

Constrained Nonlinear Predictive Control of Gasoline Low-Temperature Combustion

K. Hoffmann^{1*} and D. Abel²

¹ Formerly², Today Daimler AG, 70546 Stuttgart - Germany

² Institut für Regelungstechnik, RWTH Aachen University, 52056 Aachen - Germany

e-mail: kai.hoffmann@rwth-aachen.de - d.abel@irt.rwth-aachen.de

* Corresponding author

Résumé — Réglage d'une prédictive non-linéaire avec réalisation des limitations pour la combustion de l'essence à basse température

— La combustion de l'essence avec une grande quantité de gaz d'échappement recirculés suscite de plus en plus d'intérêt. Dans de telles conditions de charge partielle, on peut atteindre une température maximale de combustion faible, ce qui donne moins d'émissions mais provoque des instabilités du processus et un comportement fortement non-linéaire. Ces propriétés rendent nécessaire une commande en boucle fermée pour le fonctionnement en transitoire, dont la synthèse constitue un véritable défi. Le papier présente une commande NMPC (prédictive non-linéaire basée sur modèle) innovante pour réguler la PME (Pression Moyenne Efficace) et le CA50 (angle du vilebrequin où 50 % de la masse de carburant est brûlée), tout en tenant compte des contraintes sur la dpmax (montée en pression maximale). La mise en œuvre du régulateur est présentée dans un cadre de prototypage rapide de la commande (RCP) permettant à l'utilisateur de mettre en place, dans un court délai, un régulateur complexe pour le fonctionnement en transitoire. Le potentiel de l'approche est démontré par application sur le moteur réel.

Abstract — Constrained Nonlinear Predictive Control of Gasoline Low-Temperature Combustion

— Combustion of a gasoline fuel with a high amount of recirculated exhaust gas is increasingly gaining interest. Such in part load conditions, a low combustion peak temperature can be achieved, which yields lowest emissions but suffers from instabilities of the process and a highly nonlinear behavior. These properties make a closed-loop control a requirement for transient operation but also a challenge. The paper presents an innovative Nonlinear Model-based Predictive Controller (NMPC) for controlling the Indicated Mean Effective Pressure (IMEP) and Crank Angle of 50% burnt mass fraction (CA50) while accounting for constraints on the maximum pressure rise (dpmax). The implementation of the controller is presented in a framework for Rapid Control Prototyping (RCP) enabling the user to set up a complex controller for transient operation in a short time frame. The ability of the approach is proved by application to the real engine.

Definitions, Acronyms, Abbreviations

BDC _{GE}	Bottom Dead Center in the Gas Exchange cycle
BDC _{HP}	Bottom Dead Center in the High Pressure cycle
CA	Crank Angle
CA50	Crank Angle of 50% burnt mass
CAI	Controlled Auto-Ignition
CCR	Combustion Chamber Recirculation
dpmax	Maximum pressure rise per degree Crank Angle of a whole cycle
ED	Energizing Duration
EE	End of Energizing
EKF	Extended Kalman Filter
EMVT	Electro-Mechanical Valve Train
EVC	Valve event Exhaust Valve Close
EVO	Valve event Exhaust Valve Open
HIL	Hardware-In-the-Loop
IMEP	Indicated Mean Effective Pressure
IVC	Valve event Intake Valve Close
IVO	Valve event Intake Valve Open
MIL	Model-In-the-Loop
MIMO	Multiple-Input-Multiple-Output
MISO	Multiple-Input-Single-Output
MLP	Multi-Layer-Perceptron
MPC	Model-based Predictive Controller
n_{rev}	Revolution speed
NMPC	Nonlinear Model-based Predictive Controller
NNSS(IF)	Neural Network StateSpace (Innovations Form)
NO _x	Nitrogen Oxides
NVH	Noise, Vibration, and Harshness
PID	Proportional, Integral, Derivate controller
RCP	Rapid Control Prototyping
RWTH	Rheinisch Westfälische Technische Hochschule Aachen
SSE	Sum Squared Error
$T_{air,in}$	Temperature of the air in the intake manifold
$T_{CW,in}$	Temperature of the coolant water before the engine
T_{oil}	Temperature of the oil
TDC _{GE}	Top Dead Center in the Gas Exchange cycle
TDC _{HP}	Top Dead Center in the High Pressure cycle

INTRODUCTION

The discussion regarding carbon dioxide during recent years drew attention to this greenhouse gas. The research on more efficient combustion processes has become obligatory for

the automobile industry. The development of new combustion engines is characterized by the trade-off between different emissions and additional customer related requirements which often form contrary tasks. Especially highly homogenized combustion in combination with Exhaust Gas Recirculation (EGR) has shown promising potential. Known acronyms for this combustion process are HCCI (Homogeneous Charge Compression Ignition) or, for gasoline-fuel, CAI (Controlled Auto-Ignition) which is subject of this work.

This combustion can only be realized with a valve train which can offer a high variability. This necessity is due to the internal recirculation of exhaust gas which is needed to initiate the self-ignition. Compared to conventional gasoline engines, reduced throttling losses lead to improved efficiency and an up to 15% lower fuel consumption. Due to the exhaust gas' high thermal heat capacity, the peak temperature can be kept in a low range which causes a drop of NO_x-emissions by 90-99% [1]. In addition to this, the combustion duration is shorter which approaches the thermodynamically ideal constant volume cycle [2]. CAI combustion is initiated by auto-ignition in the compression stroke near Top Dead Center (TDC) without using a spark. The temperature level required for auto-ignition of gasoline is achieved by internal exhaust gas recirculation. The cylinder load is compressed until it self-ignites nearly simultaneously in hot gas spots in the compression volume near top dead center. However, the negative effect of rapid heat release on Noise, Vibration, and Harshness (NVH) characteristics has to be considered during the development of the combustion process and of the controller.

The self-ignition is influenced by the thermal attributes of the cylinder load and its stratification. As CAI works with a high amount of residual gas, the combustion of one cycle depends on the exhaust of the previous one. The start of combustion will be advanced with a higher compression end temperature. The auto-ignition may even misfire or at least be retarded with too low temperature at top dead center after the high pressure cycle [3]. The only way to address this problem is a controller stabilizing the combustion that has to be robust towards limiting constraints, *e.g.* the acceptable frame for the actuators or NVH requirements. Different valve timing strategies for the re-feed of the residual gas are under investigation. This paper examines the retaining of the exhaust gas in the cylinder.

Several groups have presented control approaches for CAI, which is also termed gasoline HCCI. Decoupled control of the peak pressure and the combustion timing was proposed [4], as well as a coordinated peak pressure and combustion timing controller based on the linearization of a nonlinear physical model about an operating point which is implemented in an H₂ control strategy [5]. Gain scheduled linear MPC control of IMEP and CA50 in steps between stationary operating points with constraints on dpmax was

realized using Wiener-type models for compensating for the nonlinear characteristics [6, 7]. The authors of this work remark that nonlinear models may improve the control results. Also a nonlinear observer-based feedback controller for regulating combustion timing during load transitions between static operating points was proposed [8]. Load transient control of IMEP and CA50 by means of PI control extended by a dynamic model-based pilot control was examined, too, and the need for taking noise attributes into account was emphasized [9, 10].

The contribution at hand proposes and validates a new NMPC for simultaneously controlling IMEP and CA50 with a variable valve train and direct injection while enabling constraints on dp_{max} with a highly dynamic load-profile. The implementation is set up by means of a RCP framework which allows for the realization from identification measurements to control within short time.

1 RESEARCH ENGINE

The research presented deals with a single cylinder engine and is part of the work of a major research project and a PhD Thesis published in 2010 [11, 12] and was partly presented at the E-COSM'09 conference [13]. The motor is run by the Institute for Combustion Engines VKA, while the controller is developed by the Institute of Automatic Control IRT, both at RWTH Aachen University. The engine is equipped with an Electro-Mechanical Valve Train (EMVT) which, in combination with the piston's shape, offers the possibility to open and close the valves at any desired point of time. The piezo direct injector is used for European RON 95 gasoline fuel. Further engine details are shown in Table 1. The degrees of freedom offer several ways to recirculate the

TABLE 1
Data of the research engine

Data	Value	Unit
Bore	84	mm
Stroke	90	mm
Con-rod length	159	mm
Displacement	0.499	dm ³
Compression ratio	12	[-]
Valves (intake/exhaust)	2/2	[-]
Valve lift	8	mm
Injector position	central	[-]

exhaust from a previous to the following cycle. With CAI-operation hot recirculated exhaust gas increases especially the temperature towards the end of compression which has to reach the auto-ignition temperature. Inhomogeneities of fresh air and residual gas are advantageous because the temperatures in EGR-rich zones are higher than those attained in a completely homogeneous cylinder charge. The air-fuel mixture will ignite at the boundaries of these EGR rich

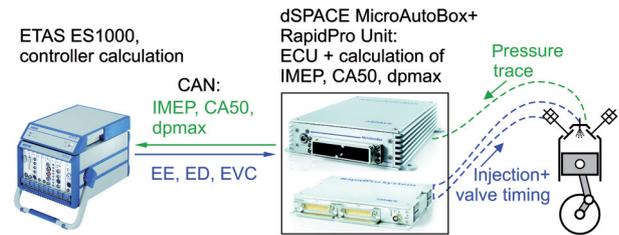


Figure 1

Schema of the test bed setup.

zones. For this reason, besides the temperature, the homogeneity of the recirculated exhaust gas is a key influence on the CAI process [14]. Therefore, the thermodynamic state and the amount of recirculated exhaust have to be balanced. The influence of EGR and temperature on the self-ignition and its use in influencing CAI was investigated within the framework of the Research Center 686 by CFD-Calculations [3].

With the Combustion Chamber Recirculation strategy (CCR) the exhaust from a previous cycle is partly held inside the combustion chamber by setting the control edge Exhaust Valve Close (EVC) before Top Dead Center of the Gas Exchange (TDC_{GE}). After the event Intake Valve Open (IVO), the necessary amount of air is aspirated. This strategy offers a hot charge and therefore can also initiate the auto-ignition towards lower loads [14]. IMEP, CA50, and dp_{max} are calculated from the pressure trace per cycle using a dSPACE MicroAutoBox programmed as a tailor-made ECU by the institute VKA at RWTH Aachen University. A schema of the test bench setup is given in Figure 1.

2 CAI CONTROL MODELING

For the development of a controller, a process model is necessary in order to set up a closed-loop simulation for design and testing purposes [15]. A closed-loop simulation of the NMPC requires two different nonlinear models of the engine. One has to serve as plant model, while the second will be implemented in the NMPC.

