

**Hochschule Karlsruhe**

University of  
Applied Sciences

Institute of

**Energy Efficient Mobility**



# Reports on Energy Efficient Mobility

Volume 2  
March 2022



Edited by:

Prof. Dr.-Ing. Dirk Feßler  
Prof. Dr.-Ing. Maurice Kettner  
Prof. Dr.-Ing. Reiner Kriesten  
Prof. Dr.-Ing. Philipp Nenninger  
Prof. Dr.-Ing. Peter Offermann

# Preface

When we released the first volume of the Reports on Energy Efficient Mobility in March 2021, our intention was to provide a stop gap before the normal conferences and congresses pick back up during that year. We wanted to keep up the communication between our peers and spark the discussions which usually take place after presentations.

The feedback we received was overwhelmingly positive and shows that the impulses provided by the papers in the Reports were indeed missing in the academic community. In addition, the process of preparing the reports turned out to be an important internal communication exercise at the Institute of Energy Efficient Mobility. So, while we all got used to returning to our offices and meeting our coworkers in person again, we decided to build on the positive experience of the first reports and release them annually. In doing so, we want to give our community a fixed point for orientation in the rapidly changing world we live in.

Building such a lighthouse is a real team exercise, and thus we would like to thank everybody who participated in it. We would like to thank Robin Iding, who had the sometimes ungrateful task of reminding us all of deadlines and keeping the layout consistent. The main attraction of course are the papers, which again show all the facets of our research fields at the Institute of Energy Efficient Mobility. Everybody in academia knows that papers are just the tip of the iceberg and we all are aware of and appreciate the immense effort required to produce six to ten pages of scientific literature. We would therefore like to especially thank all the authors who put a lot of hard work into preparing their wonderful contributions in this year's volume.

Karlsruhe, March 2022

Dirk Feßler  
Maurice Kettner  
Reiner Kriesten  
Philipp Nenninger  
Peter Offermann

# Contents

Mona Gierl Development of a CAN-Ethernet gateway communication .....	1
Marcel Rumez Development of a Vehicle Control App based on Android .....	10
Umut Can Kaya, Robin Bolz Survey on Vulnerability Report Handling in the Automotive Environment	17
Faruk Altun, Felix Müller Detecting Hardware Modifications on Vehicles .....	32
Jörn A. Judith, Maurice Kettner, Thomas Koch A Benchmark Study of Chemical Reaction Mechanisms for Ignition Delay Calculation of Natural Gas/Hydrogen Mixtures under Internal Combustion Engine Conditions .....	38
Tuyen Nguyen, Yannick Rauch Real Route Generation for Simulation Based Development .....	58
Arsema Derby Chekol, Maurice Kettner Methods of developing Model Predictive Control for Air Conditioning systems (A Literature review) .....	65
Weber Michael, Weiß Tobias, Nötzold Marlene Approaches of a Human-Machine-Interface for Augmented Reality in Automotive Systems .....	83
Rudolf Schnee How Significant are Business Models for Innovations and Technologies?	98
Tanju Gofran, Maurice Kettner IEEM-CMCNN model accuracy on different public datasets .....	109
Youssef Beltaifa, Maurice Kettner, Erik Kärcher, Marc Reutter, Peter Eilts, Bosse Ruchel Piston Bowl Design for Highly Efficient CHP Gas Engines: An Approach to Reducing Wall Heat Losses without Deteriorating the Turbulent Flame Propagation .....	119

# Development of a CAN-Ethernet gateway communication

Mona Gierl

*Institute of Energy Efficient Mobility*

*University of Applied Sciences*

Karlsruhe, Germany

mona.gierl@h-ka.de

## **Abstract**

While moving towards autonomous driving, there is an increased need for advanced driving features. Advanced features require adapted networking architectures which are able to handle the data and software demands. In order to replicate the requirements on today's in-vehicle networks, the Institute of Energy Efficient Mobility (IEEM) is in the process of developing a simulated vehicle network. Within this paper, the IEEM in-vehicle network architecture and the implementation of a cross-domain communication between CAN and Ethernet is presented.

## **Index Terms**

automotive gateway, cross-domain communication, in-vehicle network

## I. Introduction

Moving towards autonomous driving, the demand for advanced driving features is continuously growing. The driving tasks require multiple sensors to provide real time information of the environment to the vehicle's control units. Combined with the increasing number of functions, the in-vehicle network faces challenges in terms of bandwidth, cross-domain communication, and security [1].

In order to cope with the increasing complexity, the car's Electric/Electronic Architecture (E/E-Architecture) has evolved over the years as shown in Figure 1. Distributed functions were integrated into domains to segregate Electronic Control Units (ECUs) according to their functionality and facilitate communication between ECUs with similar functionalities [2]. Each domain is managed by a master ECU (domain controller) with advanced computing power to enable a flexible data exchange within and across domain boundaries. In the future, Advanced Driver Assistance Systems (ADAS) will require more computing power for processing sensor data. Thus, the trend continues towards a more centralized approach leading to fusion of ECUs and domains and resulting in centralized computers processing the core functionalities of the vehicle [2].



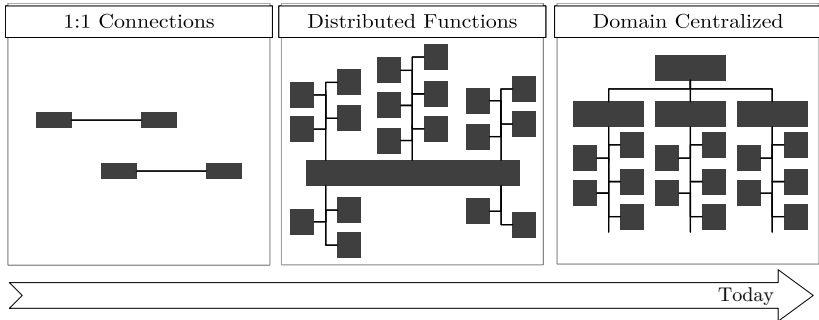


Fig. 1: Advances in e/e-architectures towards a domain architecture [4]

As technology evolves, the test effort to ensure the safe and reliable operation of the vehicle also increases. In order to investigate and test ADAS systems under realistic conditions, a domain architecture is being developed at the Institute of Energy Efficient Mobility (IEEM). Vehicle domains such as ADAS, Powertrain, Body, and Comfort are integrated directly or via models into a networking model which approximates the entire vehicle communication behavior. The software tool CANoe [3] is used as a simulation environment. Altogether, this enables the integration and research of safety and security concepts.

As of today, several ECUs are already implemented within the IEEM network including, e.g., Airbag, Electronic Stability Control (ESC), Anti-Lock Braking System (ABS), Adaptive Cruise Control (ACC), and further. The corresponding domains communicate over the Controller Area Network (CAN) bus. A cross-domain communication is planned via an Ethernet backbone which interconnects existing domains over gateways. An excerpt of the IEEM network architecture is given in Figure 2.

This paper demonstrates the efforts to put in place a cross-domain communication via CAN and Ethernet. In the end, different strategies to encapsulate CAN traffic over the Ethernet link are discussed and a tunneling approach is presented. Further, domain gateways and a central gateway are implemented to take over the task of traffic translation and routing. Finally, test results analyze the delay and bus utilization of the given gateway communication.

## II. Background Central Gateway

The central gateway is responsible for the cross-domain data exchange between ECUs [5]. Communication takes place through different communication protocols. The gateway serves as a translating unit and is connected to two or more buses. According to [6], the essential tasks and functions of an automotive gateway can be summarized as follows:

- 1) Protocol translation
- 2) Data routing
- 3) Firewall
- 4) Intrusion detection
- 5) Key management
- 6) Over-the-air (OTA) updates
- 7) Diagnostic routing
- 8) Traffic scheduling
- 9) Offboard communication

In the context of this paper, the steps to implement a domain centralized networking architecture for the existing IEEM network is shown. To enable the cross-domain communication, the focus is limited to the two main features protocol translation and data routing.

### III. Requirements Gateway Communication

Within the simulated IEEM network, the data communication in each domain is realized by the CAN Bus, while Ethernet serves as a backbone to interconnect the domain gateways with a central gateway (CGW). The architecture is shown in Figure 2. The aim is to establish a communication between the Powertrain, Body and Dynamics domain. In order to so, the gateway has to translate CAN messages into Ethernet frames and route the frames to the target domain. As of now, the used protocols are restricted to the data link layer of the ISO/OSI reference model, as the simulation solely addresses the in-vehicle data exchange, thus, the overhead for higher level protocols is avoided. Higher protocol layers are going to be relevant in case of external communication or diagnostics. The requirements differentiate between the translation and the routing task.

