OPTIMAL CONTROL OF WASTE HEAT RECOVERY SYSTEMS APPLYING NONLINEAR MODEL PREDICTIVE CONTROL (NMPC)

Philipp Petr[1]*, Christian Schröder[1], Jürgen Köhler[1], Manuel Gräber[2]

[1] TU Braunschweig, Institut für Thermodynamik
Hans-Sommer-Str. 5, 38106 Braunschweig, Germany
p.petr@tu-braunschweig.de
christian.schroeder@tu-braunschweig.de
juergen.koehler@tu-braunschweig.de

[2] TLK-Thermo GmbH
Hans-Sommer-Str. 5, 38106 Braunschweig, Germany
m.graeber@tlk-thermo.de

* Corresponding Author

ABSTRACT

This article describes an attempt for a real-time optimization of the net power output of an add-on Organic Rankine Cycle (ORC) system of a vehicle applying Nonlinear Model Predictive Control (NMPC). Therefore, a Modelica model library for satisfactorily accurate, fast vehicle and ORC models was developed. By means of the developed tool chain involving the optimizer MUSCOD II by the IWR Heidelberg, virtual simulation experiments of a waste heat recovery system for a long-distance bus could be realized. Results show an increase of the net power output of 7 % in part load engine operation in the European Transient Cycle compared to a conventional controller with optimum operation points optimized at steady-state conditions.

1. MOTIVATION FOR THE HOLISTIC OPTIMIZATION OF THE OPERATION OF A WASTE HEAT RECOVERY SYSTEM IN VEHICLES

In internal combustion engines, only a small part of fuel’s chemical energy is transformed into mechanical energy. The residual is mainly wasted as thermal energy via the cooling system or as exhaust gas into the environment. Especially the exhaust gas still has a high exergetic potential, which can be converted into mechanical energy by means of an Organic Rankine Cycle (ORC). The recovered energy reduces the amount of mechanical energy provided by the combustion engine and therefore reduces the total fuel consumption. Virtual test drives of vehicles with waste heat recovery systems (WHRS) based on ORC promises a high fuel saving potential [1], [2], [3] and [4].

1.1 Interactions of the WHRS with Energy Systems in a Motor Vehicle

The integration and operation of an ORC in such a complex energy system represented by a motor vehicle affects many other subsystems (see Fig. 1). Interaction between these systems have to be considered in the design stage as well as in the development of operation strategies.

The combustion engine is affected by higher exhaust backpressure caused by pressure losses in the evaporator resulting in a slightly lower engine efficiency. Moreover, mechanical energy provided by the ORC reduces the load demand of the engine. This leads to a shift in the engine efficiency and the exhaust temperature, which has repercussions on the ORC itself. Furthermore, the condensing heat of the ORC significantly increases the load of cooling system. Under certain conditions, additional fans have to ensure target coolant temperatures. Admittedly, the power consumption of the additional fans exceed the additional power provided by the ORC, which results in a shutdown of the waste heat recovery system. Not least, the additional weight of the ORC increases substantially the work load of the combustion engine.

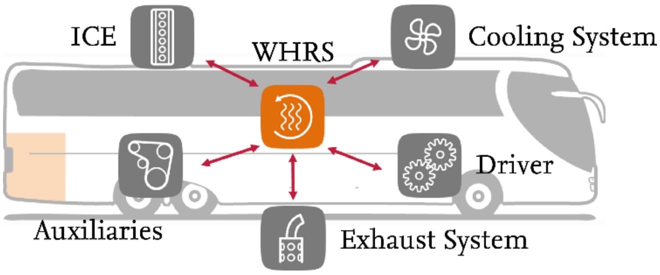


Figure 1: Interactions of the WHRS with other systems in mobile applications could influence the objective function (minimization of fuel consumption) substantially (i.e. power demand of additional fans).

In order to maximize the benefit of WHRS, interactions of different subsystems and their time constants should be integrated in the operation strategy in order to minimize shut down times of the WHRS and to maximize the recovered amount of exhaust exergy.

1.2 Variable Operation Points of the ORC Enables Waste Heat Recovery at Part Load Conditions
Focusing on the operation of the WHRS, car manufacturers currently consider a control approach with only one static operation point [1], [5], which is optimized for a representative driving cycle. Depending on actual exhaust enthalpy flow rate, the chosen operation point mostly differs from the current optimum operation point. Figure 2 shows a T-h diagram of an optimized operation point and one operation point which was designed for higher exhaust gas temperatures. It can be seen that the exhaust exergy used is significantly lower in the design operating point, which cannot compensate higher thermodynamic efficiencies. Thus, the second law efficiency for the adapted optimum operation point is 20 percentage points higher than for the fixed operation point. In the worst case, the WHRS is been shut due to low working fluid mass flow rates in part load conditions despite an still existing exhaust exergy potential.