The test bench is restricted to a constant revolution speed which limits the operation to load transient conditions. In the following, an approximately constant revolution speed of 2000 rpm is implied. Thus, the first controlled variable is IMEP because at constant rpm the power output of the engine is strictly dependent thereof. For best operation characteristics, an optimal Crank Angle position for the center of combustion can be defined dependent on the load level by means of CA50 which gives the second controlled variable. A desirable NVH characteristic can be defined by a limit

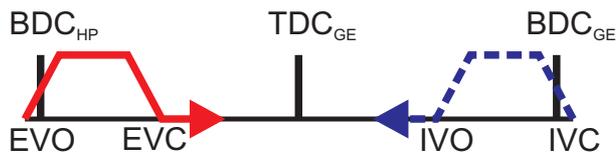


Figure 2

Schema of the valve lift and timing events for CCR with the direction of change for reducing residual gas.

on dp_{max} which is a bounded but not controlled variable. Hence, in total three outputs for both models are given.

The actuation offers a too big degree of freedom for the development of a controller. It has to be reduced to a manageable size, *e.g.* by limitation of the injection timing to a single injection which is describable by the two parameters End of Energizing (EE) and Energizing Duration (ED). The two actuation parameters EVC and IVO are operated symmetrically to TDC_{GE} in order to prevent backflow into the intake manifold. The events Exhaust Valve Open (EVO) and Intake Valve Close (IVC) are kept constant close to the Bottom Dead Center positions after the High Pressure cycle (BDC_{HP}) and the Gas Exchange (BDC_{GE}). Only if the opening duration reduces under a certain limit, the valve events EVO and IVC are shifted so that a minimum distance between EVO/EVC and IVO/IVC is granted. Therefore, the degree of freedom in the actuation of the valves can be reduced to the event EVC, see Figure 2. If the valve events are shifted in the direction indicated by the arrows, the rate of recirculated exhaust gas is reduced. The actuated variables EE, ED, and EVC define the inputs for the controller's model. All other influences are addressed by an Extended Kalman Filter (EKF) which is a computational fast and approved choice for noisy signals. To allow for additional influences and to attain inputs for disturbance replication in closed-loop simulations, the simulation model's inputs are not only the same but also the temperatures of the coolant water and of the aspirated air, the valve event IVO and the actual revolution speed. The latter is subject to dynamic deviations from the stationary value, when load transient conditions are enforced.

Different types of models need to be developed for the controller implementation. First of all, a general model is required for the virtual substitution of the process "CAI engine". The model is built in order to replace the real-world controlled plant. The controller is tuned using this model in a virtual environment, a procedure called Model-In-the-Loop (MIL). Ideally the model is also a candidate for the Hardware-In-the-Loop (HIL) test, in which the target control unit hardware calculates the fully developed controller and exchanges signals with another real-time computer which simulates the engine combustion.

For controlling the low temperature combustion in a model-based sense, on the one hand a detailed model of the process itself has to be found which is able to reflect the instabilities as well as the nonlinear characteristics of the process. On the other hand, a model is needed which is simple enough to be executed in a model-based controller in real-time [16].

2.1 Multilayer Perceptron Networks

In case no analytical or physical description of the process is available, Artificial Neural Networks (ANN) offer the desirable attributes of a good model performance besides relatively low computational requirements. A model serving as plant for tuning the controller in closed-loop simulations and additionally an internal model for the controller have to be established. In the following, a model will be needed for multiple tasks. As physical modeling is desirable but also a time consuming challenge, here a fast and reliable way to obtain a model is selected [17]. However, the use of a physically based model is highly desirable because by physical knowledge the generalization capabilities of the model can be improved as pointed out in [14]. Examples for a detailed physical engine model with elaborate chemistry which is derived from a more complex model and is suitable as a plant model for closed-loop simulations can be found in *e.g.* [18] or [19].

A very promising neural network class with many good attributes for the described purpose is the well-known Multi-Layer-Perceptron Network (MLP). This type of neural network consists of different layers. It has been proved that the standard feed-forward Multi-Layer-Perceptron with a single hidden layer can approximate any continuous function to any desired degree of accuracy [20, 21]. Consequently, the MLP has been termed an universal approximator. The very common two layer structure with n_r inputs, one hidden layer with n_h nodes and one output layer with m outputs takes the general form 1 with $i = 1, \dots, m$.

$$\hat{y}_i = F_i \left(\sum_{j=1}^{n_h} V_{i,j} f_j \left(\sum_{l=1}^{n_r} w_{j,l} r_l + w_{j,0} \right) + V_{i,0} \right) \quad (1)$$

Here the regression vector r with n_r entries represents the inputs to the net in general. These inputs can consist of any combination of system inputs as well as past outputs of the process. The system is determined by the hidden and output weights and biases of the network $w_{i,j}$ and $V_{i,j}$ [22].

The corresponding network structure can be visualized as in Figure 3. In this formulation, biases are realized by the input value one which is weighted, *e.g.* with $w_{1,0}$ in Figure 3.

Typically, the activation functions f_j of the hidden layer are hyperbolic tangent, while the ones of the output layer F_i are linear, see Figure 3, but any combination is possible. The general form of an MLP does not contain any dynamical information. Therefore, the temporal information has

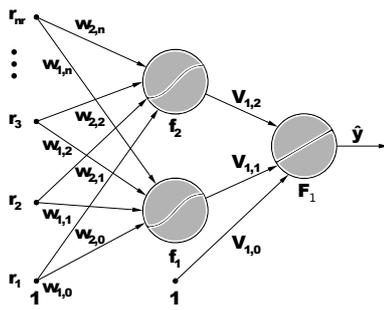


Figure 3
General structure of a Multi-Layer-Perceptron network with a hidden layer with two hyperbolic tangent nodes and an output layer with one linear output unit.

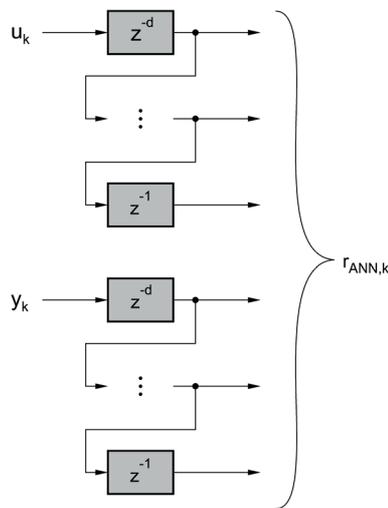


Figure 4
Regressor of dynamic Multi-Layer-Perceptron network nets contain the temporal information by the appropriate number of time shifts within the regressor vector. These retarded values represent the system's dynamics.

to be fed into the net by implementing the dynamics into the regressor vector r by retarding the corresponding values with a discrete sample time in multiple steps which in turn have to represent the systems dynamics. A system of *e.g.* third order at least has to be fed with the three last sample steps of the system's output, see Figure 4. Note that these steps not necessarily all need to have the same temporal duration, *e.g.* with varying revolution speeds.

Obviously, the order of the system to be modeled has to be estimated for defining the neural net. Several structures for the input vector r are known. In order to receive a model which is suitable for simulation, the regressor vector r must not contain any feedback from the modeled process, neither directly the output of the process nor indirectly by inserting the modeling error ϵ into r . So the order to be estimated

reduces from the system's in- and output only to the output. For NMPC implementation, the model is to be described in a discrete state space. Therefore, a network structure is chosen that easily can be linearized to obtain a discrete-time state space model which in turn is applicable to linear controllers.

2.1.1 Neural Network State Space Innovations Form

The MIMO identification of state space models out of input/output measurements from an arbitrary plant is over parameterized [23]. The suggested overlapping described by Ljung in this publication was extended to a neural network of the MLP form with some additions [22]. Nørgaard *et al.* hence call the created MLP network Neural Network State Space Innovations Form (NNSSIF). The structure of this NNSSIF net differs from the standard MLP approach in terms of the system's dynamics. The MLP net serves as the nonlinear description of a state space model. The output of the network is the nonlinear state vector x_k at time instant k . The output of the plant is modeled by the multiplication of the state vector with a fix matrix C which selects the system's outputs from the states. Here the construction of the time base varies from the standard form as in Figure 4. The temporal attributes are considered by inserting the necessary amount of states corresponding to the process' order. The state vector itself becomes part of the regressor. The original form also sets the process input signal u_k and the modeling error ϵ_k as parts of the regressor r_k [22]. All of these signals can also be multidimensional which leads to the desired MIMO identification. An example of the resulting structure is shown in Figure 5. The state vector itself is composed out of two different forms of states. Obviously, states which are not an output signal are also present. These actually serve as retarding units comparable to Figure 4. The predicted nonlinear state space behavior is achieved by adding the retarding state of current time step $\hat{x}_{(i+1),k}$ to the retarded state $\hat{x}_{i,(k+1)}$, see the example case in Figure 5.

The retarding states are only influenced by the system's input u_k and the modeling error ϵ_k . If, for instance, two

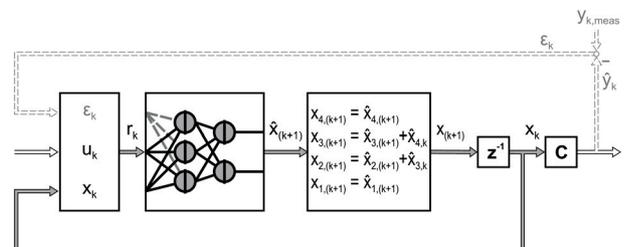


Figure 5
Example for the structure of the Neural Network State Space Innovations Form NNSSIF. The detailed MLP net is given in Figure 6.

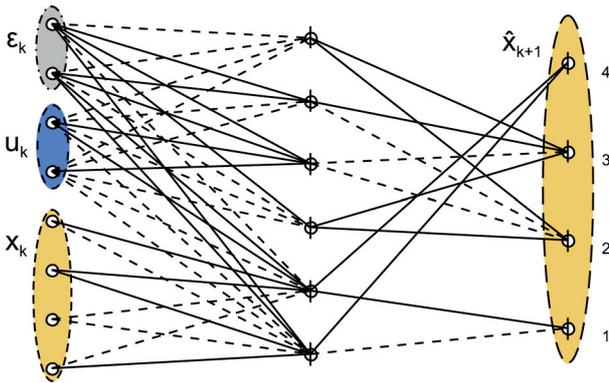


Figure 6

Example for the MLP net as part of a NNSSIF with one output of first and one of third order dynamics resulting in four states.

system outputs are modeled and one of these is of first order and the second of third order, the corresponding MLP would look like Figure 6. Straight lines represent weights, circles nodes, while the vertical lines crossing the circles denote biases. The first state in a retardation chain is the input to the queue and is influenced by the whole regression vector r_k , see state $x_{4,k+1}$ in Figure 6. The corresponding output state is the last sum in the queue, state $x_{2,k+1}$ in the example case.

2.1.2 Linearization of NNSSIF

The resulting system has properties similar to the identification scheme proposed in [23]. This fact becomes evident if the NNSSIF structure in Figure 5 is linearized. The general linearization of a two layer MLP net with linear and hyperbolic tangent activation functions has to regard four combinations of activation functions as shown in Table 2.

TABLE 2

Possible combinations for the linearization of a two layer MLP net with a linear/hyperbolic tangent activation functions.