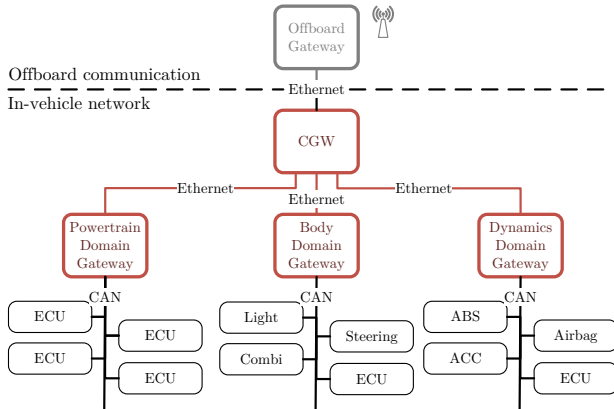


Fig. 2: Extraction of the IEEM network architecture. The developed components of the cross-domain communication are highlighted in red.

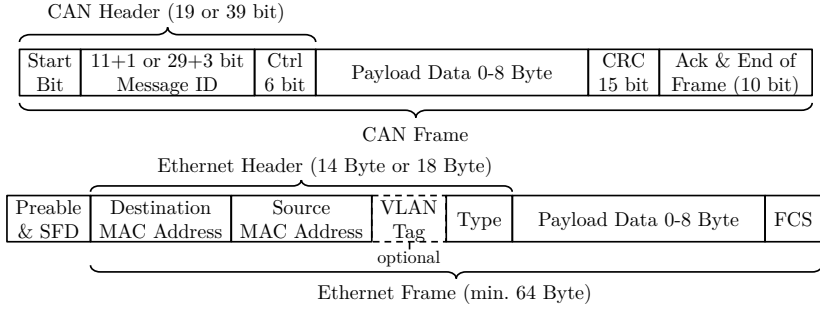


Fig. 3: Structure of a CAN and an Ethernet frame according to [7]

#### A. Translation requirements

CAN messages of one domain must be transferred timely to another domain so that all information is provided without information loss. This forwarding should work with CAN as well as CAN FD. Furthermore, special frame formats, e.g., remote frames, have to be considered for translation. In addition, standard (11 bits) and extended CAN identifiers (29 bits) should be recognized and translated correctly.

The CAN bus is designed as a broadcast system sending max. 8 Byte of data up to 1 megabit per second [7]. In comparison, Ethernet frames consist of min. 64 Byte to max. 1500 Bytes data length and common data rates are up to 100 megabit per second [7] (both frame formats are given in Figure 3). The translation mechanism should provide different packing strategies while considering both buffering delay and the amount of unused data per message.

#### B. Routing requirements

Routing the network traffic requires the gateways to know the destination of each message. This can be done by using routing tables which assign a destination domain to each CAN identifier. A routing table is also used in the context of this work. The table should be saved as a file and should be editable independently of the simulation.

Further, the central gateway has to be independent of the strategy chosen in the domain gateway for packing CAN frames into Ethernet frames. The forwarding must therefore also work if several CAN frames are packed in one Ethernet frame (see Section IV)

### IV. Gateway Strategy

To establish a communication across the domains as shown in Figure 2, the messages on the individual CAN buses must first be packed into Ethernet frames in the domain gateways and unpacked again at the target

ECU. In the following, different strategies on how CAN messages can be transmitted via Ethernet are explained.

With the goal of forwarding CAN messages from domain to domain, a tunneling gateway is suitable. In tunnel gateways, messages from one technology are transported in the payload data part of messages from the other technology. Based on the given frame formats in Figure 3, either a single CAN frames can be packed into a single Ethernet frame (Strategy 1:1), or multiple CAN frames (Strategy n:1) are encapsulated.

Aggregating multiple CAN frames within an Ethernet frame is called Pooling [8]. According to [8], various Pooling methods exist. The simplest variant is the use of buffers with a fixed buffer size. To avoid needless long delays, a timeout is added. Whenever the buffer size or the timeout is reached, the tunnel gateway forwards the Ethernet message.

The first variant (Strategy 1:1) offers the advantage of reduced waiting time of a CAN message, since each CAN frame is sent immediately. However, this variant generates a very high proportion of unused data. If, for example an 8 byte long CAN frame is transmitted, the minimum packet size for Ethernet (64 Byte) must still be reached and therefore the missing bytes must be padded with unused data.

The second variant (Strategy n:1) reduces the amount of unused data per message and requires less bus load. A disadvantage of this variant results from the fact that several messages arrive at the same time on the destination bus. This generates data bursts, which not only lead to an increase in end-to-end delay of the buffered frames, but also increases the delay of messages on the destination bus.

To improve the delays and the burst behavior, there are further approaches in the area of message prioritization and traffic scheduling. For the extended strategies, the Time Sensitive Networking Standards IEEE 802.1 [9] and the following literature can be consulted [10]–[12], however, these approaches are not discussed further, since they are not content of this work.

## V. Implementation

For the implementation, the communication between the Body, Powertrain and Dynamics domains is realized. Each domain has a domain gateway, and the domains are connected via the central gateway (see Figure 2). A functional overview of the gateways is given in Figure 4 and 5.

Within the routing tables, the CAN messages with their message identifiers and their target domain are saved. The routing table file is read in the domain gateways and the central gateway before simulation start. According to the routing table, the domain gateway decides which messages must be translated to Ethernet. To transport the required CAN or CAN FD message in the user data part of the Ethernet frame, a tunnel protocol is implemented as shown in Figure 6. The first payload byte of the tunnel protocol contains the amount of consecutive CAN frames and an indicator whether CAN or

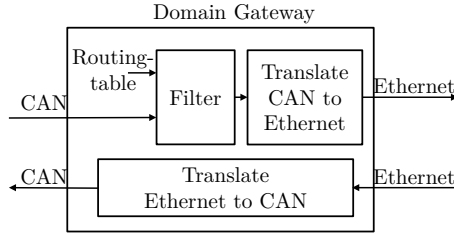


Fig. 4: Functional overview of a domain gateway. The main task is to filter and translate outgoing CAN frames and to translate incoming Ethernet frames.

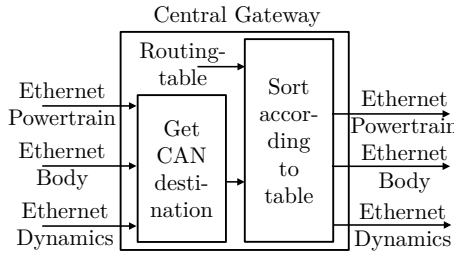


Fig. 5: Functional overview of a domain gateway. The main task is to filter and translate outgoing CAN frames and to translate incoming Ethernet frames.

CAN FD is transmitted. For the Ethernet communication the MAC address of the respective gateway is used. After translation, the Ethernet frames are forwarded to the central gateway. The central gateway transfers the incoming Ethernet frames to the respective domain on the basis of the message identifiers contained in the routing table. If the n:1 strategy is used, the CAN frames contained in an Ethernet message are unpacked and buffered, since each CAN message can have its own destination. The CAN messages are then passed individually to the corresponding output buffer. Whenever the message arrives at its destination domain gateway, the corresponding gateway translates the package back to CAN.

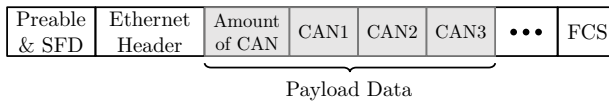


Fig. 6: Tunnel protocol to transmit CAN messages within an Ethernet frame.

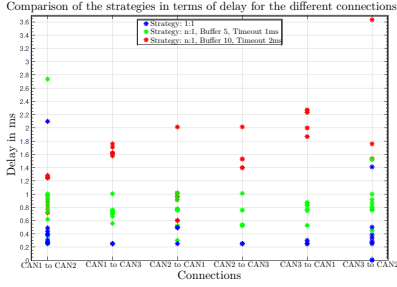


Fig. 7: Delays of the tunneling strategies in the connections determined over 10 test runs.

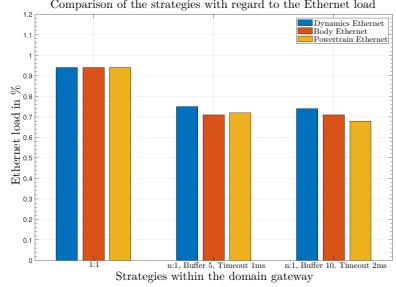


Fig. 8: Comparison of the utilization of the Ethernet connections for the different strategies

## VI. Test results

In the context of this work, requirement-based testing is applied to test the correct translation between the three domains. A comparison between the messages in the destination and source domain verified the correct cross-domain transmission. In addition, the routing table was also tested by sending undefined CAN IDs, which were ignored successfully. Beside the correct implementation, delay and bus utilization are determining factors for future enhancements. The following three strategies were considered:

- 1) Strategy: 1:1
- 2) Strategy: n:1, buffer size 5, timeout 1ms
- 3) Strategy: n:1, buffer size 10, timeout 2ms

In order to assess the delays caused by the gateway communication from domain to domain, 10 test runs were performed for each connection (see Figure 7 with CAN1=Body, CAN2=Dynamics, CAN3=Powertrain). On average, the 1:1 strategy results in a delay of 0.2516 ms, whereas strategy 2) requires three times longer with 0.7628 ms, and strategy 3) has an average delay of 1.5260 ms. This leads to the conclusion, that the 1:1 strategy should be preferred for the current status of the IEEM network while the other strategies are of interest for future expansions.