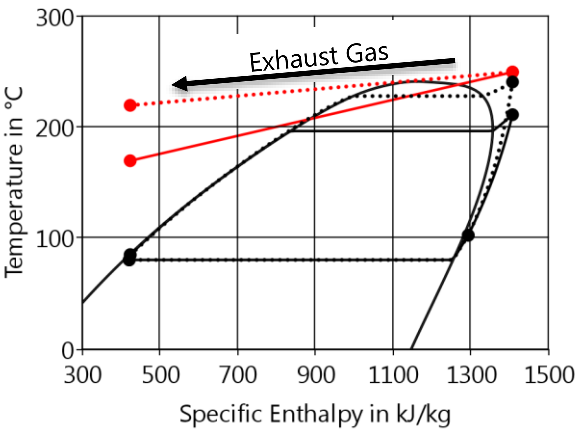


Figure 2: T-h-Diagram of two ORCs at different operation points with corresponding exhaust temperature course in the heat exchanger for an exhaust inlet temperature of 250 °C. Operation point optimized for higher loads (·····) shows poor usage of exhaust exergy compared to currently optimal operation point (**—**)

Enabling variable ORC operation points increases the benefit of the WHRS also in part load conditions, but requires the consideration of the dynamic behavior of the system (step response or time constants on excitation of the system). Optimized operation points calculated under steady state conditions differ from actual optimum control variables the more, the larger the time constants of the WHRS and the more dynamic the changes in the exhaust gas enthalpy flow rate are.

1.3 NMPC as One Promising Method to Minimize the Fuel Consumption for Real Drives

In order to maximize the benefit of a WHRS, the control strategy has to utilize variable operation points with respect to the system's dynamic behavior, future states of the ORC and dependencies of interacting systems. Conventional control strategies such as heuristic controls cannot fulfill these criteria. Nonlinear model predictive control (NMPC) is one method to meet this challenge. Based on a simplified nonlinear mathematical model of the system, an optimizer computes the ideal course of controlled variables for a certain prediction horizon. NMPC uses relevant information about the actual state of the vehicle and the predicted variables for real-time control. Relevant predicted data for the optimization are for example:

- Route including gradients and traffic
- Style of driving
- Comfort demand of passengers
- Ambient conditions (solar radiation, winds, temperature)

Due to fluid properties, the vehicle with its subsystems, especially the WHRS, is characterized by a high grade of nonlinearity. Hence, linear optimization approaches, as they are discussed in [6] and [7] for following given set points, are not applicable for set point optimization in large operation intervals. Therefore, efforts regarding fast and accurate nonlinear simulation models have to be done.

This article describes a first attempt of the development of an NMPC for a WHRS for a long distance bus. Therefore, a transient full vehicle simulation model of a long distance bus with a WHRS was set up. The WHRS model is based on a conventional ORC using Ethanol as working fluid and a Scroll expansion machine, which enables expansion in the two-phase region. The basis for the simulation study is the exhaust gas enthalpy flow rate of the part load section of the European Transient Cycle (ETC). As a first step, the chosen optimum criterion was set to the maximization of the ORC power output. To compare NMPC with heuristic control methods, optimum power points in steady state condition were calculated for each exhaust gas inlet state, which are adjusted by conventional Single-Input-Single-Output (SISO) controllers.

2. MODEL DESCRIPTION AND STEADY-STATE OPTIMIZED OPERATION POINTS FOR BENCHMARKING A CONVENTIONAL CONTROL STRATEGY

Basis for the development of the advanced control strategy is a conventional Organic Rankine Cycle, which is described in Fig. 3. The working fluid (Ethanol) leaves the feed pump in pressurized liquid state (1-2) and then enters in the evaporator, where it is heated and vaporized (2-3). In the expander (3-4), potential energy of the working fluid is converted into mechanical energy, which is used in the vehicle to drive auxiliaries. After the expansion, the working fluid is liquefied in the condenser. With re-entering the feed pump, the cycle starts again.

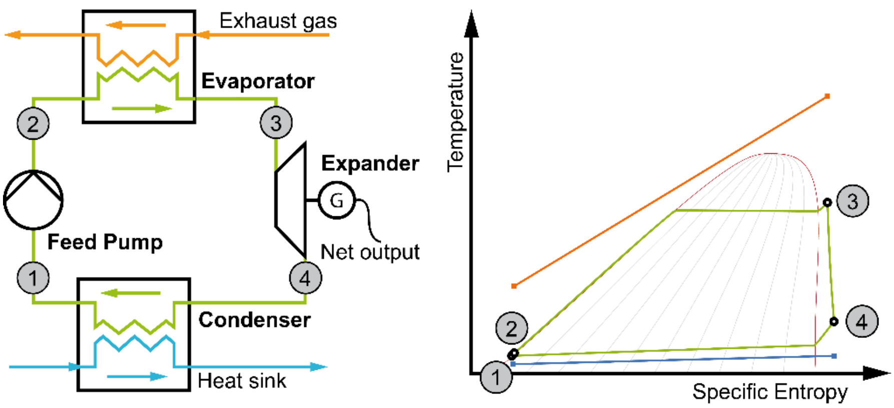


Figure 3: Investigated process configuration and T-s-Diagram. The working fluid is evaporated directly in the evaporator without using a secondary loop system

