) variables are assumed to be constants.

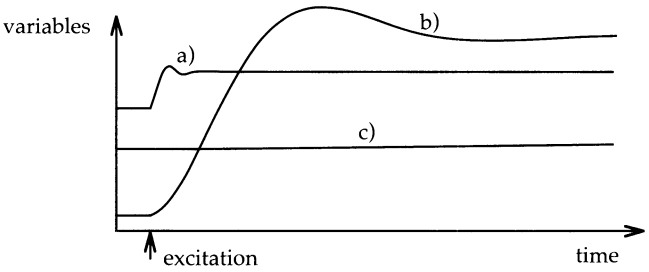


Fig. A.1. Classification of variables into "algebraic" (a), "dynamic" (b) and "static" (c).

The first four steps are shown here in further detail by deriving a model of a cylindric water-tank system as depicted in Fig. A.2:

1. The input is $\dot{m}_{in}(t)$, i.e., the inflowing water mass flow.
   The output is $h(t)$, i.e., the measured height of the water in the tank.
2. The only relevant reservoir is the mass $m(t)$ of water in the tank.
   The reservoir effects in the measuring device are very fast and can therefore be neglected.
   The water temperature (and therefore its density) is assumed to change very slowly such that it may be assumed to be constant.
3. The resulting differential equation is shown in (A.2).

$$\frac{d}{dt}m(t) = \dot{m}_{in}(t) - \dot{m}_{out}(t) \tag{A.2}$$

4. The input mass flow is assumed to be the control signal and can therefore be set arbitrarily by the controller to be designed.
   The outgoing mass flow can be expressed using Bernoulli's law, shown in

(A.3), assuming incompressible and frictionless flow conditions, resulting in (A.4).

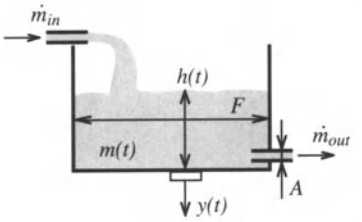


Fig. A.2. Water tank system, $m(t)$ mass of water in tank, and corresponding height $h(t)$, F tank-floor area, A = orifice area.

$$\dot{m}_{out}(t) = A\rho v(t), \quad v(t) = \sqrt{2\Delta p(t)/\rho}, \quad \Delta p(t) = \rho g h(t) \tag{A.3}$$

$$\frac{d}{dt}m(t) = \rho F \frac{d}{dt}h(t) = \dot{m}_{in}(t) - A\rho\sqrt{2gh(t)} \tag{A.4}$$

The models obtained thus are (usually) nonlinear and have non-normalized variables. Unfortunately, there is in general no "neat" controller design method for nonlinear systems, and non-normalized variables can lead to numerical problems and are hard to compare with one another.

For these reasons, the model (A.4) must first be normalized using constant nominal values ("set points"), i.e.,

$$h(t) = h_n x(t), \qquad \dot{m}_{in}(t) = \dot{m}_{in,n} u(t) \tag{A.5}$$

such that the new variables $x(t)$ and $u(t)$ will be dimensionless and around 1 in magnitude. Using the definitions (A.5), (A.4) can be rewritten as

$$\rho F h_n \frac{d}{dt}x(t) = \dot{m}_{in,n}u(t) - A\rho\sqrt{2gh_n}\sqrt{x(t)} \tag{A.6}$$

Notice that if the set point is chosen to be an equilibrium point (which is usually the case), the nominal amounts of in-flowing and out-flowing water must be equal, i.e.,

$$\dot{m}_{in,n} = A\rho\sqrt{2gh_n} \tag{A.7}$$

In order to linearize the system, only small deviations from its set points are analyzed, and the following new variables

$$x(t) = 1 + \delta x(t) \text{ with } |\delta x| \ll 1, \qquad u(t) = 1 + \delta u(t) \text{ with } |\delta u| \ll 1 \tag{A.8}$$

are introduced. Expanding (A.6) into a Taylor series and neglecting all the terms of second and higher order yields a linear differential equation

$$\rho F h_n \frac{d}{dt}\delta x(t) \approx \dot{m}_{in,n}(1+\delta u(t)) - A\rho\sqrt{2gh_n}(1+\frac{1}{2}\delta x(t)) \tag{A.9}$$

which, using (A.7), can be simplified to

$$\rho F h_n \frac{d}{dt}\delta x(t) = \dot{m}_{in,n}\delta u(t) - A\rho\sqrt{2gh_n}\frac{1}{2}\delta x(t) \tag{A.10}$$

or, by rearranging terms and again using (A.7)

$$\tau\frac{d}{dt}\delta x(t) = 2\delta u(t) - \delta x(t) \tag{A.11}$$

where the *time constant* τ is defined by

$$\tau = \frac{F}{A}\sqrt{\frac{2h_n}{g}} \tag{A.12}$$

This equation shows that the system becomes slower when either the ratio F/A or the set point h_n is increased. The *gain* of the system is 2 in its normalized form and equal to $2h_n/\dot{m}_{in,n}$ in its non-normalized representation.

Example Cruise-Control Problem

The objective of a series of examples is to design a robust cruise-control system,[1] i.e., a control loop that keeps the vehicle speed at its set point despite unobservable disturbances (grading, wind, etc.) and which is robust with respect to changing system parameters (speed set point, vehicle mass, gear ratio, etc.) and neglected dynamic effects (engine torque delays, sensors, etc.). In this section a first model of the dynamic behavior of a simplified vehicle (see Fig. A.3) is derived.

Nonlinear System Description

First, the nonlinear and nonnormalized system equations are derived. The following assumptions are made in this step:

1. The clutch is engaged such that the gear ratio γ is piecewise constant.
2. No drivetrain elasticities and wheel slip effects need be considered, i.e., the following relations are valid: $v(t) = r_w \cdot \omega_w(t)$ and $\omega_w(t) = \gamma \cdot \omega_e(t)$.
3. The vehicle must overcome rolling friction (the force acting on the vehicle is $F_r = c_r \cdot m \cdot g$) and aerodynamic drag ($F_a(t) = 1/2 \cdot \rho \cdot c_w \cdot A \cdot v^2(t)$).

[1] Of course, due to space limitations many important aspects of a real cruise-control system cannot not be addressed in this example.

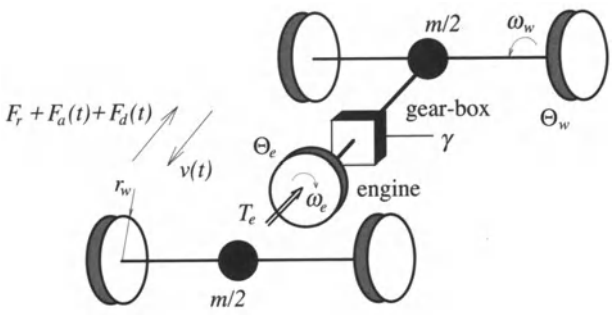


Fig. A.3. Simplified vehicle structure, vehicle mass m (wheel masses included) concentrated in the two axles.

4. All other forces are lumped into an unknown disturbance $F_d(t)$.
5. The mechanical energy of a moving part is $1/2 \cdot m \cdot v^2$ (pure translation) and $1/2 \cdot \Theta \cdot \omega^2$ (pure rotation).

The procedure outlined in Sect. A.1 is followed now.

Step 1: The input to the system is the engine torque T_e, which can be assumed to be arbitrarily controllable (the time delay caused by the engine dynamics is very small compared to the typical time constants of the vehicle speed and is therefore assumed to be an algebraic variable, see Fig. A.1). The system output is its speed $v(t)$.

Step 2: The relevant reservoirs are the kinetic energies stored in the vehicle's translational and rotational degrees of freedom, i.e.,

$$E_{tot} = \frac{1}{2} m v^2(t) + 4 \frac{1}{2} \Theta_w \omega_w^2(t) + \frac{1}{2} \Theta_e \omega_e^2(t) \tag{A.13}$$

Using the assumptions listed above (A.13) can be simplified as follows

$$E_{tot} = \frac{1}{2} \left[m + \frac{4\Theta_w}{r_w^2} + \frac{\Theta_e}{\gamma^2 r_w^2} \right] v^2(t) \tag{A.14}$$

Obviously, the level variable is the vehicle speed $v(t)$.

Step 3: The flows acting on the system are the mechanical powers affecting the system, i.e.,

$$P_+(t) = T_e(t) w_e(t), \quad \text{and} \quad P_-(t) = (F_r + F_a(t) + F_d(t)) v(t) \tag{A.15}$$

The energy conservation equation is therefore

$$\frac{d}{dt} E_{tot}(t) = P_+(t) - P_-(t) \tag{A.16}$$

Inserting (A.14) and (A.15) yields

$$\frac{1}{2}\left[m + \frac{4\Theta_w}{r_w^2} + \frac{\Theta_e}{\gamma^2 r_w^2}\right] 2v(t)\frac{d}{dt}v(t) = T_e(t)v(t)\frac{1}{r_w\gamma} - (F_r + F_a(t) + F_d(t))v(t) \tag{A.17}$$

Step 4: This step simply consists of rearranging (A.17) to obtain

$$M(\gamma, m)\frac{d}{dt}v(t) = \frac{1}{r_w\gamma}T_e(t) - (c_r mg + \frac{1}{2}\rho c_w A v^2(t) + F_d(t)) \tag{A.18}$$

where

$$M(\gamma, m) = \left[m + \frac{4\Theta_w}{r_w^2} + \frac{\Theta_e}{\gamma^2 r_w^2}\right] \tag{A.19}$$

is the total inertia of the system.

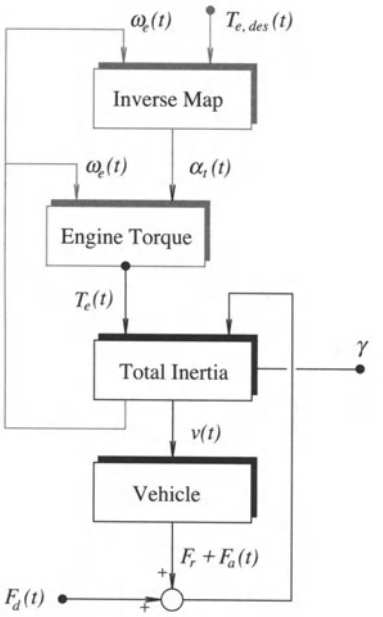


Fig. A.4. Cause-and-effect diagram of the longitudinal system dynamics. Explanations of the gray-shaded blocks will follow.

Normalization and Linearization

The next steps are to choose a meaningful set point and to normalize and linearize the obtained equations. The set point is $\{v_n, T_{e,n}, 0\}$ and the normalized system variables are

$$v(t) = v_n x(t), \qquad T_e(t) = T_{e,n} u(t), \qquad F_d(t) = \rho c_w A v_n^2 d(t) \tag{A.20}$$

Note that since the disturbance set point is zero, the disturbance force is normalized by twice the nominal aerodynamic force (normalization with rolling friction is also possible). The normalized version of (A.18) thus has the form

$$M(\gamma, m) v_n \frac{d}{dt} x(t) = \frac{T_{e,n}}{r_w \gamma} u(t) - (c_r mg + \frac{1}{2} \rho c_w A v_n^2 x^2(t) + \rho c_w A v_n^2 d(t)) \tag{A.21}$$

The set point $\{v_n, T_{e,n}, 0\}$ is again assumed to be an equilibrium point such that

$$\frac{T_{e,n}}{r_w \gamma} = c_r mg + \frac{1}{2} \rho c_w A v_n^2 \tag{A.22}$$

Linearizing the system using the small deviations

$$x(t) = 1 + \delta x(t), \quad |\delta x| \ll 1 \tag{A.23}$$

$$u(t) = 1 + \delta u(t), \quad |\delta u| \ll 1 \tag{A.24}$$

$$d(t) = \delta d(t), \quad |\delta d| \ll 1 \tag{A.25}$$

yields

$$M(\gamma, m) v_n \frac{d}{dt} \delta x = \frac{T_{e,n}}{r_w \gamma} (1 + \delta u) - c_r mg - \frac{1}{2} \rho c_w A v_n^2 \left[1 + 2\delta x + \delta x^2 + 2\delta d\right] \tag{A.26}$$

(for reasons of space, the time dependencies have been omitted). Neglecting the second-order terms and inserting (A.22) yields the desired linear differential equation

$$M(\gamma, m) v_n \frac{d}{dt} \delta x(t) \approx \frac{T_{e,n}}{r_w \gamma} \delta u(t) - \rho c_w A v_n^2 \left(\delta x(t) + \delta d(t)\right) \tag{A.27}$$

This equation can be put into its normal form by dividing all terms by $\rho c_w A v_n^2$ and using equations (A.22) and (A.19)

$$\tau(\gamma, m, v_n) \frac{d}{dt} \delta x(t) = k(m, v_n) \delta u(t) - \delta x(t) - \delta d(t) \tag{A.28}$$

where

$$\tau(\gamma, m, v_n) = \frac{1}{\rho c_w A v_n} \left(m + \frac{4\Theta_w}{r_w^2} + \frac{\Theta_e}{\gamma^2 r_w^2} \right) \tag{A.29}$$

and

$$k(m, v_n) = \frac{1}{2} + \frac{c_r mg}{\rho c_w A v_n^2} \tag{A.30}$$

Both τ and k are functions of the system parameters mass m and gear-ratio γ (the other parameters will vary much less) and of the vehicle speed set point. This is emphasized by using the explicit notation $...(\gamma, m, v_n)$.

System Parameter Variations

Of course, the physical system parameters are not known precisely and vary within some reasonable limits during normal operation of the vehicle. Table A.1 gives examples of such limits.

Table A.1. System parameters, masses correspond to empty vehicle, one passenger and fully loaded, gear ratios correspond to third, fourth and fifth gears (cruise control is assumed to be disabled in first and second gear).

$r_w = 0.25\ m$	$\Theta_w = 1.0\ kg\,m^2$	$\Theta_e = 0.25\ kg\,m^2$
$\rho = 1.2\ kg/m^3$	$c_r = 0.012\ -$	$c_w A = 0.75\ m^2$
$m = 1500\ kg$	$m = 1600\ kg$	$m = 2000\ kg$
$\gamma_3 = 0.19\ -$	$\gamma_4 = 0.25\ -$	$\gamma_5 = 0.37\ -$
$v_n^* = 20\ m/s$	$v_n^* = 30\ m/s$	$v_n^* = 40\ m/s$

The parameters τ and k of the control-oriented model (A.28) depend upon all of these physical parameters in a complex way. Therefore, it would be interesting to know what the limits

$$\tau_{min} \leq \tau \leq \tau_{max}, \qquad k_{min} \leq k \leq k_{max} \tag{A.31}$$

are in which they vary. This problem is best solved by writing a small program that does a reasonable "gridding" in the parameter space $\{m, \gamma, v_n\}$ and that finds (estimates) of the model parameter intervals. Note that this approach, in general, is *not* guaranteed to give exact results (readers interested in this are referred to [1]).

Remark: The nominal vehicle speed v_n substantially influences the system's dynamic behavior. In order to avoid a poor closed-loop system behavior, the controllers are usually scheduled with regard to nominal vehicle speed (gain scheduling will be discussed in Sect. A.7.1). For the sake of simplicity, only three scheduling speeds v_n^* are assumed in the following. Each value will be active while the vehicle is in a 10 m/s interval, e.g., for $v \in [25, 35]\ m/s$ a v_n^* of 30 m/s will be chosen (cruise control is assumed to be disabled for speeds below 15 m/s, which corresponds to approximately 50 km/h or 30 mph, and vehicle top-speed is assumed to be at 45 m/s, which corresponds to 160 km/h or 100 mph).

Table A.2 shows the intervals of the system parameters τ and k obtained for the three nominal speeds mentioned above.

Table A.2. Extreme values of the control model parameters for the scheduling speeds v_n^*.

$v_n^* = 20\ m/s$	$\tau_{min} = 68.7\ s$	$\tau_{max} = 155.\ s$	$k_{min} = 0.813$	$k_{max} = 1.663$
$v_n^* = 30\ m/s$	$\tau_{min} = 49.0\ s$	$\tau_{max} = 92.8\ s$	$k_{min} = 0.660$	$k_{max} = 0.919$
$v_n^* = 40\ m/s$	$\tau_{min} = 38.2\ s$	$\tau_{max} = 66.3\ s$	$k_{min} = 0.596$	$k_{max} = 0.714$

Some Comments

The maximum variation intervals shown in Table A.2 are not independent. As Fig. A.5 shows, in general control-oriented parameters like τ and k depend in a nonlinear way, upon the physical parameters (mass, gear ratio, nominal speed, etc.). Assuming that τ and k are independent, i.e., assuming that *all* values inside the rectangle shown in Fig. A.5 must be expected, yields sufficient, but not necessary, conditions for stability and performance levels. Accordingly, such designs will be "conservative," i.e., they will not realize some of the available potential.

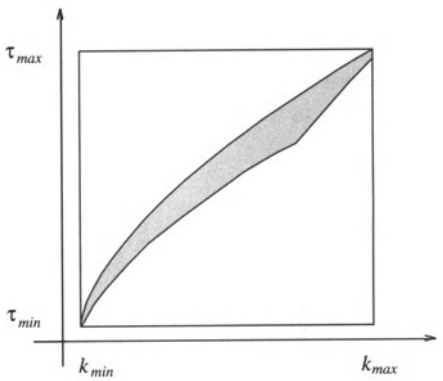


Fig. A.5. Set of possible control-oriented parameters τ and k for varying physical parameters m, γ and v_n^*.

Another comment concerns the chosen input signal. Of course, the assumption that the engine torque is the *input* to the system is not physically correct:

1. The real system input is the throttle angle α_t, as shown in Fig. A.4.
2. The *static* engine torque depends on the engine speed and on the throttle angle in a nonlinear way $T_e = f(\omega_e, \alpha_t)$, as shown in Fig.A.6.

3. The *dynamic* engine torque reacts with some delay to changes in the throttle position, mainly due to the intake manifold dynamics and the induction to power stroke delays.

However, the nonlinearity can be compensated by a static inverse map, as shown with gray lines in Fig. A.4, such that $T_e(t) \approx T_{e,des}(t)$ and the dynamic effects are, as mentioned, small, and will be dealt with by designing sufficient robustness into the system.

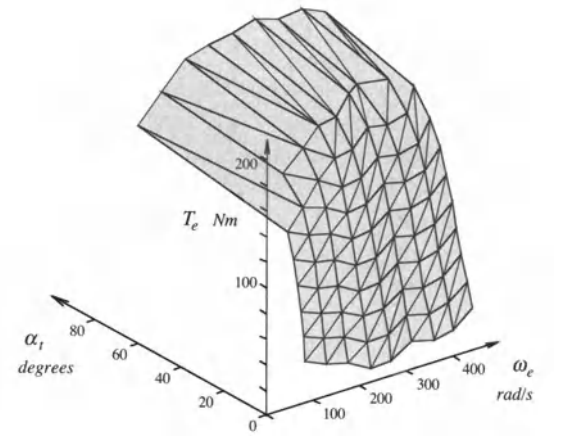


Fig. A.6. Static relation between engine speed ω_e, throttle valve opening angle α_t and engine torque T_e (2-liter SI engine).

A.2 System Description and System Properties

General Remarks

This section summarizes several ways to describe and analyze SISO dynamic systems. One important description for dynamic systems was used in the last section. In fact, ordinary differential equations, in general, can be written as follows

$$\frac{d}{dt}x(t) = f(x(t), u(t), t), \quad x(t) \in \Re^n, \quad u(t) \in \Re, \quad x(0) = x_0 \tag{A.32}$$

are the first way to describe the dynamic behavior of a system. Once the level variable $x(t)$ is known, the output $y(t)$ depends in an algebraic way on this variable and on the input signal $u(t)$

$$y(t) = g(x(t), u(t)), \quad y(t) \in \Re \tag{A.33}$$

Normalizing and linearizing (A.32) and (A.33) yields the following linear system equations

$$\begin{aligned} \tfrac{d}{dt}\delta x(t) &= A \cdot \delta x(t) + b \cdot \delta u(t) \\ \delta y(t) &= c \cdot \delta x(t) + d \cdot \delta u(t) \end{aligned} \tag{A.34}$$

The system matrices $\{A, b, c, d\}$ are defined by

$$\begin{aligned} A &= \left.\frac{\partial f(x)}{\partial x}\right|_{x_0,u_0} \in \Re^{n\times n}, \quad b = \left.\frac{\partial f(x)}{\partial u}\right|_{x_0,u_0} \in \Re^{n\times 1} \\ c &= \left.\frac{\partial g(x)}{\partial x}\right|_{x_0,u_0} \in \Re^{1\times n}, \quad d = \left.\frac{\partial g(x)}{\partial u}\right|_{x_0,u_0} \in \Re^{1\times 1} \end{aligned} \tag{A.35}$$

where x_0 is the system's equilibrium point associated to the input u_0 at which the system is linearized. Usually, the prefix δ is omitted when using this "state-space form."

The solution for (A.32) can be found for given initial conditions x_0 and for the input signal u using *numerical* routines running on digital computers. This process, denoted by system simulation, has become a routine operation since easily programmable and reliable software tools exist today.

While simulations are very useful to *analyze* the system's behavior, they cannot be used for controller *synthesis*. The basic problem is that, in general, closed-form solutions for (A.32), or even for its linearized version (A.34), are not easy to find.

However, at least for linear or linearized systems (A.34), such solutions can be found in the frequency domain using *Laplace transformation* techniques. The main idea of Laplace transformation is to replace the independent (real) variable t by another (complex) variable s such that in the frequency domain differentiation $\frac{d}{dt}$ is transformed to a multiplication with s. Three simple rules are important:

- Variables $x(t)$ are transformed to variables $X(s)$.
- Linear multiplications are invariant under Laplace transformation, i.e., $k \cdot x(t)$ is transformed to $k \cdot X(s)$.
- Derivatives of variables are transformed according to the rule $\frac{d}{dt}x(t) \rightarrow sX(s) - x(0)$.

Applying these rules to the water-tank equation

$$\tau \frac{d}{dt}\delta x(t) = 2\delta u(t) - \delta x(t) \tag{A.36}$$

yields

$$\tau\left(s\delta X(s) - \delta x(0)\right) = 2\delta U(s) - \delta X(s) \tag{A.37}$$

which is an algebraic equation that can now be solved easily

$$\delta X(s) = \frac{2}{\tau s + 1} \cdot \delta U(s) + \frac{\tau}{\tau s + 1} \cdot \delta x(0) \tag{A.38}$$

The initial condition $\delta x(0)$ is often zero. Specifically for feedback control design, the important component is the *input-output* behavior characterized by the plant's *transfer function*

$$P(s) = \frac{\delta X(s)}{\delta U(s)} = \frac{2}{\tau s + 1} \tag{A.39}$$

This representation, which in general has the form

$$P(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \ldots + b_1 s + b_0}{s^n + a_{n-1} s^{n-1} + a_{n-2} s^{n-2} + \ldots + a_1 s + a_0} \quad , \tag{A.40}$$

is useful for system analysis and synthesis. The zeros of the denominator polynomial (the transfer function's poles) are the eigenvalues of the original linear system, which describe the qualitative behavior of the system (unstable/stable, damping, etc.). The zeros of the numerator polynomial (the transfer functions zeros) also play an important role (e.g., they limit the achievable performance of any feedback controller).