Looking at the bus load in Figure 8, all strategies use less than 1 % of the available bandwidth. Hence, even though strategies 2) and 3) reduce the bus load, it has no influence on the overall performance, since a high bus capacity remains.

## VII. Conclusion

In this paper, a simulated vehicle network was extended by gateways to enable a cross-domain communication. The aim was to facilitate the in-vehicle data exchange between the Powertrain, Body, and Dynamics domain

of the IEEM vehicle network. A central gateway was developed to route messages to their destination domain. In addition, domain gateways were implemented to translate messages from CAN to Ethernet. To do so, the CAN data from the domains is packed into an Ethernet frame. For this purpose, a tunnel protocol was designed that specifies the form in which the data is encapsulated in the Ethernet frame.

In total, three different strategies were implemented to compare the delay and bus load of the simulated network. Test results showed that the current IEEM network has the lowest delays in case of a 1:1 translation from CAN to Ethernet.

In summary, the vehicle network now consists of an Ethernet backbone and is able to exchange data between three different domains. Thus, the current project concentrated on the in-vehicle communication as well as the translation and data routing tasks of the gateways. In future projects, the further development of the IEEM network will be pursued. In addition to the further expansion of the in-vehicle network, it is planned to enable an external network communication by developing an offboard gateway.

### Acknowledgment

A special acknowledgment to P. Herzog who was the main contributor to the research and implementation of the cross-domain communication and who gave his permission to report the findings. Another special acknowledgment goes out to F. Sommer for his inspiring high-quality review comments for this paper.

### References

- [1] AVnu Alliance, “Automotive Market” online: <https://avnu.org/automotive/>, last accessed: 26-11-2021.
- [2] V. Navale, K. Williams, A. Lagospiris, M. Schaffert, M.-A. Schweiker,“(R)evolution of E/E Architectures”, in: SAE International Journal of Passenger Cars - Electronic and Electrical Systems, Volume 8, Issue 2, pp. 282-288, 2015, doi: <https://doi.org/10.4271/2015-01-0196>, ISSN: 1946-4622.
- [3] Vector Informatik GmbH, “CANoe”, online: <https://www.vector.com/int/en/products/products-a-z/software/canoe/>, last accessed: 06-12-2021
- [4] McKinsey & Company, “Automotive software and electrical/electronic architecture: Implications for OEMs”, 2019, online: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/automotive-software-and-electrical-electronic-architecture-implications-for-oems>, last accessed: 26-11-2021.
- [5] R. Hvanth, D. Valli, K. Ganesan, “Design of an In-Vehicle Network (Using LIN, CAN and FlexRay), Gateway and its Diagnostics Using Vector CANoe”, in: American Journal of Signal Processing 1 (2012), No. 2, p. 40–45, doi: <https://doi.org/10.5923/j.ajsp.20110102.07>, ISSN: 2165–9354.
- [6] NXP Halbleiter Deutschland GmbH, “Automotive Gateway: A Key Component to Securing the Connected Car”, Version 2018, online: <https://www.nxp.com/docs/en/white-paper/AUTOGWDEVWPUS.pdf>, last accessed: 17-11-2021.

- [7] W. Zimmermann, R. Schmidgall, "Bussysteme in der Fahrzeugtechnik" (Bus systems in automotive engineering), Springer Vieweg, Wiesbaden, 2014, doi: <https://doi.org/10.1007/978-3-658-02419-2>, ISBN: 978-3-658-02418-5.
- [8] T. Steinbach, "Ethernet-basierte Fahrzeugnetzwerkarchitekturen für zukünftige Echtzeitsysteme im Automobil" (Ethernet-based vehicle network architectures for future real-time systems in the automobile), Springer Fachmedien Wiesbaden, 2018, doi: <https://doi.org/10.1007/978-3-658-23500-0>, ISBN: 978-3-658-23499-7.
- [9] TSN Task Group, "Time-Sensitive Networking (TSN) Task Group", online: <https://1.ieee802.org/tsn/>, last accessed: 30-11-2021.
- [10] A. Kern, D. Reinhard, T. Streichert, J. Teich, "Gateway Strategies for Embedding of Automotive CAN-Frames into Ethernet-Packets and Vice Versa" in: Architecture of Computing Systems - ARCS 2011, Springer, doi: [https://doi.org/10.1007/978-3-642-19137-4\\_22](https://doi.org/10.1007/978-3-642-19137-4_22), ISBN: 978-3-642-19137-4.
- [11] A. Nacer, K. Jaffres-Rünser, J.-L. Scharbarg, C. Fraboul, "Strategies for the interconnection of CAN buses through an Ethernet switch", in: 2013 8th IEEE International Symposium on Industrial Embedded Systems (SIES), p. 77-80, IEEE, doi: <https://doi.org/10.1109/SIES.2013.6601474>, ISBN 978-1-4799-0658-1.
- [12] D. Thiele, J. Schlatow, P. Axer, R. Ernst, "Formal timing analysis of CAN-to-Ethernet gateway strategies in automotive networks", in: Real-Time Systems 52 (2016), No. 1, p. 88-112, doi: <http://dx.doi.org/10.1007/s11241-015-9243-y>, ISSN: 1573-1383.



# Development of a Vehicle Control App based on Android

Marcel Rumez

*Institute of Energy Efficient Mobility*

*University of Applied Sciences*

Karlsruhe, Germany

marcel.rumez@h-ka.de

## **Abstract**

At the Institute of Energy Efficient Mobility (IEEM), a prototypical automotive electrical and electronical architecture (E/E architecture) consisting out of various Electronic Control Units (ECUs) and software functions has been developed by different student projects to provide a research and development platform. The aim of this work is to develop an Android app to provide a remote control of vehicle functions of the E/E architecture. Furthermore, this app shall serve as a platform for other project activities at the IEEM (e.g., control of an electric scooter, information security investigations).

## **Index Terms**

App-Development, Android, E/E architecture, Vehicle Simulation Network

## **I. Introduction**

The automotive industry is undergoing a transformation, driven by the megatrends of electric mobility, shared mobility and automated driving [1]. As a result, the software used in vehicles as well as the degree of connectivity is continuously increasing. Therefore, more than 470 million connected vehicles will be on the road by 2025 [2]. Furthermore, up to 300 million lines of code will be implemented on ECUs in 2030 [3]. The connectivity enables a data exchange between vehicles or the environment in order to optimize control functions [4]. Moreover, vehicle apps for mobile devices are available to provide remote comfort and control functions for customers such as unlocking the vehicle, starting heating or displaying the state of charge. However, the increasing degree of connectivity increases the risk of information security vulnerabilities and corresponding cyber attacks [5]. Therefore, future automotive systems must be secured by different counter-measures. To analyze the effectiveness of specific measures, prototypical setups of real hardware or simulations are useful for research purposes. The following sections describe the development of a vehicle app, which serves as a foundation for future investigations and development activities. A main focus will be the investigation regarding the attack security of wireless

vehicle interfaces (e.g., WLAN) to external network nodes (e.g., Original Equipment Manufacturer (OEM)-backend). The app communicates with a vehicle simulation network, which has been implemented via the software tool CANoe [6] at the IEEM.

## II. Network & CANoe Simulation

For the connection of the vehicle simulation network with the control app, different network components as well as technologies are used (see figure 1). An Odroid-XU4 board [7] serves as the access point, providing a WLAN connection for the mobile device (app). In addition, the connection to the vehicle simulation (CANoe) is established via Ethernet.

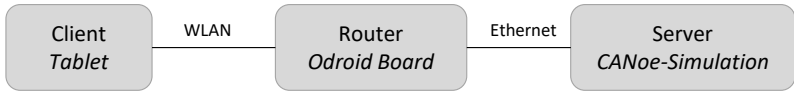


Figure 1. Schematic network architecture for the connection between vehicle network simulation and control app.

### A. Configuration Odroid-XU4 & WiFi-module

In order to configure the Odroid board, a suitable operating system has to be selected. The two most used operating systems for the Odroid-XU4 board are LineageOS [8] and Ubuntu MATE V.18.04 [9]. LineageOS is a simple and user-friendly Android-based operating system. However, it is not suitable for the required application because it cannot be used to install an access point. This kind of system is automatically rebuilt every time the system is started. For this reason, Ubuntu MATE is selected, which has already been established in recent years.

After starting Ubuntu and logging into the system as an administrator, the access point settings and configurations are made via the Ubuntu terminal based on [10].

## III. Android-App

### A. Connection Setup

The connection setup for the message transmission between vehicle control app (client) and vehicle simulation (server) was divided into different development steps. First, a unidirectional connection is established between an app and a laptop via an WLAN router (see figure 2) without the vehicle simulation.