Combination #	Hidden layer	Output layer
1	Linear	Linear
2	Linear	Hyperbolic tangent
3	Hyperbolic tangent	Linear
4	Hyperbolic tangent	Hyperbolic tangent

These four resulting cases lead to linearizations of Equation (1) as follows. With an NNSSIF structure the number of outputs m equals the number of modeled states, e.g. $m = 4$ in Figure 6, therefore $i = 1, \dots, 4$.

Case 1:

$$\frac{\partial g_i(r_k)}{\partial r_k} = \sum_{j=1}^{n_h} \sum_{l=1}^{n_r} V_{i,j} \cdot w_{j,l} \quad (2)$$

Case 2:

$$\frac{\partial g_i(r_k)}{\partial r_k} = \sum_{j=1}^{n_h} \sum_{l=1}^{n_r} V_{i,j} w_{j,l} \times \left(1 - \tanh \left(\sum_{a=1}^{n_h} V_{i,a} \left(\sum_{b=1}^{n_r} w_{a,b} r_{b,0} + w_{a,0} \right) + V_{i,0} \right) \right)^2 \quad (3)$$

Case 3:

$$\frac{\partial g_i(r_k)}{\partial r_k} = \sum_{j=1}^{n_h} \sum_{l=1}^{n_r} V_{i,j} w_{j,l} \times \left(1 - \tanh \left(\sum_{b=1}^{n_r} w_{j,b} r_{b,0} + w_{j,0} \right) \right)^2 \quad (4)$$

Case 4:

$$\frac{\partial g_i(r_k)}{\partial r_k} = \sum_{j=1}^{n_h} \sum_{l=1}^{n_r} V_{i,j} w_{j,l} \times \left(1 - \tanh \left(\sum_{b=1}^{n_r} w_{j,b} r_{b,0} + w_{j,0} \right) \right)^2 \times \left(1 - \tanh \left(\sum_{a=1}^{n_h} V_{i,a} \left(\sum_{b=1}^{n_r} \tanh(w_{a,b} r_{b,0} + w_{a,0}) \right) + V_{i,0} \right) \right)^2 \quad (5)$$

Here n_r and n_h represent the number of scalars in r_k and the number of hidden neurons, respectively. The linearization leads to a matrix of dimension $(m \times n_r)$, where m is indicating the total number of modeled states $x_{i,k}$. As all combinations of linear and hyperbolic tangent units are possible in each layer, the linearization consists of the sum of all four matrices resulting from each case or Equations (2) to (5), respectively.

The resulting matrix contains three parts which match the parts in the regressor r_k , namely x_k , u_k , and ϵ_k . These result in the state space matrices **A** and **B** following the common nomenclature of a state space representation and a third matrix **K**. This matrix **K** can be interpreted as Kalman-gain. The retarded states $\hat{x}_{2,(k+1)}$ and $\hat{x}_{3,(k+1)}$ in Figure 6 lead to rows in **A** with entries equal to zero, when linearizing the MLP of Figure 6 as they only depend on u_k and ϵ_k . The NNSSIF form as in Figure 5 leads to an entry equal to one in matrix **A** at the position right to the trace. The linearized state space obtained from the shown example case in Figure 5 and Figure 6 takes the form (6).

$$\mathbf{A} = \begin{bmatrix} * & * & * & * \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ * & * & * & * \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} * & * \\ * & * \\ * & * \\ * & * \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} * & * \\ * & * \\ * & * \\ * & * \end{bmatrix} \quad (6)$$

Within the training of the neural net, it is possible to reduce the NNSSIF structure by the error feedback ϵ_k and therefore to set the resulting Kalman-gain matrix \mathbf{K} to $\mathbf{0}$ before the training. This leads to a pure simulation model which is desired for the implementation in an observer structure. The neglecting of the error feedback yields an identified Neural Network StateSpace model and hence the acronym reduces to NNSS. However, an observer is characterized by the feedback of measurements and the NNSSIF net already shows such a structure.

It is possible to select the best linear representation out of a set with amount q if a measurement of q discrete-time steps is available. By neglecting the matrix \mathbf{K} , the best fitting linear state space for the representation of the test case can be found when evaluating the sum squared error of the linear model output over the whole measurement data set.

2.2 Process Identification

For the identification of the neural net models a special test program is set up. The aim is to achieve measurements of the modeled process outputs as a result of an excitation of the combustion process by means of the constituted manipulated variables. A common choice for this purpose is the step response experiment, which is often used for the identification of relevant models for controller application.

Because often test signals are used to superpose the inputs of the open-loop process to influence the process around a working point, the mean value of that test signal should be zero [24]. A Pseudo-Random Binary Sequence is a periodic, deterministic signal with white-noise-like properties. It excites all frequencies equally well. The signal used here is generated using an $n = 9$ bit shift register with feedback through an exclusive-OR logic. The shift register of the PRBS source is based on the engine cycles. The amplitude is kept constant with zero mean. While appearing random in time, actually the sequence repeats every $2 \times n - 1 = 17$ values or engine cycles, respectively. A property of PRBS is that variations in response signals between two periods of the stimulus can be attributable to noise due to the periodic nature of the signal. It is therefore very well suited for identification purposes and a common signal for linear identification and has been reported for identification experiments with an HCCI engine before [25]. For nonlinear identification the PRBS with two modulated values is not sufficient as for this also a modulation of the signal's amplitude is a requirement [26].

Because Controlled Auto-Ignition combustion is very sensitive to the actuators especially at the boundaries of the operational window, a random variation of the exciting signal's amplitude was not applied for safety reasons. Within the possible actuation frame of two actuators, not every combination leads to auto-ignition. For example, a valve timing with a very high amount of residual gas cannot be combined with a large injected fuel mass. Therefore,

a combination of two excitation signals is chosen for the identification of the system. The revolution speed n_{rev} is approximately kept constant at 2000 rpm as the test bench is not suitable for transient revolution conditions. Thus, the resulting sample time is around 0.06 seconds. Because the test bench's brake is not capable of enforcing an exactly constant revolution speed, but shows a distinct dynamic during load steps for instance, this is monitored during the experiment. Further influencing effects are the coolant temperature $T_{CW,in}$ and the temperature of the aspired air $T_{air,in}$ which are thus logged, too.

The identification measurements are carried out for different load operation points, namely 1 bar, 2 bar, 3 bar, and 4 bar IMEP. The actuators End of Energizing EE, Energizing Duration ED, and Exhaust Valve Close EVC are operated such that best combustion properties are achieved in the basis setting in terms of static combustion stability. Steps in the manipulated variables are carried out from this optimum condition while the actuator signals are superposed with the PRBS. The steps are defined to reach half and afterwards full way to the extreme actuator values, with which operation is barely possible. The PRBS is of low amplitude, *i.e.* under 5% of the size of the identification step. As the experiments are carried out over the whole load operation range, the non-linear influence of the actuators on the process output variables IMEP, CA50, and dp_{max} can be measured and hence be identified in the relevant operation range. Since n_{rev} and $T_{air,in}$ are conditioned values which can hardly be actuated, they only are recorded but not excited. $T_{air,in}$ is conditioned to approximately 50°C. The remaining influencing variable $T_{CW,in}$ is accounted for by measurements at coolant temperatures of approximately 80°C and 100°C. Figure 7 shows the steps in EVC from the optimum calibration full way to the operation boundaries. Note that the overlaying PRBS are shifted against each other such that the sequences do not coincide.

For identification purposes a significant data set is needed. As two different models will be identified in the following and an additional data set is necessitated for the validation of the models, the complete surveying is carried out triply. Thereby for each model identification a different data set can be used, while the validation can be carried out using one of the two spare data sets. Subsequently the data is revised concerning runaway values. The last remaining step is a preprocessing of the data for the identification. Each training data set is freed from means and averaged over the standard deviation. Note that consequently the identified model as well as all elements of the controller, which are based on this model, operate in this scaled environment. Before the further processing, all manipulated and measured values have to be unscaled or scaled, respectively. All outputs modeled in the following are assumed as second order nonlinear discrete-time differential equations. This order gave best fitting results for CAI

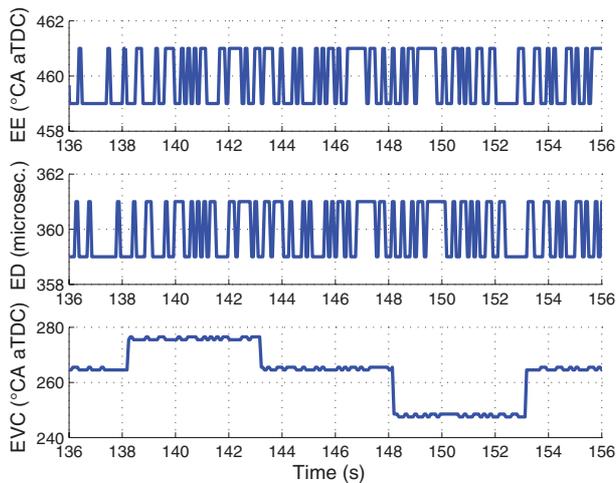


Figure 7

Detail from the signal sequence used for the identification. Shown is a step in the valve timing EVC while EE and ED are kept at a constant mean value.

combustion, which was reported for the related diesel HCCI combustion, too [27].

2.2.1 Simulation Model

In the following, the controller tuning will be carried out by evaluation of the response of this model instead of tuning the controller at the real engine. Thus, the accuracy of this model is inherently important for the controller performance. Because the coeval identification of all outputs of a MIMO system is quite complex, this is reduced to multiple MISO identifications here in order to achieve highest accuracy of the model. Each of the system's outputs IMEP, CA50, and dpmax is modeled separately by one NNSS net. To allow for the perturbation of the model, while it is used for the controller layout and review, this model receives six inputs in total. These are the manipulated and influencing variables mentioned before. All three part models were trained as an NNSS model with 12 hidden hyperbolic tangent and 2 linear output units which results in 68 weights per part-model.

The complete composed model is given by the parallel simulation of the three MISO models which all receive the same six inputs. All outputs are modeled as second order NNSS. This model also is used as an HIL model in the following. The modeling results are demonstrated by the evaluation of an actuated load step sequence in IMEP, namely 3 bar - 4 bar - 3 bar - 2 bar - 3 bar. The model of second order is not able to reproduce the noise present in the measurements solely on basis of the six inputs. Therefore, this attribute of the engine is addressed by separate modeling. The outputs of the model are superposed with white noise

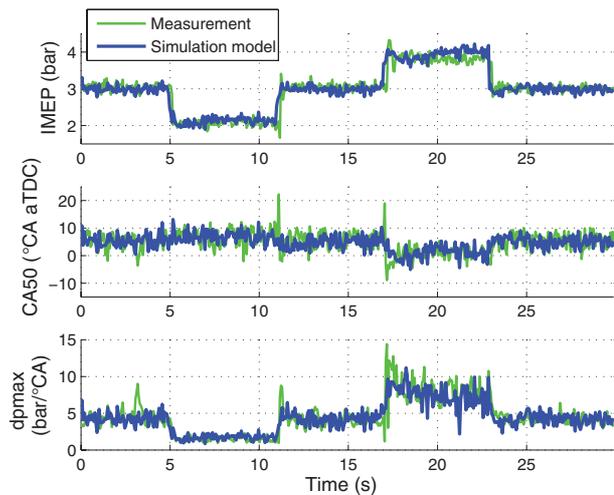


Figure 8

Composed CAI simulation model with superposed noise.

which is approximately of the same magnitude as seen in the measurements. As the noisy behavior of dpmax increases linearly with the load, this is addressed by a load-dependent amplification of the noise. Results from the previously described load steps with the complete simulation model are shown in Figure 8.