In synthesis, another property of transfer functions is especially useful, *viz.* the series interconnection of two dynamic systems results in a simple multiplication of their transfer functions, e.g., for

$$\tau_1 \frac{d}{dt}\delta x(t) = k_1 \delta u(t) - \delta x(t), \quad \tau_2 \frac{d}{dt}\delta z(t) = k_2 \delta x(t) - \delta z(t) \tag{A.41}$$

the frequency domain representation is

$$P(s) = \frac{\delta Z(s)}{\delta U(s)} = \frac{k_1}{\tau_1 s + 1} \cdot \frac{k_2}{\tau_2 s + 1} = \frac{k_1 k_2}{\tau_1 \tau_2 s^2 + (\tau_1 + \tau_2) s + 1} \tag{A.42}$$

Transfer functions are completely equivalent to differential equations (no information is lost going from one domain to the other), i.e., they describe the full dynamic behavior of the system. If one is interested only in the system's response to harmonic excitations, and, moreover, if only the steady-state component of the response is considered, then the simpler (but very useful) description as a *frequency response* is sufficient.

In the steady-state component, i.e., after the transients due to initial conditions have vanished, the system's response is also a sinusoidal curve having exactly the same frequency as the excitation. However, the *amplitude* of the output signal is different and its *phase* usually lags behind the input signal.

$$u(t) = cos(\omega t) \quad \rightarrow \quad y(t) = A(\omega) \cdot cos(\omega t + \varphi(\omega)) \tag{A.43}$$

Since these two quantities depend on the excitation frequency, it is useful to display this dependency in amplitude/frequency and phase/frequency diagrams which are called *Bode Diagrams.* (Fig. A.8 shows the Bode diagram of the water-tank system.)

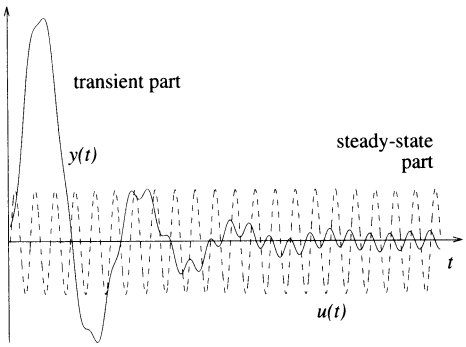


Fig. A.7. Response $y(t)$ (solid) of an asymptotically stable system to a sinusoidal input $u(t)$ (dashed).

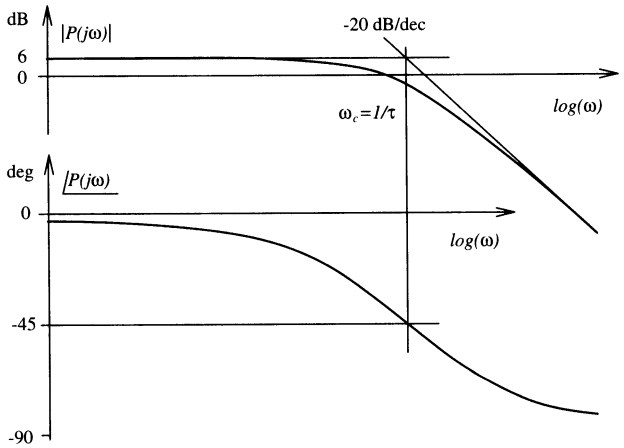


Fig. A.8. Bode diagram of the water-tank system, ω_c is the *corner frequency* defined by $\omega_c = 1/\tau$.

To facilitate series interconnections of systems, Bode diagrams are usually plotted with a logarithmic scale in the amplitude and a linear scale in the phase, such that series connections like (A.41) can be visualized by simply adding the two Bode diagrams. Often the phase is given in degrees and the amplitude in dB defined as

$$A_{dB} = 20 \cdot log_{10}(A) \tag{A.44}$$

Although very useful for quantitative reasonings, Bode diagrams have the drawback that the phase and the magnitude information is separated. This problem is avoided in *Nyquist diagrams* which combine this information by using a frequency-implicit representation (see Fig. A.9).

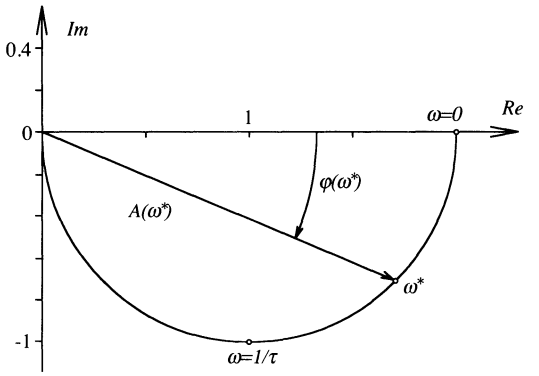


Fig. A.9. Nyquist diagram of the water-tank system.

Notice that the frequency response is obtained by substituting the complex variable s in the transfer function $P(s)$ by the purely imaginary variable $j\omega$ ($j = \sqrt{-1}$). For instance assume that the transfer function of a system is given by

$$P(s) = \frac{bs + 1}{\gamma s^2 + \beta s + 1} \tag{A.45}$$

then its frequency response is simply

$$P(j\omega) = \frac{bj\omega + 1}{\gamma(j\omega)^2 + \beta j\omega + 1} = \frac{1 + j\omega b}{(1 - \gamma\omega^2) + j\omega\beta} \tag{A.46}$$

The maximum value of the magnitude of $P(j\omega)$ over all frequencies ω is an important parameter. This variable is therefore denoted by a special symbol

$$||P(s)||_\infty = \max_{\omega \in \Re} \{|P(j\omega)|\} \tag{A.47}$$

If $P(s)$ is a transfer function that describes the influence of a disturbance d on the system output y, then minimizing $||P(s)||_\infty$ corresponds to minimizing the maximum influence over all frequencies that d can have on y in steady state.

The three mentioned system representations, i.e., results of simulations, Bode diagrams and Nyquist diagrams, are the "language" with which all the following controller design and analysis ideas will be explained.

The last point discussed in this section is the question of system stability. The stability of a linear and time-invariant system can be checked by computing (or measuring) its response to impulse inputs. In this case, the system output $y(t)$ can have three forms:

- $\lim_{t\to\infty} |y(t)| = 0$ in which case the system is *asymptotically stable*
- $\lim_{t\to\infty} |y(t)| = c < \infty$ in which case the system is *stable* (but not *asymptotically* stable)

- $\lim_{t\to\infty} |y(t)| = \infty$ in which case the system is *unstable*.

It can be shown that a system with transfer function $P(s)$ is asymptotically stable if and only if all of the poles of $P(s)$ have negative real parts.

One of the most useful stability analysis tools is the Nyquist criterion, which will be presented below for the case where both the plant and the controller are asymptotically stable. The Nyquist criterion is able to predict the stability of a closed-loop system based on the *frequency response* of the *open-loop* system. In other words, the steady-state behavior of the open loop is sufficient to predict the transient behavior of the closed loop. This rather surprising fact is a very powerful system analysis and synthesis tool, as will be shown in the following sections.

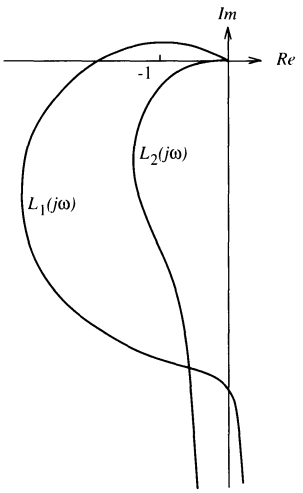


Fig. A.10. Open-loop systems which will be unstable ($L_1(j\omega)$) and stable ($L_2(j\omega)$) when the loop is closed with unity feedback.

Nyquist Stability Criterion (simplified) Assume that the controller $C(s)$ and the plant $P(s)$ have no poles in the complex right-half plane. Then the closed-loop system $T(s) = L(s)/(1 + L(s))$ will be asymptotically stable if and only if the frequency response $L(j\omega)$ of the open-loop transfer function $L(s)$ does not encircle the critical point $z = -1$.

Examples of a stable, $L_2(s)$, and an unstable, $L_1(s)$, closed-loop system are shown in Fig. A.10.

Example Cruise-Control Problem

As an example, the Bode and the Nyquist diagrams of the cruise-control system for all possible values of the parameters and the scheduling speed

$v_n^* = 30\ m/s$ are computed here. Fig. A.11 shows the resulting Bode sets for "all" possible parameter values listed in Table A.1. The analogous Nyquist set is shown in Fig. A.13. Of course, this plot has been derived by "gridding" the (three-dimensional) uncertain parameter space and therefore there is no guarantee that the exact limits are found in this way. Notice that – in general – it is not a vertex in the parameter space (i.e., a combination of extreme parameter values) that produces the boundary of the Bode sets and that the nominal parameters are not the mean values of the interval. For more details see [1].

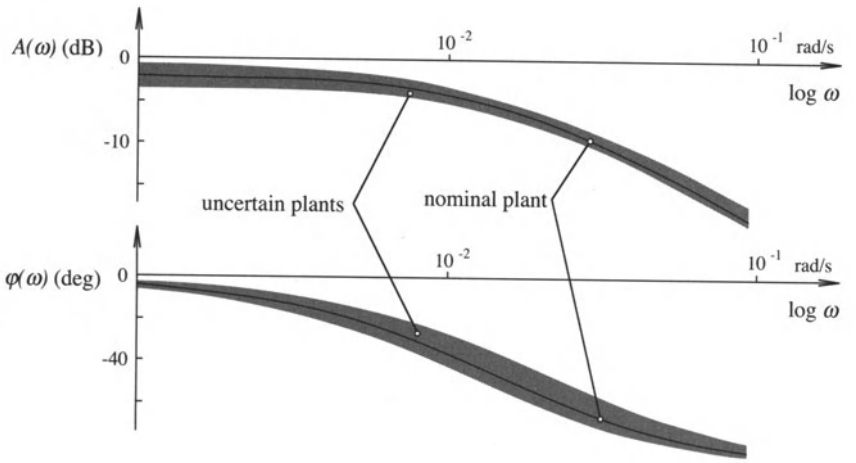


Fig. A.11. Bode sets corresponding to all possible parameter values as indicated in Table A.1 and for the scheduling speed $v_n^* = 30\ m/s$ (speed interval $25\ m/s \leq v_n < 35\ m/s$).

Notice that the implicit assumption when using linear stability theory is that a nonlinear and parameter-varying system can be described by a set of linear time-invariant equations. This assumption is by no means obvious and can be justified only if the set points and the parameter changes are substantially slower than the relevant dynamic effects that are to be controlled.

For the cruise-control problem, this is indeed the case, i.e., the vehicle mass remains constant during a transient (i.e., a grading step) and vehicle speed set points and gear ratios remain piecewise constant for a finite interval.

A.3 Model Uncertainty

General Remarks

In the last section it was shown that uncertain systems can be described as families of linear systems centered around a nominal model. However, the

resulting sets are in general much too complicated to be analyzed directly. Therefore, a description of the model uncertainty must be found that is both simple enough to work with and sufficiently complex to permit useful conclusions about the system behavior when feedback controls are designed.

There are many approaches that satisfy these requirements. In this text only the most simple method will be presented. Simplicity will always induce conservativism, i.e., the uncertainty descriptions will exaggerate the actual uncertainty and the resulting controllers will be able to cope with larger sets of uncertain systems than the one at hand, at the price of not fully exploiting the possibilities offered by the uncertain system.

The starting point of all uncertainty descriptions is always a nominal system that must be centered in the family such as to minimize conservativism. Here, this is equivalent to the nominal plant being in the center of the Nyquist set of the uncertain system (black curve in Fig. A.13). The uncertain system (grey in Fig. A.13) is now "overbounded" by the black area set, which has to be as simple as possible in its description.

An obvious idea is to parameterize the uncertainty by the set

$$\Pi = \{P(s)(1 + W(s)\Delta(s)) \mid \quad |\Delta(j\omega)| \leq 1, \mathrm{Arg}(\Delta(j\omega)) \in \{-\pi, \pi\}\} \quad \text{(A.48)}$$

where:

- $P(s)$ is the nominal plant (exactly known, centered in the uncertain set, and used below to design the nominal controller).
- $W(s)$ represents the a-priori knowledge about the uncertainty; essentially it is an upper bound on the modulus of the uncertainty for different frequencies.
- $\Delta(s)$ is a unit disc centered at the origin.

The set Π is best visualized in a Nyquist plot as shown in Fig. A.12. For each fixed frequency ω^*, the set $\Pi(\omega^*)$ is a *disc* with the radius $|P(j\omega^*)W(j\omega^*)|$ and which is centered at $P(j\omega^*)$. By choosing an appropriate weighting $W(s)$, the actual uncertain set is covered, whereby $W(s)$ has to be chosen such that as little overlap as possible occurs. Note that only the magnitude of $W(s)$ is important. Its phase information is neutralized by Δ.

The uncertainty parametrization shown here is one of the simplest available. Other parametrization are known, most being more useful but also more complex. The most important limitation of the multiplicative parametrization is the restriction that the number of unstable poles may not change for the whole uncertain set.

Fig. A.12 also shows the fictitious true plant $P_t(s)$ that below will be assumed to exist. The design rules discussed in the next section will guarantee some desired properties for all possible true plants as long as they are completely inside the set Π. Of course, such a true plant is only an idealization (linear, time invariant) and therefore the "guarantees" obtained should be used with the necessary caution.

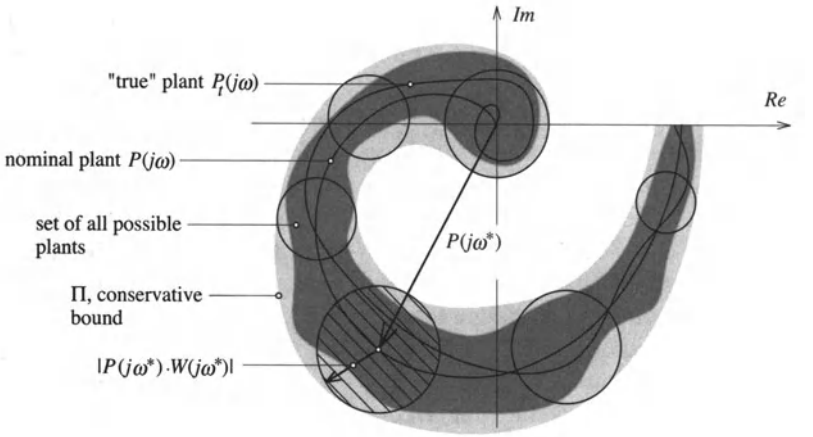


Fig. A.12. Geometric interpretation of the set Π.

Example Cruise-Control Problem

In this example, the two quantities $P(s)$ and $W(s)$, as introduced above, will be defined for the cruise-control example. The nominal system $P(s)$ is derived first. This problem is solved iteratively by finding a well-centered nominal system. Sometimes physical intuition is helpful, but usually it is not the case that the average physical parameter values define a good nominal system. Note that there might be many parameter combinations that produce good nominal systems. In this case, the following parameters were chosen: $m = 1850\ kg$, $\gamma = 0.28\ -$ and $v_n = 29\ m/s$. The resulting Nyquist curve is shown in Fig. A.13.

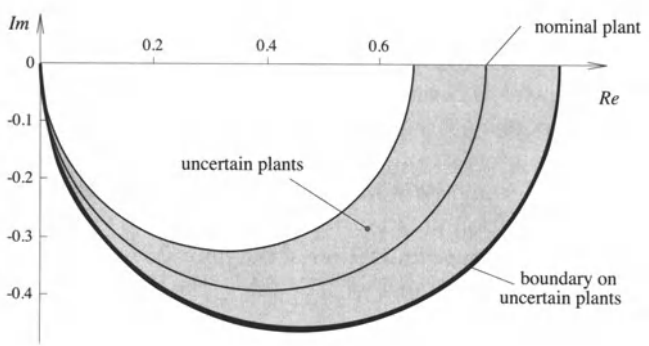


Fig. A.13. Nyquist sets of the cruise-control example, nominal, parametric uncertain (gray) and conservative boundary set (black).

The second step is to find a tight uncertainty bound $W(s)$. Again, this step involves several iterations in order to find a satisfactory solution. As a

starting point, the gain of the weighting $W(s)$, i.e., its value at $s = 0$ must be found. If the nominal system is well centered at $s = 0$, the value of $W(0)$ is minimized and corresponds to half the extension of the uncertain set at $s = 0$. As can be seen in Fig. A.13, a gain of approximately $k = 0.165$ is a good starting point.

This information provides the first choice possible for $W(s) = k$. This is, of course, the simplest possible uncertainty and its results are not very good. Since the weighting $W(s)$ is constant for all frequencies, it overestimates the uncertainties at higher frequencies.

To reduce that effect, an all-pass element

$$W(s) = 0.165 \cdot \frac{20s + 1}{60s + 1} \tag{A.49}$$

can be added such that the weight $W(s)$ now better fits the plant uncertainty. The result of this choice is shown in Fig. A.13 as well. All of the following considerations will use this uncertainty bound $W(s)$.

Remark: Often a system is not modeled from physical first principles, but its input-output behavior is measured using devices such as frequency analyzers. If these measurements are repeated several times (e.g., for different operating points), a set of Nyquist plots is obtained that is very similar to those presented here. Fitting a (fictitious) nominal system and a conservative uncertainty bound $W(s)$ is then completely analogous to the approach introduced in the previous sections. All of the following arguments will be applicable to this case as well.

A.4 Control System Design for Nominal Plants

General Remarks

Control System Structure

This section discusses the *nominal*, controller design step, i.e., the procedure for finding a good controller $C(s)$ for the nominal plant $P(s)$. In this step it is very important to distinguish between feedforward and feedback controllers. Figure A.14 shows the setup that is analyzed in this section.

In addition to the input signal $u(t)$, measurable and unmeasurable disturbances influence the system output through the corresponding transfer functions. The system output is not available. Only the variable $y(t)$, which is affected by an unknown noise, is available to the control system. The four variables that are accessible to the combined controller are the observable disturbance $d_o(t)$, the measured output $y(t)$, the reference signal $r(t)$ which are the inputs to the controller and $u(t)$, i.e., the input to the plant.

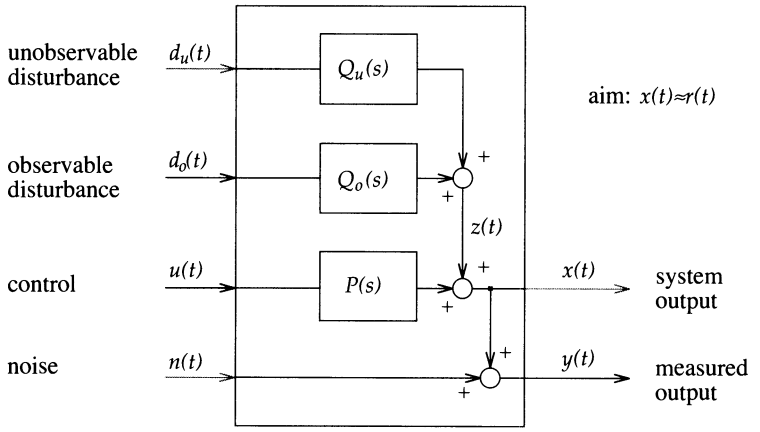


Fig. A.14. System setup for controller design.

A controller will therefore have the structure shown in Fig. A.15. The main idea in the feedforward component is to use all available knowledge, i.e., the nominal plant model $P(s)$ and the nominal disturbance model $Q_o(s)$, to design the filters $K_o(s)$ and $F(s)$. The design of these feedforward components is relatively straightforward for plant and disturbance models that are minimum phase (no zeros in the right-half complex plane). Essentially, this step requires the plant dynamics to be "inverted," i.e., a filter must be designed such that when connected to the input of the plant, it produces a near-unity transfer function over the frequency range of interest. The feedforward filter must include some low-pass elements with high corner frequencies in order to be realizable and in order to not amplify sensor noise.

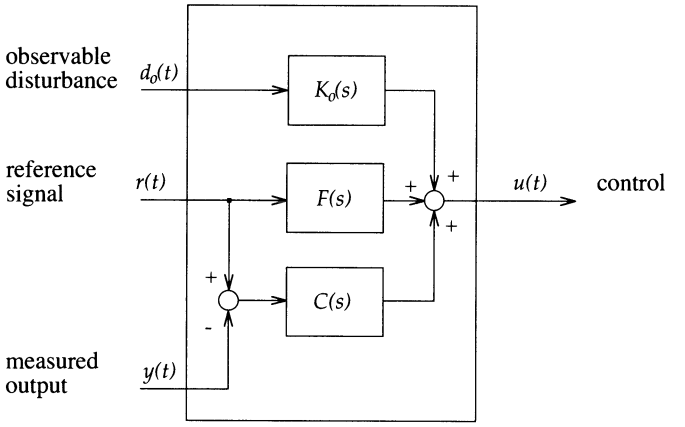


Fig. A.15. General control system structure.

For the controller design process, the following general guidelines are important:

- Use feedforward (open-loop) control whenever possible. It is fast, does not cause any additional stability problems and it does not introduce the measurement noise into the system.
- All models are uncertain and plants are always affected by not measurable disturbances. The sole purpose of feedback control is to reduce the uncertainty introduced by these unobservable disturbances and modeling errors.
- All feedback controllers must be designed such that stability and performance are guaranteed as long as the actual disturbances remain inside the bounds that have been determined a priori.

Design Guidelines

The design of the nominal feedback controller $C(s)$ is an interesting task for which a large amount of literature is available. In this section, only single-input-single-output (SISO) systems are discussed which, moreover, are assumed to be "nice" in the sense that they have no unstable poles and no "unstable" zeros (i.e., no zeros in the right complex plane, a system feature which is often called *minimum-phase*). A plant that does not have all three of these properties requires more caution, and the straightforward methods presented below will not work properly.

The following definitions will be useful

$$\begin{aligned} L(s) &= C(s) \cdot P(s) && \text{loop gain} \\ S(s) &= 1/(1 + L(s)) && \text{sensitivity} \\ T(s) &= L(s)/(1 + L(s)) && \text{complementary sensitivity.} \end{aligned} \tag{A.50}$$

These transfer functions describe how the exogenous inputs affect the closed-loop system output when only feedback control is applied, i.e., $F(s) = K_0(s) = 0$, $Q_u(s) = 1$, $d_o(t) = 0$ (cf. Figs. A.14 and A.15)

$$y(s) = S(s) \cdot d_u(s) + T(s) \cdot r(s) \tag{A.51}$$

The starting point for many controller designs are the observations that:

- Most plants are low-pass elements, i.e., they have a finite gain at very low frequencies and attenuate input signals of higher frequencies.
- High loop gains are useful to reduce the effects of unmeasurable disturbances and improve the set-point tracking behavior, as expressed by eq. (A.50).
- Phase lag, model uncertainty and noise limit the frequency range where high loop gains are permitted; otherwise an unstable or badly damped closed-loop system behavior results.

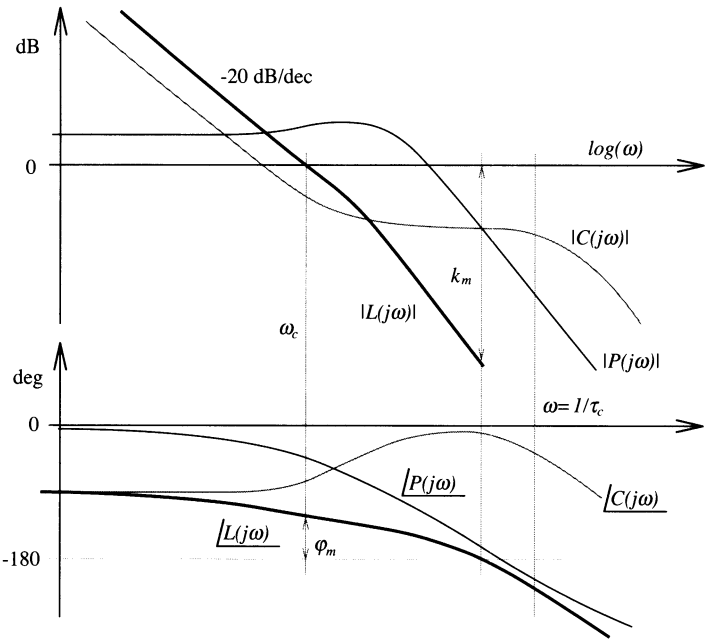


Fig. A.16. Typical result of a PI controller design for a "nice" plant.