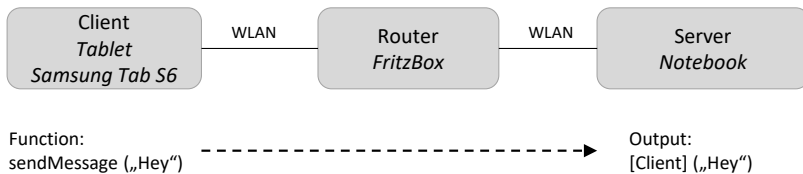


Figure 2. Schematic network architecture for the connection between notebook and control app.

## B. Sending Messages

1) *Background Information:* In apps there is generally a main thread, which is responsible for the processing of inputs and outputs in the user interface and should not be overloaded by computational operations, such as data exchange in a network. Otherwise, it can lead to a "frozen screen", making it impossible to operate with the app. Separate threads are required to operate data exchange over a network in applications with a user interface. Threads are sub-processes, which enable simultaneous execution of several programs. The main thread is able to call other threads and terminated only the app is closed compared to additionally created threads, which are executed once and then terminated.

2) *Implementation:* The following app is programmed in Java by using Android Studio (development environment). Starting with the user interface, a button calls the connection establishment function via the main thread. In this function the client class is started. The destination address for the connection setup is passed to the client class so that it can connect to the server. The class *client* runs on a newly created thread. As soon as a message is to be sent from the client to the server, another thread is started for sending. The basic functions required for sending messages are the following:

- **Socket():** Opens an unconnected socket. The method *connect(Socket-Address endpoint)* can be used to connect the client to a server. The *SocketAddress* contains the address of the server (see listing 1).

Listing 1. Code extract for establishing a connection between app and client

```

1 public void run() {
2     try {
3         socket = new Socket();
4         socket.connect(host, 1000);
5     } catch (Exception e){
6         e.printStackTrace();
7     }}

```

- **DataOutputStream():** Allows writing various data types. The method *writeUTF(String str)* enables to send strings based on UTF-8 encoding. With *DataInputStream()* the data can be read again.

The server (notebook) can receive messages from the client (tablet) and output them in the console of the development environment. The server was programmed with Java in Eclipse (development environment). The basic functions required for receiving messages are as follows:

- **Socket():** Opens an unconnected socket.
- **ServerSocket():** Opens an unconnected *ServerSocket* on a selected port on the device. The *accept()* method waits until a client logs on to the port on the laptop. When a connection is established by the client, the method returns the *socket* type.
- **DataInputStream():** Allows to read messages from a *socket*. The *readUTF()* method can read the received messages from the *socket* and store them in a string (see listing 2).

Listing 2. Code extract for receiving messages on the server

```

1  public void sendMessage(String msg){
2  new Thread(new Runnable() {
3      @Override
4      public void run () {
5          try {
6              if(socket.isConnected()) {
7                  dos = new DataOutputStream(socket.getOutputStream());
8                  dos = writeUTF(msg);
9              }
10             catch (IOException e) {
11                 e.printStackTrace();
12                 closeConnection();
13             }
14         }
15     }
16 }

```

### C. Extension for Receiving Messages

After the implementation of the unidirectional connection, this is extended to a bidirectional one (see figure 3). When the client sends a message, the server sends a message back to the client, which outputs the message via the user interface.

In the app a new thread is created from the class *client* which permanently listens for new messages using *DataInputStream()*. The output via the user interface is realized by using the function *runOnUiThread()*. On the server (notebook), the function is extended to send messages. When a message is received from the app, a new message is sent from the server to the client (app) via *DataOutputStream()*. The app displays each received message on a list via the user interface. The list is continuously updated by the function *runOnUiThread()*.

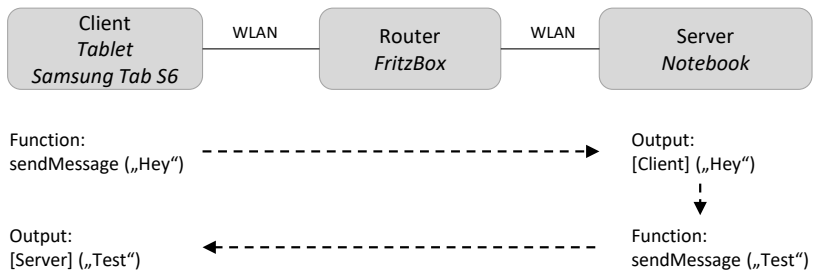


Figure 3. Schematic network architecture of a bidirectional communication between notebook and control app.

#### D. Communication - Vehicle Simulation & App

In the following, the notebook used is now replaced by the CANoe vehicle simulation for data exchange with the app. Therefore, the router is replaced by the Odroid board, which establishes an access point for data exchange. The vehicle simulation is connected to the Odroid Board via Ethernet and serves as a server. Furthermore, the vehicle control app is also connected with the Odroid board via WLAN. For the communication, it has to be considered that the CANoe simulation implemented by the CAPL programming language uses char arrays instead of strings. However, in order to output the data in the user interface, strings are more suitable. Therefore a function is implemented, which converts the char arrays into a string based on a specified protocol.

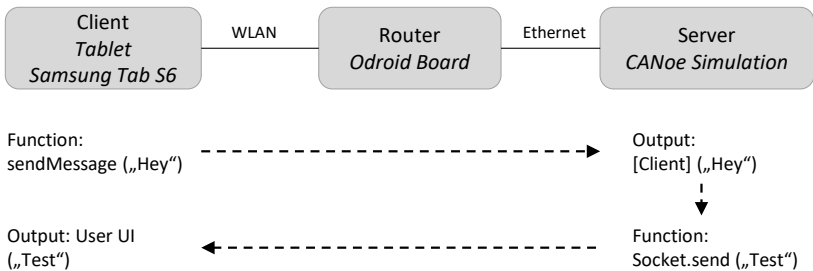


Figure 4. Schematic network architecture of a bidirectional communication between vehicle network simulation and control app.

### E. App Design & Usability

A high priority is assigned to the app design in order to achieve a good usability. One of our main requirements is an appealing design that could be used intuitively (see figure 5). In order to provide visual as well as haptic feedback when the state of the vehicle changes, the colors of the buttons change or vibrations are triggered.

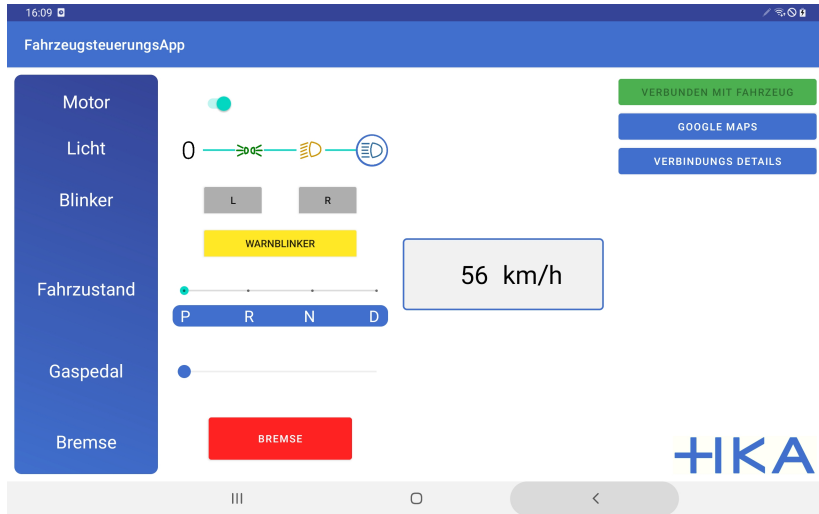


Figure 5. User interface of the developed vehicle control app.

While the app is not connected to the vehicle or server, the connection button is orange and all vehicle control functions are disabled. As soon as a connection is established, the button changes to green and the text content changes to *Connected to Vehicle*. Below is the option to show the current location via the Google Maps button. Furthermore, the button *Connection Details* activates the settings, which allow the connection to any address by using IP and port. Incoming messages from the vehicle are displayed directly under IP and Port.

The light position currently selected by the user is permanently framed by a ring, so it can be clearly recognized whether the daytime running light, low beam or high beam is switched on. Like the other functions, the light can only be set while a connection is established. The functions for active vehicle control have been programmed in such a way that either braking or accelerating is possible. Accelerating the vehicle is only available within the driving modes *D* and *R*. The throttle represents a real vehicle behavior. When releasing the button on the seekbar, it is restored to the default value.

For activating the brake, the button has to be pressed permanently, while the color remains orange.

#### IV. Conclusion & Future Work

The vehicle control app allows to communicate with the simulated IEEM E/E architecture via a WLAN access point. This enables the execution of various control functions of the vehicle remotely and different sensor information can be accessed, which serves as a foundation for future research and development activities. A future work will address the implementation of security measures. Therefore, the network will be secured by using well-established protocols (e.g., IPSec, TLS). Moreover, firewalls and access controls will also be integrated.

#### Acknowledgment

The vehicle app was developed within a student project in the summer semester of 2021 at Karlsruhe University of Applied Sciences (HKA) by the students Alex Schink and Tim Schmidt. This report is based on the associated project report.