2.2.2 Observer Models

The main difference between simulation model and the created observer models is that at least the controlled variables IMEP and CA50 are modeled as a MIMO system and the model only receives the manipulated variables. All further influences will be accounted for by an Extended Kalman filter described in Section 3.

The controlled variables of the MPC to be developed will be IMEP and CA50. If no further conditions are considered, a model calculating these two values as a function of the manipulated variables is needed. This yields a 3×2 -MIMO structure which is modeled using an NNSS with 4 linear and 4 hyperbolic tangent neurons in the hidden layer and 4 linear nodes in the output layer which results in 33 weights. Validation Data is presented in Figure 9. The linearized model has consequently 4 states and 2 outputs of second order. If additional inputs or outputs are required, these have to be added to the structure of the neural net that has to be trained. The net should be extended by additional weights in the hidden and output layer in a way that yields the desired accuracy. Adding an output will consequently also add an output and state to the linearized model. For every order of the added output, one additional state will arise, compare Section 2.1. Adding further inputs may cause an increase of weights in the two layers. However, the linearized model

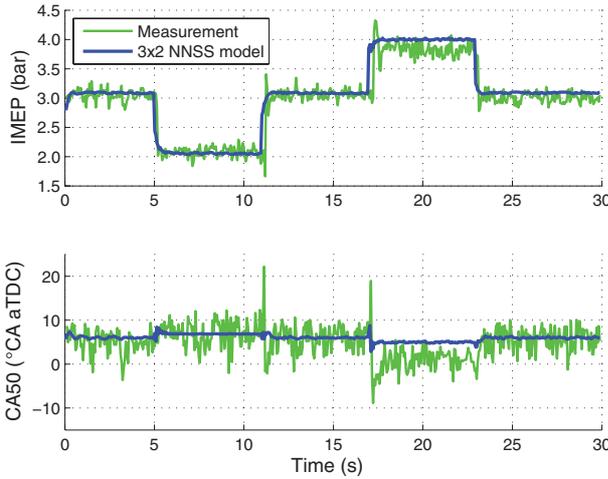


Figure 9

Identified nonlinear observer model in 3×2 -MIMO structure. Shown is the same experimental data as in Figure 8.

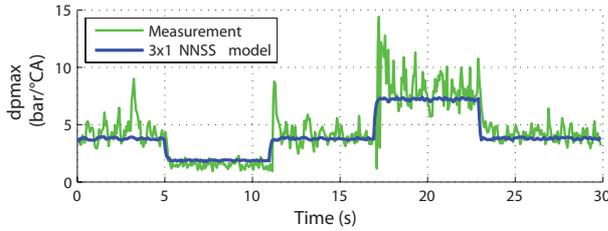


Figure 10

Identified nonlinear and linear observer model in 3×1 -MISO structure for the bounded variable dpmax. Shown is the same experimental data as in Figure 8.

will just show a further column added to matrix \mathbf{B} in that case.

In Section 4, the optimization problem is split into a controlled and a constrained part. The controlled part is addressed by the 3×2 model, the bounded part has to be covered by a separate 3×1 model. This is composed out of 3 linear and 3 tangent hyperbolic activation functions in the hidden layer and two linear units in the output layer. This in total gives 17 weights. Figure 10 demonstrates the nonlinear observer model for dpmax.

3 OBSERVER DESIGN

The further steps of the RCP-procedure are controller design, coding, and controller testing. The controller has to fulfill multiple tasks in parallel. In order to allow for the best possible results despite strong noise in the measured signal, the controller is based on an EKF that holds the described NNSS model.

This nonlinear EKF was proved to be a valuable observer for determining the system’s states [28]. The most important part for the tuning of the EKF is the correct estimation of the covariance matrices and of the disturbance model. The latter is laid out as a linear state disturbance model affecting all states and outputs of the nonlinear model. The states are augmented by integrating states equal to the number of measured outputs. It can be shown that the total number of integrating states for disturbance rejection must be smaller or equal to the number of measured outputs [29]. The disturbance model is revealed by the linear matrices \mathbf{B}_d and \mathbf{D}_d forming the nonlinear augmented system (7).

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + \mathbf{B}_d d_k \\ d_{k+1} &= d_k \\ y_k &= \mathbf{C}x_k + \mathbf{D}_d d_k \end{aligned} \quad (7)$$

For detectability in the linearized case postulate (8)

$$\text{rank} \begin{bmatrix} \mathbf{I} - \mathbf{A}_k & -\mathbf{B}_d \\ \mathbf{C} & \mathbf{D}_d \end{bmatrix} = \dim(x) + \dim(d) \quad (8)$$

has to be fulfilled [29], with $\dim(x)$ as the number of states and $\dim(d)$ as the number of integrating disturbances. \mathbf{A}_k is the linear system matrix derived from the nonlinear system at time k . The observer structure is given in Figure 11. Here the resulting EKF gain (9) is updated in every sample.

$$\mathbf{L}_{EKF,k} = \begin{bmatrix} \mathbf{L}_{EKFx,k} \\ \mathbf{L}_{EKFd,k} \end{bmatrix} \quad (9)$$

Details on the used form of EKF can be found in several publications, e.g. [30]. In order to preserve the identified NNSS dynamics, \mathbf{B}_d and \mathbf{D}_d are set to the zero and the identity matrix for the calculation of the observer’s output \hat{y}_k . However, for the iterative calculation of $\mathbf{L}_{EKF,k}$ they are tuned automatically assuming the green structure in Figure 11.

Each of the two controller model parts requires the tuning of two separate observers for the controlled and the

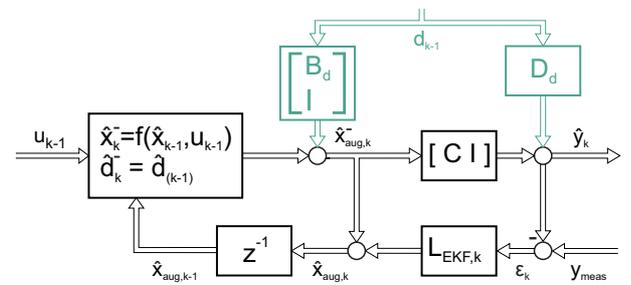


Figure 11

Structure of the implemented EKF.

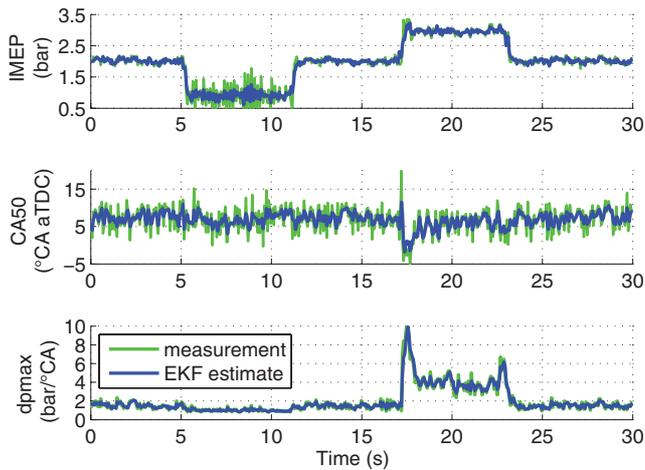


Figure 12

Outputs of the automatically tuned EKFs. The observers are based on the same models as in Figures 9 and 10.

bounded variables. The observer-tuning can be carried out by an automated MATLAB/SIMULINK[®] environment, in which the Sum Squared Error (SSE) between measurement and observed system output is minimized by the adjustment of the disturbance model and the covariance matrices of the EKF. Odelson *et al.* [31] developed an autocovariance least-squares method for estimating noise covariances which is applicable strictly to linear systems though. Here a custom-made search algorithm was implemented for finding the entries of matrices \mathbf{B}_d and \mathbf{D}_d and the covariances simultaneously. Postulate (8) is checked afterwards. In this manner, the tuning of the EKF can be inserted as a part of the automation chain along the RCP-procedure. Figure 12 demonstrates the observers' performance and the filtering of too strong measurement noise.

4 CONTROLLER DESIGN

The principle of Model-based Predictive Control is founded on the minimization of a cost function which weights the deviation of the predicted controlled variables from predicted set point trajectories. Figure 13 depicts the components of the NMPC and their interaction. The deviation between predicted set point \hat{w}_k and predicted system output \hat{y}_k is regarded in a time frame from lower to upper prediction horizon, from N_1 to N_2 . As the engine combustion is a very fast process without any delays, the choice $N_1 = 1$ is adequate. The superscript “ $\hat{\cdot}$ ” indicates that these future values are calculated predictions. The change of the actuated variable Δu_k is considered from the actual time step until the control horizon N_u and held constant for all predictions exceeding N_u , see Figure 14. Both terms are quadratically

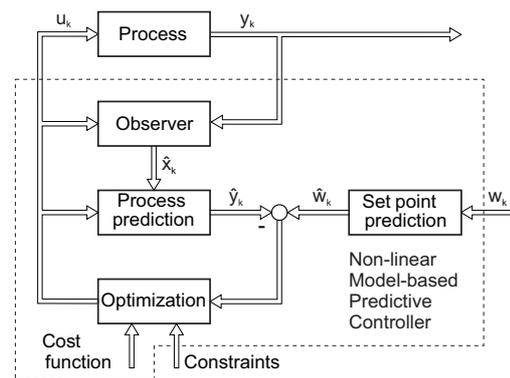


Figure 13

Schema of the NMPC.