PI Controllers

Therefore, at lower frequencies the controller must have high gains (they may increase without limits with frequencies tending to zero) and at higher frequencies the controller must have finite or even decreasing gains to guarantee a quick *roll-off*. Since PI controllers satisfy these conditions and are relatively easy to implement and tune, the majority of control loops are realized with such controllers

$$C(s) = k_p \cdot (1 + \frac{1}{\tau_i \cdot s}) \cdot \frac{1}{(\tau_c \cdot s + 1)^\rho}, \quad \rho \geq 1 \tag{A.52}$$

Figure A.16 shows a typical PI controller design in a Bode diagram. In general, the following characteristics are important:

- the crossover frequency ω_c is that frequency at which the loop gain magnitude $|L(j\omega)|$ crosses the 0 dB line;
- the phase margin φ_m is the the distance of the loop gain phase at ω_c to the -180 deg line; and
- the gain margin k_m is the distance of the loop gain magnitude to the 0 dB line at that frequency at which the phase of the loop gain is -180 deg.

The design of a PI controller consists of finding controller parameters $\{k_p,\ \tau_i,\ \tau_c\}$ such that the desired crossover frequency and phase and gain margins are attained. Useful rules of thumb are:

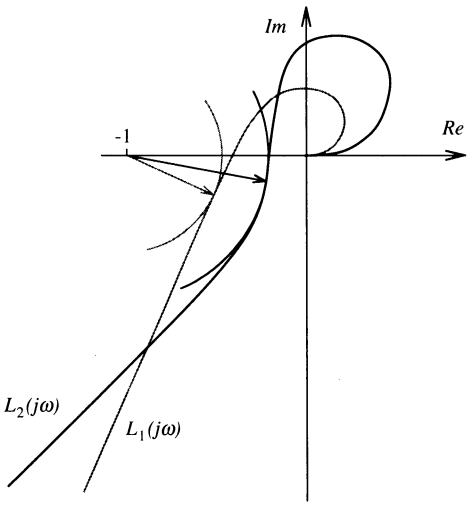


Fig. A.17. Loop shaping to increase minimum return difference, $L_2(j\omega)$ is the better design than $L_1(j\omega)$.

- $\omega_c \approx 2.3/t_{des}$ where t_{des} is the desired settling time of the closed-loop system after a step disturbance;
- $\varphi_m > 70 \ldots 80$ *deg* is required if no overshoot to disturbance steps is acceptable, a smaller phase margin can be acceptable, but at least 45 *deg* should be guaranteed for the nominal plant; and
- $k_m > 4$ *dB* is required which corresponds to a loop-gain error tolerance of approximately 50% .

The best approach to assessing the robustness of a specific design is to look at the Nyquist plot of the loop gain and to try to maximize the minimum distance d_{min} (A.53) to the critical point

$$d_{min} = \min_{\omega}\{|1 + L(j\omega)|\} \tag{A.53}$$

Figure A.17 shows as an example two Nyquist curves obtained with two sets of parameters $\{k_p,\ \tau_i,\ \tau_c\}$ and the corresponding d_{min}. Section A.5 introduces lead/lag elements (A.58) that can be used to improve a PI control system design in specific frequency bands.

Experimental Tuning Methods

If no mathematical model is available, the parameters of a PI controller (A.52) can be determined using one the following two experimental methods. Note that these approaches work well in *some* cases; however, there is no guarantee that a satisfactory design is obtained in *every* case.

The setup used in the first method is illustrated in Fig. A.18. The integrator time constant τ_i of the controller (A.52) is set to infinity and the gain k_p is

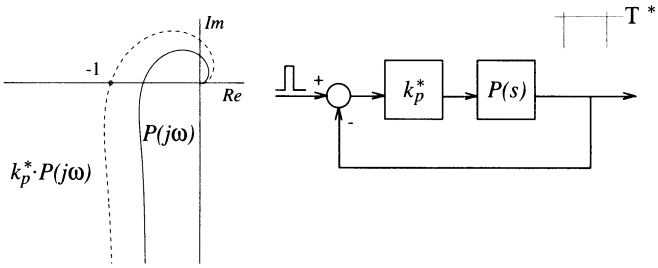


Fig. A.18. Illustration of the Ziegler-Nichols PI tuning method.

increased until the loop oscillates, i.e., until the poles of the closed-loop system are on the stability boundary. The corresponding critical gain k_p^* and critical oscillation period τ^* are used to derive the parameters of the PI controller as indicated in Table A.3.

Table A.3. Experimental design of a PI controller.

method	k_p	τ_i
Ziegler-Nichols	$0.45 \cdot k_p^*$	$0.85 \cdot \tau^*$
Chien-Hrones-Reswick, 0% overshoot	$0.6/\alpha$	$4.0 \cdot \delta$
Chien-Hrones-Reswick, 20% overshoot	$0.7/\alpha$	$2.3 \cdot \delta$

The second approach is illustrated in Fig. A.19. Here, a step response is the starting point for the estimation of the parameters of the PI controller. First, time t^* must be found where the second derivative $\partial^2 y/\partial t^2$ is equal to zero. The intersection of the tangent at $y(t)$ in that point with the axes yields the two necessary parameters: δ a delay time and α the strating point of the affine tangent. Again, once these two parameters are known, the PI controller gains can be found using the rules indicated in Table A.3.

The Ziegler-Nichols approach usually yields the more precise results. Unfortunatly, its application is sometimes not possible or too time consuming. The Chien-Hrones-Reswick method is easy to apply but estimating the parameters is not always easy, especially for noisy output signals. More information on these and other design methods for PI control systems can be found in [12].

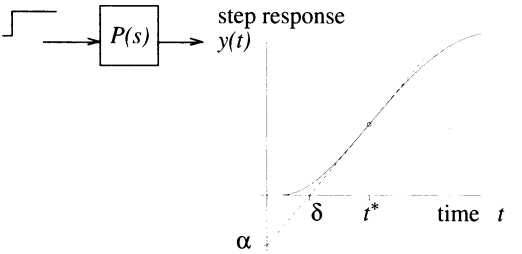


Fig. A.19. Illustration of the Chien-Hrones-Reswick PI tuning method.

Example Cruise-Control Problem

In this section of the cruise-control example, a nominal controller and a feedforward filter will be designed. The linearized and normalized plant and disturbance model of the cruise-control system has the following form

$$\delta y(s) = \frac{k(m, v_n)}{\tau(\gamma, m, v_n)s + 1} \cdot \delta u(s) + \frac{-1}{\tau(\gamma, m, v_n)s + 1} \cdot \delta d(s) \qquad \text{(A.54)}$$

and therefore

$$P(s) = \frac{k(m, v_n)}{\tau(\gamma, m, v_n)s + 1}, \quad \text{and} \quad Q_u(s) = \frac{-1}{\tau(\gamma, m, v_n)s + 1} \qquad \text{(A.55)}$$

Following the rules of thumb listed above, a crossover frequency of 0.4 rad/s is aimed at. This corresponds to approximately 6 s of settling time, which is much slower than the neglected torque dynamics, but sufficiently fast to satisfy customer demands. Overshoot should be avoided (at least in the nominal case) such that a phase margin of at least 70 deg must be achieved. Gain margin is not a problem in the cruise-control case since the dynamics of the system are essentially first order only.

After some iterations, the values $k_p = 33$ and $\tau_i = 10$ s are found. The additional low-pass filter is chosen at $\tau_c = \tau_i/20 = 0.5$ s such that it attenuates any neglected dynamics (torque dynamics, drive train ringing, etc.). As Fig. A.20 shows, the closed-loop system is stable since the Nyquist criterion is satisfied. This figure also shows that the return difference (A.53) has a minimal magnitude of approximately 0.87.

The disturbance response is the relevant time-based information since set-point tracking can be influenced by feedforward filter design (see below). Figure A.21 shows the closed-loop system response for a disturbance step $\delta d = 0 \rightarrow 1$, which corresponds to a step in demanded power of $17kW \rightarrow 40$ kW, or a disturbance force of $0 \rightarrow 756$ N. The settling time, i.e., the time needed for the vehicle speed to stop decreasing, as expected, is around 5 s. Also, the speed trajectory satisfies the "no overshoot" condition (torque slightly overshoots around $t = 20$ s which, of course, could be perceived by the driver as a small jerk).

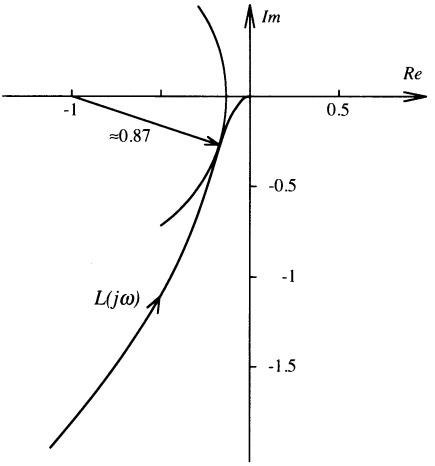


Fig. A.20. Nyquist diagram of a design of a PI controller for the cruise-control problem, also indicated is the minimum value of the *return difference*.

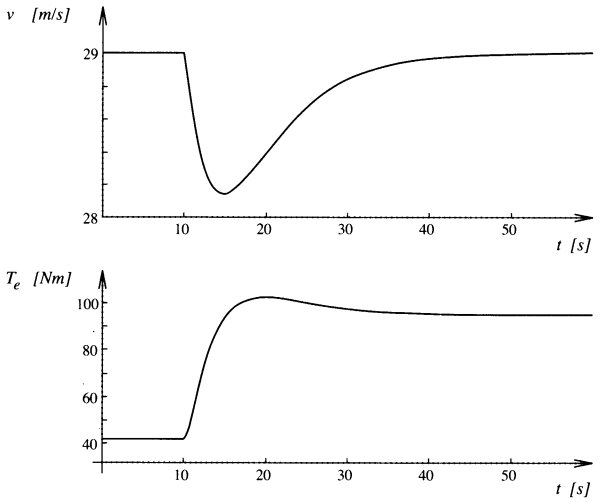


Fig. A.21. Closed-loop system response for a disturbance step of $\delta d = 0 \rightarrow 1$.

Figure A.21 shows both the simulation results obtained when applying the PI controller to the original nonlinear plant (A.18) and to its linearization (A.28). Although the disturbance step magnitude is quite large, the two trajectories of the original nonlinear plant and of the linearized plant are indistinguishable. This is a consequence of the fact that vehicle speed does not change much during this transient (less than 5%). No engine torque limits have been considered yet such that the system's trajectories would not change substantially for larger disturbance steps. In Sect. A.7, however, it will be shown

that (unavoidable) engine torque limitations will have very important consequences.

As mentioned, the overall control system performance can be improved substantially by adding feedforward action. In the cruise-control example, such a feedforward filter $F(s)$ will have the form

$$F(s) = \frac{\tau \cdot s + 1}{k \cdot (\tau_F \cdot s + 1)^\rho}, \qquad \rho \geq 1 \tag{A.56}$$

The available knowledge about the system is "inverted" and used to derive the driving signal, i.e., the desired engine torque. Since for low-pass systems a full inversion is not possible, additional low-pass filters with time constants τ_F are used that must be chosen as a compromise between fast system behavior and low control efforts and noise sensitivity.

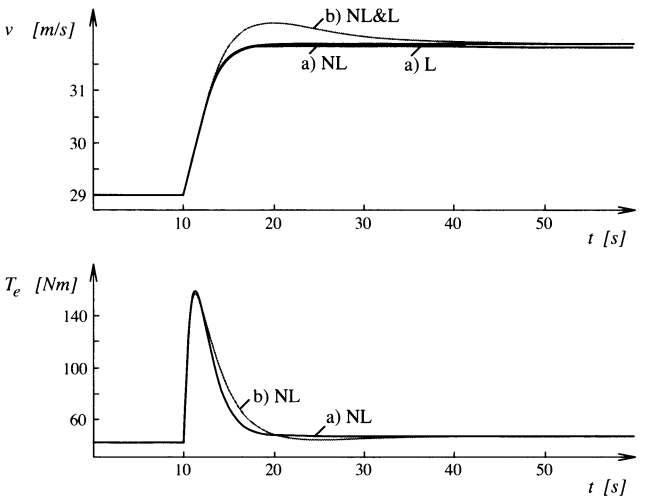


Fig. A.22. Speed and engine torque response to a set-point step for a) feedforward action only and b) feedback action only (NL: nonlinear plant model, L: linear plant model).

For the cruise-control example, a time constant $\tau_F = \tau/60$ turns out to be a good choice. Figure A.22 shows the set-point tracking behavior of the system in case a) with feedforward only ($C(s) = 0$) and case b) with feedback only ($F(s) = 0$). In case a), the response of the system to a step change in desired speed is clearly better without requiring higher control efforts. Notice, however, that in case a) (no feedback control), the linear and nonlinear plant's behavior differ substantially (an unavoidable consequence of the missing feedback correction action). In case b), a clear tendency to speed overshoot is observed. If this overshoot is eliminated by reducing the controller gains, the

disturbance rejection deteriorates, which, of course, is not desirable. A combination of the two control blocks, i.e., a feedforward and feedback part, can achieve good disturbance rejection *and* good tracking behavior simultaneously.

A.5 Control System Design for Uncertain Plants

General Remarks

In this section, the controller designed for the nominal system is tested with respect to its stability properties for the uncertain system. The main objective will be to "guarantee" stability of the closed-loop system for all possible true plants in the set defined by (A.48). The problem of how to improve a given design with respect to its stability properties is discussed below. However, the *robust performance* problem, i.e., how to guarantee that even in the uncertain case a design performs according to the chosen specifications, is not. The interested reader will find this information and many other important points in [44].

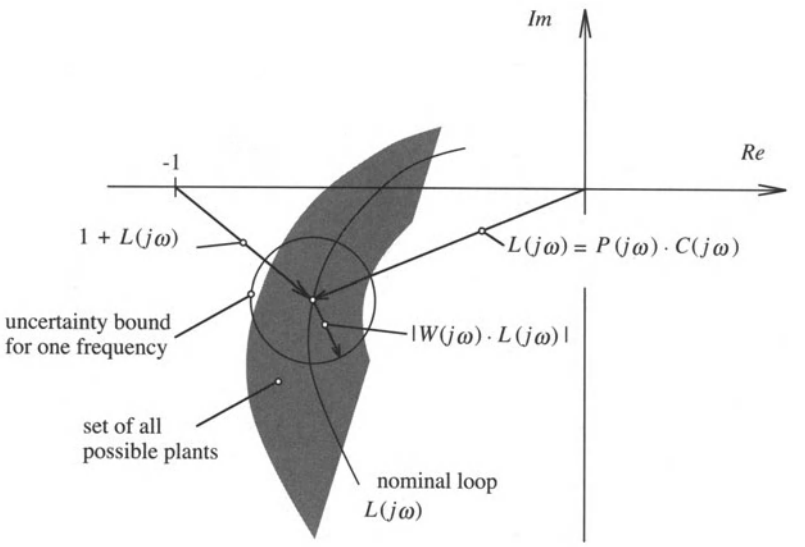


Fig. A.23. Interpretation of the robust stability condition in a Nyquist diagram.

The key idea of *robust stability* is to use the conservative bound $W(s)$, described in Sect. A.3, to test the uncertain system for closed-loop stability. If the following assumptions are satisfied:

- a nominal plant $P(s)$, a conservative bound on its uncertainty $W(s)$, and a fixed controller $C(s)$ are known,

- the nominal closed-loop system is asymptotically stable,
- the following inequality is satisfied for all frequencies ω

$$|1 + C(j\omega) \cdot P(j\omega)| \geq |C(j\omega) \cdot P(j\omega) \cdot W(j\omega)| \quad , \tag{A.57}$$

then the controller $C(s)$ guarantees the closed-loop system to be asymptotically stable for all true plants $P_t(s)$ in the set Π.

This robust stability theorem is very useful since it involves only one test on a known transfer function. Its main limitation is the assumption that a true plant exists. Figure A.23 shows a graphic interpretation of the condition (A.57).

If an insufficient minimum robust return difference is obtained, an iteration in the controller design is required. First attempts will be to optimize the controller parameters (k_p, τ_i and τ_c in the PI case). If this does not produce the desired results (insufficient stability margins or a too slow nominal system behavior), then the order of the controller must be increased. Often this is done by adding *lead/lag* elements which shape the open loop in specific frequency intervals such that the design specifications are reached.

These lead or lag elements are defined by

$$C_{ll} = \frac{\tau \cdot s + 1}{\alpha \cdot \tau \cdot s + 1} \tag{A.58}$$

where $\tau = 1/\tilde{\omega}$ is the corner frequency (the frequency where the loop-shaping action is centered) and α is the element's gain which determines the character of the element ($\alpha < 1$ for a lead element with differentiating behavior, $\alpha > 1$ for a lag with integrating behavior). As Fig. A.24 shows, a lead element can be used to shift the loop phase towards positive values (e.g., to increase the return difference in some frequency band). The drawback of that is an increased loop gain at higher frequencies. The lag element is used to decrease the loop gain (e.g., to improve the roll-off behavior), but increases the phase lag in a certain frequency band. Because of these partially contradicting effects, in general several design iterations must be made before a satisfactory result is obtained. More systematic procedures exist, but they require more evolved mathematical concepts. The interested reader is referred to [44].

Of course, the ultimate goal of all control system designs is not to merely achieve robust stability, but robust *performance*, i.e., the closed-loop system is expected to perform as specified as long as the true plant is included in the set (A.48). To achieve this goal, an additional sufficient stability margin must be maintained by the controller $C(s)$.

Example Cruise-Control Problem

Figure A.25 shows, the same quantities depicted in the schematic representation of Fig. A.23 for the cruise-control example. Clearly, the controller $C(s)$ designed in the last section satisfies the robust stability condition (A.57). The

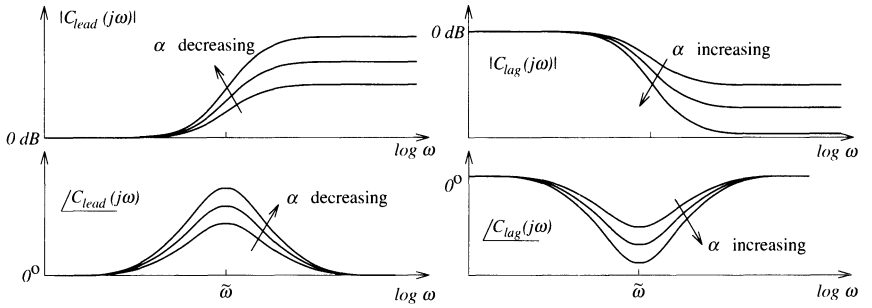


Fig. A.24. Bode diagrams of lead and lag element.

minimum distance of the critical point to the uncertainty bounding set can be interpreted as a *robust return difference* which has to be larger than zero in order to guarantee robust stability. For the cruise-control example, a minimal robust return difference of 0.84 is obtained with the PI controller designed in the last section. As was shown in Fig. A.20, the minimal *nominal* return difference was 0.87, so the deterioration caused by the uncertainty has been limited by the chosen controller to a small percentage.

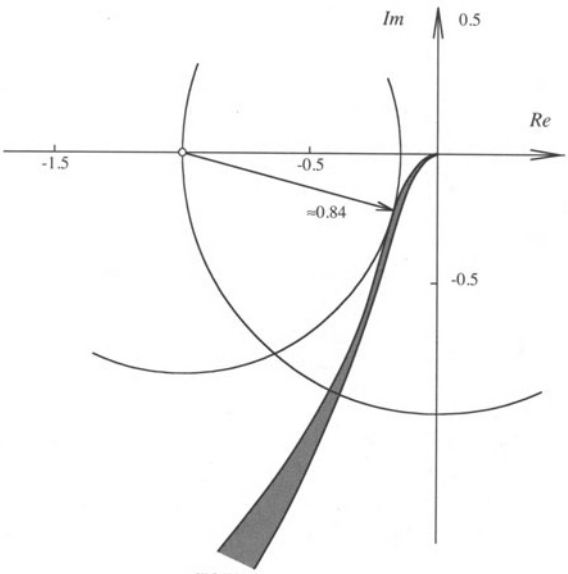


Fig. A.25. Nyquist diagram of the uncertain cruise control loop.

As mentioned above, the controller designed in Sect. A.4 meets the robust stability requirement (A.57) with a large additional stability margin. This reserve in stability indicates that the *performance* of the closed-loop system remains close to the nominal performance as long as the plant remains within

the set Π of (A.48). This assertion is supported by the results of a series of numerical simulations of the linear controller applied to the nonlinear plant model. As Fig. A.26 shows, the disturbance rejection properties of the controller $C(s)$ designed in Sect. A.4 remain satisfactory for the full range of expected nominal vehicle speeds and parameter variations.

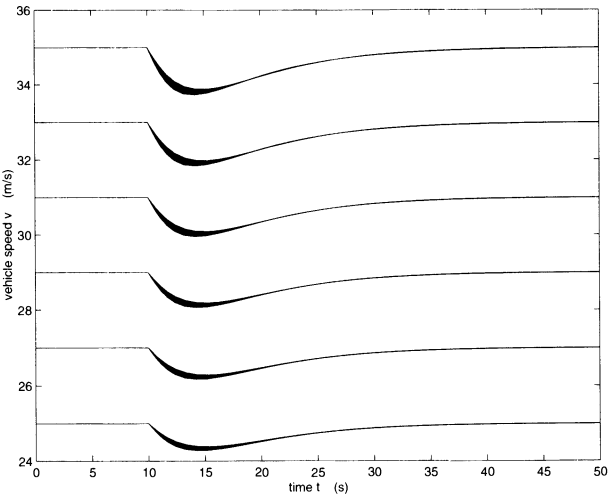


Fig. A.26. Disturbance rejection behavior for a scheduling speed $v_n^* = 30\ m/s$, parameter intervals as shown in Table A.2.

Note that these simulations use the nonlinear and non-normalized plant model. The interface between this plant and the controller that was designed for the normalized and linearized plant is shown in Fig. A.27. This structure is similar to the one encountered in the later control system realization. Some of the most important additions will be discussed in the next section.

A.6 Controller Discretization

General Remarks

Structure of a Discrete-Time Control System

Almost all control systems in passenger cars are implemented using digital computers. These electronic devices can only handle information using a *sampled-data* approach. The schematic layout of such a control system is shown in Fig. A.28.

The analog-to-digital converter (AD) samples the continuous-time signal ($e(t)$) and transforms the sampled value to a binary electronic signal ($e^*(k)$)

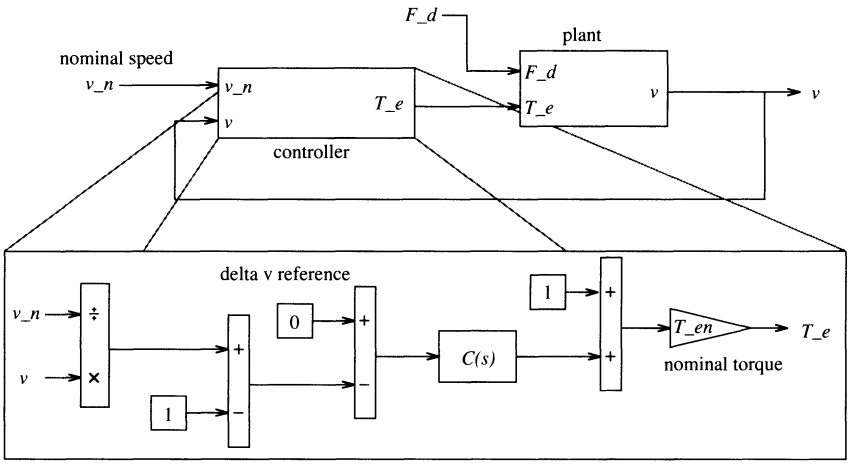


Fig. A.27. Interface structure of the closed-loop system.

that can be processed in the microcontroller. Usually, the amplitude quantization involved in that step can be neglected. Correctly handling *overflow problems*, which are due to the fixed-comma arithmetic of low-cost microprocessors, is often more important.