#### References

- [1] "The 2019 strategy & digital auto report: Fast and furious: Why making money in the roboconomy is getting harder." [Online]. Available: <https://www.strategyand.pwc.com/media/file/2017-Strategyand-Digital-Auto-Report.pdf>
- [2] "The 2017 strategy & digital auto report," 2017. [Online]. Available: <https://www.strategyand.pwc.com/media/file/2017-Strategyand-Digital-Auto-Report.pdf>
- [3] M. Staron, *Automotive Software Architectures: An Introduction*. Springer, 2017.
- [4] M. A. Rahim, M. A. Rahman, M. M. Rahman, A. T. Asyhari, M. Z. A. Bhuiyan, and D. Ramasamy, "Evolution of iot-enabled connectivity and applications in automotive industry: A review," *Vehicular Communications*, vol. 27, p. 100285, 2021.
- [5] E. Aliwa, O. Rana, C. Perera, and P. Burnap, "Cyberattacks and countermeasures for in-vehicle networks," *ACM Computing Surveys*, vol. 54, no. 1, pp. 1–37, 2021.
- [6] "Canoe." [Online]. Available: <https://www.vector.com/us/en-us/products/products-a-z/software/canoe/>
- [7] "Odroid-xu4 einplatinen-computer," 2021. [Online]. Available: <https://www.pollin.de/p/odroid-xu4-einplatinen-computer-samsung-exynos-5422-2-gb-2x-usb-3-0-810409>
- [8] "Lineageos android distribution," 2021. [Online]. Available: <https://lineageos.org/>
- [9] "Ubuntu mate 18.04 lts (v4.0)," 2021. [Online]. Available: [https://wiki.odroid.com/odroid-xu4/os\\_images/linux/ubuntu\\_4.14/20180501](https://wiki.odroid.com/odroid-xu4/os_images/linux/ubuntu_4.14/20180501)
- [10] "Wireless access point." [Online]. Available: [https://wiki.odroid.com/accessory/connectivity/wifi/wlan\\_ap](https://wiki.odroid.com/accessory/connectivity/wifi/wlan_ap)

# Survey on Vulnerability Report Handling in the Automotive Environment

1<sup>st</sup> Umut Can Kaya, B. Eng.

*Institute of Energy Efficient Mobility*

*University of Applied Sciences*

Karlsruhe, Germany

Umut.Can.Kaya@outlook.de

2<sup>nd</sup> Robin Bolz, M.Sc.

*Institute of Energy Efficient Mobility*

*University of Applied Sciences*

Karlsruhe, Germany

robin.bolz@h-ka.de

## Abstract

This paper deals with the handling of reports of automotive cyber vulnerabilities found in the field. Processes for disclosing reported vulnerabilities within the automotive industry are examined and compared. Both, the world's most important vehicle manufacturers (OEM) and relevant suppliers are considered. In addition, all published vulnerabilities (in scope) are taken from the two databases NVD and ADD and analyzed. From this, insights can be gained into the quality and quantity of reporting opportunities as well as downstream disclosure processes for vulnerabilities within the automotive industry. In connection with this, the extent to which information on resolved vulnerabilities is made available to the public is examined. Furthermore, it is recorded whether or not companies are present on a service provider platform for the coordination of security vulnerabilities. At the end of the elaboration, conclusions are drawn on how to deal with security vulnerabilities in the automotive industry. It can be shown that an in-depth analysis is difficult due to insufficient data. However, it is clear that more transparency and a stronger disclosure culture regarding automotive security vulnerabilities could improve the quality of vulnerability management in the after-production phase. In our view, one of the desirable steps would be to establish independent and trustworthy bodies in disclosure processes.

## Index Terms

Automotive Security, Vulnerability Handling, Vulnerability Disclosure

## I. Introduction

Due to increasing networking and automation, more and more software and computer-based solutions are being integrated into vehicles [1]. In addition to many advantages, however, there is also an increasing risk of possible security vulnerabilities that can be exploited for attacks on a vehicle or entire vehicle fleets [2][3]. Corresponding trends have been observed for several years. For example, the number of reported attacks in the automotive environment increased sixfold from 2010 to 2018 (circa 60



in 2018). At the same time, in 2018, the proportion of malicious attacks exceeded that of well-intentioned ones for the first time [1]. Both trends, the number of cyber incidents, and the dominance of malicious attacks has been continuously confirmed [4]. This development highlights the need to massively strengthen the community of benign hackers. This can be supported by intensifying incentives, as well as expanding low-threshold reporting options and trustworthy disclosure processes. To counter these threats, companies need to drive appropriate security mechanisms. Given the long product lifetimes of vehicles, vulnerability management in the operational phase plays a particularly important role in the automotive sector. Involving the knowledge of well-intentioned hackers outside the industry can make a helpful contribution [4]. In order to achieve the best possible cooperation between this group and the industry, incentives and fair disclosure processes are essential, in addition to trusted and published reporting facilities. Bug bounty [5] and vulnerability disclosure (VD) programs [6] address these requirements and have become widely established. The aim of a VD program is the efficient and coordinated fixing of a vulnerability before it is exploited for an attack. Publishing selected information about the fixed vulnerability or advisories for end users can provide additional value. Bug Bounty programs aim to recruit as many well-intentioned hackers as possible by issuing additional incentives such as fixed rewards. This paper examines how stakeholders within the automotive industry deal with reports of cyber incidents in the field. For this purpose, the current implementation status of measures such as vulnerability disclosure and bug bounty programs in the automotive environment is considered in section II. For this purpose, the scope of consideration as well as assessment criteria for the investigation are defined. This is followed in Section III by an analysis of the information obtained. Here, both, vulnerability related parameters like quantity and severity of vulnerabilities and process related parameters are taken into account. Finally, the results are summarized and an outlook on possible proposals for improvements is given in section IV and V.

## II. Status of Implementation

### A. Definition of Scope

In the research for this work, certain pre-selections were made in order to limit the scope of the work without compromising the informative value. The ten best-selling passenger cars of the last two years (2018 & 2019) worldwide [7][8], in Europe [9][10] and in Germany [11] were used as reference for the selection of the automobile manufacturers (OEM) to be investigated. Here it was assumed that these vehicles, due to their age, are not only the most widespread, but additionally the most digitized vehicles. They therefore offer the largest attack surface for potential hackers. In addition, the two manufacturers BMW and Tesla were included in the research.

BMW was included because their model “X6” lists itself as the most stolen passenger car in Germany [12]. Tesla, in turn, was included in the list as a new and all-electric brand to provide a comparison to the established companies. Furthermore, the company’s strong focus on autonomous driving features and smart cars played a major role in the selection. In addition to OEMs, the largest suppliers to the automotive industry by revenue were included in the research. The regions of the world [13], Europe [14] and Germany [15] were also considered for the suppliers and a corresponding selection was made. Suppliers that exclusively produce components that have no relevance in terms of cybersecurity (e.g.: tires) were excluded. For the research of security vulnerabilities, the public databases NVD [16] and AAD [17] were used. Published security vulnerabilities are listed on these databases. It should be noted that the data in AAD only goes up to 2018 and includes vulnerabilities that have not been published by the companies themselves. The information comes from published scientific papers. It should also be noted that the NVD database only lists vulnerabilities that have a CVE-ID [54]. Thus, both databases do not contain all published vulnerabilities, which is why a combination of both data sources was used for this work. Additionally, if used by the companies, the service provider platforms “HackerOne” [18] and “Bugcrowd” [19] were used as sources. These platforms allow good hackers to establish contact with the companies as easily as possible. They also give companies the opportunity to have their products and services tested by a large and trustworthy number of hackers. The platforms offer the possibility to start either a VD or a bug bounty program. The first VD program was launched in 2014, which is why the research period extends from the beginning of 2014 to the end of 2020. When evaluating the published vulnerabilities, no pre-selection was made. The aim of this work is to make a general statement about the handling of the automotive industry with this topic. However, the respective implementation processes in the supply chain could not be taken into account, as these are not publicly accessible.

## *B. Evaluation Criteria*

In order to obtain both a comparability of the companies and a statement about the overall behavior of the automotive industry, some criteria for evaluation were defined in advance. The assessment was made in principle with regard to the existence of reporting opportunity or platforms as well as appropriate disclosure conditions and guidelines (disclosure policy). For both categories, more in-depth criteria were examined to provide further insight into the quality of existing efforts. The companies were analyzed according to the following criteria.

1) *Vulnerability disclosure policy*: By publishing its disclosure policy a company can provide transparency for their expectations by communicating, for example, the scope of their program or restrictions and conditions for the disclosure of submitted vulnerabilities. The following points are considered during the research to assess the quality of disclosure policies found to be publicly accessible:

- Is there a company disclosure policy available to the public?
- Is it the company's own disclosure policy or has it been adopted from a 3rd party service provider?
- What is the scope of the program regarding domains and product areas (out-of-scopes)?
- Is any information provided on the expected procedure of the process after a notification has been submitted (disclosure timeline)?