2.1 Transient Simulation Model in Modelica

For the evaluation of investigated control strategies, a transient simulation model of the ORC was set up in the programming language Modelica by means of the self-developed model libraries TIL and TILMedia. The evaporator is a physical model of a fin-and-tube heat exchanger, which is designed to a steady-state pinch point temperature of 10 K at the highest occurring exhaust gas enthalpy flow rate. The expansion machine model is efficiency based and represents a volumetric expander (scroll), which enables expansion in two-phase region. The condenser is connected to the vehicle cooling system, which supplies constant coolant temperature of 90 °C.

2.2 Optimum Operation Points in Steady State Conditions

In order to evaluate NMPC, a conventional control strategy based on steady-state optimized operation points was set up. Two degrees of freedom govern the net power output: expander inlet pressure and expander inlet enthalpy. The optimum operation point correlates not necessarily with the maximum cycle efficiency, but with the second law efficiency. A gradient-based optimization algorithm applied on a steady-state model of the ORC finds the optimum operation point for each boundary condition. Although the heat transfer coefficients depend on the mass flow rate of the exhaust gas and the working fluid as well as on the states in the heat exchanger, a constant pinch point temperature of 10 K in the evaporator was assumed. The expander inlet pressure was limited to 5 MPa.

Figure 4 shows the net power output of described ORC as a function of the expander inlet state. For low exhaust gas enthalpy flow rates (Fig. 4, left), optimum expander inlet state is located in the two-phase region, whereas it is in superheated region for high enthalpy streams (Fig. 4, right).

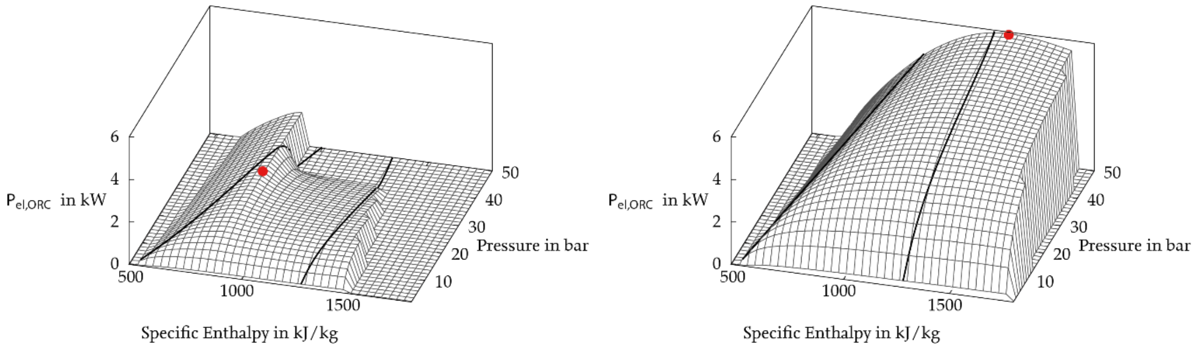


Figure 4: Net power output of ORC dependent on expander inlet pressure and enthalpy for low (left) and high (right) exhaust gas enthalpy flow rates. Projected fluid dew and bubble lines indicate the properties of mapped states. Working fluid: Ethanol, expander type: scroll, maximum pressure: 5 MPa

3. INTRODUCTION TO NMPC

This section introduces the theory of NMPC. The following NMPC scheme is derived based on the formulation and numerical solution approach of an Optimal Control Problem (OCP). Additionally, the key parameters influencing both solution quality and speed are discussed.

3.1 Optimal Control Problem (OCP)

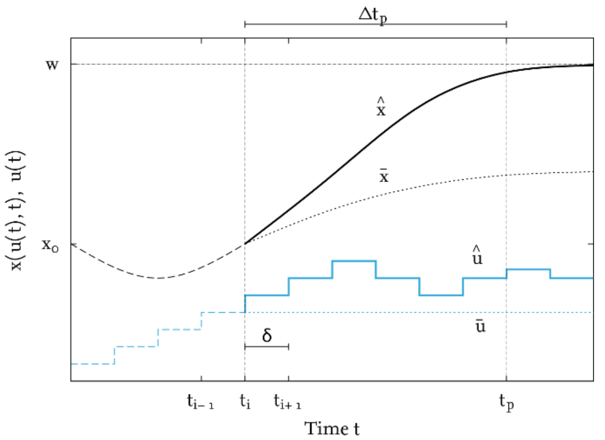
The concept of this method is to calculate a specific control variable trajectory, which influences the dynamic behaviour of a plant in a way that it maximizes a given objective function. As a first step, the objective functions is the maximization of the net power output of the ORC over the considered horizon:

$$\max_{n_{Pump},n_{Expander}} P_{el,ORC} = P_{el,Generator} - P_{el,Pump} \qquad 4.1$$