The discrete-time controller $C(z)$, whose design will be discussed below, computes the control signal $u^*(k)$. This signal is transformed by the digital-to-analog converter to a continuous time signal $u(t)$. The DA converter always includes a hold element. In most cases, this will be a zero-order hold (ZOH) defined by

$$u(t) = u^*(k), \quad \forall\, t \in [k \cdot \tau_s, (k+1) \cdot \tau_s) \tag{A.59}$$

where τ_s is the sampling time.

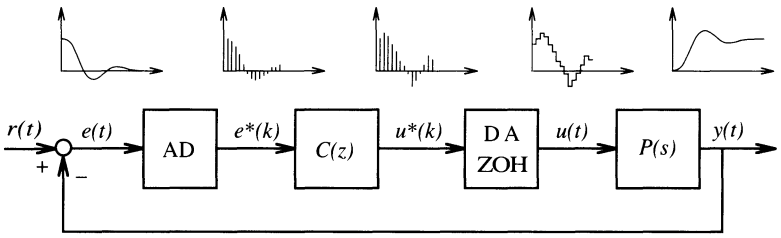


Fig. A.28. Discrete-time control system structure.

As the commutation diagram in Fig. A.29 shows, there are two approaches for the design of discrete-time control systems. The direct approach, denoted by "A" in Fig. A.29, consists of first discretizing the plant and then designing a discrete-time controller using the discrete-time plant model $P(z)$. This ap-

proach is mandatory for cases where the sampling time τ_s is large compared to the plant's time constants. The problem of discretizing a plant will be discussed below, including some basic remarks on the $\mathcal{Z}$ transformation that is the discrete-time counterpart of the Laplace transformation.

In the alternative approach, denoted by "B" in Fig. A.29, a continuous-time controller $C(s)$ is designed as a first step. This system is then transformed to an equivalent discrete-time transfer function $\widetilde{C}(z)$ that emulates, as well as possible, the closed-loop behavior of $C(s)$ in the sampled-data control loop shown in Fig. A.28. This approach makes no guarantees with respect to stability, robustness, etc. However, for relatively small sampling times τ_s, it works well in most cases.

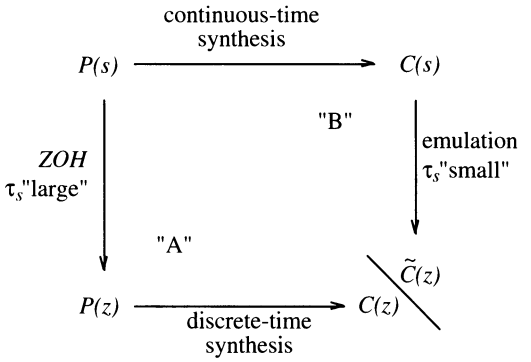


Fig. A.29. The possible design procedures to obtain a discrete-time control system for a continuous-time plant.

Plant Discretization

The starting point of the direct approach "A" is the discretization of the continuous-time plant

$$\begin{aligned} \tfrac{d}{dt}x(t) &= A \cdot x(t) + b \cdot u(t), \quad x \in \Re^n,\ u \in \Re \\ y(t) &= c \cdot x(t) + d \cdot u(t), \quad y \in \Re \end{aligned} \tag{A.60}$$

using the fact that the input $u(t)$ is piecewise constant (A.59). The result of this discretization is the discrete-time plant

$$\begin{aligned} x((k+1) \cdot \tau_s) &= F \cdot x(k \cdot \tau_s) + g \cdot u(k \cdot \tau_s) \\ y(k \cdot \tau_s) &= c \cdot x(k \cdot \tau_s) + d \cdot u(k \cdot \tau_s) \end{aligned} \tag{A.61}$$

where

$$F = e^{A \cdot \tau_s}, \quad g = \int_0^{\tau_s} e^{A\sigma}\, d\sigma\, b = A^{-1} \cdot \left[e^{A \cdot \tau_s} - I\right] \cdot b \tag{A.62}$$

The second expression for g is, of course, only applicable for plants with no poles in the origin ($\det(A) \neq 0$). The term $e^{A \cdot \tau_s}$ denotes the matrix exponential. Several reliable numerical software packages are available for its computation.

For first-order systems, scalar exponential functions are sufficient, i.e., such a first-order system can be represented in the following four forms (for the sake of brevity, the independent variable $k \cdot \tau_s$ is replaced by k):

1. continuous-time, time domain ($\{a, b, c, d\}$ real scalars)

$$\begin{aligned} \tfrac{d}{dt} x(t) &= a \cdot x(t) + b \cdot u(t) \\ y(t) &= c \cdot x(t) + d \cdot u(t) \end{aligned} \tag{A.63}$$

2. continuous-time, frequency domain

$$Y(s) = \frac{d \cdot s + [c \cdot b - a \cdot d]}{s - a} \cdot U(s) \tag{A.64}$$

3. discrete-time, time domain ($a \neq 0$)

$$\begin{aligned} x(k+1) &= e^{a \cdot \tau_s} \cdot x(k) + \tfrac{(e^{a \cdot \tau_s} - 1) \cdot b}{a} \cdot u(k) \\ y(k) &= c \cdot x(k) + d \cdot u(k) \end{aligned} \tag{A.65}$$

4. discrete-time, frequency domain

$$Y(z) = \frac{d \cdot z + [c \cdot (e^{a \cdot \tau_s} - 1) \cdot b/a - e^{a \cdot \tau_s} \cdot d]}{z - e^{a \cdot \tau_s}} \cdot U(z) \tag{A.66}$$

In the fourth representation, the $\mathcal{Z}$ transformation of the discrete-time signal $x(k)$ has been used[2]

$$X(z) = \mathcal{Z}\{x(k)\} = \sum_{k=0}^{k=\infty} x(k) \cdot z^{-k} \tag{A.67}$$

The $\mathcal{Z}$ transformation is the counterpart of the Laplace transformation and may be used to solve a difference equations using the rules

$$z \cdot X(z) - z \cdot x_0 \leftrightarrow x(k+1), \quad z^{-1} \cdot X(z) \leftrightarrow x(k-1) \tag{A.68}$$

The first relation can be used to solve the difference equation (A.61) in the frequency domain

[2] The signal $x(k)$ is assumed to be equal to zero for all $k < 0$.

$$z \cdot X(z) = F \cdot X(z) + g \cdot U(z) + z \cdot x(0) \tag{A.69}$$

$$Y(z) = c \cdot X(z) + d \cdot U(z) \tag{A.70}$$

where the variables $X(z)$ and $U(z)$ represent the $\mathcal{Z}$ transformation of their corresponding counterparts $x(k)$ and $u(k)$. Assuming a zero initial condition, i.e., $x(0) = 0$, the *transfer function*

$$P(z) = \frac{Y(z)}{U(z)} = c \cdot [z \cdot I - F]^{-1} \cdot g + d \tag{A.71}$$

can be derived. This transfer function plays the same role as its continuous-time counterpart $P(s)$, i.e., it describes how any input u is transformed to the system output y.

Once $P(z)$ is known a controller, $C(z)$ can be designed with any of the many discrete-time design methods available. The general idea behind all these methods is the same as in the continuous-time case: guarantee a desired closed-loop system performance *and* a sufficient robustness to modeling errors. The interested reader is referred to the classical texts on this subject, e.g., [54].

Note that, compared to the corresponding continuous-time formulation, some systems can be described much better using a discrete-time approach. One interesting example of such an element is a pure delay that can be represented by in the following four forms

$$y(t) = u(t - \tau_s) \tag{A.72}$$

$$Y(s) = e^{-\tau_s \cdot s} \cdot U(s) \tag{A.73}$$

$$y(k) = u(k - 1) \tag{A.74}$$

$$Y(z) = z^{-1} \cdot U(z) \tag{A.75}$$

The two frequency domain representations can be used to derive the important correspondence

$$z = e^{\tau_s \cdot s} \tag{A.76}$$

Emulation Methods

In practice, another approach is often chosen that consists in designing the controllers as shown in Sect. A.4, and to transform these systems into their discrete-time equivalents using an approximation. One of the most popular is the *Tustin transformation* defined by

$$s \approx \frac{2}{\tau_s} \cdot \frac{z - 1}{z + 1} \tag{A.77}$$

where τ_s is the sampling time. Using this relation, a continuous-time linear controller $C(s)$ can be transformed into its approximate discrete-time equivalent $C(z)$ by substituting the variable s by the variable z.

These rules allow any given controller $C(z)$ to be transformed into a series of commands which are executable on a digital computer. This is best explained with the simple example of a low-pass controller. First, the system is Tustin-transformed

$$C(s) = \frac{c}{\tau \cdot s + 1} \quad \rightarrow \quad C(z) \approx \frac{c \cdot (z+1)}{(2\frac{\tau}{\tau_s} + 1) \cdot z + (1 - 2\frac{\tau}{\tau_s})} = \frac{U(z)}{E(z)} \tag{A.78}$$

This can be written as

$$U(z) \cdot \left[(2\frac{\tau}{\tau_s} + 1) \cdot z + (1 - 2\frac{\tau}{\tau_s}) \right] = E(z) \cdot c \cdot (z+1) \tag{A.79}$$

or

$$U(z) \cdot \left[(2\frac{\tau}{\tau_s} + 1) + (1 - 2\frac{\tau}{\tau_s}) \cdot z^{-1} \right] = E(z) \cdot c \cdot (1 + z^{-1}) \tag{A.80}$$

The corresponding time-domain equation is obtained using the transformation rules (A.68)

$$u_k \cdot (2\frac{\tau}{\tau_s} + 1) + u_{k-1} \cdot (1 - 2\frac{\tau}{\tau_s}) = c \cdot (e_k + e_{k-1}) \tag{A.81}$$

Solving this equation for the unknown variable u_k, a recursive formulation of the controller is found

$$u_k = (2\frac{\tau}{\tau_s} + 1)^{-1} \left[c \cdot (e_k + e_{k-1}) - u_{k-1} \cdot (1 - 2\frac{\tau}{\tau_s}) \right] \tag{A.82}$$

This expression can then be transformed in a computer-executable program.

Notice that this approach is *not* guaranteed to yield satisfactory results. Specifically, if the sampling time τ_s is relatively large compared to the systems time constants, a badly damped or even unstable closed-loop system can result (see the example below).

Frequency Domain Representation of Discrete-Time Systems

The frequency response $P_c(j\omega)$ of a continuous-time system $P_c(s)$ has been shown to contain much valuable information on the dynamic properties of that system. The graphic representations of $P_c(j\omega)$ in Nyquist or Bode diagrams are standard tools in control systems analysis and design. It is therefore interesting to discuss what form a discrete-time frequency response will have. Unfortunately, the sampling and holding process has several unexpected effects on the frequency response such that a complete discussion is beyond the scope of this appendix. Only the most basic effects will be discussed below using intuitive but not precise arguments.

As indicated in Fig. A.30, a sampled-data control loop can always be broken at two different points yielding two representations of the same system.

Breaking the loop at the point 1 yields a *continuous-time* loop transfer function

$$L_c(s) = \frac{Y(s)}{E(s)} \quad , \tag{A.83}$$

while breaking the loop at the point 2 yields a *discrete-time* loop transfer function

$$L_d(z) = \frac{V(z)}{U(z)} \tag{A.84}$$

The transfer function $L_c(s)$ describes the behavior of the system for *each* time t, while $L_d(z)$ only gives information on the system behavior at the discrete points in time $t = k \cdot \tau_s$. Not surprisingly, it is more difficult to correctly formulate the mathematical form of $L_c(s)$ than those of $L_d(z)$.

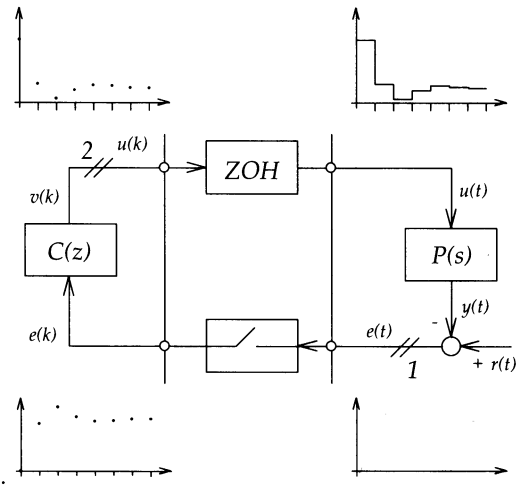


Fig. A.30. Sample-data system.

Continuous-Time Interpretation

In the continuous-time formulation, i.e., using $L_c(s)$, the sampling process is modeled with a impulse sampling (see [54]). This mechanism introduces spectral aliasing.[3] Moreover, the zero-order hold element, which can be modeled using the transfer function

$$ZOH(s) = \frac{1 - e^{-\tau_s \cdot s}}{\tau_s \cdot s} \quad , \tag{A.85}$$

[3] To avoid spurious effects caused by aliasing, sampled-data systems usually include continuous-time anti-aliasing filters at the inputs of the control system.

introduces zeros at frequencies equal to integer multiples of the sampling frequency. The Bode diagram of the frequency response $ZOH(j\omega)$ of this element is shown in Fig. A.31.

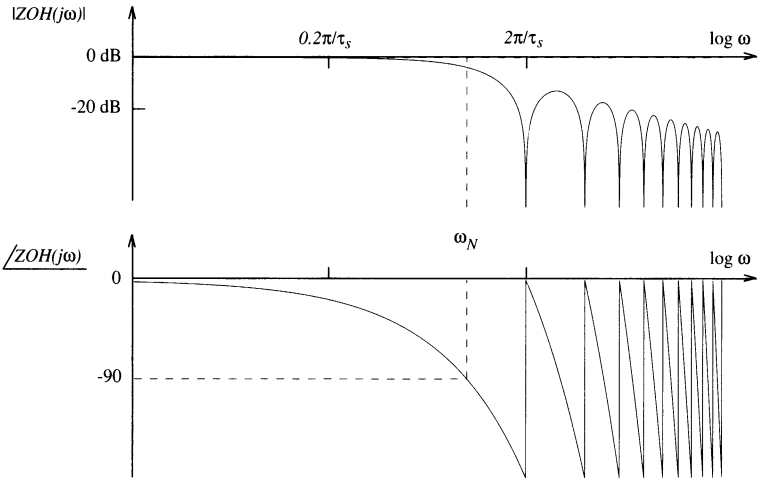


Fig. A.31. Bode diagram of the ZOH system; $\omega_N = \pi/\tau_s$ is the Nyquist frequency.

Discrete-Time Interpretation

In this setting, the discrete-time transfer function $L_d(z)$ is obtained using one of the approaches shown above. Equation (A.76) shows that the imaginary axis $j\omega$, which is used in frequency responses as an independent variable, is transformed to the unit circl $e^{j\omega}$ when changing from a continuous-time to a discrete-time system representation

$$s \rightarrow j\omega \qquad \Rightarrow \qquad z = e^{\tau_s \cdot s} \rightarrow e^{j\omega\tau_s} \tag{A.86}$$

Accordingly, the Bode and Nyquistdiagrams of a discrete-time system $L_d(z)$ are obtained by plotting $L_d(e^{j\omega\tau_s})$. For frequencies higher than the Nyquist frequency, the function $e^{j\omega\tau_s}$ is not one-to-one. Therefore, the frequency range in these plots must be limited to $\omega_N = \pi/\tau_s$, where ω_N is usually referred to as the Nyquist frequency.

Figure A.32 shows the Bode diagrams of the continuous-time system

$$L_c(s) = \frac{1}{\tau \cdot s + 1} \tag{A.87}$$

and of the two discrete-time systems obtained for a sampling time of τ_s

$$L_{d,ZOH}(z) = \frac{1 - e^{-\tau_s/\tau}}{z - e^{-\tau_s/\tau}} \tag{A.88}$$

$$L_{c,Tustin}(z) = \frac{z+1}{(2 \cdot \frac{\tau}{\tau_s} + 1) \cdot z + (1 - 2 \cdot \frac{\tau}{\tau_s})} \tag{A.89}$$

with the ZOH direct method and the Tustin emulation approach.

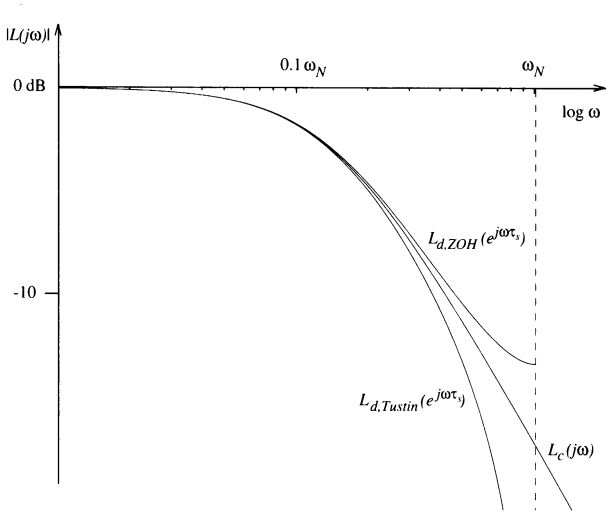


Fig. A.32. Bode diagram of the example (A.87); $\tau = 7\ s$.

Example Cruise-Control Problem

In a series-production vehicle, the PI controller designed in Sect. A.5 will be realized as a discrete-time control system. Accordingly, the continuos-time controller

$$C(s) = k_p \cdot \left(1 + \frac{1}{\tau_i \cdot s}\right) \tag{A.90}$$

must be transformed into an equivalent discrete-time controller $C(z)$ using either an emulation method or designing it based on the discretized plant model $P(z)$. Here, the first approach is chosen.

This controller $C(z)$ will be inserted in a structure similar to the one shown in Fig. A.27, with the lower component replaced by the system shown in Fig. A.33.

In this example, the discrete-time controller $C(z)$ is found using a Tustin emulation (A.77) of the continuous-time controller $C(s)$ (A.90)

$$C(z) = k_p \cdot \frac{z \cdot (\frac{\tau_s}{2\tau_i} + 1) + (\frac{\tau_s}{2\tau_i} - 1)}{z - 1} \tag{A.91}$$

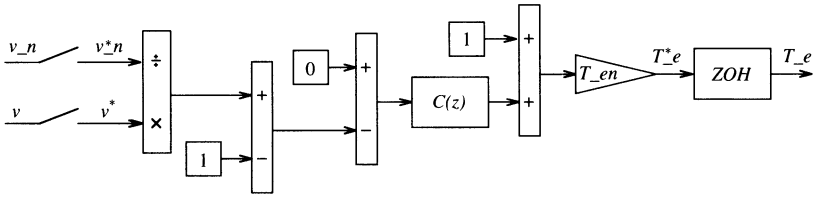


Fig. A.33. Discrete-time control system that replaces the controller in Fig. A.27.

The results shown in Fig. A.34 are obtained by inserting the numerical values for the parameters $\{k_p, \tau_i\}$ of the PI controller found in Sect. A.4 and simulating the closed-loop system for different sampling times τ_s.

The response obtained with a very fast sampling $\tau_s = 0.1\ s$ is almost the same as the one obtained with the original continuous-time controller. Increasing the sampling time to $\tau_s = 1, \ldots, 4\ s$ yields a different, but still acceptable, closed-loop system response.

For sampling times $\tau_s \approx 4.7\ s$, the closed-loop system starts to oscillate and eventually reaches its stability limit at $\tau_s \approx 5.63$.

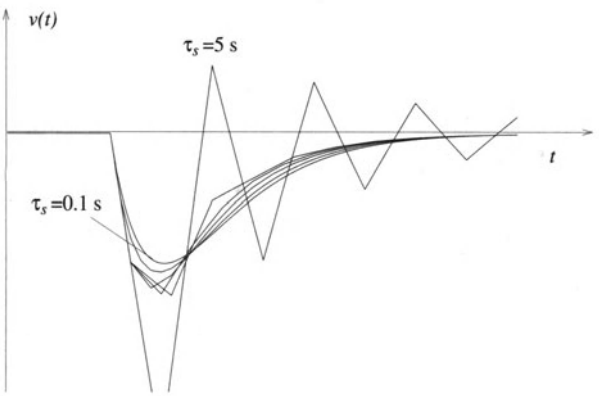


Fig. A.34. Disturbance rejection behavior of the closed-loop cruise-control system. Controller (A.91) realized with sampling times $\tau_s = \{0.1, 1, 2, 3, 4, 5\}\ s$.

A.7 Controller Realization

Control system realization is, of course, a vast subject that includes many problem areas. This section briefly discusses only two important theoretic aspects of controller realization, namely:

- controller gain scheduling, i.e., the adaptation of the controller gains to changing operating conditions; and
- controller anti reset windup, i.e., the potential problems that arise when plant input saturation and integral controller action are combined

A.7.1 Gain Scheduling

Gain scheduling is a straightforward method to adapt the controller parameters to changing operating conditions. The key idea is to parameterize the controller with respect to a *slowly* changing operating point around which the plant is linearized and normalized. This scheduling point is then assumed to be representative for a certain set of operating points in its neighborhood. The compromise to be made is between a finer "gridding" (large design and implementation efforts, but better controller performance) and a coarser "gridding" (less design efforts, but more conservative and therefore slower controllers).

Figure A.35 shows a block representation of such a controller gain scheduling scheme. The controller parameters (gains, time constants, etc.) are denoted as p and the scheduling variable as o. The block "linearization and normalization" is the interface between the nonlinear and the linear components, as shown in the inset in Fig. A.27.

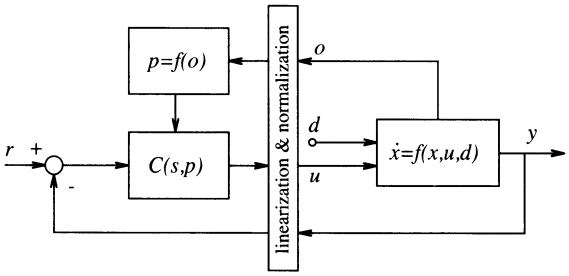


Fig. A.35. Gain scheduling controller structure.

In the cruise-control example, the scheduling variable is the desired vehicle speed (gear ratio and mass have been shown to have less influence on plant behavior). This variable changes only when the driver changes the reference values, but remains constant during a finite interval.

A major challenge in designing a gain scheduling system is the bumpless and smooth transition between one controller setting and the next. Interpolation of controller values, ramped transitions and several other "tricks" have to be used here. Also, some type of hysteresis must be implemented in order to avoid a limit-cycle behavior at the boundaries between two scheduling zones.

Note that there is no simple way to guarantee that a gain scheduling system is asymptotically stable, even if this is true for all of the closed loops with a

fixed controller. It is therefore important to monitor the system behavior in order to detect any instabilities induced by the scheduling process.