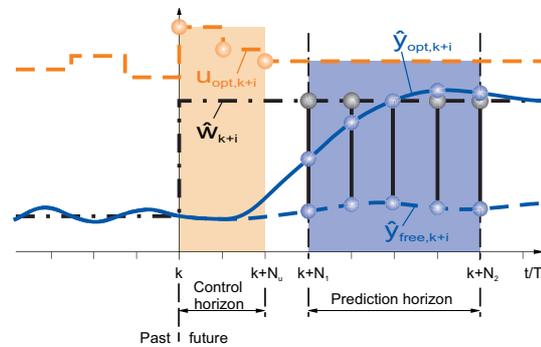


Figure 14

Principle of Model-based Predictive Control.

weighted with the matrices Γ or Λ respectively. The cost function for the case presented is given by (10).

$$J_k = \sum_{i=1}^{N_2} \left((\hat{w}_{k+i} - \hat{y}_{k+i})^T \Gamma (\hat{w}_{k+i} - \hat{y}_{k+i}) \right) + \sum_{j=1}^{N_u} \left(\Delta u_{k+j}^T \Lambda \Delta u_{k+j} \right) \quad (10)$$

4.1 Predictions

The controller is tuned by the adjustment of the time horizons and of the weighting matrices. The first term of (10) is dependent on the prediction of the system's behavior which is calculated based on the EKF's estimation of the system's states. In the case presented, the prediction is determined with the nonlinear NNSS model. The observed disturbance of the current time instant k , on which the predictions are based, is assumed to be constant over the prediction horizon, $d_k = \dots = d_{k+N_2}$, see Figure 11. By means of a linearization

of the nonlinear function $f(x_{k-1}, u_{k-1})$ with every predicted time step from $(k+1)$ to $(k+N_2)$ the process' nonlinearities are accounted for and new matrices $\widetilde{\mathbf{A}}_k$ and $\widetilde{\mathbf{B}}_k$ are calculated following (11).

$$\widetilde{\mathbf{A}}_k = \frac{\partial f(x_{k-1}, \widehat{u}_{k-1})}{\partial x_{k-1}}, \quad \widetilde{\mathbf{B}}_k = \frac{\partial f(x_{k-1}, \widehat{u}_{k-1})}{\partial \widehat{u}_{k-1}} \quad (11)$$

Details on the linearization of the used model can be found in [17]. In the following, the system will be augmented for the calculation of the control moves as in (12) in order to achieve a cost function with zero minimum which is desirable for the optimization.

$$\begin{aligned} \widehat{x}_k &= \begin{bmatrix} x_k \\ \widehat{u}_k \end{bmatrix}, \mathbf{A}_k = \begin{bmatrix} \widetilde{\mathbf{A}}_k & \widetilde{\mathbf{B}}_k \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \mathbf{B}_k = \begin{bmatrix} \widetilde{\mathbf{B}}_k \\ \mathbf{I} \end{bmatrix}, \\ \mathbf{C}_c &= [\mathbf{C} \ \mathbf{0}] \end{aligned} \quad (12)$$

Over the whole control horizon a new set of actuated variables is calculated of which only the first set is applied in a receding horizon manner. The remaining of the calculated control moves can be used for the prediction of the future system's behavior. Because in the previous sample these actions were determined to be optimal, they will be a better assumption for the future than simply keeping the control moves constant which was proposed for the linear case [32]. The term $\widehat{u}_{(k+i|k-1)}$ addresses the vector of actuated variables which was calculated for step $(k+i)$ at the previous sample $(k-1)$. The dependency of the predicted process output on the control moves is derived to (13), (14) and (15). The moves $\Delta u_{(k+i)}$ are expected to be small and superposition of the nonlinear prediction and the time variant linear influence of $\Delta u_{(k+i)}$ is assumed to be valid. For all $i > N_u$ the controller output is held constant.

$$\widehat{x}_{(k+i)} = f(x_{(k+i-1)}, \widehat{u}_{(k+i-1|k-1)}) = f(\widehat{x}_{(k+i-1)}) \quad (13)$$

$$\approx \mathbf{A}_{(k+i-1)} \widehat{x}_{(k+i-1)} + \mathbf{B}_{(k+i-1)} \Delta u_{(k+i-1)}$$

$$\widehat{y}_{(k+i)} = \mathbf{C} \cdot x_{(k+i)} + d_k = \mathbf{C}_c \cdot \widehat{x}_{(k+i)} + d_k \quad (14)$$

$$\approx \mathbf{C}_c \mathbf{A}_{(k+i-1)} \widehat{x}_{(k+i-1)}$$

$$+ \mathbf{C}_c \mathbf{B}_{(k+i-1)} \Delta u_{(k+i-1)} + d_k$$

$$= \mathbf{C}_c \mathbf{A}_{(k+i-1)} \mathbf{A}_{(k+i-2)} \widehat{x}_{(k+i-2)}$$

$$+ \mathbf{C}_c \mathbf{A}_{(k+i-1)} \mathbf{B}_{(k+i-2)} \Delta u_{(k+i-2)}$$

$$+ \mathbf{C}_c \mathbf{B}_{(k+i-1)} \Delta u_{(k+i-1)} + d_k$$

⋮

$$\widehat{y}_{(k+i)} = \mathbf{C}_c \left(\prod_{n=1}^i \mathbf{A}_{(k+i-n)} \right) \widehat{x}_k + \mathbf{C}_c \Theta_i + d_k \quad (15)$$

$$\Theta_i = \begin{cases} \sum_{n=1}^{i-1} \left(\left(\prod_{m=1}^n \mathbf{A}_{(k+i-m)} \right) \cdot \right. \\ \quad \left. \cdot \mathbf{B}_{(k+i-n-1)} \Delta u_{(k+i-n-1)} \right) + \\ \quad + \mathbf{B}_{(k+i-1)} \Delta u_{(k+i-1)}, & \text{if } i > 1 \\ \mathbf{B}_k \Delta u_k, & \text{if } i = 1 \end{cases}$$

The first term of the right hand side in (15) is also referred to as "free system response" as it reflects an assumed system

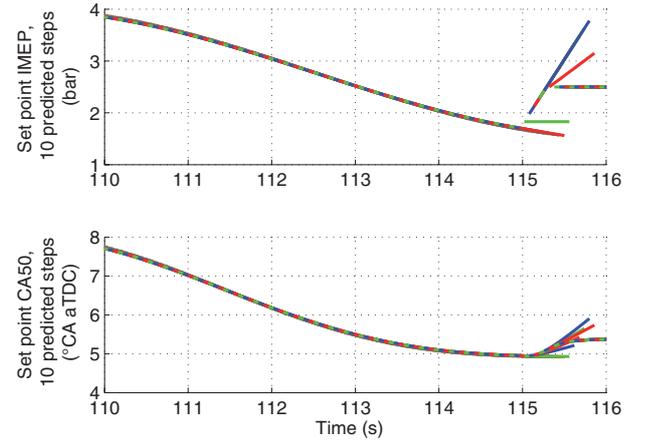


Figure 15

Exemplary predictions of the set points during the sine wave and ramp from profile shown in Sections 5 and 6. A bar depicts the set point predicted at the time instant before its begin.

behavior which arises without further control actions. Nevertheless, here it is no longer truly "free" because it depends on predicted control moves. This first part of the prediction is re-substituted by the nonlinear function $f(x_k, u_k)$ as in (16).

$$\widehat{y}_{(k+i)} = \overbrace{\mathbf{C}_c f^{(i)}(\widehat{x}_{(k)})}^{\widehat{y}_{free,(k+i)}} + d_k + \mathbf{C}_c \Theta_i \quad (16)$$

The exponent (i) declares the embedding of the function in a *for*-loop with i runs which grants a nonlinear prediction of the states. In the case presented, it is based on the optimal control moves $\widehat{u}_{(k+i|k-1)}$ calculated previously in the last sample $(k-1)$.

For completion of Equation (10), a prediction of the set point values \widehat{w}_k is necessary. The set point trajectory for CA50 is constituted by a quadratic function determining the optimal value for CA50 dependent on the desired IMEP. The gradient in the requested load is limited to ± 2.5 bar/second or ± 0.15 bar/cycle for proper engine operation. In consequence the gradient of the set point for CA50 is limited, too. The changes in the set points are predicted using IT_1 -dynamics with small time constants. Figure 15 shows a detail from the set point profile shown later in Sections 5 and 6 with an exemplary ten-steps-ahead prediction which equals the upper limit implemented for the automated determination of N_2 . A revolution speed of 2000 rpm results in a sample interval of 0.06 seconds. The maximum predictions therefore last 0.6 seconds. This can be recognized by the bars of different color for each sample step in Figure 15. Their length symbolizes the prediction horizon. Obviously the prediction fits nicely future set points as long as these are a nearly continuous function since the

bars overlay each other during the sine wave before 115 s in Figure 15. But even in the case of discontinuities, like at 115 s in Figure 15, this manner leads to a reasonable deviation between predicted and applied set point.

4.2 Cost Function Minimization

The actual control moves are found from the solution of a quadratic minimization problem derived from the cost function (17) subject to constraints on the actuators and on dpm_{\max} .

$$\begin{aligned} \min_{\Delta U_k} J_k &= \frac{1}{2} \Delta U_k^T \mathbf{H}_{qp,k} \Delta U_k + \mathbf{G}_{qp,k} \Delta U_k \quad (17) \\ \mathbf{H}_{qp,k} &= \mathbf{H}_k^T \Gamma_{qp} \mathbf{H}_k + \Lambda_{qp} \\ \mathbf{G}_{qp,k} &= 2 \mathbf{H}_k^T \Gamma_{qp} (\widehat{W}_k - \widehat{Y}_{free,k}) \\ &\text{s.t.} \\ \mathbf{A}_{qp,k} \cdot \Delta U_k &\leq \mathbf{B}_{qp,k} \end{aligned}$$

In (17) the capital letters for the vectors \widehat{W}_k , $\widehat{Y}_{free,k}$ and ΔU_k imply that they contain all vectors \widehat{w}_{k+i} , $\widehat{y}_{free,k+i}$ with $i \in \{1, \dots, N_2\}$ or Δu_{k+i} with $i \in \{1, \dots, N_u\}$, respectively. Γ_{qp} and Λ_{qp} contain the weights Γ or Λ from (10) as diagonal entries N_2 -times or N_u -times, respectively. The matrix H_k sums up the linearized time variant dependencies of the predicted system response from (16) on ΔU_k to every predicted time step as revealed by (18).

$$\begin{aligned} \mathbf{H}_k &= \begin{bmatrix} \Phi_{1,1} & \cdots & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \Phi_{N_2-1,1} & \cdots & \cdots & \Phi_{N_2-1,N_u-1} \\ \Phi_{N_2,1} & \Phi_{N_2,2} & \cdots & \Phi_{N_2,N_u} \end{bmatrix} \quad (18) \\ \Phi_{i,j} &= \begin{cases} \mathbf{C}_c \cdot \left(\prod_{m=1}^{i-j} \mathbf{A}^{(k+i-m)} \right) \cdot \\ \quad \cdot \mathbf{B}^{(k+j-1)}, & \text{if } i > j \\ \mathbf{C}_c \mathbf{B}^{(k+j-1)}, & \text{if } i = j \\ \mathbf{0}, & \text{if } i < j \end{cases} \end{aligned}$$

In [16] an NMPC was demonstrated which used a 3×3 NNSS model with only two controlled outputs, while the third was bounded. This output was not weighted in the cost function but only used for the calculation of constraints $\mathbf{A}_{qp,k} \cdot \Delta U_k \leq \mathbf{B}_{qp,k}$. Therefore, the dimension of the stated quadratic problem was higher than necessary. As the optimization code contains iterative loops, it is most important for fast calculation times to keep the dimension of the problem as low as possible. The time available for the calculation of the observations and the solution of (17) is limited by the revolution speed. The calculation of controlled and limited output variables were separated by means of two individual observers to reduce the optimization problem to the dimension of the controlled output variables. Hence, Equation (16) has to be formulated for each of the observers. In this way,

the SSE could be improved in closed-loop simulations compared to the conventional approach [12].