2) *Vulnerability disclosure platforms or reporting opportunity*: Providing low-threshold and trusted channels for communicating and submitting vulnerability information to a company is a key cooperation incentive for security researchers. Therefore, the research also focuses on the following points:

- Does a way exist to communicate vulnerability information to the company?
- If there is a reporting option, does the company use a service provider platform (e.g., HackerOne or BugCrowd)?
- How transparent is the existing platform with regard to the publication of vulnerability information after their remediation?

### C. Implementation Status

Most of the investigated OEM offer the possibility to report security vulnerabilities directly. In doing so, they also all specify what type of vulnerabilities may be researched and reported (in scope). In connection with this, almost all companies that offer the possibility of reporting security vulnerabilities also have a vulnerability disclosure policy. It is noticeable, however, that specifying information is often missing. Nine companies specify what is in scope, but only six of them specify what exactly is out of scope. However, these specifications can prevent misunderstandings and provide "good hackers" with orientation for legally compliant action. Furthermore, there is no discernible standard in the policies of the companies. Some companies are specific about the process and the information they want to publish, while others only specify how the vulnerabilities should be sent. The comparison also shows that only those companies have a Hall of Fame that are also present on an external platform. The Hall of Fame lists hackers which have already successfully found a security vulnerability (and helped to solve it). Such a public list is an indication of transparency and fair dealing and is an incentive for cooperation for security researchers [20]-[32]. The results of the investigation of the implementation status within

the group of OEM are shown in Figure 1. Among the relevant suppliers, it

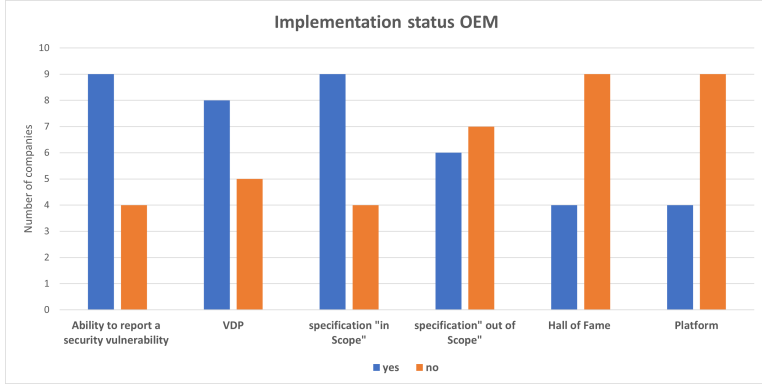


Fig. 1. Implementation status of the disclosure options - OEM

can be observed that most of them do not offer any possibility to report a security vulnerability. Only three companies [36]-[39] provide such an option. Particularly striking is Bosch, which offers the greatest transparency of all the companies considered in this paper through its own platform for reporting security vulnerabilities. Detailed information is provided about the VD program as well as about the publication process itself [33]-[35]. A possible explanation for the fact that only a few suppliers offer a reporting option is given in Section III-C. The results of the investigation of the implementation status within the group of suppliers are shown in Figure 2.

### III. Analysis of Success Grade

#### A. Evaluation Criteria

At the beginning, the number of published security vulnerabilities is considered. The increasing or decreasing willingness to publish them should give an indication of the trend to be expected in the future. In connection with this, the number of published security vulnerabilities is to be put in relation to the security vulnerabilities that have been solved but not made available to the public. However, for the industry as a whole this can only be estimated, as information on this can only be found for individual companies on service provider platforms. The success of companies in dealing with security vulnerabilities is also to be assessed by looking at the patching times for published security vulnerabilities. Relevant vulnerabilities are those whose threat level is rated as high or critical (according to the rating metric of the respective platform). In addition, we will examine whether the (additional) presence on service provider platforms correlates with the

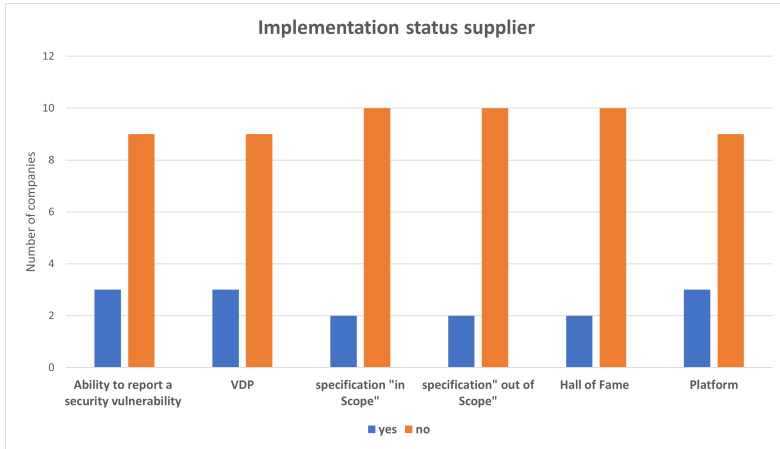


Fig. 2. Implementation status of the disclosure options - suppliers

frequency with which information about fixed vulnerabilities is published. Furthermore, a statement is to be made as to whether further positive or negative effects occur as a result of a service provider platform presence.

### B. Analysis of OEM

It has already been mentioned that the success of the vehicle manufacturers can only be analyzed in its entirety due to the insufficient data available. First of all, the total number of vulnerabilities published during the period under review is taken into account. This shows that although the number of published security vulnerabilities has risen sharply in percentage terms over the period under review, this is not significant when looking at the total numbers (see Figure 3). What is striking, however, is the abrupt increase in 2018 and the steep drop in the numbers in the following year. This can be explained by the fact that several security vulnerabilities were published in 2018 that impacted several vehicles and companies at the same time. In connection with the total published security vulnerabilities, there are also clear differences between the manufacturers. Three brands in particular stand out. BMW has by far the most publications (28), followed by Tesla (21) and Mercedes (15). For all other companies, there are comparatively few published security vulnerabilities (see Figure 4). This could indicate either a lower willingness to publish or a higher level of cyber security of the vehicles of these manufacturers. The attractiveness and public perception of a brand can also have an influence on the number of attacks carried out on its vehicles. This potentially infers a higher number of vulnerability reports as well as corresponding publications. Looking at

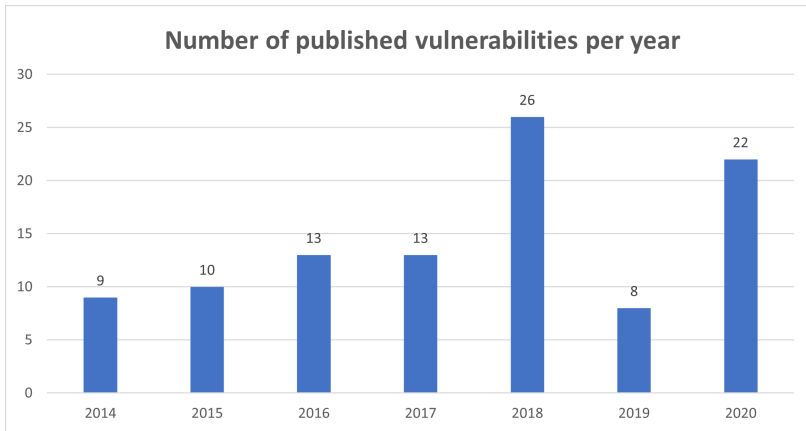


Fig. 3. Number of published vulnerabilities per year

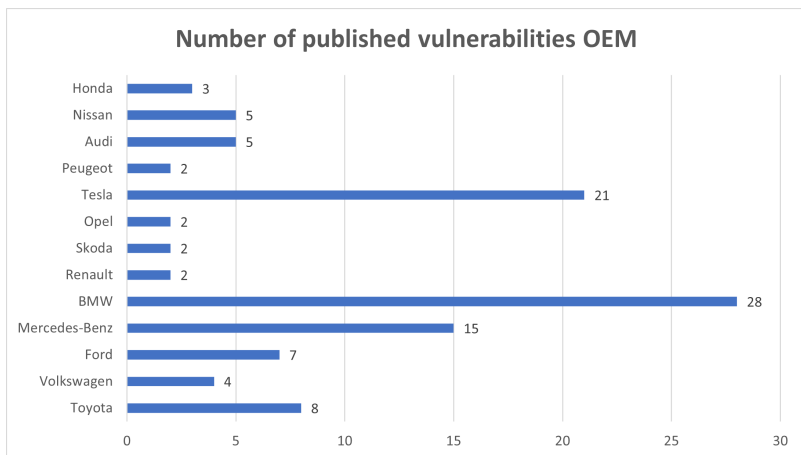


Fig. 4. Number of published vulnerabilities per brand

the publications per brand in relation to the presence on a service provider platform, we see a positive correlation. This could be an indication that cooperation with service provider platforms leads to an increased number of publications (see Figure 5). Over a period of six years, it can be seen that companies with a service provider platform presence (with an average of 12.8 vulnerabilities) have published far more than those without (5.6 publications per brand). In turn, companies that do not offer the option of

reporting security vulnerabilities have published only 12 security vulnerabilities (3 per brand) in six years. Before a conclusion can be drawn, however, it must be remembered that the security vulnerabilities mentioned were not primarily reported via the platforms, but that this is only intended to make a statement about the willingness of the companies to be transparent. In addition, it must be taken into consideration when the companies started their respective platform presence (cf. Figure 5). It can be seen that, in