A.7.2 Anti Reset Windup

A very important nonlinearity, which in most cases is present at the plant's input, is a control input saturation. For instance, in the cruise-control problem, the engine torque will always be limited by $T_{e,min}(\omega_e) \leq T_e \leq T_{e,max}(\omega_e)$. This saturation can cause problems if the controller has any components which are not asymptotically stable. Note that this is always the case if the controller includes integral action (which is stable, but not asymptotically stable). This combination can lead to an unexpected overshoot for large disturbance or set-point changes, even when the linear (small-signal) behavior is perfectly damped.

In order to suppress such phenomena (often called *reset windup*), a special controller structure must be implemented. Figure A.36 shows a possible structure or discrete-time systems (similar structures exist for continuous-time controllers).

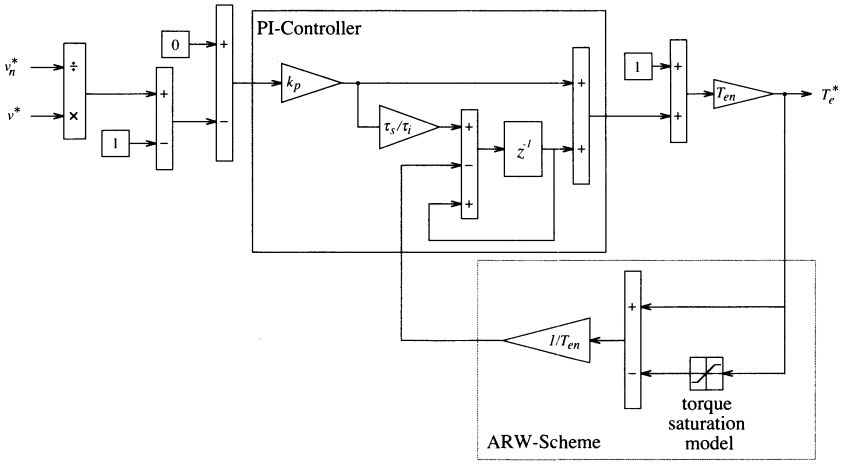


Fig. A.36. Discrete-time PI controller with anti reset windup circuit.

A.8 Further Reading

A good first bookfor those who wish to enter the field of control systems analysis and design is [53]. An extension of that first text that contains a classical treatment of digital (discrete-time) control systems is the following [54]. The text [44] extends the classic concepts to the more modern robust

control perspective. A good discussion of the design of PID controllers is found in [12]. The subject of Anti Reset Windup is discussed in detail in [59]. A good overview on gain-scheduling techniques can be found in [153].

B

Appendix B

The control loop discussed in this case study is the well-known idle-speed control system for which a rich literature is available (see [136] for an early work or [82] for an overview paper). The objective of this appendix is not to show all steps of the design and realization of such a control system. For instance the control problem here will be formulated as an SISO problem (see [128] and Sect. 4.2.2 for a true MIMO formulation) and only feedback action will be investigated (see [29] for a design that includes both feedback and feedforward parts). Instead the focus will be on showing how some of the basic building blocks introduced in the main text can be combined to form a mathematical model and on how the parameters of this model can be identified using measurements. Then, using this mathematical model, a simplified control system is designed and tested using the methods introduced in Appendix A.

To be more specific, the following four subproblems will be discussed in this case study:

1. *Synthesis of a mathematical model of the dynamic behavior of the relevant engine parts using some of the building blocks introduced in Chapter 2.*
2. *Parameter identification and model validation using data obtained on a dynamic test bench with a 2.8-liter SI engine.*
3. *Model linearization and delay approximation to obtain a finite-dimensional linear system description.*
4. *Synthesis of a simplified idle speed control system (air path actuation only) using the mathematical model derived in step 3. Implementation of this controller on a digital signal processor and test of the closed-loop system on an engine test bench.*

B.1 Modeling of the Idle Speed System

B.1.1 Introduction

Modern SI engines have rather low idle speed limits (around 700 rpm) in order to minimize fuel consumption and pollutant emission. The drawback of such low limits is that sudden changes in engine torque (electric loads on the alternator, AC compressor, etc.) may stall the engine. A fast idle speed control system (ISCS) is therefore mandatory.

The basic structure of a modern ISCS for SI engines is shown in Fig. B.1. The engine speed is the main input signal, other engine variables (oil temperature, intake air pressure and temperature, etc.) are used to compute the reference speed at which the engine should run.

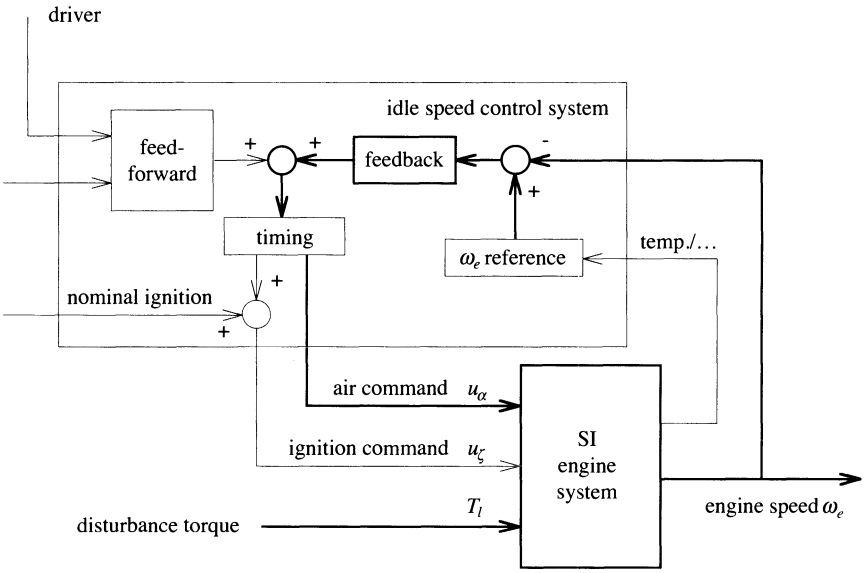


Fig. B.1. Complete ISCS structure; parts discussed in this case study are drawn with thick black lines.

As Fig. B.1 shows, the disturbance torque is compensated in a coordinated way using both the air command and the spark command. This approach permits a very fast reaction to unmeasurable load disturbances. In fact, as shown in Sect. 3.2.1, the ignition input is able to change the engine torque almost instantaneously albeit at the price of a reduced engine efficiency and higher pollutant emission. Therefore the "timing" block in the ISCS utilizes the spark channel in a first phase in which the air command has not yet produced the desired torque change. As soon as that command starts to produce the desired action, the ignition command is phased back to its nominal value.

In conventional engine systems with mechanical throttle valves the air command actuates a parallel idle speed control bypass valve. In engine systems with electronically controlled throttle valves the idle speed controller usually directly commands that element. Fig. 4.1 shows the layout of such a modern engine control system.

The design of a complete ISCS is well beyond the scope of this case study. The following simplifications are therefore made:

- The engine speed reference value is kept constant (warmed engine, etc.).
- Only the air command path is considered, i.e., only those parts in Fig. B.1 that are drawn with thick lines.
- The injection control system is assumed to be able to keep the air/fuel ratio at its stoichiometric value of $\lambda = 1$.
- The ignition angle is not changed from its value as defined by the nominal ignition map (see Fig. 4.6). It turns out that this is a rather critical simplification for the performance of the system.
- The disturbance torque is assumed not to be measurable, no feedforward control is therefore possible and only the feedback part will be considered.

The main signal measured will be the engine speed. Four other signals (intake manifold pressure, temperature, air mass flow and load torque) are available for the parameter identification and model validation parts.

The engine used in this case study is a 2.8-liter V6 SI engine with conventional port injection. The ignition and the injection are actuated using the basic maps of the ECU. Only the air loop will be redesigned and implemented in a digital signal processor real-time system. The output of that control system is the throttle valve position command signal u_α.

B.1.2 System Structure

The system to be modeled is shown in Fig. B.2. It is a simplified version of the general SI engine system as introduced in Sect. 2.2.1. According to the assumptions listed in the previous section, the fueling and the ignition parts may be omitted. The measured variables and the two input signals are explicitly shown in Fig. B.2.

As introduced in Chapter 2, the engine is modeled as a mean-value volumetric pump with the intake manifold and the intake air throttle upstream of the intake manifold. The torque produced by the engine is linked to the amount of air aspirated by the engine and acts positively on the constant engine inertia. The engine is assumed to be decoupled from the rest of the power train (this is the worst-case situation because then the inertia is minimal) and the load torque is presumed to act on the engine flywheel directly.

One problem is the generation of known load disturbances for test purposes. The dynamometer cannot be used since its inertia would change the dynamic behavior of the system considerably. A viable alternative is to utilize

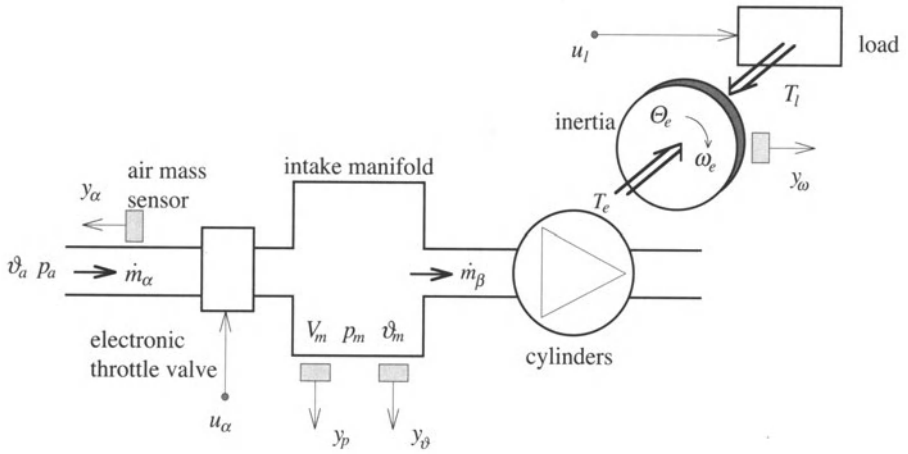


Fig. B.2. Mean-value model structure of the simplified ISCS engine system.

the engine's alternator as a variable load source by connecting it to an electronically controllable load. In this case a 0 to 1 kW load is used, which yields a load disturbance of up to 15 Nm (more on that will be said in Sect. B.2). The electronic load and alternator subsystem has its own dynamics that will have to be modeled.

The corresponding cause and effect diagram of the combined ISCS is shown in Fig. B.3. As in the main text, while the shaded boxes indicate dynamic modules, the non-shaded boxes represent algebraic modules. The control input is u_α, i.e., the command signal for the electronic throttle valve. The disturbance input u_l, which in the series application is assumed not to be measurable, is the only exogenous disturbance signal. In order to test the performance of the ISCS,it is to be designed in this case study.

The measured system variables are:

- y_α: the air mass flowing through the throttle into the intake manifold;
- y_ω: the engine speed;
- y_p: the intake manifold pressure;
- y_ϑ: the intake manifold temperature;
- y_U: the voltage at the alternator; and
- y_I: the current flowing through the alternator.

The intake manifold temperature, $y_\vartheta(t)$, is measured. Its dynamics thus need not be modeled.

B.1.3 Description of Subsystems

Most of the subsystems indicated by boxes in Fig. B.3 have been discussed in Chapter 2. Here those results are adapted to the ISCS problem and a

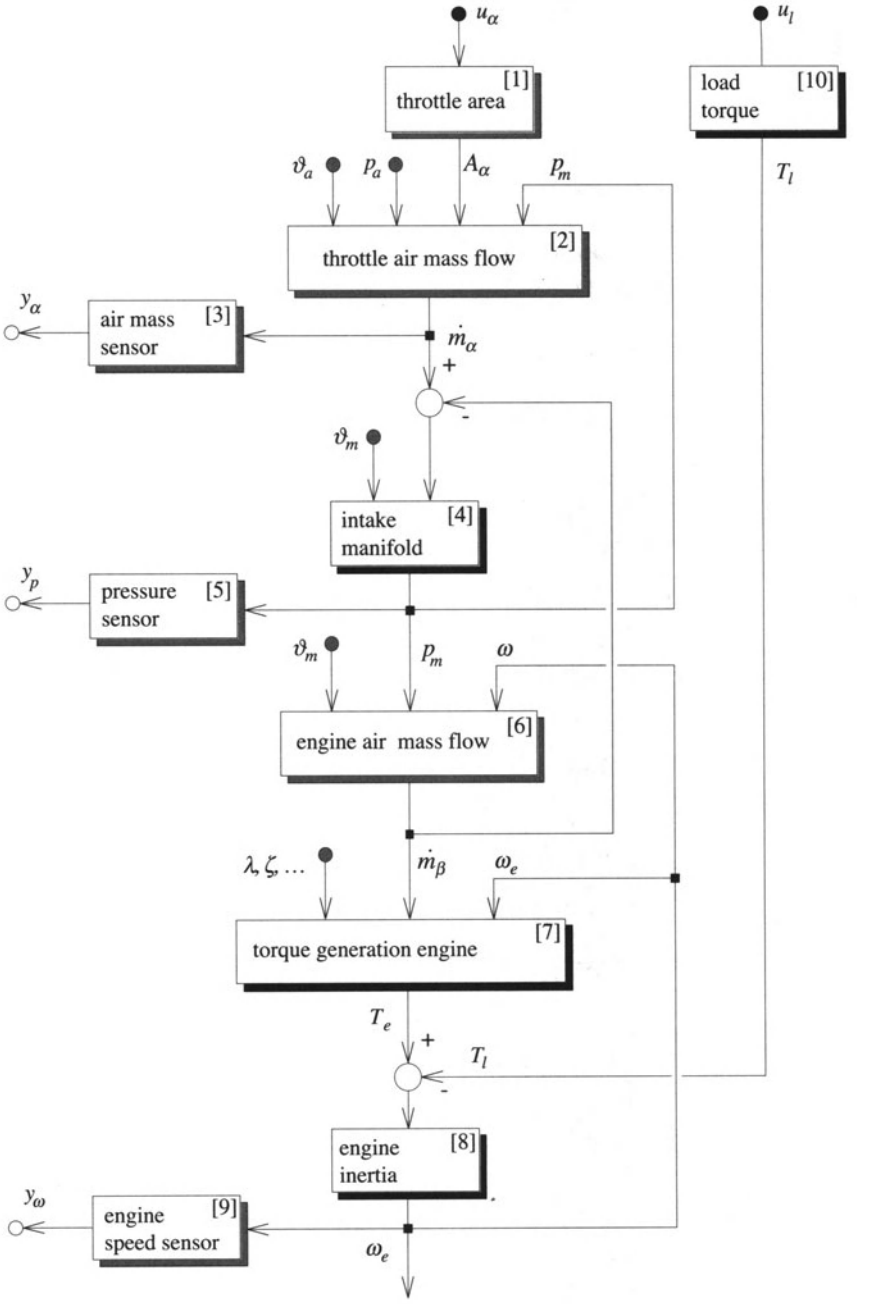


Fig. B.3. Cause and effect diagram of the simplified ISCS engine system.

formal description of each block is given to facilitate the subsequent parameter identification and control system synthesis steps.

Block [1]: *Throttle Area*

Purpose: Models the open area as a function of the input signal $u_\alpha(t)$

Signals: Input: Control signal $u_\alpha(t)[0 \ldots 1]$
Output: Open area $A_\alpha[m^2]$

Assumptions: The dynamics of the actuator may be neglected.

Equations: $\alpha_{th} = \alpha_{th,0} + (\frac{\pi}{2} - \alpha_{th,0}) \cdot u_\alpha$

$A_\alpha(\alpha_{th}) = \frac{\pi d_{th}^2}{4} \cdot \left(1 - \frac{\cos \alpha_{th}}{\cos \alpha_{th,0}}\right) + A_{th,leak}$

where $\alpha_{th,0}$: angle when throttle fully closed,
where $A_{th,leak}$: throttle opening area when $\alpha_{th} = \alpha_{th,0}$

Block [2]: *Throttle Mass Flow*

Purpose: Models the air mass flow through throttle area and pressure drop.

Signals: Input: Open area A_α $[m^2]$
Input: Pressure upstream, ambient pressure p_a $[Pa]$
Input: Temperature upstream ϑ_a $[K]$
Input: Pressure downstream, manifold pressure $p_m(t)$ $[Pa]$
Output: Air mass flow $\dot{m}_\alpha(t)$ $[kg/s]$

Assumptions: Air a perfect gas, throttle isenthalpic (see Section 2.3.2).

Equations:

$$\dot{m}_\alpha(t) = \begin{cases} A_\alpha(t) \frac{p_a}{\sqrt{R\vartheta_a}} \frac{1}{\sqrt{2}} & \text{if } \frac{p_m(t)}{p_a} < 0.5 \\ A_\alpha(t) \frac{p_a}{\sqrt{R\vartheta_a}} \sqrt{2 \frac{p_m(t)}{p_a} \left[1 - \frac{p_m(t)}{p_a}\right]} & \text{else} \end{cases}$$

Block [3]:	*Air Mass Sensor*
Purpose:	Produces signal proportional to air mass flow
Signals:	Input: Inflowing air mass $\dot{m}_\alpha(t)$ $[kg/s]$ Output: Sensor signal $y_\alpha(t)$ $[V]$
Assumptions:	Sensor static linear system with known characteristics.
Equations:	$y_\alpha(t) = k_a \cdot \dot{m}_\alpha(t)$

Block [4]:	*Intake Manifold*
Purpose:	Describes emptying and filling dynamics of the manifold
Signals:	Input: Difference of mass flows $\dot{m}_\alpha(t) - \dot{m}_\beta(t)[kg/s]$ Input: Air temperature in the intake manifold ϑ_m $[K]$ Output: Manifold air pressure $p_m(t)$ $[Pa]$
Assumptions:	Lumped-parameter model, isothermal conditions.
Equations:	$\frac{d}{dt}p_m(t) = \frac{R \cdot \vartheta_m}{V_m} \cdot [\dot{m}_\alpha(t) - \dot{m}_\beta(t)]$

Block [5]:	*Pressure Sensor*
Purpose:	Models the behavior of the intake manifold pressure sensor.
Signals:	Input: Air pressure $p_m(t)$ $[Pa]$ Output: Sensor output signal $y_p(t)$ $[V]$
Assumptions:	Sensor is a static linear system
Equations:	$y_p(t) = k_p \cdot p_m(t)$

Block [6]: *Engine Air Mass Flow*

Purpose: Models the mass flow the engine inducts which is equal to the air mass flow that exits the intake manifold.

Signals: Input: Manifold pressure $p_m(t)$ $[Pa]$
Input: Engine speed ω_e $[rad/s]$
Input: Air temperature in manifold ϑ_m $[K]$
Input: Exhaust manifold pressure p_e $[Pa]$
Output: Air mass flow exiting manifold $\dot{m}_\beta(t)$ $[kg/s]$

Assumptions: No pressure drop in the intake runners.
The temperature drop from the evaporation of the fuel is canceled out by the heating effect of the hot intake duct walls. The temperature in the cylinder is thus the same as in the intake manifold.
The air/fuel ratio is constant and equal to stoichiometry.
Volumetric efficiency depends on manifold pressure as shown in Section 2.3.3.
The exhaust manifold pressure is constant.

Equations: Mixture entering cylinders:
$\dot{m}_e(t) = \frac{p_m(t)}{R\vartheta_m(t)} \cdot \lambda_l(\omega_e(t), p_m(t)) \cdot V_d \cdot \frac{\omega_e(t)}{4\pi}$
Air exiting intake manifold:
$\dot{m}_\beta(t) = \frac{\dot{m}_e(t)}{1+1/(\lambda\,\sigma_0)}$ with the stoichiometric air/fuel ratio $\sigma_0 \approx 14.7$
Volumetric efficiency:
$\lambda_l(\omega_e, p_m) = \lambda_{l\omega}(\omega_e) \cdot \eta_{lp}(p_m)$
$\lambda_{l\omega}(\omega_e) = \gamma_0 + \gamma_1 \cdot \omega_e + \gamma_2 \cdot \omega_e^2$
$\lambda_{lp}(p_m) = \frac{V_c+V_d}{V_d} - \frac{V_c}{V_d} \cdot \left(\frac{p_e}{p_m(t)}\right)^{1/\kappa}$

Block [7]: *Torque Generation Engine*

Purpose: Estimation of torque produced by the engine.

Signals: Input: Air mass flow entering cylinders $\dot{m}_\beta(t)$ $[kg/s]$
Output: Engine torque $T_e(t)$ $[Nm]$

Assumptions: The engine may be simplified as a Willan's machine, see (2.105).
The delay $\tau_{inj\to TC}$ is calculated according to the definition shown below.
As the engine system is equipped with a fast, research- oriented control unit, the calculation time $\tau_{calculation}$ and the sampling time t_{sample} can be neglected.

Equations:

$$
\begin{aligned}
T_e(t) &= p_{me}\frac{V_d}{4\pi} \\
&= (e \cdot p_{m\varphi} - p_{me0f} - p_{me0g}) \cdot \frac{V_d}{4\pi} \\
e \cdot p_{m\varphi} &= (\eta_0 + \eta_1 \cdot \omega_e) \cdot \frac{H_l \cdot m_\varphi}{V_d} \\
&= (\eta_0 + \eta_1 \cdot \omega_e) \cdot H_l \cdot \frac{\dot{m}_\varphi \cdot 4\pi}{\omega_e \cdot V_d} \\
&= (\eta_0 + \eta_1 \cdot \omega_e) \cdot H_l \cdot \frac{\dot{m}_\beta(t - \tau_{inj\to TC}) \cdot 4\pi}{\lambda \cdot \sigma_0 \cdot \omega_e(t - \tau_{inj\to TC}) \cdot V_d} \\
p_{me0f} &= (\beta_0 + \beta_2 \cdot \omega_e^2) \cdot \frac{4\pi}{V_d} \\
p_{me0g} &= p_a - p_m \\
\tau_{inj\to TC} &\approx \tau_{Uinj\to TC} \\
&= \tau_{inj} + \tau_{IVO\to IC} + \tau_{IC\to TC}
\end{aligned}
$$

Remark: As the delay occurs within the feedback loop, it has a strong influence on the dynamic behavior of the idle speed control system.

Block [8]:	*Engine Inertia*
Purpose:	Models engine speed dynamics.
Signals:	Input: Engine torque $T_e(t)$ $[Nm]$ Input: Load torque $T_l(t)$ $[Nm]$ Output: Engine speed $\omega_e(t)$ $[rad/s]$
Assumptions:	The engine has a constant inertia and all internal friction is included in T_e.
Equations:	$\frac{d}{dt}\omega_e(t) = \frac{1}{\Theta_e}\left[T_e(t) - T_l(t)\right]$
Block [9]:	*Engine Speed Sensor*
Purpose:	Models the behavior of the engine speed sensor.
Signals:	Input: Engine speed $\omega_e(t)$ $[rad/s]$ Output: Sensor signal $y_\omega(t)$ $[V]$
Assumptions:	The sensor is linear and very fast.
Equations:	$y_\omega(t) = k_\omega \cdot \omega_e(t)$
Block [10]:	*Load Torque*
Purpose:	Models the dynamic behavior of the load torque unit.
Signals:	Input: Desired load torque $u_l(t)$ $[V]$ Output: Actual load torque $T_l(t)$ $[Nm]$
Assumptions:	The alternator and electronic load dynamics may be lumped into one first-order system.
Equations:	$\frac{d}{dt}T_l(t) = \frac{1}{\tau_l} \cdot \left[-T_l(t) + k_l \cdot u_l(t)\right]$

All of these blocks form a mathematical description that approximates the dynamic behavior of the idle speed system. The engine speed is limited to the range of interest (say 70 to 120 rad/s). With the exception of the equations of block [7], all system descriptions are ordinary differential equations that could be combined into one large vector-valued ODE. However, this is not recommended for such complex systems. Here the alternative approach of encoding these equations using a CACSD software tool will be chosen. This representation is then used for all subsequent analysis and synthesis steps.