Over the whole prediction horizon the compliance of the solution with constraints on the actuators and constraints on the system's outputs has to be ensured. Thus, the matrix $\mathbf{A}_{qp,k}$ and the vector $\mathbf{B}_{qp,k}$ can be segmented into two parts (19).

$$\mathbf{A}_{qp,k} = \begin{bmatrix} \mathbf{A}_{qp,\Delta u,k} \\ \mathbf{A}_{qp,y,k} \end{bmatrix}, \quad \mathbf{B}_{qp,k} = \begin{bmatrix} \mathbf{B}_{qp,\Delta u,k} \\ \mathbf{B}_{qp,y,k} \end{bmatrix} \quad (19)$$

As the cost function (17) depends on ΔU_k , all constraints in the absolute value of U_k have to be reformulated dependent on ΔU_k . The minimization of the cost function depends on the controlled variables assumed for the prediction in terms of $\widehat{Y}_{free,k}$. Because superposition is assumed to be valid, the control command $U_{control,k}$ is set by (20).

$$U_{control,k} = \widehat{U}_{(k|k-1)} + U_{\Delta,k} \quad (20)$$

Here the first term regards the vector of control moves from the previous optimization. The second term addresses the future control moves. As the system is realigned, the absolute signal $u_{\Delta,k+i}$ has to be summed up (21).

$$u_{\Delta,k+i} = 0 + \Delta u_{k+1} + \dots + \Delta u_{k+i} \quad (21)$$

The limits on the actuators are claimed to be constant over the prediction horizon. The matrices \mathbf{I} and $\mathbf{0}$ are the identity or the zero matrix of dimension $\text{dim}(u)$ equal to the number of actuated variables. This yields the formulation of $\mathbf{A}_{qp,\Delta u,k}$ and $\mathbf{B}_{qp,\Delta u,k}$ as in (22) and (23).

$$\mathbf{A}_{qp,\Delta u,k} = \begin{bmatrix} \overbrace{\mathbf{I} \ \mathbf{0} \ \cdots \ \mathbf{0}}^{\text{dim}(u) \cdot N_u} \\ \mathbf{I} \ \mathbf{I} \ \cdots \ \mathbf{I} \\ \vdots \ \ddots \ \ddots \ \mathbf{0} \\ \mathbf{I} \ \mathbf{I} \ \cdots \ \mathbf{I} \\ -\mathbf{I} \ \mathbf{0} \ \cdots \ \mathbf{0} \\ -\mathbf{I} \ -\mathbf{I} \ \cdots \ \mathbf{I} \\ \vdots \ \ddots \ \ddots \ \mathbf{0} \\ -\mathbf{I} \ -\mathbf{I} \ \cdots \ -\mathbf{I} \end{bmatrix} \quad (22)$$

$$\mathbf{B}_{qp,\Delta u,k} = \begin{bmatrix} u_{\max} - \widehat{u}_{(k+1|k-1)} \\ u_{\max} - \widehat{u}_{(k+2|k-1)} \\ \vdots \\ u_{\max} - \widehat{u}_{(k+N_u|k-1)} \\ -u_{\min} - \widehat{u}_{(k+1|k-1)} \\ -u_{\min} - \widehat{u}_{(k+2|k-1)} \\ \vdots \\ -u_{\min} - \widehat{u}_{(k+N_u|k-1)} \end{bmatrix} \quad (23)$$

The second part of the constraints concerns the system's outputs. With (16) the correlation of future system output

and optimization variable ΔU_k is given. $\mathbf{A}_{qp,y,k}$ and $B_{qp,y,k}$ are stated by (24) and (25) referring to (18) and (16).

$$\mathbf{A}_{qp,y,k} = \begin{bmatrix} \mathbf{H}_k \\ -\mathbf{H}_k \end{bmatrix} \quad (24)$$

$$B_{qp,y,k} = \begin{bmatrix} Y_{max} - \widehat{Y}_{free,k} \\ -Y_{min} - \widehat{Y}_{free,k} \end{bmatrix} \quad (25)$$

Concluding the cost function (17) and its constraints are formulated. Although not only the predictions are nonlinear but also the impact of a change in the actuated variables is modeled as nonlinear in time, the resulting optimization problem is convex due to the assumption of superposition to be valid. For the actual minimization of the cost function, a custom-made optimization routine is implemented. The minimization of convex functions subject to constraints is a well-known problem for which different approaches exist [33]. Following the “barrier-function”-method, a convex solver was programmed which gives very good results. To ensure feasibility, the constraints are defined as “soft”. This means they can be violated but every disrespecting leads to an increase of the costs weighted with the penalty factor ρ . As the code has to be compiled for usage on the RCP system, it has to be processed by the corresponding compiler. Hence, the observer, cost function, and optimizer are coded in C and embedded into a SIMULINK[®] C-s-function. Details on the optimization routine are given in [12].

4.3 Controller Tuning

The controller s-function is coded in a way that the tuning parameters N_2 , N_u , Γ , Λ , ρ , the time constants for the prediction of the set points, and additional tuning parameters for the optimizer are adjustable. These are tuned in an automated loop-shaping procedure. The SIMULINK[®] model is simulated in iterative loops and contains the engine simulation model, the controller, and a dynamic set point profile like shown in Sections 5 and 6 which is used as a standard test case. During the Model-In-the-Loop (MIL) test, the simulation model is overlaid with white Gaussian noise as found with the measurements. The MIL-test is simulated for 160 s. The SSE between the set point profile and the simulated engine output is evaluated as performance index. For the constrained variable, there is no set point. Therefore, if the limit on dp_{max} is exceeded, this violation is recognized in the calculation of the SSE by adding the value of dp_{max} weighted by a penalty factor. In this way, the compliance with the constraint on dp_{max} is advantaged but not enforced. The only required user input for the automated tuning procedure is the definition of disturbances, a set of start parameters for the tuning process, and a search interval per tuning parameter.

5 CONTROLLER IMPLEMENTATION

The different controller setups were tested in Model-In-the-Loop tests (MIL) so far. The real-time capabilities of the controllers were not considered.

For this investigation, the controllers have to be tested on the target hardware system, an ETAS ES1135 simulation board. A common and accepted requirement in the automotive industry is that a Hardware-In-the-Loop test (HIL) has to prove the applicability of the hardware which has the complete software functionality implemented. For that an HIL test simulator is needed.

HIL Test Bed

As the custom-made ECU at the test bench is realized on a dSPACE MicroAutoBox, the same hardware will be used as HIL simulator. Therefore, the data transfer can be kept exactly the same as with the engine application. The data connection is based on the same CAN protocol mentioned in Section 1. The HIL simulator runs the composed nonlinear model with superposed noise presented in Section 2.2.1 and is to substitute the whole engine test bed. Figure 16 shows the implementation of the HIL simulation which is compiled onto the dSPACE hardware using the appropriate MATLAB[®] real-time workshop target. The model receives the manipulated variables and sends the engine model’s response in IMEP, CA50 and dp_{max} . All three outputs are superposed with suitable white noise as before. The sample time was chosen to 0.06 seconds corresponding to a revolution speed of 2 000 rpm.

All controllers have to fulfill the specification for a cycle-based control. The resulting calculation time in this HIL test has to be fast enough to guarantee the calculation of a control action at least within the time frame given by the reception of the next CAN-message containing the next set of measured variables. For the current work, a static revolution speed of 2 000 rpm is considered which results in a 60 milliseconds time frame for the calculation of the controller.

For controlling the cycle following directly to that cycle of which the controlled variables were calculated from, the

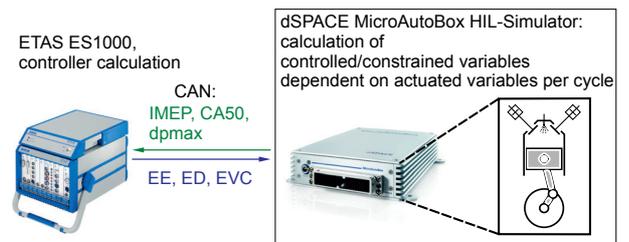


Figure 16

Schema of the HIL test setup.

calculation has to be faster than 5.5 milliseconds. This reduced time frame is caused by the the time required for calculating the values of the controlled variables by the custom-made ECU and transmitting the required signals. Hence, in the following for every controller the maximum time $t_{\text{turn,max}}$ is given that was required by the real-time platform to calculate the controller action. This allows the evaluation of the approaches and their further ranking as $t_{\text{turn,max}}$ has to be less than 60 milliseconds, see Table 3. For comparison, gain-scheduled PID-control was developed which does not respect constraints. ED is used to control IMEP, while EE and EVC control coevally CA50. The scheduling variable is the set point for IMEP. The PID-controller is based on the same but filtered inputs as the NMPC and tuned in the same loop-shaping procedure. The PID-controller is simulated with exactly the same influencing conditions as with the NMPC. Details on the PID-controller can be found in [12]. The result achieved with this gain-scheduled PID-controller is given for the test results in the following.

However, with true test bed conditions some influences are changing randomly which causes disturbances affecting the closed-loop control. Therefore, in the following the evaluation conditions of the controllers are changed to a setup more demanding than the test bench standards should reach. For this test, the controllers are not tuned again. The temperatures of air and coolant are overlaid with a sine wave and white noise. Such $T_{\text{air,in}}$ is modulated by 10% from 45°C to 55°C, while $T_{\text{CW,in}}$ ranges from 80°C to 100°C. By the identification of the brake's dynamic the revolution speed can be modeled as a time-discrete function of the current load. This test is close to the real application, but the revolution speed depends on the controlled variable IMEP and therefore on the controller performance which therefore is harder to compare. Moreover, the simulated control-loop including a model of the brake can easily become unstable and the controller has to meet the set point profile even more exactly. The results involving the simulation of the engine brake's dynamics demonstrate that the revolution speed is an important influence on the CAI process.

The influencing variables of Figure 17 have to be kept in mind when judging the result shown in Figures 18 and 19. A perturbation of the aspired air's and the coolant water's temperature in the assumed magnitude and dynamic is unlikely. The test is set up drastically in order to find out the limitations of the controllers' capabilities.