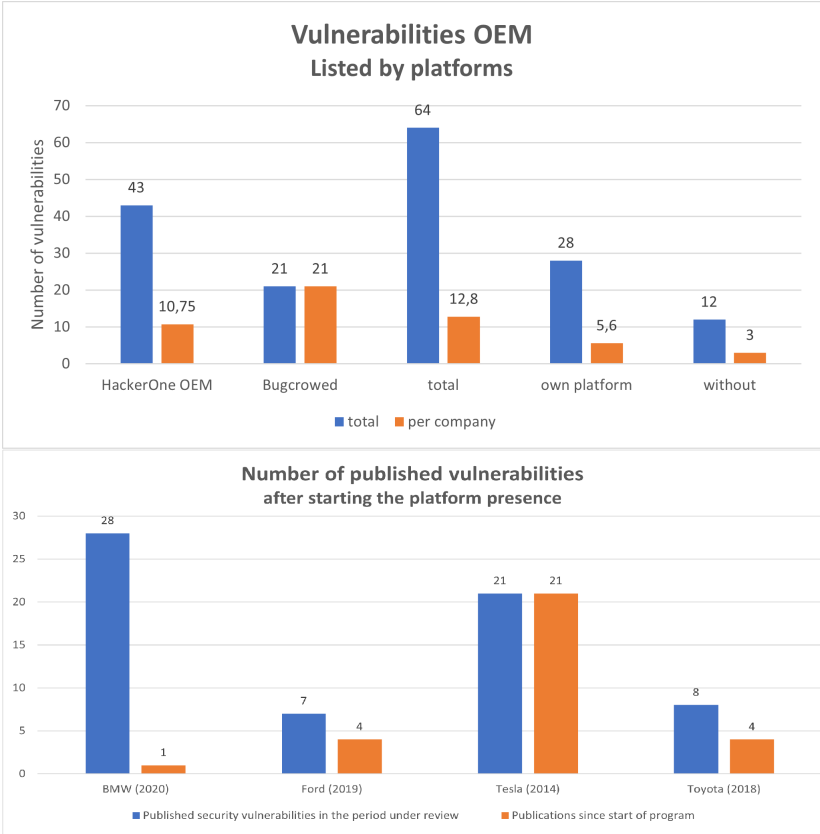


Fig. 5. Number of product related vulnerabilities published, by platform use (top); number of publications, by platform presence launch (below)

connection with Ford and Toyota, comparatively more security vulnerabilities were published after the start of their platform presence than before. Tesla cannot be included here, as the company has a platform presence throughout the entire period under review. BMW, on the other hand, per-

forms worse in this comparison. They have only published one more security vulnerability since the start of their BugBounty program in 2020. However, this can be explained by the fact that some security vulnerabilities have a long patching time [44]-[50], so that a first comparison should not be made until the end of 2021. Nevertheless, the comparison in Figure 5 shows that companies with a platform presence tend to publish more security vulnerabilities. However, this statement must be made cautiously due to the paucity of data. Publications not only reveal the security vulnerability itself, but often also the associated scientific work, the patch time and the severity of the security vulnerability. It is striking that comparatively few "critical" and "high-risk" security vulnerabilities have been reported or published (see Figure 6). The reason for this could be that companies

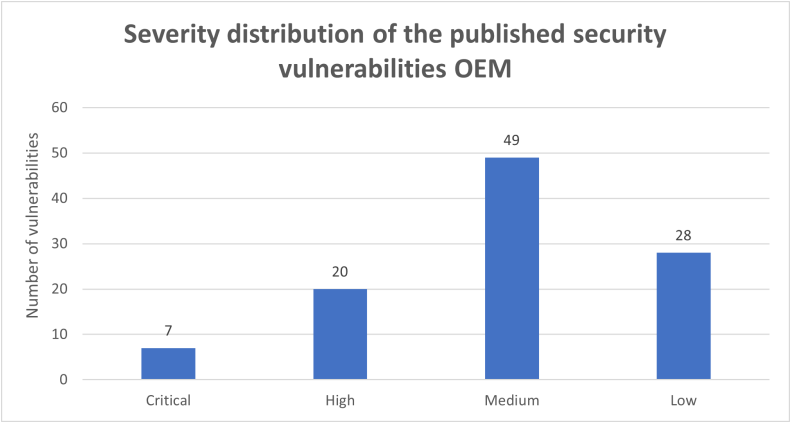


Fig. 6. Severity distribution of the published security vulnerabilites (OEM)

are less interested in particularly critical security vulnerabilities becoming public. It is also possible that they hope this will prevent criminal attackers from finding out about similar high-security vulnerabilities at other companies or in other vehicles. Data that could provide information on this was not collected as part of this work. A large variance can be observed for the patching time of these high-security vulnerabilities. The patching time for security vulnerabilities with severity level High ranges from 5 to 1095 days, whereas for severity level Critical it ranges from 3 to 364 days. Based on the presence of individual vehicle manufacturers on service provider platforms, it can be seen that many more security vulnerabilities are reported and resolved than are published. Ford is an example of this. In the period under consideration, Ford published seven security vulnerabilities, whereas the HackerOne platform announces that the company has solved 1726 security vulnerabilities (since January 2019) [22]. Of course, it must be kept in



mind here that these are security vulnerabilities in other domains, such as webpages, and are thus not relevant to vehicles themselves. However, this very clearly shows the lack of transparency in dealing with security vulnerabilities within the industry. Ford or HackerOne does not provide any information about the exact nature of the security vulnerabilities, but only names the finders.

### *C. Analysis of Suppliers*

An analysis of suppliers is not possible due to the insufficient data available. Publications on security vulnerabilities were only found for components from the manufacturers Bosch and Continental during the period under review. One reason for this may be that hackers do not specifically contact the manufacturers of the various components, but rather the vehicle manufacturers directly, as reporting options are much more widespread among OEM.

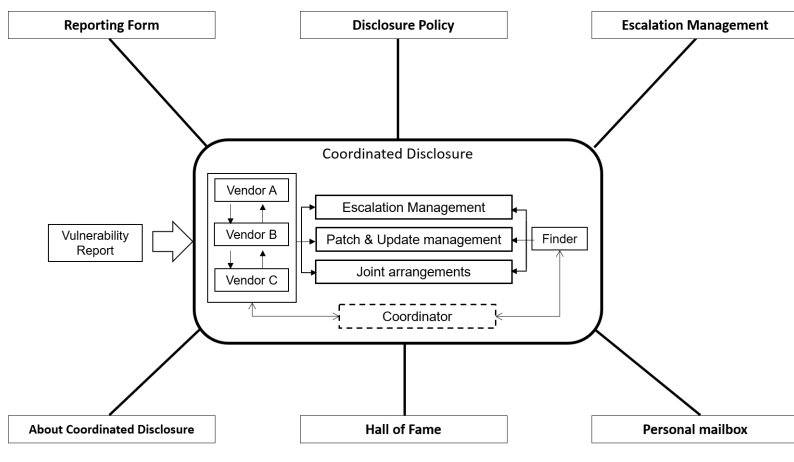
## IV. Conclusion

In summary, there is little transparency in the automotive industry with regard to security vulnerabilities that occur. Neither the exact number nor the success rates, nor the quality of the disclosure processes can be accurately determined. Despite the low level of data, however, this work has shown that there are also companies that stand out positively and have already taken the path toward greater transparency and security. It is also positive to see that most OEMs offer the possibility of reporting security vulnerabilities in encrypted form. Even if the subsequent process mentioned above is often very opaque. It is also clear that companies that are willing to work with external service providers to eliminate security vulnerabilities are more willing to disclose them. However, it also emerges negatively from the research that patching times for high-risk security vulnerabilities can be very variant and long. However, more data would have to be available to make a statistically relevant statement on this. It is also not possible to make a relevant statement about the success of the companies. Only the companies that work with a service provider platform provide an insight into their percentage success rate on the platform. With regard to Tier 1 suppliers, it is apparent that not only do they for the most part not provide the opportunity to report security vulnerabilities, but they also make almost no disclosures. In addition, most of the suppliers analyzed are members of the "Auto-ISAC" association [40]. Through this association, the member companies commit to share emerging security vulnerabilities with the other members and thus identify related or the same security vulnerabilities. Thus, suppliers could feel additionally secured and thus not launch their own programs. Compared to the IT industry, the automotive industry still has a very long way to go before a relevant statement can be made about your success. A search on the NVD database has shown that over 3000[41],

Apple over 7000 [42] and Microsoft over 8000 [43] security vulnerabilities have been published for products from Intel. A central, independent reporting and coordination body in Europe to coordinate and record security vulnerabilities is necessary to better counter the development of potentially increasing cyber incidents, even in the operational phase of vehicles. Even though individual companies place a comparatively high value on the topic of security, the transparency of the entire industry is very weak. As a result, high-risk security breaches may not be resolved for several hundred days and pose a strong threat to the safety of vehicle occupants. The reporting office should be designed in such a way that companies are forced to put an appropriate amount of effort into solving the problem, depending on the severity of the security gap. Government monitoring is also conceivable in this context. Furthermore, in view of the ever-closer autonomous driving and the entry of other technology giants such as Apple [52] and Google [53] into the industry, the issue should probably attract more attention from the public.