B.2 Parameter Identification and Model Validation

These two steps are closely linked. Some measurements are utilized to estimate the as yet unknown parameters of the model and other measurements are to assess the quality of the model in predicting the system behavior. In this section the static and the dynamic behavior of the plant will be analyzed in two different steps. This helps to cope with the large number of unknown parameters.

B.2.1 Static Behavior

The plant has two inputs (u_α and u_l) and three[1] outputs (y_α, y_p, and y_ω). The static relations between these quantities are investigated first.

The engine operating conditions are chosen such that the corresponding input signals represent the widest possible range of meaningful values.

For the identification of the model, the load torque has to be measured. This is not a trivial task, especially because a dynamometer cannot be used since it would increase the rotational inertia. Instead, the alternator is used to produce the desired load torque. This variable is inferred from the easily measurable electric variables current and voltage.

$$T_{alt} = \frac{P_{alt}}{\omega_e} = \frac{U_{alt}\, I_{alt}}{\omega_e\, \eta_{alt}} = T_l \tag{B.1}$$

The electric voltage U_{alt}, the current I_{alt} and the speed ω_e can be measured. The efficiency η_{alt} is estimated using the efficiency map shown in Fig. B.4. As this figure shows, an assumed average efficiency of 0.6 will produce errors of less than 10% for all operating points of interest here (low speeds, zero to full torque).

The steady-state behavior of the plant is defined by setting all time derivatives to be equal to zero, i.e., the following two non-trivial conditions have to be satisfied

$$\dot{m}_\alpha(t) = \dot{m}_\beta(t) \tag{B.2}$$

$$T_e(t) = T_l(t) \tag{B.3}$$

Some of the system parameters introduced in the last section are not relevant for the steady-state behavior of the system and cannot be derived using the last equations. Moreover, a closed-form solution of the coupled nonlinear static equations following from (B.2) and (B.3) is not possible such that numerical methods will have to be applied.

The set of unknown parameters influencing the steady-state behavior is

$$\begin{aligned}\hat{\Pi}_s = \{&\alpha_{th,0}, A_{th,leak}, p_a, \vartheta_a, R, \vartheta_m, k_p, V_d, V_c,\\ &\gamma_0, \gamma_1, \gamma_2, \eta_0, \eta_1, \beta_0, \beta_2, k_\omega, k_l\}\end{aligned} \tag{B.4}$$

[1] Since the temperature ϑ_m is measured, no model is needed to predict the behavior of this variable.

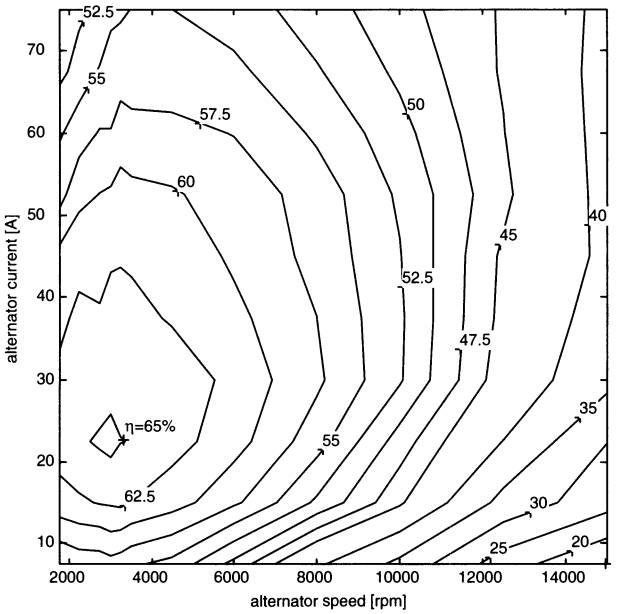


Fig. B.4. Alternator efficiency map. Alternator speed is twice the engine speed.

Some of these parameters can be determined easily (from direct measurements, physical handbooks or manufacturer data sheets) such that only a subset will have to be estimated using the measured data. In principle, this could be done using nonlinear least-squares methods with the error to be minimized defined by

$$E = \sum_{i=1}^{N} \left[k_1 \left(p_{m,i} - \tilde{p}_{m,i} \right)^2 + k_2 \left(\dot{m}_{\alpha,i} - \tilde{\dot{m}}_{\alpha,i} \right)^2 + k_3 \left(\omega_{e,i} - \tilde{\omega}_{e,i} \right)^2 \right] \quad \text{(B.5)}$$

Here the variables $\{p_{m,i}, \dot{m}_{\alpha,i}, \omega_{e,i}\}$ are the *measured* and $\{\tilde{p}_{m,i}, \tilde{\dot{m}}_{\alpha,i}, \tilde{\omega}_{e,i}\}$ are the *predicted* system outputs, whereas $i = 1, \ldots, N$ are the individual measurements taken. The constants k_i are weighting factors that are chosen to normalize the three different parts in the sum.

However, such a direct approach is not very likely to succeed, especially in the noisy environment around SI engines. Usually a modular approach is more promising, i.e., the first subsystems in the causality chain are identified first using intermediate system outputs. Moreover, whenever possible the problems have to be reformulated in order to obtain equations that are linear in the parameters (lip) in which case a linear LS estimation becomes possible. This approach is used below starting with the throttle valve as a first element.

Identification of the Throttle Behavior

Since the manifold pressure p_m is smaller than 0.5 bar in all operating conditions relevant for the idle-speed control problem, this simplifies the identification problem because in this case the nonlinear function Ψ is constant.

Second, as the diameter of the throttle d_{th} enters in a nonlinear way in the model equations, the following substitution has to be used in order to formulate a lip problem

$$\delta = \frac{\pi \cdot d_{th}^2}{4} \tag{B.6}$$

The following definitions simplify the subsequent derivations

$$\tilde{y}_\alpha = \begin{bmatrix} \dot{\tilde{m}}_{\alpha,1} \\ \dot{\tilde{m}}_{\alpha,2} \\ \cdots \\ \dot{\tilde{m}}_{\alpha,N} \end{bmatrix} \in \Re^{N\times 1},\ M_\alpha = k_\alpha \cdot \begin{bmatrix} 1 & 1 - \frac{\cos\alpha_{th}}{\cos\alpha_{th0,1}} \\ 1 & 1 - \frac{\cos\alpha_{th}}{\cos\alpha_{th0,2}} \\ \cdots & \cdots \\ 1 & 1 - \frac{\cos\alpha_{th}}{\cos\alpha_{th0,N}} \end{bmatrix} \in \Re^{N\times 2} \tag{B.7}$$

where $k_\alpha = \frac{p_a}{\sqrt{R\vartheta_a}} \frac{1}{\sqrt{2}}$.

The system behavior as expressed in the equations of block [1] and [2] can now be compactly written as follows

$$\tilde{y}_\alpha = M_\alpha \cdot [A_{th,leak}, \delta]^T \tag{B.8}$$

The solution $[A_{th,leak}, \delta]^T$ of this equation which minimizes the error

$$E_{LS} = \sum_{i=1}^{N} \left(\dot{m}_{\alpha,i} - \dot{\tilde{m}}_{\alpha,i}\right)^2 \tag{B.9}$$

between the measured air mass flow $\dot{m}_{\alpha,i}$ and the predicted air mass flow $\dot{\tilde{m}}_{\alpha,i}$ will then be

$$[A_{th,leak}, \delta]^T = \left[M_\alpha^T \cdot M_\alpha\right]^{-1} \cdot M_\alpha^T \cdot y_\alpha \tag{B.10}$$

where y_α is the measurement vector that collects all measured $\dot{m}_{\alpha,i}$.

Once δ is identified with (B.10), the substitution (B.6) can be inverted and the diameter of the throttle thus becomes

$$d_{th} = 2 \cdot \sqrt{\frac{\delta}{\pi}} \tag{B.11}$$

The results of this identification can be found in Fig. B.5.

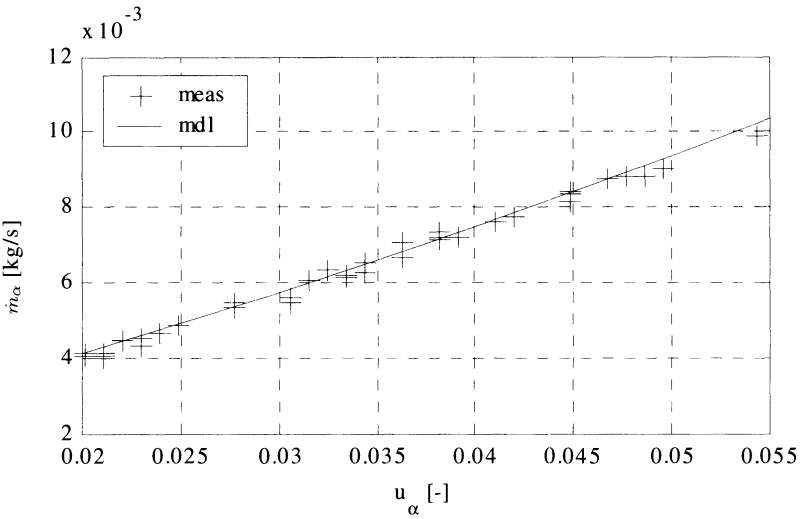


Fig. B.5. Comparison of measured and modeled mass flow $\dot{m}_\alpha$ after the identification of the throttle behavior.

Identification of the Volumetric Efficiency

The coefficients γ_i of the speed-dependent part of the volumetric efficiency are found using the same approach (notice that in steady-state operation the equation $\dot{m}_{\beta,i} = \dot{m}_{\alpha,i}$ is fulfilled). The starting point is given by the equations of block [6]. If the air mass flow $\dot{m}_{\alpha,i}$, the intake manifold pressure $p_{m,i}$, and the engine speed $\omega_{e,i}$ are known (i is the counter that indicates the i-th measurement), the volumetric efficiency can be written as

$$\gamma_0 + \gamma_1 \cdot \omega_{e,i} + \gamma_2 \cdot \omega_{e,i}^2 = \psi(\omega_{e,i}, p_{m,i}, \dot{m}_{\alpha,i}) \tag{B.12}$$

where

$$\psi(\omega_{e,i}, p_{m,i}, \dot{m}_{\alpha,i}) = \\ \frac{R\vartheta_m}{p_{m,i}} \cdot \left(\frac{V_c+V_d}{V_d} - \frac{V_c}{V_d} \left(\frac{p_a}{p_{m,i}} \right)^{1/\kappa} \right)^{-1} \cdot \dot{m}_{\alpha,i} \cdot \left(\frac{1+\sigma_0}{\sigma_0} \right) \cdot \frac{4\pi}{\omega_{e,i} V_d} \tag{B.13}$$

Analogously to the definitions introduced above for the throttle parts, the following new variables are defined here

$$\tilde{y}_\eta = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \dots \\ \psi_N \end{bmatrix} \in \Re^{N\times 1}, \quad M_\eta = \begin{bmatrix} 1 & \omega_{e,1} & \omega_{e,1}^2 \\ 1 & \omega_{e,2} & \omega_{e,2}^2 \\ & \dots & \dots \\ 1 & \omega_{e,N} & \omega_{e,N}^2 \end{bmatrix} \in \Re^{N\times 3} \tag{B.14}$$

The solution of this equation which minimizes the prediction error in the LS sense is

$$\left[\gamma_0\ ,\ \gamma_1\ ,\ \gamma_2\right]^T = \left[M_\eta^T \cdot M_\eta\right]^{-1} \cdot M_\eta^T \cdot \tilde{y}_\eta \tag{B.15}$$

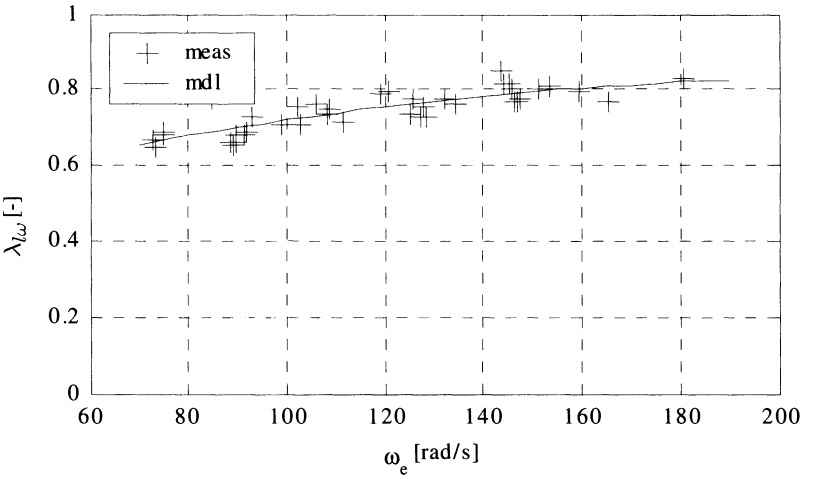


Fig. B.6. Comparison of the measured and the modeled speed-dependent factor $\lambda_{l\omega_e}$ of the volumetric efficiency λ_l.

Identification of the Torque Generation Module

The parameters of this subsystem can be identified using the outputs of the previously identified subsystems as inputs. However, this approach yields less precise results than the one used below. In this alternative approach the throttle valve command u_α is used as input and the following error criterion is minimized

$$E = \sum_{i=1}^{N} \left[\left(\omega_{e,i} - \tilde{\omega}_{e,i}\right)^2\right] \tag{B.16}$$

The variable $\omega_{e,i}$ is the measured and the variable $\tilde{\omega}_{e,i}$ is the engine speed predicted by the model. The index $i = 1, \ldots, N$ indicates the individual measurements.

B.2.2 Dynamic Behavior

Several parameters cannot be determined using steady-state measurements. This set of parameters is

$$\Pi_d = \{V_m, \Theta_e, \tau_{inj \to TC}, \tau_l\} \tag{B.17}$$

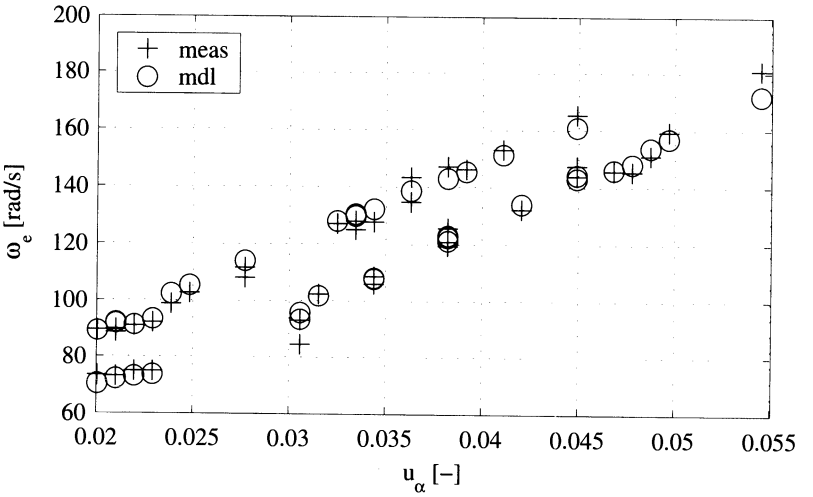


Fig. B.7. Comparison of measured and modeled correlation between the throttle signal (u_α) and the resulting engine speed (ω_e).

The time constant of the load can be estimated very well by separate experiments or theoretical considerations. The remaining unknown "dynamic" parameters are therefore the intake manifold volume V_m, the engine inertia Θ_e and the transport delay from injection to torque center $\tau_{inj \to TC}$. These parameters are identified using measured frequency responses. For both of the following experiments, the input signal is the throttle valve command u_α. In a first experiment the engine is connected to a dynamometer and the engine speed is kept constant. The measured output signal is the manifold pressure p_m.

The only parameter that has an influence on the transfer function $u_\alpha \rightarrow p_m$ is the volume of the manifold V_m. This parameter is adapted until the measured and the predicted frequency responses match well. The result shown in Fig. B.8 indicate a good agreement up to a frequency of 10 rad/s.

The delay from injection to torque center ($\tau_{inj \to TC}$) is identified according to (3.5). Figure B.9 shows the phase plotted linearly over the excitation frequency. After higher frequencies, where the influence of the finite-dimensional dynamic subsystems are not relevant, the phase rolls off linearly. The gradient of this line is proportional to the unknown delay.

A second frequency analysis is carried out with the engine decoupled from the dynamometer. This allows to measure the transfer function from the throttle angle to the engine speed.

With the identified transport delay at hand, the phase in the Bode plot can be corrected for that effect. Figure B.10 shows the corresponding transfer function from the throttle angle to the speed of the engine.

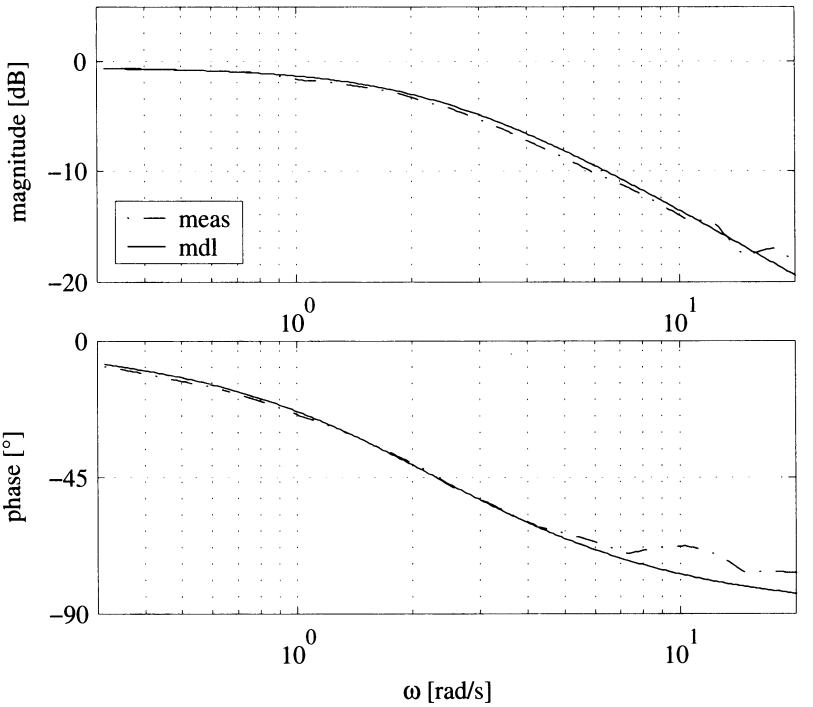


Fig. B.8. Comparison of the measured and modeled transfer function from the throttle position to the pressure in the intake manifold ($\omega_e = 105\ [rad/s]$, $T_l = 0\ [Nm]$).

The result of this identification is evident in Fig. B.11, where the measured dynamic response of the engine is compared to the simulation.

B.2.3 Numerical Values of the Model Parameters

The results of the dynamic and the static parameter identification process is a set of model parameters. For the sake of completeness these parameters are all listed below.

Block [1]:	*Throttle Area*		
Parameter:	Angle $\alpha_{th,0}$	[°]	{7.9}
	Diameter d_{th}	$[m]$	$\{58.7 \cdot 10^{-3}\}$
	Leakage area $A_{th,leak}$	$[m^2]$	$\{5.3 \cdot 10^{-6}\}$
Block [2]:	*Throttle Mass flow*		
Parameter:	Gas constant air R	[J/kg K]	{287}
	Ambient temperature ϑ_a	[K]	{298}
	Ambient pressure p_a	[Pa]	$\{0.98 \cdot 10^6\}$

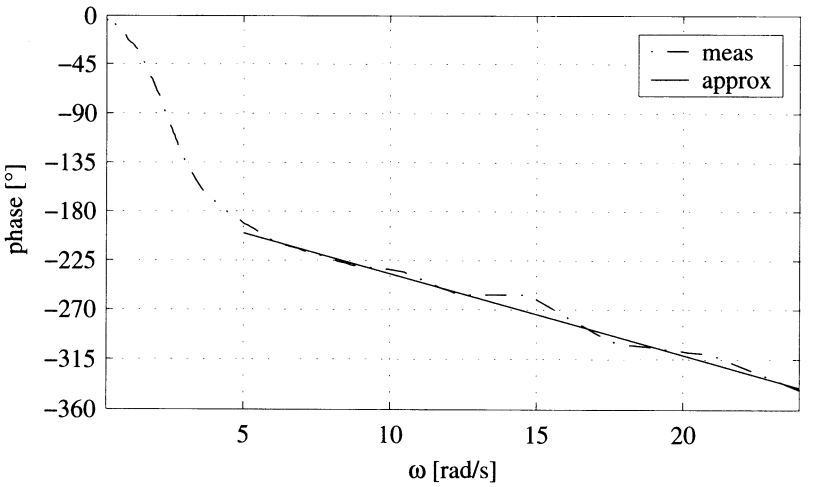


Fig. B.9. Phase response from the throttle (u_α) to the engine speed (ω_e). By plotting the phase linearly over the excitation frequency, the transport delay can be identified directly ($u_{\alpha 0} = 0.023[-]$, $T_l = 0[Nm]$)

Block [3]:	*Air Mass Sensor*	
Parameter:	Static gain k_a	[V s/kg] {1}
Block [4]:	*Intake Manifold*	
Parameter:	Volume intake manifold V_m	$[m^3]$ $\{5.8 \cdot 10^{-3}\}$
	Temperature air in manifold ϑ_m	[K] {340}
Block [5]:	*Pressure Sensor*	
Parameter:	Gain k_p	[V/Pa] $\{0.1 \cdot 10^{-3}\}$
Block [6]:	*Engine Air Mass Flow*	
Parameter:	Coefficient 1 γ_0	[-] {0.45}
	Coefficient 2 γ_1	[s] $\{3.42 \cdot 10^{-3}\}$
	Coefficient 3 γ_2	$[s^2]$ $\{-7.7 \cdot 10^{-6}\}$
	Displacement V_d	$[m^3]$ $\{2.77 \cdot 10^{-3}\}$
	Compression volume at TDC V_c	$[m^3]$ $\{0.277 \cdot 10^{-3}\}$
	Isentropic exponent air κ	[-] {1.35}
	Back pressure exhaust manifold p_e	[Pa] $\{1.08 \cdot 10^5\}$
Block [7]:	*Torque Generation Engine*	
Parameter:	Willansparameter 1 η_0	[J/kg] {0.16}
	Willansparameter 2 η_1	[J s/kg] $\{2.21 \cdot 10^{-3}\}$
	Willansparameter 3 β_0	[Nm] {15.6}
	Willansparameter 4 β_2	[Nm s^2] $\{0.175 \cdot 10^{-3}\}$
	Transport delay $\tau_{inj \to TC}\|_{\omega_e = 90 rad/s}$	[s] {0.125}

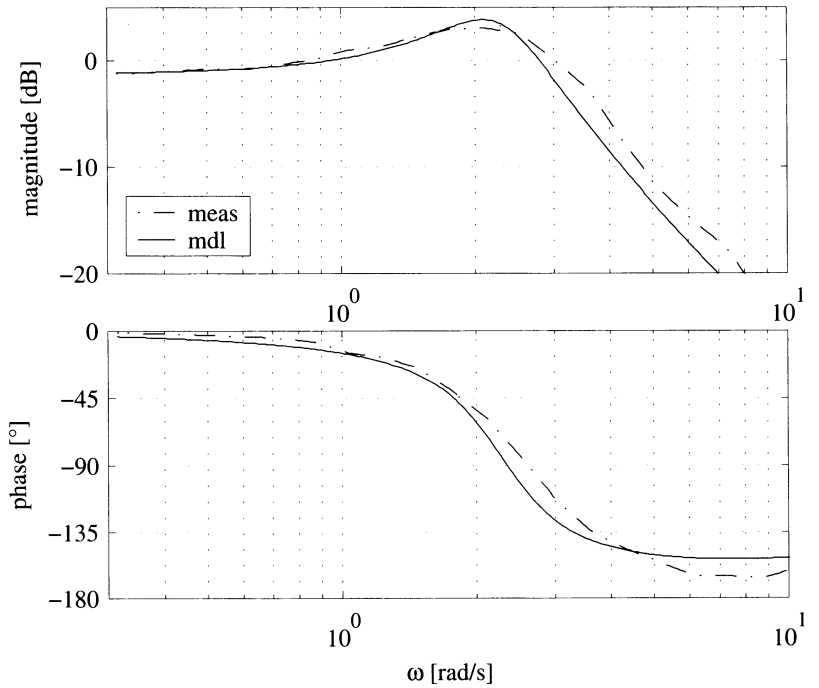


Fig. B.10. Transfer function from the throttle position to the engine speed. The phase is corrected by the amount of transport delay $\varphi = \omega \cdot \tau_{inj \to TC}$ ($u_\alpha = 0.23$ [−], $T_l = 0$ [Nm]).