The presented NMPC approach achieves a more persuasive control result than the benchmark PID-controller. The HIL robustness test shows that the automated tuning of the NMPC with the proposed standard test-case leads to a controller that is robust against disturbances. The same automated tuning of gain-scheduled PID-control including a pilot control based on the identification experiments leads to a good control result for the training case, but achieves poor robustness against further disturbances. This result indicates

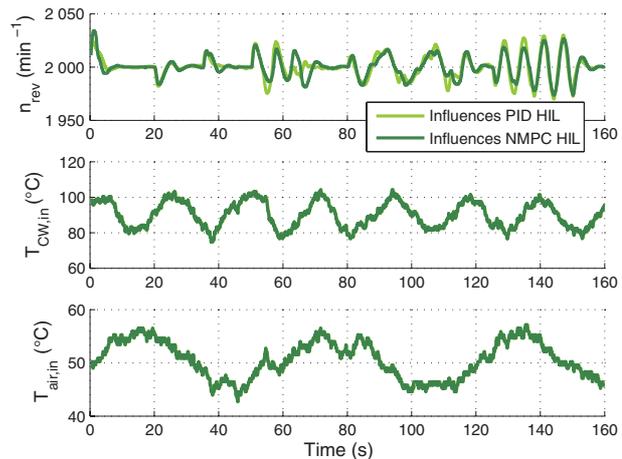


Figure 17

Influencing variables for the robustness test of the controller. The revolution speed is enforced as a DT_1 -dynamic dependent on IMEP for all controllers individually. Shown is the simulated brake's response to the control result in IMEP of Figures 18 and 19.

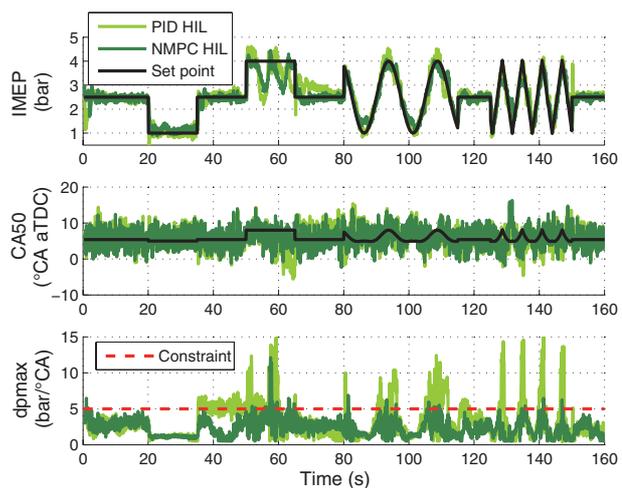


Figure 18

HIL robustness test results, controlled variables. The result of gain-scheduled PID-control is given for comparison.

TABLE 3

HIL-results of the NMPC and gain-scheduled PID controller as in Figures 17 through 19. The SSE is given separately for the controlled and the constrained variables

Controller	SSE controlled	SSE constrained	$t_{\text{turn,max}}$ (ms)
NMPC	1 843.0	421.3	2.783
PID	2 545.0	5 222.7	0.037

the superiority of NMPC compared to linear conventional control methods in this application, even without considering the ability to respect constraints.

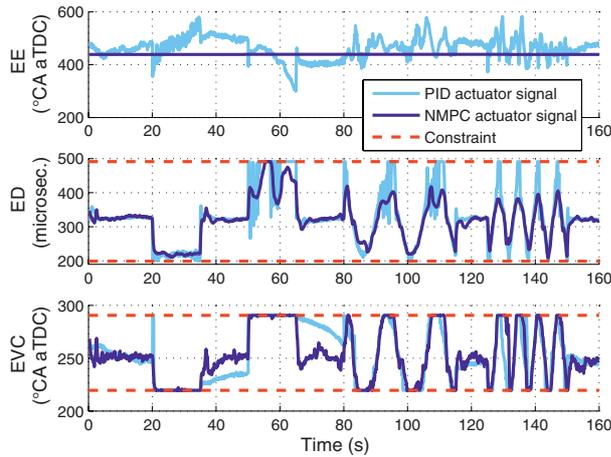


Figure 19

HIL robustness test results, manipulated variables. The result of gain-scheduled PID-control is given for comparison.

6 ENGINE CONTROL RESULTS

The last step of the RCP V-model is the application to the real process. As the conditions cannot be kept the same for a benchmark, an HIL test is carried out for which the measured conditions are enforced on the HIL simulation model. The benchmark PID-controller is evaluated in this HIL test. Consequently, the measured NMPC result can be compared to the result that gain-scheduled PID-control would have reached under the same test-bench conditions.

CAI is a very sensitive combustion mode which requires certain arrangements for starting a closed-loop control test and can only be started directly without a previous SI combustion if strong preheating of the intake air is used. This is not desirable for an engine application. Therefore, the CAI mode is switched on from conventional SI combustion. For the given test results the engine was manually switched to CAI at a medium load of 3 bar IMEP. From this initial situation the test was started. The transition from SI to CAI operation or back is not part of the measurements.

Figure 20 shows the results achieved using the created RCP-environment. The shown set point for CA50 is calculated dependent on the IMEP set point. The limit for d_{pmax} is set to 5 bar/CA.

The conditioning system for the aspired air holds strong electric heaters which are actuated by the test bench control in this experiment for preventing the temperature to fall. Consequently, strong disturbances arise at around 80, 130, and 150 seconds in Figure 20. Hence, the control result has to be reviewed with respect to the circumstances. The result shows misfires to the same time instances, at which the disturbances in $T_{air,in}$ arise. As pointed out in the introduction, one key influence of CAI is the temperature level at the end of the compression stroke which is obviously

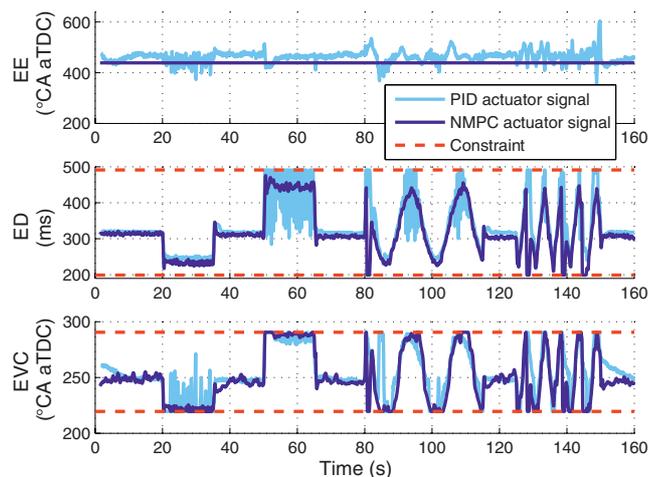
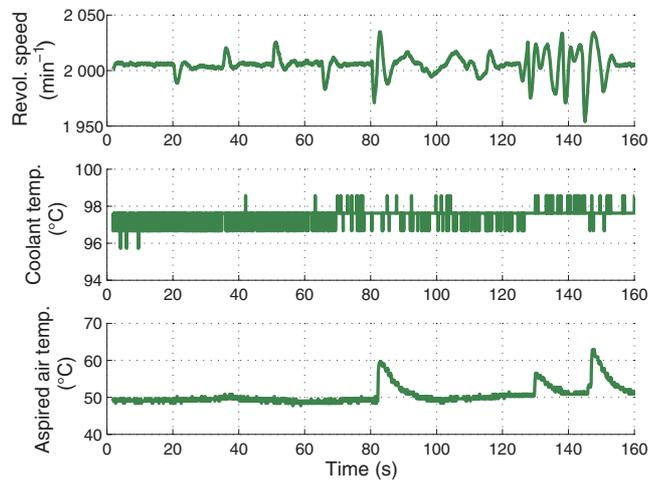
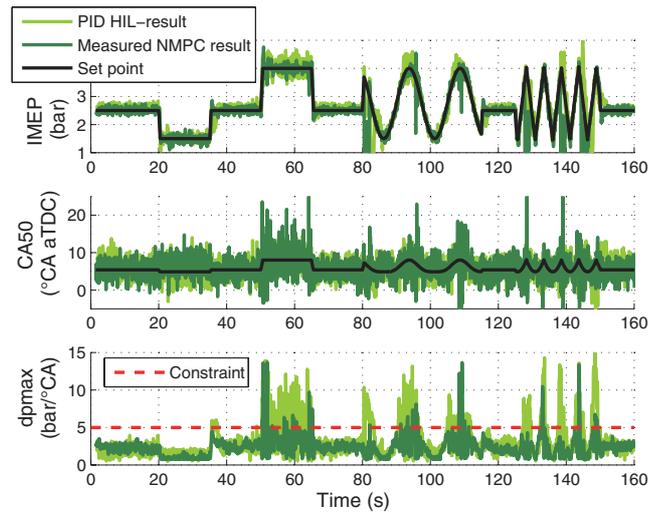


Figure 20

Test bench measurements from controller test. Controlled and bounded variables with set point profile and d_{pmax} limit, test bench conditions, and controller output with limits.

dependent on $T_{air,in}$. At 80 seconds also a load step occurs, and the set point for CA50 is shifted back. With higher load temperature the combustion is advanced. The offset between set point and measurement in CA50 causes the controller to reduce the injected fuel or ED and to adjust the EGR by decreasing EVC. This retards the combustion but also causes a collapse of IMEP which in turn makes the controller readjust ED and EVC. Similar actions can be seen at the points of time at which the disturbance in $T_{air,in}$ arises again. It is noted that the main influence of the end of enrgizing EE is on NO_x -formation, which could be shown from detailed testing [34]. Therefore, the NMPC only makes little use of this control edge as it “knows” of the little impact of this variable while as well the pre-control as the gain-scheduling of the PID-controller alter all of the manipulated variables regardless the influence. This would change if *e.g.* the NO_x formation would be considered, too.

Though, despite the disturbances, the controllers were able to keep the process alive without using a spark, while only the NMPC could offer a limitation of the NVH characteristics. This clearly demonstrates the ability of the NMPC approach since the soft constraint was just rarely violated with higher loads, *e.g.* around 50 s. With the setup used in the experiments, CA50 is subject to strong noise which is caused by different aspects. On the one hand, the combustion process itself is subject to deviations in the pressure trace from cycle to cycle. This effect can possibly be reduced by closed loop control. On the other hand, a measurement noise is obviously always present. Further noise on the CA50 signal arises from its calculation. For this a model of the compression and the combustion in the cylinder is necessary which has to be quite basic for real time calculation on the ECU and might severely deviate from reality. The determination of CA50 on the tailor-made ECU was still under investigation when the experiments were carried out. This noise is not fully reproduced by the observer but filtered by the EKF. This results in the fact that only the filtered value of CA50 can be driven to the set point with the set up used; the noise is the same in open and closed loop operation. IMEP is weighted stronger in the controller because this variable is filtered less; accordingly the control error is recognized by the controller. Since high dpmax is caused by early CA50, a lower constraint on CA50 at TDC could help to further reduce dpmax. The automatically tuned controller parameters are given in Table 4.

Figure 20 demonstrates that the PID-controller is able to control IMEP and CA50 to the set points but the control

result is of lower quality. Moreover, the PID-controller cannot respect the constraint on dpmax while the NMPC can. This proves the ability of the NMPC approach. It is noted that also linear MPC was investigated but the results were not promising [12]. Either the dpmax-constraint was violated intensively or the set point profiles were not adequately met. To sum up, the experiment demonstrates a good robustness of the NMPC to disturbances because it could keep the combustion alive without the use of a spark despite strong disturbances.