The optimal control trajectory is determined based on a mathematical model of the plant, which would be a system of semi-explicit differential algebraic equations of index 1. To achieve an optimal control trajectory, an objective function is needed as a benchmark of the system's performance. Therefore, Lagrange-type terms can be used to maximize or minimize it over the entire time horizon or a Mayer-type term limits the target value to a set endpoint. In addition, a combination of the two known as a Bolza type objective function can be utilized. This leads to the general formulation of an OCP subject to several constraints. Besides the already described DAE system, we find path-constraints, e.g. a maximum temperature limit, which must not be exceeded at all times and point-constraints, defining boundaries only at definite instants. At last, the trajectories of the state variables have to be a solution to the initial value problem (IVP) with the given start values. Thus, the optimal control function must satisfy these constraints in order to be a valid solution of the OCP.

3.2 Nonlinear Model Predictive Scheme

In order to control a plant in a closed loop rather than an open loop setting, the OCP has to be solved repeatedly with up-to-date measurements of the system's states as initialization (Fig. 5). As proposed in [8], the real-time iteration approach is used, that performs only one Sequential Quadratic Programming (SQP) iteration per NMPC sample. Starting with a preparation phase, all functions and derivatives are evaluated, that do not depend on the current state x_0. As soon as the new measurement of x_0 is available, the feedback phase starts with the computation of the control function $u(t)$, which is limited to a single Quadratic Programming (QP) due to the computation during preparation. Eventually, $u(t)$ can be given to the plant in the transition phase. Afterwards the new time sample starts according to the three phases.



Δt_p	Prediction horizon
w	Set value
δ	Sample time
$\bar{u}$	Control variable at t_i
$\bar{x}$	System response to constant $\bar{u}$
$\hat{u}$	Optimized control variable
$\hat{x}$	System response to $\hat{u}$

Figure 5: Nonlinear Model Predictive Control is a repeated solving of OCPs. At t_{i+1}, solving OCP restarts for the prediction horizon with current measured or estimated input values

3.3 Key Parameters in NMPC

The optimality of the solution as well as the computation time can be influenced not only by the complexity of the model, which is used for NMPC. A great impact also derives from the number of shooting knots N, hence the discretization of the NLP, and the prediction horizon Δt_P. Higher values lead to significantly increased numerical effort, but also to a better solution. A key parameter for online NMPC is the sample rate at which measurement data is provided and new control values are passed on to the plant. The optimization problem has to be solved fast enough to satisfy the required sample interval. Otherwise, the control values of the previous synchronization are used, which are usually not minimizing the objective function of the progressed system state. Table 1 shows the key parameters for the performed study:

Table 1: Key parameters for the performed study

Prediction horizon	4 s
Shooting knots	4
Sample rate	3,3 Hz

In this study, a higher prediction horizon shows only a slight increase in the ORC power output, but also in missing the real-time criterion (Intel Core™ i7-4770K CPU, 3,50GHz, 32GB RAM, Windows 8 - 64 bit**)**

3.4 Approach for a Prototype Setup

In this section, the software framework for a convenient setup of simulated and real-world NMPC loops is presented. The basic principle is to use different specialized software for each task and to couple it to a co-simulation master. Using Functional Mockup Interface (FMI) ensures integrity of the underlying plant model that is used in several places, and avoids error-prone and time-consuming model transformations.

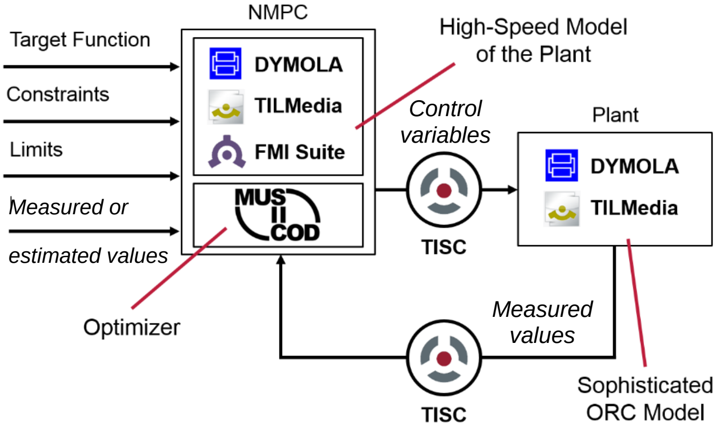


Figure 6: Used software framework for prototype NMPC attempt

The software framework for the NMPC prototype environment is described in Figure 6. For data exchange, the co-simulation platform TISC [9] is used to set-up a NMPC prototyping environment. TISC acts as master and manages data exchange between different clients. There already exist interfaces between TISC and a variety of simulation, visualization and measurement tools, e.g. Dymola, LabView, and Simulink. The user has to define types and names of variables to be sent and received for each client. Data routing between clients is automatically managed by matching variable types and names. The TISC type of time is Double, whereas all other variables are of TISC type Double Array. This definition enables the exchange of components of an NMPC loop. Hence, the virtual plant modeled in Modelica can be replaced easily by a real plant interfaced with LabView [8]. As optimizer, MUSCOD II of IWR at Heidelberg University, Germany as described in [10], [11] and [12], is proven to be appropriate.