Block [8]:	*Engine Inertia*	
Parameter:	Engine inertia Θ_e	[kg/m^2] {0.2}
Block [9]:	*Engine Speed Sensor*	
Parameter:	Gain k_ω	[V s/rad] {0.1}
Block [10]:	*Load Torque*	
Parameter:	Gain k_l	[Nm/V] {1}
	Time constant τ_l	[s] {0.02}

B.3 Description of Linear System

To design the idle speed control system a linear system representation is needed. The main problem encountered in this step is that the system derived in the last sections contains a substantial delay. In the time domain such delays cannot be represented exactly by any finite-dimensional system description such that some sort of approximation will be needed. The key idea is to use Padé all-pass elements to approximate these delays

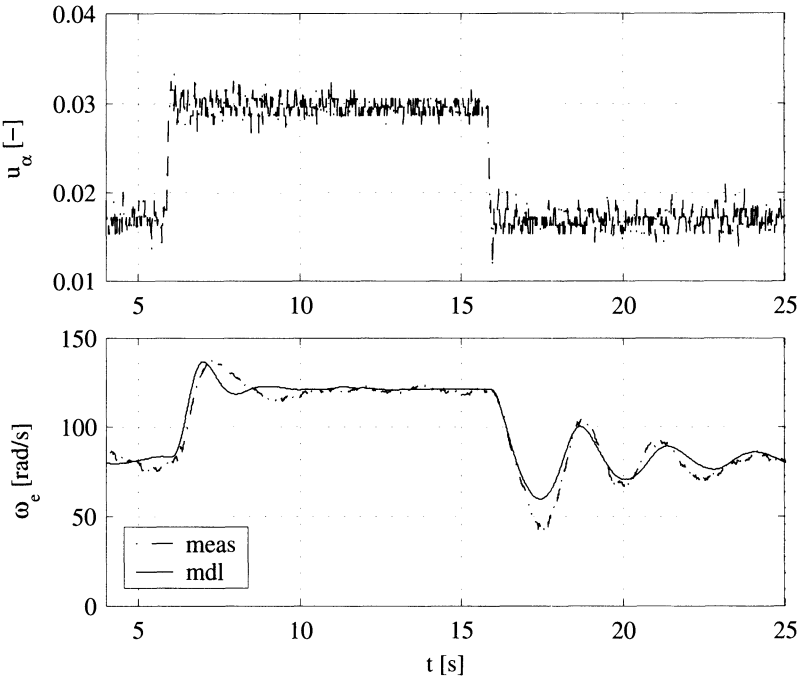


Fig. B.11. Comparison of the measured and simulated step response.

$$e^{-s\cdot\delta} \approx \frac{\sum_{j=0}^{n_{apx}} a_j \cdot (-s\cdot\delta)^j}{\sum_{j=0}^{n_{apx}} a_j \cdot (s\cdot\delta)^j} \qquad a_j = \frac{(2\cdot n_{apx}-j)!\cdot n_{apx}!}{j!\cdot(n_{apx}-j)!\cdot(2\cdot n_{apx})!} \tag{B.18}$$

Usually a small order n is sufficient, while for larger system orders other, numerically more reliable forms are recommended. Matlab offers appropriate software tools for those computations (see `help pade`)

Notice that Padé approximations introduce non-minimum phase system zeros which are known to limit the achievable closed-loop bandwidth. This is not a problem of the chosen delay approximation but is a limitation imposed by the unavoidable IPS delay itself.

For this case study a sixth-order Padé element is chosen to approximate the IPS delay ($\tau_{inj\to TC}$) in block [7]. The variable delay is approximated by the nominal one

$$\tau_{inj\to TC,0} = \frac{\Delta}{\omega_0} \tag{B.19}$$

where ω_0 is the reference idle speed. Figure B.12 shows the effect of these approximations by comparing the step responses of the system model with variable delay and with fixed-delay Padé approximations.

The linearization of the system equations can be done using a purely numerical approach by approximating the partial derivatives by finite differ-

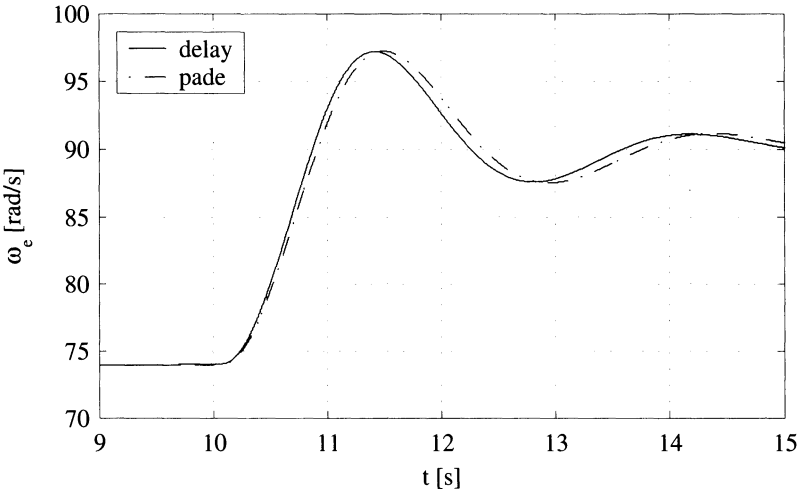


Fig. B.12. Comparison of the system step response with variable delay and with fixed Padé approximation.

ences. Again, Matlab offers appropriate software tools for that step (see `help linmod2`). Notice that the system can be linearized (in closed form or numerically) only if its differential equations are sufficiently "smooth" (continuous first partial derivatives in all variables) around the nominal point.

When using the numerical approach the differences have to be chosen small enough to avoid nonlinear effects and, at the same time, large enough to avoid numerical problems. Some iterations are therefore usually necessary to find the optimum differences.

Fig. B.13 shows the Bode diagram of the control disturbance channel. Also shown is the influence of the IPS delay as approximated by a sixth-order Padé element. Its contribution to the phase lag of the system will have a substantial influence on the control system design process.

B.4 Control System Design and Implementation

In this section a control system is designed using the model derived above. The specifications that have to be met by this control system are listed below:

- The reference speed is constant at 90 rad/s. No substantial changes in this reference signal have to be expected.
- Load disturbances are the main phenomenon for which the controller has to be tuned.
- The steady-state error caused by a constant load disturbance has to be zero.

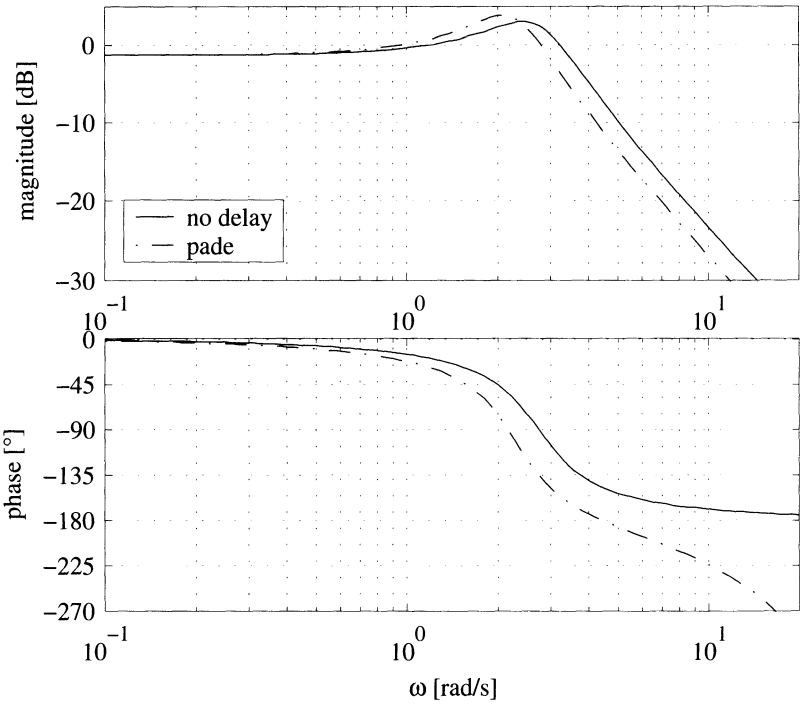


Fig. B.13. Bode diagram of the control channel of the linearized system ($u_\alpha = 0.02$ [-], $u_l = 0$ [-]), dashed with Padé approximation, full without any delays.

- The engine speed may not fall below 60 rad/s in order to avoid engine stall.
- The torque ripple induced by the reciprocating engine behavior may not excite the closed-loop system.
- The settling time to load disturbances has to be around 5 s.
- Overshoot of engine speed after a load disturbance is acceptable but should be limited to 10% of the reference value.
- A minimum return difference of 0.5 has to be achieved in order to cope with the large model uncertainties and neglected nonlinearities.

Obviously, to satisfy these specifications the controller has to include an integral part. The engine has a very low damping rate around $2.5\,rad/s$ (see Fig. B.15). This resonance turns the design of the idle speed controller into a nontrivial task.

The controller is designed using an observer-based pole-placement approach.

In order to avoid a steady-state error, the state feedback controller is augmented by an integrator. Figure B.14 shows a detailed signal flow chart of the controller.

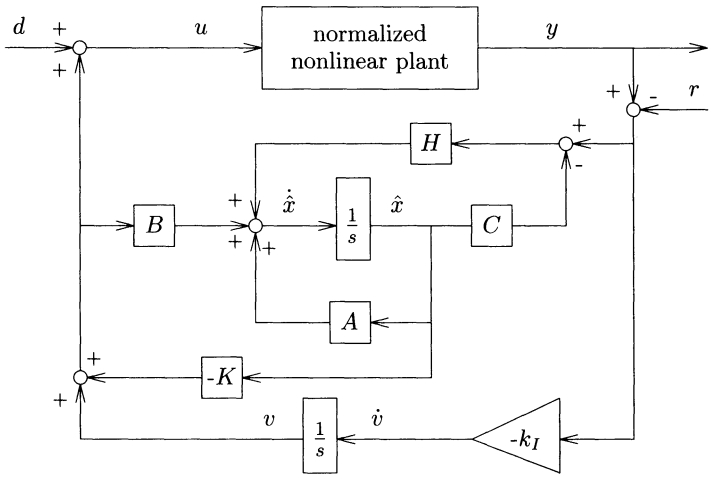


Fig. B.14. Detailed signal flow chart of the idle-speed controller.

The differential equations of the augmented observer-based state feedback controller can be written as follows.

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} A - BK - HC & B \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \hat{x} \\ v \end{pmatrix} + \begin{bmatrix} 0 & H \\ 0 & -k_I \end{bmatrix} \begin{pmatrix} d \\ y - r \end{pmatrix} \tag{B.20}$$

$$[u] = \begin{bmatrix} -K & 1 \end{bmatrix} \begin{pmatrix} \hat{x} \\ v \end{pmatrix} + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{pmatrix} d \\ y - r \end{pmatrix} \tag{B.21}$$

Figure B.15 shows the Bode diagrams of the plant $P(s)$, the controller $C(s)$ and the open loop $L(s) = C(s) \cdot P(s)$). The controller damps the frequencies with a critical phase shift and corrects the phase where the amplification of the gain is critical.

The gain margin of the implemented controller is 2.3 and the phase margin is 48°.

The last step is to test the closed-loop system on an engine test bed. Figure B.16 shows the results. The controller fulfills all the required specifications, but cannot completely eliminate the steady-state oscillations of the system. Adding a second fast control loop that utilizes the spark advance as an additional control action would considerably reduce these oscillations.

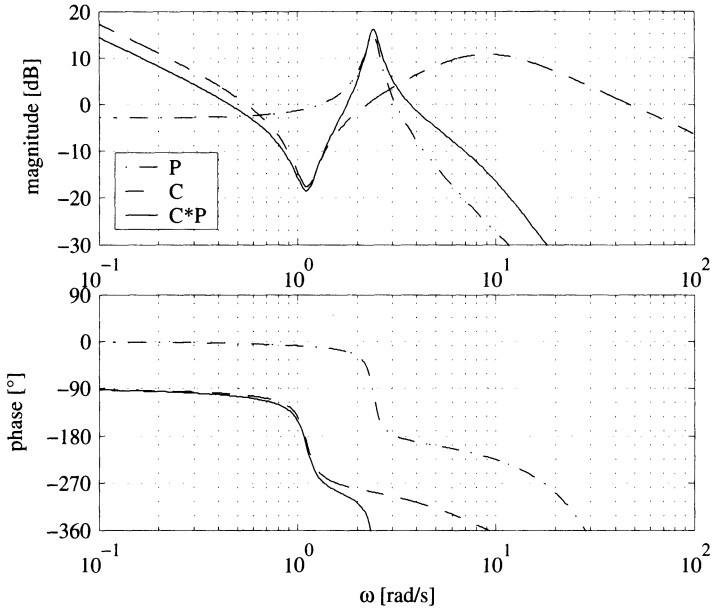


Fig. B.15. Bode plots, plant $P(s)$, controller $C(s)$ and open loop gain $L(s) = C(s) \cdot P(s)$.

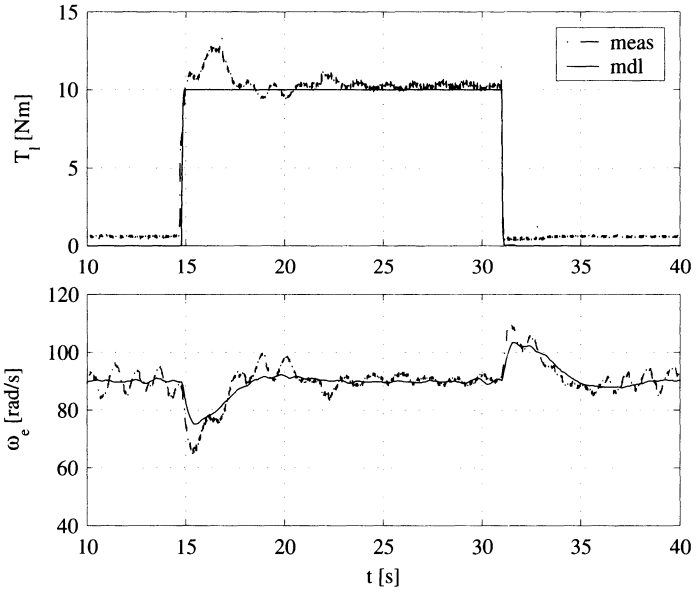


Fig. B.16. Comparison of the closed-loop behavior (solid = simulated and dash-dot = measured).

References

1. Ackermann J. (1993) Robuste Regelung. Springer Verlag, Berlin
2. Alkidas A.C. (1997) Intake-Valve Temperature and the Factors Affecting It. SAE paper 971729
3. Alkidas A.C. (2001) Intake-Valve Temperature Histories During S.I. Engine Warm-Up. SAE paper 2001-01-1704
4. Allmendinger K., Guzzella L., Seiler A., and Loffeld O. (2001) A Method to Reduce the Calculation Time for an Internal Combustion Engine Model. SAE paper 2001-01-0574
5. Amstutz A. and Guzzella L. (1997) Fuel Cells for Transportation – An Assessment of their Potential for CO_2-Reduction. Conference on Greenhouse Gas Control, Interlaken, Switzerland
6. Andersson P. and Eriksson L. (2004) Cylinder Air Charge Estimators in Turbocharged SI-Engines. SAE paper 2004-01-1366
7. Anonymous (1997) Bosch Automotive Handbook. Robert Bosch GmbH, Stuttgart
8. Anonymous (2001) Emission Measurement Standards of the European Union. *http : //europa.eu.int/eurlex/en/search/search_lif.htm*
9. Aquino C.F. (1981) Transient A/F Control Characteristics of a 5 Liter Central Fuel Injection Engine. SAE paper 810494
10. Arici O., Johnson J. H. and Kulkarni A. J. (1999) Cooling System Simulation; Part 1 - Model Development. SAE paper 1999-01-0240
11. Arsie I., Pianese C., and Rizzo G. (1998) Models for the Prediction of Performance and Emissions in a Spark Ignition Engine. SAE paper 980779
12. Aström K. J. and Hägglund T. (1992) PID Controllers: Theory, Design and Tuning. Instrument Society of America, NC
13. Atkins P.W. (1988) Physikalische Chemie. VCH Verlag, Weinheim
14. Atkinson C.M., Long T.W., and Hanzevack E.L. (1998) Virtual Sensing: A Neural Network-Based Intelligent Performance and Emissions Prediction System for On-Board Diagnostics and Engine Control. SAE paper 980516
15. Auckenthaler T., Onder C. and Geering H. (2002) Modelling of a Solid-Electrolyte Oxygen Sensor SAE paper 2002-01-1293
16. Auckenthaler T., Onder C. and Geering H. (2004) Online Estimation of the Oxygen Storage Level of a Three-Way Catalyst. SAE paper 2004-01-0525

17. Auckenthaler T., Onder C. and Geering H. (2004) Aspects of Dynamic Three-Way-Catalytic Converter Behaviour Including Oxygen Storage. IFAC Symposium "Advances in Automotive Control," Salerno
18. Balenovic M., Backx T. and de Bie T. (2002) Development of a Model-Based Controller for a Three-Way Catalytic Converter. SAE paper 2002-01-0475
19. Balluchi A., Benvenuti L., Di Benedetto M.D., Villa T., Wong-Toi H. and Sangiovanni-Vincentelli A.L. (2000) Hybrid Controller Synthesis for Idle Speed Management of an Automotive Engine. Proc. 2000 IEEE American Control Conference, Chicago, vol. 2, pp. 1181–5
20. Balluchi A., Benvenuti L., Di Benedetto M.D. and Sangiovanni-Vincentelli A.L. (2003) Idle Speed Control Synthesis Using an Assume-Guarantee Approach. In Nonlinear and Hybrid Systems in Automotive Control, Springer-Verlag, Berlin
21. Beard A. and Smith G. J. (1971) A Method of Calculating the Heat Dissipation from Radiators to Cool Vehicle Engines. SAE paper 710208
22. Bejan A. (1996) Heat Transfer. John Wiley and Sons, New York
23. Bianchi G. M., Falfari S., Parotto M. and Osbat G. (2003) Advanced Modeling of Common Rail Injector Dynamics and Comparison with Experiments. SAE paper 2003-01-0006
24. Bosch H. and Janssen F. (1988) Catalytic Reduction of Nitrogen Oxides - A Review on the Fundamentals and Technology. Catalysis Today, vol. 2, no. 4, pp. 369–531
25. Bozek J.W., Evans R., Tyree C.D. and Zerafa K.L. (1992) Operating Characteristics of Zirconia Galvanic Cells (Lambda Sensors) in Automotive Closed-Loop Emission Control Systems. SAE paper 920289
26. Brailsford A.D., Yussouff M., and Logothetis E.M. (1996) Theory of Gas Sensors: Response of an Electrochemical Sensor to Multi-Component Gas Mixtures. International Solid-State Sensors and Actuators Conference, B: Chemical, Stockholm, vol. B34, pp. 407–11
27. Brand D. (2004) Control-Oriented NO Models for SI Engines. Dissertation ETH Zürich no. 11948
28. Brandt E., Wang Y., and Grizzle J. (2000) Dynamic Modeling of a Three-Way Catalyst for SI Engine Exhaust Emission Control. IEEE Transactions on Control Systems Technology, vol. 8, no. 5, pp. 767–776
29. Butts K., Sivashankar N. and Sun J. (1995) Feedforward and Feedback Design for Engine Idle Speed Control Using l_1 Optimization. Proceedings of the American Control Conference, vol. 4, pp. 2587–2590
30. Cassidy J. (1977) A Computerized On-Line Approach to Calculating Optimum Engine Calibrations. SAE paper 770078
31. Cavina N., Corti E., Minelli G. and Serra G. (2002) Misfire Detection Based on Engine Speed Time-frequency Analysis. SAE paper 2002-01-0480
32. Chang C.F., Fekete N.P., and Powell J.D. (1993) Engine Air-Fuel Ratio Control Using an Event-Based Observer. SAE paper 930766
33. Chin Y.K. and Coats F.E. (1986) Engine Dynamics: Time-Based Versus Crank-Angle Based. SAE paper 860412
34. Christen U. (1996) Engineering Aspects of H_∞-Control. Dissertation ETH Zürich no. 11433

35. Christen U., Vantine K.J., and Collings N. (2001) Event-Based Mean-Value Modeling of DI Diesel Engines for Controller Design. SAE paper 2001-01-1242
36. Cipollone R. and Sciarretta A. (2001) The Quasi-Propagatory Model: A New Approach for Describing Transient Phenomena in Engine Manifolds. SAE paper 2001-01-0579
37. Cook J.A., Grizzle J.W. and Sun J. (1996) Engine Control. In The Controls Handbook, edited by W. S. Levine, CRC Press-Times Mirror Books
38. Cortona E., Onder Ch. H., and Guzzella L. (2002) Engine Thermal Management with Components for Fuel Consumption Reduction. International Journal of Engine Research, IMechE, vol. 3, no. 3, pp. 157–170
39. Cortona Kanne Elena Engine (2003) Thermomanagement For Fuel Consumption Reduction. Dissertation ETH Zürich no. 13862
40. Cowart J., Cheng W. (1999) Intake Valve Thermal Behavior During Steady-State and Transient Engine Operation. SAE paper 1999-01-3643
41. Csallner P. (1981) Eine Methode zur Vorausberechnung der Änderung des Brennverlaufes von Ottomotoren bei geänderten Betriebsbedingungen. Dissertation TU München, 1981
42. Curtis E.W., Aquino C.F., Trumpy D.K. and Davis G.C. (1996) A New Port and Cylinder Wall Wetting Model to Predict Transient Air/Fuel Excursions in a Port Fuel Injected Engine. SAE paper 961186
43. Curtis E.W., Aquino C.F., Plensdorf W.D., Trumpy D.K., Davis G.C., Lavoie G.A. (1997) Modeling Intake Warm-Up. ASME Fall Technical Conference 29-1 (97-ICE-40)
44. Doyle J., Francis B. and Tannenbaum A. (1992) Feedback Control Theory. Macmillan, New York
45. Dyntar D., Onder C. and Guzzella L. (2002) Modeling and Control of CNG Engines. SAE paper 2002-01-1295
46. Eastwood P. (2000) Critical Topics in Exhaust Gas Aftertreatment. Engineering Design Series, Research Studies Press Ltd., Baldock, Hertfordshire, England
47. Ellman A. and Piché R. (1999) A Two Regime Orifice Formula for Numerical Simulation. ASME J-DSMC, vol. 121, pp. 721–724
48. Emmerling G., Zuther F. (1999) Motorische Verbrennung - aktuelle Probleme und moderne Lösungsansätze. Berichte zur Energie- und Verfahrenstechnik, Editor A. Leiperz, Heft 99.1, Erlangen
49. Engh G.T., Chiang C. (1970) Correlation of Convective Heat Transfer for Steady Intake-Flow Through a Poppet Valve. SAE paper 700501
50. Eriksson L., Nielsen L., and Glavenius M. (1997) Closed Loop Ignition Control by Ionization Current Interpretation. SAE paper 970854
51. Eriksson L. (2002) Mean Value Models for Exhaust Temperatures. SAE paper 2002-01-0374
52. Ferguson C. R. and Kirkpatrick A. T. (2001) Internal Combustion Engines. John Wiley & Sons, New York
53. Franklin G., Powell J. D. and Emami-Naeini A. (1986) Feedback Control of Dynamic Systems. Addison-Wesley, Reading (MA)
54. Franklin G., Powell J. D. and Workman M. (1990) Digital Control of Dynamic Systems. Addison-Wesley, Reading (MA)
55. Friedrich A. (1997) The Assessment of Fuel Cells in Transport from the Environmental Point of View. Umweltbundesamt, Berlin