SUMMARY AND CONCLUSIONS

A novel NMPC-approach for simultaneously regulating IMEP and CA50 while limiting dpmax of Gasoline Low-Temperature Combustion (Controlled Auto-Ignition) was developed. This work contributes to the state of technology by extending the complexity of the controllers, the control objective, and the enforced dynamics. The manipulated variables were reduced to the parameters of the injection EE, ED, and of the valve train which could be reduced to the event EVC. At the begin of the RCP procedure, identification experiments were carried out using the target real-time platform intended for the controller calculation. A combination of an PRBS and steps in the manipulated variables from an optimal operation condition to the operational limits were carried out while the process’ responses were measured. On this basis, the identification of Neural Network StateSpace models was carried out.

For the development of the observer, an automated tuning and identification environment with a nonlinear search algorithm was developed. The disturbance model, which is necessary for offset-free tracking MPC, was simultaneously estimated alongside the covariance matrices of the Extended Kalman filter. The prediction of the future set point trajectory was introduced by the interpretation of the gradient assuming a first order dynamic. The separation of bounded and controlled variables by means of two separate observers yielded advantages for the output-constrained control result.

The robustness of the machine-tuned controller was investigated by the implementation of an HIL test. Finally, the approach was justified by an exemplary application to the single-cylinder research engine. The highly dynamic load transient set point profile was enforced while constraints on dpmax were realized. For comparison, gain-scheduled PID-control was tested in an HIL test in which the boundary conditions of the measured controller test were impressed. The limitation of dpmax was successfully implemented just with the NMPC. This demonstrated the practical usage of the proposed methods which, in the end, all intend to speed up the controller application and to increase the quality of the optimization result.

Therefore, the MATLAB/SIMULINK® RCP environment was justified which enables the user to set up the controller

TABLE 4

Machine-tuned controller parameters

N_u	N_2	$\Gamma_{(IMEP,CA50)}$	$\Lambda_{(EE,ED,EVC)}$	ρ
5	7	diag(4.03 1)	diag(97001.7 0 1)	9.1009e+018

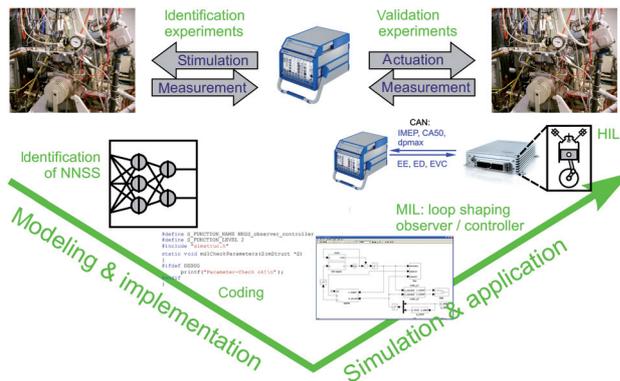


Figure 21

RCP V-model of the realized controller development.

from identification measurements within short time. Figure 21 illustrates the RCP V-model and its realization in the developed tool chain [12]. It will be extended by means of pollutant control and of parallel computing methods in future work [11].

ACKNOWLEDGMENTS

The authors thank D. Seebach, K.-G. Stapf and Prof. S. Pischinger from the Institute for Combustion Engines VKA, RWTH Aachen University, for fruitful discussions on the CAI-mechanisms and for providing a sophisticated test bench environment. Dipl.-Inform. Dipl.-Math. Kai Bollue is given thanks for his help in formulating compact matrix equations.

REFERENCES

- Lang O., Salber W., Hahn J., Pischinger S., Hortmann K., Buecker C. (2005) Thermodynamical and mechanical approach towards a variable valve train for the controlled auto ignition combustion process, *SAE paper* 2005-01-0762.
- Adomeit P., Sehr A., Weinowski R., Stapf K.G., Seebach D., Pischinger S., Hoffmann K., Abel D., Fricke F., Kleeberg H., Tomazic D. (2009) Operation strategies for controlled auto ignition gasoline engines, *SAE Int. J. Engines* **2**, 164-172.
- Stapf K.G., Seebach D., Fricke F., Pischinger S., Hoffmann K., Abel D. (2007) CAI-engines: Modern combustion system to face future challenges, *Proc. SIA Conference, Strasbourg, 28-29 November 2007*.
- Shaver G.M. (2005) Physics-Based Modeling and Control of Residual-Affected HCCI Engines Using Variable Valve Actuation *PhD Thesis*, Stanford University.
- Shaver G.M., Roelle M., Gerdes J.C. (2006) A two-input two-output control model of hcci engines, *Proc. American Control Conference, Minneapolis, Minnesota, USA, 2006*.
- Bengtsson J., Strandh P., Johansson R., Tunestl P., Johansson B. (2006) Hybrid control of homogeneous charge compression ignition (HCCI) engine dynamics, *Int. J. Control* **79**, 5, 422-448.
- Bengtsson J., Strandh P., Johansson R., Tunestl P., Johansson B. (2007) Hybrid modelling of homogeneous charge compression ignition (HCCI) engine dynamic - A survey, *Int. J. Control* **80**, 11, 1814-1848.
- Chiang C.J., Stefanopoulou A.G., Jankovic M. (2007) Nonlinear observer-based control of load transitions in homogeneous charge compression ignition engines, *IEEE Trans. Control Syst. Technol.* **15**, 438-448.
- Kulzer A., Fischer W., Karlemeyer R., Sauer C., Wintrich T., Benninger K. (2009) Homogeneous Charge Compression Ignition on gasoline engines, *MTZworldwide* **70**, 1, 32-39.
- Karlemeyer R., Graf G., Scherrer D., Fischer W., Hathout J.P. (2009) Control concepts for a gasoline HCCI combustion engine, *Automatisierungstechnik* **57**, 3-13.
- RWTH Aachen University and Bielefeld University (since 2006) SFB 686 - Modellbasierte Regelung der homogenisierten Niedertemperatur-Verbrennung. <http://www.sfb686.rwth-aachen.de>.
- Hoffmann K. (2010) Nonlinear Model-based Predictive Control of a Low-Temperature Gasoline Combustion Engine, *PhD Thesis*, RWTH Aachen University, VDI Verlag. <http://darwin.bth.rwth-aachen.de/opus3/volltexte/2010/3220/pdf/3220.pdf>.
- Hoffmann K., Abel D. (2009) Nonlinear Model-based Predictive Control with Constraints for Controlled Auto-Ignition, *Proc. E-COSM'09, IFAC Workshop on Engine and Powertrain Control, Simulation and Modeling, Rueil-Malmaison, France, 30/11-2/12 2009*.
- Stapf K.G., Seebach D., Pischinger S., Hoffmann K., Abel D. (2009) Aspects of gasoline Controlled Auto Ignition Development of a controller concept, *MTZworldwide* **70**, 4, 18-24.
- Abel D., Bollig A. (2006) *Rapid Control Prototyping: Methoden und Anwendungen*, Springer-Verlag.
- Hoffmann K., Drews P., Seebach D., Stapf K.G., Pischinger S., Abel D. (2008) Optimized Closed Loop Control of Controlled Auto Ignition CAI, *Proc. of FISITA 2008 World Automotive Congress, Munich, Germany, 14-19/09 2008*.
- Hoffmann K., Seebach D., Pischinger S., Abel D. (2007) Neural networks for controlling future low temperature combustion technologies, in *Proc. 3rd IFAC Workshop on Advanced Fuzzy and Neural Control 2007*, 29-30 October 2007, Valenciennes, France.
- Albrecht A., Grondin O., Berr F.L., Sollic G.L. (2007) Towards a Stronger Simulation Support for Engine Control Design: A Methodological Point of View, *Oil Gas Sci. Technol.* **62**, 437-456.
- Hoffmann K., Drews P., Abel D., Felsch C., Vanegas A., Peters N. (2009) A cycle-based multi-zone simulation approach including cycle-to-cycle dynamics for the development of a controller for PCCI combustion, *SAE Int. J. Engines* **2**, 511-526.
- Cybenko G. (1989) Approximation by Superpositions of a Sigmoidal Function, *Math. Control Signals Syst.* **2**, 303-314.
- Hornik K., Stinchcombe M., White H. (1989) Multilayer Feed-forward Networks are Universal Approximators, *Neural Netw.* **2**, 359-366.

- 22 Nørgaard M., Ravn O., Poulsen N.K., Hansen L.K. (2003) *Neural Networks for Modelling and Control of Dynamic Systems*, Springer-Verlag.
- 23 Ljung L. (1987) *System Identification: Theory for the User*, Prentice-Hall.
- 24 Zhu Yucai (2001) *Multivariable system identification for process control*, Elsevier.
- 25 Tunestål P., Johansson B., Zhao H. (ed.) (2007) *HCCI ad CAI engines for the automotive industry*, Woodhead Publishing.
- 26 Nelles O. (2001) *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*, Springer-Verlag.
- 27 Bengtsson J., Strandh P., Johansson R., Tunestål P., Johansson B. (2004) Closed-loop Combustion Control of Homogeneous Charge Compression Ignition (HCCI) Engine Dynamics in: *Int. J. Adapt. Control Signal Process.* **18**, 167-179.
- 28 Hoffmann K., Seebach D., Pischinger S., Abel D. (2008) Control of future low temperature combustion technologies with nonlinear model based predictive control based on neural networks, *Proc. 17th IFAC World Congress*, The International Federation of Automatic Control, Seoul, Korea, July 6-11, 2008, pp. 9467-9472.
- 29 Muske K.R., Badgwell T.A. (2002) Disturbance modeling for offset-free linear model predictive control, *J. Process Control* **12**, 617-632.
- 30 Welch G., Bishop G. (2006) *An introduction to the Kalman filter. Technical Report TR 95-041*, University of North Carolina, Department of Computer Science, Chapel Hill, NC 27599-3175.
- 31 Odelson B.J., Rajamani M.R., Rawlings J.B. (2006) A new autocovariance least-squares method for estimating noise covariances, *Automatica* **42**, 303-308.
- 32 Rossiter J.A. (2003) *Model-based Predictive Control: A Practical Approach*, CRC Press.
- 33 Boyd S., Vandenberghe L. (2008) *Convex Optimization*, Cambridge University Press.
- 34 Seebach D., Stapf, K.-G., Pischinger S., Hoffmann K., Abel D. (2009) Modelling and Actuation of the Controlled Auto Ignition, *Proc. Haus der Technik, IX. Tagung "Motorische Verbrennung - aktuelle Probleme und moderne Lösungsansätze"*, Munich, Germany, 18-20 März 2009.
- 35 Hoffmann K., Seebach D., Pischinger S., Abel D. (2008) Mehrgrenregelung der kontrollierten Selbstzündung CAI in einem Einzylinder Forschungsmotor, *Proc. of AUTOREG 2008*.

*Final manuscript received in March 2011
Published online in October 2011*

Copyright © 2011 IFP Energies nouvelles

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than IFP Energies nouvelles must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee: Request permission from Information Mission, IFP Energies nouvelles, fax. +33 1 47 52 70 96, or revueogst@ifpen.fr.