4. NONLINEAR HIGH-SPEED MODEL FOR ONLINE OPTIMIZATION

The NMPC requires a nonlinear high-speed model, which maps real systems or, like presented in this study, a sophisticated model of the ORC accurately and which fulfils the demand of real-time optimization. Every component has been developed against certain requirements that need to be fulfilled in order to guarantee an efficient usage in the optimization process. The first key aspect is based on the solving process of the NLP. Like described in section 3, the SQP method is used in this case, which obligates that both the objective function and the constraints, thus all variables of the NMPC plant-model, are at least twice continuously differentiable with respect to the input $u(t)$ [10]. Any discontinuity in the state trajectories leads to a less efficient optimization and hence should be avoided. One key component for optimization of the transient operation of the ORC is the evaporator. Therefore, big effort is being made to map the physical characteristics of the real evaporator. For calculating the heat transfer, the NTU approach is used. To ensure continuous differentiability, the term

$$\dot{Q} = kA'\Delta T_o \left(\frac{1 - e^{-\theta L}}{\theta}\right) \tag{4.2}$$

$$\text{with } \theta = kA' \left(\frac{1}{\dot{m}_1 c_{p1}} - \frac{1}{\dot{m}_2 c_{p2}}\right) \tag{4.3}$$

is transformed with the help of the hyperbolic sine function. The term is now defined for equal heat capacity flow rates [13]:

$$\dot{Q} = kA'\Delta T_o e^{-\frac{\theta L}{2}} * sinhc\left(\frac{\theta L}{2}\right) \qquad 4.4$$

To take account of the different heat transfer coefficients of the working fluid in the different states, the heat exchanger is divided into three NTU sections (subcooled, two-phase and superheated). The mass balance in the three volume elements is calculated by

$$\dot{m}_{in} - \dot{m}_{out} = \left[\left(\frac{\partial\rho}{\partial p}\right)_h \frac{dp}{dt} + \left(\frac{\partial\rho}{\partial h}\right)_p \frac{dh}{dt}\right] V, \qquad 4.5$$

whereas the energy balance is given by

$$m\frac{dh}{dt} = \dot{m}_{in}(h_{in} - h) - \dot{m}_{out}(h_{out} - h) + \dot{Q} + \frac{dp}{dt}V \qquad 4.6$$

The thermal capacity of the heat exchanger wall is modeled by one differential state, which determines the transferred heat to the working fluid. Figure 7 shows a scheme of the model approach for the evaporator.

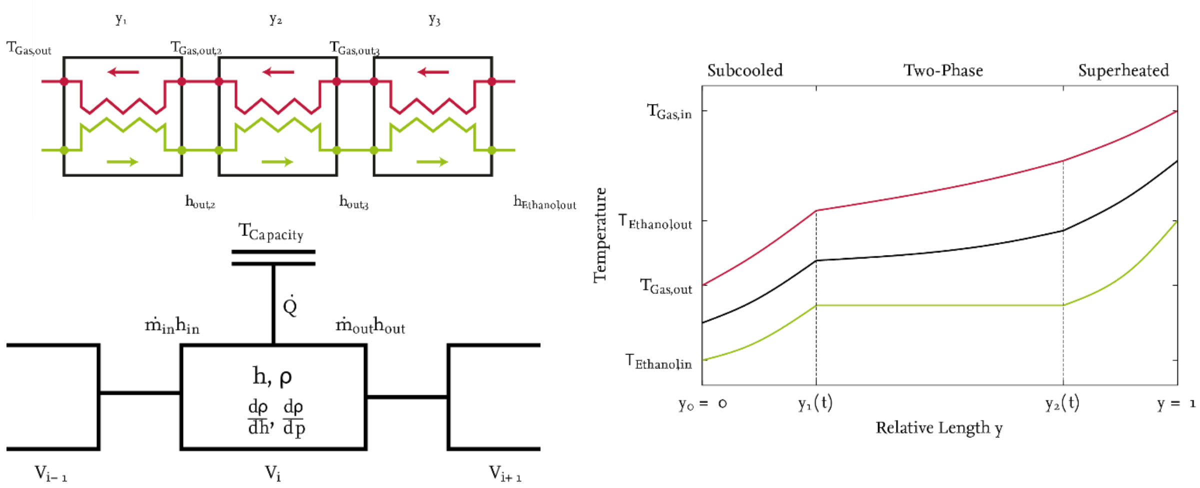


Figure 7: Evaporator Model. Consists of one cell for each working fluid phase. A modified NTU approach is used to calculate heat transfer coefficients. Transient conduct is modeled by finite volume elements

The total high-speed model is illustrated in Figure 8 (left). Efficiencies of the pump and the expander are mapped; component inertias are represented by first order elements.