56. Geering H., Onder C. H., Roduner C. A., Dyntar D. and Matter D. (2002) ICX-3 — A Flexible Interface Chip for Research in Engine Control. Proc. of the FISITA 2002 World Automotive Congress, Helsinki, Paper no. FV02V315
57. Gerhardt J., Hönninger H. and Bischof H. (1998) A new Approach to Functional and Software Structure for Engine Management Systems – BOSCH ME7. SAE paper 980801
58. Germann H., Tagliaferri S. and Geering H. (1996) Differences in Pre- and Post-Converter Lambda Sensor Characteristics. SAE International Congress, SAE paper 960335
59. Glattfelder A.H. and Schaufelberger W. (2003) Control Systems with Input and Output Constraints. Springer Verlag, Berlin
60. Göpel W., Wiemhöfer H.D. (2000) Statistische Thermodynamik. Spektrum Akademischer Verlag, Heidelberg
61. Gorille I. (1980) MOTRONIC - ein neues elektronisches System zur Steuerung von Ottomotoren. MTZ Motortechnische Zeitschrift 41, vol. 51, pp. 203–215
62. Gravdahl J. and Egeland O. (1999) Centrifugal Compressor Surge and Speed Control. IEEE Transactions on Control Systems Technology, vol. 7, no. 5, pp. 567–579
63. Greene A. B. and Lucas G. G. (1969) The Testing of Internal Combustion Engines. English University Press Ltd.
64. Guzzella L., Geering H. and Hirzel H. (1987) Anwendungsspezifische Chips für die Regelung von Ottomotoren. Technische Rundschau, vol. 79, no. 1/2, p. 46–49
65. Guzzella L. and Amstutz A. (1998) Control of Diesel Engines. IEEE Control Systems Magazine, vol. 8, No. 9, pp. 55–71
66. Guzzella L. and Amstutz A. (1999) CAE-Tools for Quasistatic Modeling and Optimization of Hybrid Powertrains. IEEE Transactions on Vehicular Technology, vol. 48, no. 6, pp. 1762–1769
67. Guzzella L., Wenger U. and Martin, R. (2000) IC-Engine Downsizing and Pressure-Wave Supercharging for Fuel Economy. SAE Technical paper, no. 2000-01-1019
68. Gyarmathy G. (1983) How does the COMPREX Pressure Wave Supercharger Work. SAE paper 830234
69. Haluska P. and Guzzella L. (1998) Control Oriented Modeling of Mixture Formation Phenomena in Multi-Port Injection SI-Gasoline Engines. SAE paper 980628
70. Harigaya Y., Toda F. and Suzuki M. (1993) Local Heat Transfer on a Combustion Chamber Wall of a Spark-Ignition Engine. SAE paper 931130
71. Harmsen J. (2001) Kinetic Modelling of the Dynamic Behaviour of an Automotive Three-Way Catalyst under Cold-Start Conditions. PhD Thesis, Technical University of Eindhoven
72. Hasegawa Y., Akazaki S., Komoriya I., Maki H., Nishimura Y., and Hirota T. (1994) Individual Cylinder Air-Fuel Ratio Feedback Control Using an Observer. SAE paper 940376
73. Hendricks E. (1986) A Compact, Comprehensive Model of Large Turbocharged, Two-Stroke Diesel Engines. SAE paper 861190
74. Hendricks E. and Sorenson S.C. (1990) Mean-Value Modeling of SI Engines. SAE paper 900616

75. Hendricks E., Jensen M., Chevalier A., and Vesterholm T. (1994) Conventional Event Based Engine Control. SAE paper 940377
76. Hendricks E., Jensen M., Chevalier A., and Vesterholm T. (1994) Problems in Event Based Engine Control. American Control Conference, Baltimore, vol. 2, pp. 1585–1587
77. Hendricks E., Chevalier A., and Jensen M. (1995) Event-Based Engine Control: Practical Problems and Solutions. SAE paper 950008
78. Hendricks E. (2001) Isothermal vs. Adiabatic SI Engine Mean Value Engine Models. Proc. 3rd IFAC Workshop on Advances in Automotive Control, Karlsruhe
79. Herden W. and Küsell M. (1994) A New Combustion Pressure Sensor for Advanced Engine Management. SAE paper 940379
80. Hertzberg A. (2001) Betriebsstrategien für einen Ottomotor mit Direkteinspritzung und NOx-Speicher-Katalysator. Ph.D. thesis, Universität Karlsruhe
81. Heywood J.B. (1988) Internal Combustion Engine Fundamentals. McGraw Hill, New York
82. Hrovat D. and Sun J. (1997) Models and Control Methodologies for IC Engine Idle Speed Control Design. Control Engineering Practice, vol. 5, no. 8, pp. 1093–1100
83. Huber E., and Koller T. (1977) Pipe Friction and Heat Transfer in the Exhaust Pipe of a Firing Combustion Engine. Proceedings of the CIMAC Convention, Tokio
84. Incropera F.P., DeWitt D.P. (1996) Fundamentals of Heat and Mass Transfer. John Wiley & Sons, New York
85. Ingram G. and Gopichandra S. (2003) On-line Oxygen Storage Capacity Estimation of a Catalyst. SAE paper 2003-01-1000
86. Inhelder J. (1996) Verbrauchs- und Schadstoffoptimiertes Ottomotor-Aufladekonzept. Dissertation ETH Zürich no. 11948
87. Jensen J. P., Kirstensen A. F., Sorenson S. C., Houba N. and Hendricks E. (1991) Mean Value Modeling of a Small Turbocharged Diesel Engine. SAE paper 910070
88. Jobson E., Hjortsberg O., Andersson S. and Gottberg I. (1996) Reactions Over a Double Layer Tri-Metal Three-Way Catalyst. SAE paper 960801
89. Johnson T. (2003) Diesel Emission Control in Review - The Last 12 Months. SAE paper 2003-01-0039
90. Jurgen R. (1995) Automotive Electronics Handbook. McGraw-Hill, New York
91. Kampelmühler F.T., Paulitsch R., and Gschweitl K. (1993) Automatic ECU-Calibration: An Alternative to Conventional Methods. SAE paper 930395
92. Kang J.M., Kolmanovsky I., and Grizzle J.W. (2001) Dynamic Optimization of Lean Burn Engine Aftertreatment. ASME J-DSMC, vol. 123, no. 2, pp. 153–160
93. Kiencke U. and Nielsen L. (2000) Automotive Control Systems. Springer, Berlin
94. Kiencke, U. (1999) Engine Misfire Detection. IFAC Control Engineering Practice, vol. 7, no.2, pp. 203–208
95. Kim Y.W., Sun J., Kolmanovsky I., and Koncsol J. (2003) A Phenomenological Control-Oriented Lean NOx Trap Model. SAE paper 2003-01-1164

96. Kladopoulou E.A., Yang S.L., Johnson J.H., Parker G.G., and Konstandopoulos A.G. (2003) A Study Describing the Performance of Diesel Particulate Filters During Loading and Regeneration: A Lumped Parameter Model for Control Applications. SAE paper 2003-01-0842
97. Klepatsch, M. (1997) Gemischbildung beim Ottomotor: Neue Möglichkeiten der Simulationsrechnung. Wiener Motorensymposium, VDI Reihe 12, Nr. 306, p. 217
98. Koch F. W. and Haubner F. G. (2000) Cooling System Development and Optimization for DI Engines. SAE paper 2000-01-0283
99. Kolmanowsky I. and Stefanopoulou A. (2001) Optimal Control Techniques for Assessing Feasibility and Defining Subsystem Level Requirements: An Automotive Case Study. IEEE Trans. Control Systems Technology, vol. 9, no. 3, pp. 524–534
100. Koltsakis G.C. and Stamatelos A.M. (1999) Modeling Dynamic Phenomena in 3-Way Catalytic Converters. Chemical Engineering Science, vol. 54, no. 20, pp. 4567–4578
101. Konstandopoulos A.G., Margaritis K., and Paraskevi H. (2001) Spatia Non-Uniformities in Diesel Particulate Trap Regeneration. SAE paper 2001-01-0908
102. Konstandopoulos A.G., Skaperdas E., and Masoudi M. (2001) Inertial Contributions to the Pressure Drop of Diesel Particulate Filters. SAE paper 2001-01-0909
103. Krause W. and Spies K. H. (1996) Dynamic Control of the Coolant Temperature for a Reduction of Fuel Consumption and Hydrocarbon Emission. SAE paper 960271
104. Krüger S. (2000) Advanced Microsystems for Automotive Applications. Springer Verlag, Berlin
105. Larsson M., Andersson L., Fast O., Litorell M., and Makuie R. (1999) NOx Trap Control by Physically Based Model. SAE paper 1999-01-3503
106. Lee, H. K. and Russell, M. F. and Bae, C. S. (2002) Mathematical model of diesel fuel injection equipment incorporating non-linear fuel injection. Proc. IMechE Part D - Journal of Automobile Engineering, vol. 216, no. D3, pp 191–204
107. Levenspiel O. (1972) Chemical Reaction Engineering. John Wiley & Sons, New York
108. Ljung L. (1987) System Identification; Theory for the User. Prentice Hall, Englewood Cliffs, NJ
109. Locatelli M., Onder C.H. and Geering H. (2003) A Rapidly Tunable Model of the Droplet Evaporation for the Study of the Wall-Wetting Dynamics. JSAE paper 20035416
110. Locatelli M., Onder C.H. and Geering H. (2004) A Rapidly Tunable Wall-Wetting Model for SI Engine Control. SAE paper 2004-01-1461
111. Malchow G. L., Sorenson S. C. and Buckius R. O. (1979) Heat Transfer in the Straight Section of an Exhaust Port of a Spark Ignition Engine. SAE paper 790309
112. Maloney P.J. (1999) An Event-Based Transient Fuel Compensator with Physically Based Parameters. SAE paper 1999-01-053
113. Maroteaux F. and Le Moine L. (1995) Modeling of Fuel Droplets Deposition Rate in Port Injected Spark Ignition Engines. SAE paper 952484

114. Metzler F., Hesse U., Rocklage G. and Schmitt M. (1996) Thermomanagement. SAE paper 1999-01-0238
115. Mladek M., Onder Ch. H. and Guzzella L. (2000) A Model for the Estimation of Inducted Air Mass and the Residual Gas Fraction using Cylinder Pressure Measurements. SAE paper 2000-01-0958
116. Mladek M. (2003) Cylinder Pressure for Control Purposes of Spark Ignition Engines. Dissertation ETH Zürich no. 14916
117. Morselli R., Corti E., and Rizzoni G. (2002) Energy Based Model of a Common Rail Injector. Proc. of the 2002 IEEE Conference on Control Applications, Glasgow, vol. 2, pp. 1195–1200
118. Moore F.K. and Greitzer E.M. (1986) A Theory of Post-Stall Transients in a Axial Compressor Systems. Journal of Engineering for Gas Turbines and Power, vol. 108, pp. 68–76
119. Moraal P. and Kolmanovsky I. (1999) Turbocharger Modeling for Automotive Control Applications. SAE paper 1999-01-0908
120. Moran M. and Shapiro H.N. (1992) Fundamentals of Engineering Thermodynamics. Wiley, New York
121. Moser F., Sams T., Cartellieri W. (2001) Impact of Future Exhaust Gas Emission Legislation on Heavy Duty Truck Engine. SAE paper 2001-01-0186
122. Müller M., Hendricks E. and Sorenson S. C. (1998) Mean Value Modelling of Turbocharged Spark Ignition Engines. SAE Technical paper, no. 980784
123. Müller N. and Isermann R. (2001) Control of Mixture Composition Using Cylinder Pressure Sensors. SAE paper 2001-01-3382
124. Müller R., Hart M., Truscott A., Noble A., Krötz G., Eickhoff M., Cavalloni C. and Gnielka M. (2000) Combustion-Pressure-Based Engine Management System. SAE paper 2000-01-0928
125. Nievergeld A.J.L. (1998) Automotive Exhaust Gas Conversion: Reaction Kinetics, Reactor Modelling and Control. PhD Thesis, Technical University of Eindhoven
126. Nowak U., Frauhammer J. and Nieken U. (1996) A Fully Adaptive Algorithm for Parabolic Partial Differential Equations in One Space Dimension. Computers and Chemical Engineering, 20,5, 547–561.
127. Onder C. (1993) Modellbasierte Optimierung der Steuerung und Regelung eines Automobilmotors. Dissertation ETH Zürich no. 10323
128. Onder C. H. and Geering H.P. (1993) Model-Based Multivariable Speed and Air-to-Fuel Ratio Control of an SI Engine. SAE Paper No. 930859
129. Onder C.H. (1995) Model-Based Engine Calibration for Best Fuel Efficiency. SAE paper 950983
130. Onder C. H., Roduner C. and Geering H. (1997) Model Identification for the Air/Fuel Path of an SI Engine. SAE paper 970612
131. Pattas K., Stamatelos A., Pistikopoulos P., Koltsakis G., Konstandinidis P., Volpi E. and Leveroni E. (1994) Transient Modeling of 3-Way Catalytic Converters. SAE paper 940934
132. Patton K.J., Nitschke R. G. and Heywood J. B. (1989) Development and Evaluation of a Friction Model for Spark Ignition Engine. SAE paper 890836
133. Peyton Jones J., Jackson R., Roberts J. and Bernard P. (2000) A Simplified Model for the Dynamics of a Three-Way Catalytic Converter. SAE paper 2000-01-0652

134. Plapp G., Klenk M. and Moser W. (1989) Methods of On-board Misfire Detection. SAE paper 900232
135. Poulikakos D. (1994) Conduction Heat Transfer. Prentice Hall, Englewood Hills
136. Powell B.K. and Cook J.A. (1987) Nonlinear low frequency phenomenological engine modeling and analysis. Proc. of the 1987 IEEE American Control Conference, vol. 1, pages 332–340
137. Powell J.D. (1987) A Review of IC Engine Models for Control System Design. Proc. of the 10th IFAC World Congress, San Franciso
138. Powell J.D. (1995) IC Engine Models for Control System Design. Proc. of the 1st IFAC-Workshop on Advances in Automotive Control, Ascona
139. Ranz W.E. and Marshall W.R. (1952) Evaporation from Drops. Chem. Engrg. Prog., pp. 141–146
140. Rao H., Cohen A., Tennant J. and Van Voorhies K. (1979) Engine Control Optimization Via Nonlinear Programming. SAE paper 790177
141. Rassweiler G. and Withrow L. (1979) Motion Pictures of Engine Flames Correlated with Pressure Cards. SAE Transactions, pp. 297–303
142. Rausen D.J., Stefanopoulou A.G., Kang J.M., Eng J.A., and Kuo T.W. (2004) A Mean-Value Model for Control of Homogeneous Charge Compression Ignition (HCCI) Engines. Proc. of the 2004 IEEE American Control Conference, Boston
143. Ribi B. (1996) Rradialverdichter im Instabilitätsbereich. Dissertation ETH Zürich no. 11717
144. Rizzoni, G. and William, B. (1990) Onboard Diagnosis of Engine Misfires. SAE paper 901768
145. Roduner Ch. A. (1997) H_∞-Regelung linearer Systeme mit Totzeiten. Fortschritt-Berichte VDI, Reihe 8, Nr. 708, VDI Verlag GmbH, Düsseldorf
146. Roduner C., Onder C. and Geering H. (1997) Automated Design of an Air/Fuel Controller for an SI Engine Considering the Three-Way Catalytic Converter in the H_∞ Approach. Proc. 5th IEEE Mediterranean Conference on Control and Systems, Cyprus
147. Schaer C. (2003) Control of a Selective Catalytic Reduction Process. Dissertation ETH Zürich no. 15221
148. Schmitz G., Oligschläger U., Eifler G., and Lechner H. (1994) Automated System for Optimized Calibration of Engine Management Systems. SAE paper 940151
149. Schöggl P., Koegeler H.M., Gschweitl K., Kokal H., Williams P., and Hulak K. (2002) Automated EMS Calibration Using Objective Driveability Assessment and Computer Aided Optimization Methods. SAE paper 2002-01-0849
150. Schubiger R.A., Boulouchous K., and Eberle M.K. (2002) Russbildung und Oxidation bei der dieselmotorischen Verbrennung. Motortechnische Zeitschrift (MTZ), no. 5, pp. 342–353
151. Schwarz Ch. (1998) Theorie und Simulation aufgeladener Verbrennungsmotoren. Habilitationsschrift, TU Hannover
152. Sellnau M.C., Matekunas F.A., Battiston P.A., Chang C.F., and Lancaster D.R. (2000) Cylinder-Pressure-Based Engine Control Using Pressure-Ratio Management and Low-Cost Non-Intrusive Cylinder Pressure Sensors. SAE paper 2000-01-0932

153. Shamma J.S. and Athans M. (1992) Gain scheduling: Potential Hazards and Possible Remedies. IEEE Control Systems Magazine, vol. 12, no. 3, pp. 101–107
154. Shafai E., Roduner C. and Geering H. (1996) Indirect Adaptive Control of a Three-Way Catalyst. SAE paper 961038
155. Shafai E., Roduner Ch. and Geering H. (1997) On-Line Identification of Time Delay in the Fuel Path of an SI Engine. SAE paper 970613
156. Shaver G.M., Gerdes J.C., Jain P., Caton P.A., and Edwards C.F. (2003) Modeling for Control of HCCI Engines. Proceedings of 2003 American Control Conference, Denver, vol. 1, pp. 749–754
157. Shaver G.M. and Gerdes J. (2003) Cycle-to-Cycle Control of HCCI Engines. Proceedings of the 2003 ASME International Mechanical Engineering Congress and Exposition, IMECE2003-41966, Washington
158. Shibata Y., Shimonosono H. and Yamai Y. (1993) New Design of Cooling System with Computer Simulation and Engine Compartment Simulator. SAE paper 931075
159. Shirawaka T., Miura M., Itoyama H., Aiyoshizawa E., and Kimura S. (2001) Study of Model-Based Cooperative Control of EGR and VGT for a Low- Temperature, Premixed Combustion Diesel Engine. SAE paper 2001-01-2006
160. Spring P., Guzzella L. and Onder C. (2003) Optimal Control Strategy for a Pressure-Wave Supercharged SI Engine. ASME Spring Technical Conference, Salzburg, paper no. ICES2003-645
161. Stobart R.K. (1997) The Demands of Cylinder Event Control. SAE paper 970617
162. Stöckli M. (1989) Reibleistung von 4-Takt Verbrennungsmotoren. LVV Technical Report, ETH Zürich
163. Stone R. (1999) Introduction to Internal Combustion Engines. SAE Publishing, Warrendale
164. Streit E. E. and Borman G. L. (1971) Mathematical simulation of a large turbocharged two-stroke Diesel engine. SAE paper 710177
165. Stuhler H., Kruse T., Stuber A., Gschweitl K., Piock W., Pfluegl H., and Lick P. (2002) Automated Model-Based GDI Engine Calibration Adaptive Online DoE Approach. SAE paper 2002-01-0708
166. Sun J., Kolmanovsky I., Brehob D., Cook J. A., Buckland J. and Haghgooie M. (1999) Modeling and Control of Gasoline Direct Injection Stratified Charge (DISC) Engines. Proc. of the 1999 IEEE Conf. on Control Applications, Hawai, vol. 1, pp. 471–477
167. Tennant J., Giacomazzi R., Powell J. and Rao H. (1979) Development and Validation of Engine Models Via Automated Dynamometer Tests. SAE paper 790178
168. Tounsi M.F., Menegazzi P., and Rouchon P. (2003) NOx Trap Model for Lean-Burn Engine Control. SAE paper 2003-01-2292
169. Trella T. (1979) Spark Ignition Engine Fuel Econoomby Control Optimization - Techniques and Procedures. SAE paper 790179
170. Treybal R.E. (1980) Mass-Transfer Operations. McGraw-Hill, 3rd Edition, New York
171. Turin R.C. and Geering H. (1993) On-line Identification of Air-to-Fuel Ratio Dynamics in a Sequentially Injected SI Engine. SAE paper 930857

172. Turin R., Casartelli E. and Geering H. (1994) A New Model for Fuel Supply Dynamics in an SI Engine. SAE paper no 940208
173. Turin R.C. and Geering H. (1995) Model-Reference Adaptive A/F-Ratio Control in an SI Engine based on Kalman Filtering Techniques. Proc. of the 1995 American Control Conference, pp. 4082–4090, Seattle, WA
174. Ulrich O., Wlodarczyk R., and Wlodarczyk M.T. (2001) High-Accuracy Low-Cost Cylinder Pressure Sensor for Advanced Engine Controls. SAE paper 2001-01-0991
175. Urlaub A. (1995) Verbrennungsmotoren. Springer Verlag, Berlin
176. Vantine K.J., Christen U., Glover K. and Collings N. (2001) Analysis of an Event Based Diesel Engine Model for Control Purposes. Proc. 3rd IFAC Workshop on Advances in Automotive Control, Karlsruhe
177. Vibe I.I. (1970) Brennverlauf und Kreiprozessrechnung. VEB Verlag Technik Berlin, 1970
178. Vidyasagar M. (1978) Nonlinear Systems Analysis. Prentice-Hall, New Jersey
179. Wang Y., Raman S., and Grizzle J.W. (1999) Dynamic Modeling of a Lean NOx Trap for Lean Burn Engine Control. Proc. of the 1999 American Control Conference, San Diego
180. Watanabe S., Machida K., Iijima K., and Tomisawa N. (1996) A Sophisticated Engine Control System Using Combustion Pressure Detection. SAE paper 960042
181. Watson N. and Janota M. (1982) Turbocharging the Internal Combustion Engine. The McMillan Press Ltd.
182. Weber F., Guzzella L., and Onder C. (2002) Modeling of a Pressure Wave Supercharger Including External Exhaust Gas Recirculation. Proc. IMechE, Part D: Journal of Automobile Engineering, vol. 216, no. 3, pp 217–235
183. Weeks R.W. and Moskwa J.J. (1995) Automotive Engine Modeling for Real-Time Control using MATLAB/SIMULINK. SAE paper 950417
184. Wen C.Y. and Fan L.T. (1975) Models for Flow Systems and Chemical Reactors. Marcel Dekker Inc., New York
185. Woermann R.J., Teuerkauf H.J., and Heinrich A. (1999) A Real-Time Model of a Common-Rail Diesel Engine. SAE paper 99010862
186. Yamamoto M., Yoneya S., Matsuguchi T., and Kumagai Y. (2002) Optimization of Heavy-Duty Diesel Engine Parameters for Low Exhaust Emissions Using the Design of Experiments. SAE paper 2002-01-1148