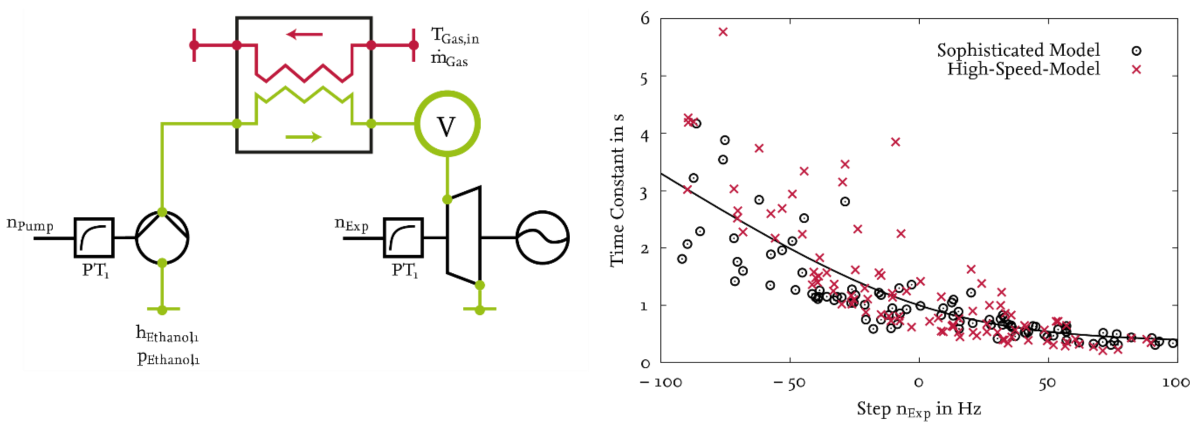


Figure 8: Scheme of high-speed model (left). Deviations in time constants on random pressure steps between high-speed and sophisticated model are acceptable (right)

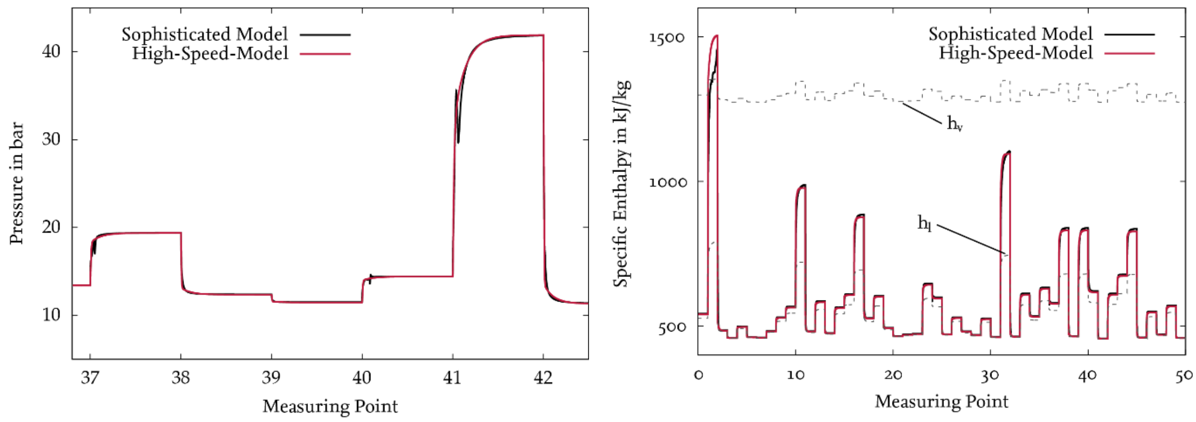


Figure 9: Step response on expander speed steps (left) and pump speed steps (right) of sophisticated and high-speed model

To evaluate the accuracy of the high-speed model, simulations with random steps on both models were carried out. Figure 9 shows only minor deviations in the step response for the sophisticated model and the high-speed model. Deviations in time constants are in accordance to low simulation times in acceptable intervals (Fig 8 right). The median error of the high-speed model for steady state target values are 5% for the expander inlet enthalpy and 1% for expander inlet pressure.

5. BENCHMARK OF NMPC AND CONVENTIONAL CONTROL IN THE URBAN PART OF THE EUROPEAN TRANSIENT CYCLE

In order to benchmark NMPC and the conventional control approach based on steady state optimized operation points in part load conditions, simulation experiments within the urban section of the European Transient Cycle (Fig. 10) were carried out for each control strategy.

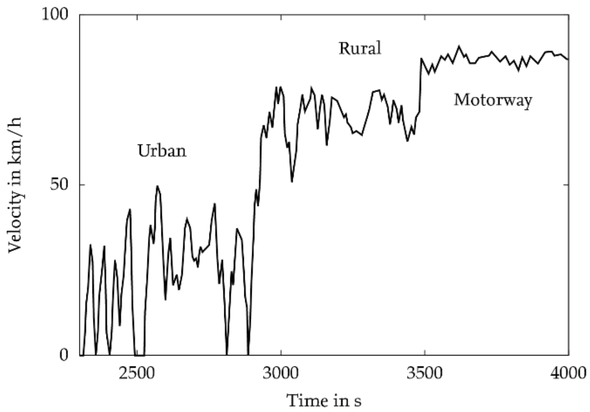


Figure 10: Velocity profile of the European Transient Cycle (ETC)

Figure 11 shows the course of expander inlet pressure and enthalpy over the cycle. It can be seen, that the conventional control cannot satisfactorily adjust the set values, which are calculated through steady state optimization. This is caused by the time constants of the ORC, which are not considered in steady state simulations. NMPC however tends to lower expander inlet pressures and higher expander inlet enthalpies.

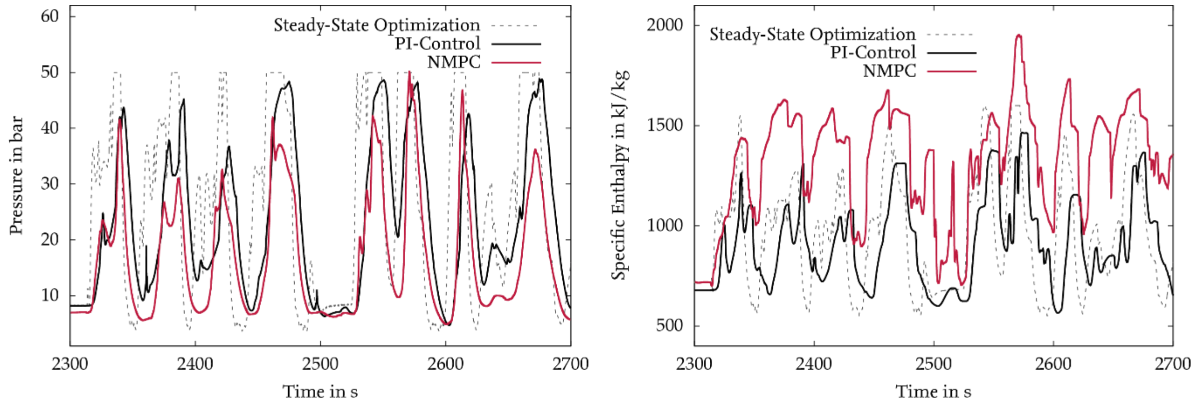


Figure 11: Course of expander inlet pressures (left) and inlet enthalpies (right). NMPC tends to higher inlet enthalpies and lower inlet pressures. Steady state operation points are not adjusted satisfactorily by PI-control

These different courses could be explained with focusing on the pump work, which affects the net output of the ORC. Because of the higher pressure steps of the PI-control, the pump work is significantly higher than of the NMPC (see Fig. 12, left) and the pump work of the steady state model.

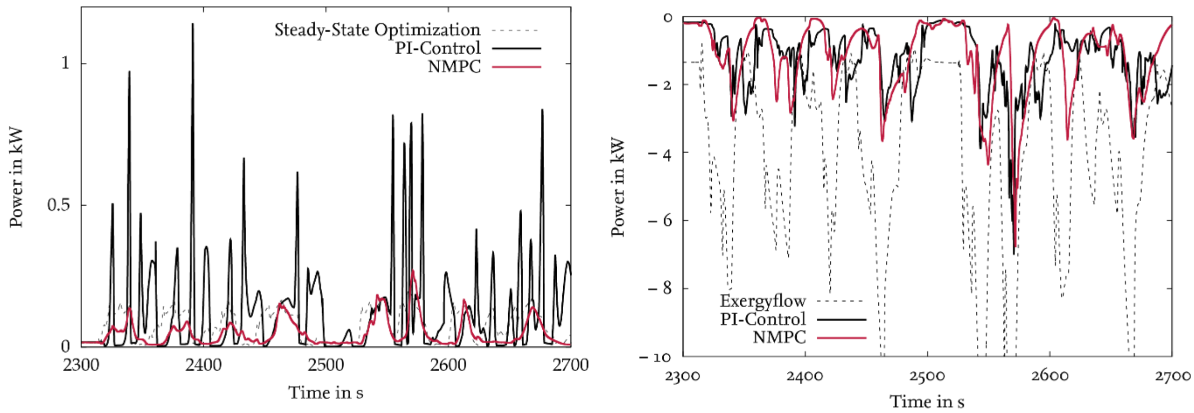


Figure 12 Pump work (left) and expander power output (right) of the ORC for both PI-Control and NMPC. NMPC show a 7 % higher average net power output

Thus, the expander power output of NMPC is not significantly higher (see Fig. 12, right), but cleared with actual pump work, NMPC provides a 7 % higher average net power output in part load operation of the internal combustion engine than the conventional control strategy based on operation points optimized in steady state conditions. Assuming that the ORC is shut down in part load operation for a conventional strategy with an operation point designed to high load operation, NMPC with variable operation points could increase the power output of the ORC by 15 %.

6. CONCLUSIONS

Optimal control approaches could increase the total system efficiency by considering interaction of several subsystems, future states and inertias of the thermal systems. NMPC is one promising method. At the TU Braunschweig and TLK-Thermo, a software tool chain was developed in order to facilitate NMPC for thermal systems. In this article, a NMPC approach for a WHRS of a long distance bus was developed. Particular challenge was the development of a high-speed model of the evaporator, which maps heat transfer coefficients and inertias accurately accomplishing the requirements of real-time optimization. In a simulation study, maximization of the WHRS net output in the part load section of the European Transient Cycle was considered. Results show an increase of the net power output of 7 % compared to a conventional controller with operation points optimized at steady-state conditions.

NOMENCLATURE

ETC	European Transient Cycle	FMI	Functional Mockup Interface
ICE	Internal Combustion Engine	IVP	Initial Value Problem
NLP	Nonlinear Programming	NTU	Number of Transfer Units
NMPC	Nonlinear Model Predictive Control	QP	Quadratic Programming
OCP	Optimal Control Problem	SQP	Sequential Quadratic Programming
WHRS	Waste Heat Recovery System		

REFERENCES

[1] Horst, T. A., Tegethoff, W., Eilts, P., Koehler, J., 2014, Prediction of dynamic Rankine Cycle waste heat recovery performance and fuel saving potential in passenger car applications considering interactions with vehicles' energy management, *Energy Conversion Management*, vol. 78, pp. 438–451.

[2] Espinosa, N., Tilman, L., Lemort, V., Quoilin, S., Lombard, B., Rankine cycle waste heat recovery on commercial trucks: approach, constraints and modelling. In: Proceedings of the international conference and exhibition on diesel engines, Rouen, France; 2010.

[3] Sprouse, C., Depcik, C., 2013, Review of organic Rankine cycles for internal combustion engine exhaust waste heat recovery, *Appl. Therm. Eng.*, vol. 51, no. 1–2, pp. 711–722.

[4] Boretti, A. A., 2012, Transient operation of internal combustion engines with Rankine waste heat recovery systems, *Appl. Therm. Eng*., vol. 48, pp. 18–23.

[5] Hartmann, A., 2014, Energie- und Wärmemanagement mit thermischer Rekuperation für Personenkraftwagen, Dissertation, TU Braunschweig.

[6] Feru, E., Willems, F., De Jager, B., Steinbuch, M., 2014, Modeling and Control of a Parallel Waste Heat Recovery System for Euro-VI Heavy-Duty Diesel Engines," *Energies*, vol. 7, no. 10, pp. 6571–6592.

[7] Luong, D., 2013, Modeling, Estimation, and Control of Waste Heat Recovery Systems, Dissertation, University of California.

[8] Gräber, M., Kirches, C., Scharff, D., Tegethoff, W., 2012, Using Functional Mock-up Units for Nonlinear Model Predictive Control, *9th International Modelica Conference*, p. 781-790.

[9] Kossel, R., Löffler, M., Strupp, N. C., Tegethoff, W., 2011, Distributed energy system simulation of a vehicle, *Vehicle Thermal Management Systems Conference*, Institution of Mechanical Engineers, SAE International.

[10] Bock, H. G., Plitt, K. J., 1984, A Multiple Shooting algorithm for direct solution of optimal control problems, *Proceedings 9th IFAC World Congress Budapest*, pages 243–247. Pergamon Press.

[11] Diehl, M., 2001, Real-Time Optimization for Large Scale Nonlinear Processes, PhD thesis.

[12] Leineweber, D. B., Bauer, I., Bock, H. G., Schlöder, J. P., 2003, An Efficient Multiple Shooting Based Reduced SQP Strategy for Large-Scale Dynamic Process Optimization (Parts I and II). Computers and Chemical Engineering, 27:157–174.

[13] Tegethoff, W., Schulze, C., Graeber, M., Huhn, M.; Stulgies, N., Kaiser, C., Loeffler, M., 2011 TEMO Thermische Echtzeitfähige Modelle, Final Project Report.