## V. Related Work

The work described shows positive trends in recent years, particularly on the side of the OEM with primary responsibility. However, it also shows that there is still potential in terms of both the quantity and quality of reporting options and disclosure processes. We also see a great need for stakeholders outside the industry. For example, there is currently still a lack of independent players in this field who can act as coordinators and mediators when disclosure processes do not proceed to everyone's satisfaction. There is also a lack of independent, trustworthy reporting points with low-threshold reporting options. Such independent actors have established themselves over decades in mature domains such as IT security as a strong and important authority in the field of security management and perform important work both within disclosure processes but also as an informative authority (e.g., CERT/CC [52]). The current work at the Institute for Energy Efficient Mobility at Karlsruhe University of Applied Sciences is derived from the recognition of this need. There, the development of an exemplary demonstrator of a web-based reporting platform for automotive vulnerabilities is in its final phase. This demonstrator is intended to simulate a practical platform that is primarily aimed at non-commercial security researchers. This demonstrator is designed to be practical and could easily be taken into real operation by appropriate instances. The goal is to lower the barrier of entry for potential independent actors and thus promote the establishment of independent coordinators for disclosure processes in the environment of automotive security vulnerability disclosure. The contents of the demonstrator are shown in Figure 7.



Reporting Form	Disclosure Policy	Escalation Management
<ul style="list-style-type: none"> <li>✓ Secure report submission</li> <li>✓ User friendly operation</li> <li>✓ Pseudonym</li> </ul>	<ul style="list-style-type: none"> <li>✓ Basic conditions</li> <li>✓ Expectations</li> <li>✓ Requirements</li> <li>✓ Out-of-Scopes</li> <li>✓ Process workflow</li> </ul>	<ul style="list-style-type: none"> <li>✓ Reasonable grace periods</li> <li>✓ Fixed deadlines</li> <li>✓ Fair &amp; transparent treatment of stakeholders</li> </ul>
About Coordinated Disclosure	Hall of Fame	Personal mailbox
<ul style="list-style-type: none"> <li>✓ Definition of terms</li> <li>✓ Description of ideal process flow</li> <li>✓ Guidance</li> </ul>	<ul style="list-style-type: none"> <li>✓ Voluntary entry</li> <li>✓ Also pseudonym</li> <li>✓ Given incentive &amp; Appreciation</li> </ul>	<ul style="list-style-type: none"> <li>✓ Secure access</li> <li>✓ Current process status</li> </ul>

Fig. 7. Illustration of the headings and contents of the Automotive Vulnerability Reporting Platform-AVRP

#### Author Contributions

Conceptualization, U.K. & R.B.; formal analysis, U.K.; investigation, U.K.; writing-original draft, U.K.; writing-review and editing, R.B.; supervision, R.B. This work was done as part of a student R&D project at the Institute for Energy Efficient Mobility at Karlsruhe University of Applied Sciences in 2021.

#### References

- [1] Upstream Security, "Global Automotive Cybersecurity Report, 2020", <https://www.upstream.auto/research/automotive-cybersecurity/?id=null>, online available 05/04/2020

- [2] S. Jafarnejad, L. Codeca, W. Bronzi, R. Frank, T. Engel, "A Car Hacking Experiment: When Connectivity meets Vulnerability", ISBN:978-1-4673-9526-7, 02/25/2016
- [3] WIRED, "Hackers Remotely Kill a Jeep on a Highway — WIRED", <https://www.youtube.com/watch?v=MK0SrxBC1xs>, online available 07/21/2015
- [4] S. Sinha, Dr. Y. Arora, "Ethical Hacking: The Story of a White Hat Hacker", ISSN: 2347-5552, 2020
- [5] S. S. Malladi, H. C. Subramanian, "Bug Bounty Programs for Cybersecurity", Digital Object Identifier: 10.1109/MS.2018.2880508, 2020
- [6] J. Neutze, "Coordinated Vulnerability Disclosure (CVD)", Brussels, 2017.
- [7] Statista GmbH, "Top 10 Automodelle weltweit nach Anzahl der Verkäufe", <https://de.statista.com/statistik/daten/studie/734092/umfrage/top-10-automodelle-weltweit-nach-anzahl-der-verkaeufe/>, online available 12/01/20
- [8] Car Sales Statistics, H. Bekker, "2018 (Full Year) International: Global Top-Selling Car Models", <https://www.best-selling-cars.com/global/2018-full-year-international-global-top-selling-car-models/>, online available 12/01/20
- [9] Car Sales Statistics, H. Bekker, "2019 (Full Year) Europe car sales and market analysis", <https://www.best-selling-cars.com/europe/2019-full-year-europe-car-sales-and-market-analysis/>, online available 12/01/20
- [10] Car Sales Statistics, H. Bekker, "2018 (Full Year) Europe: Best-Selling Car Models", <https://www.best-selling-cars.com/europe/2018-full-year-europe-best-selling-car-models-in-the-eu/>, online available 12/01/20
- [11] Car Sales Statistics, H. Bekker, "2019 (Full Year) Germany: Best-Selling Car Models", <https://www.best-selling-cars.com/germany/2019-full-year-germany-best-selling-car-models/>, online available 12/01/20
- [12] Car Sales Statistics, H. Bekker, "2019 Germany: Car Theft — Most-Often Stolen Cars", <https://www.best-selling-cars.com/germany/2019-germany-car-theft-most-often-stolen-cars/>, online available 12/01/20
- [13] C. Köllner, "Das sind die größten Automobilzulieferer", <https://www.springerprofessional.de/unternehmen---institutionen/transformation/das-sind-die-groessten-automobilzulieferer/18157500#:~:text=Bosch%2C%20Continental%20und%20Denso%20f%C3%BChren,Valeo%20komplettiert%20die%20Top%20D10>, online available 12/01/20
- [14] Statista GmbH, "Größte europäische Automobilzulieferer nach weltweitem Umsatz", <https://de.statista.com/statistik/daten/studie/1017154/umfrage/groesste-europaeische-automobilzulieferer-nach-weltweitem-umsatz/>, online available 12/01/20
- [15] S. Gelowicz, "Was sind Automobilzulieferer? Grundlagen, Ranking und Beispiele", <https://www.automobil-industrie.vogel.de/was-sind-automobilzulieferer-grundlagen-ranking-und-beispiele-a-724889/>, online available 12/01/20
- [16] National Institute of Standards and Technology, USA, <https://nvd.nist.gov/>
- [17] Institut für Energieeffiziente Mobilität, Hochschule Karlsruhe, Automotive Attack Database, <https://github.com/IEEM-HsKA/AAD>, online available 11/20/21
- [18] Plattform HackerOne, <https://www.hackerone.com/>, online available 11/20/21
- [19] Plattform Bugcrowd, <https://www.bugcrowd.com/>, online available 11/20/21
- [20] Toyota AG, "Vulnerability Disclosure Program", <https://hackerone.com/toyota?type=team>, online available 12/04/20
- [21] Volkswagen AG, "Kontakt Cyber Security Volkswagen", <https://www.volkswagen.de/de/mehr/rechtliches/kontakt-cyber-security.html>, online available 12/04/20
- [22] Ford AG, "Vulnerability Disclosure Program – Policy", <https://hackerone.com/ford?type=team>, online available 12/04/20

- [23] Mercedes Benz AG, "Vulnerability Reporting Policy", <https://www.mercedes-benz.com/en/whitehat/> , online available 12/04/20
- [24] BMW AG, "How to report vulnerabilities" , <https://www.bmwgroup.com/en/general/Security1.html> , online available 12/04/20
- [25] BMW AG, "Bug Bounty Program- Policy", [https://hackerone.com/bmwgroup?type=team&view\\_policy=true](https://hackerone.com/bmwgroup?type=team&view_policy=true) , online available 12/04/20
- [26] Renault AG, "Vulnerability Disclosure Policy", <https://group.renault.com/en/vulnerability-disclosure-policy/> , online available 12/04/20
- [27] ŠKODA AUTO Deutschland GmbH, Cyber Security, <https://www.skoda-auto.com/other/cyber-security> , online available 12/26/20
- [28] Tesla Inc., "Product Security", [https://www.tesla.com/de\\_DE/about/security](https://www.tesla.com/de_DE/about/security) , online available 12/26/20
- [29] Tesla Inc., "Bug Bounty Program", <https://bugcrowd.com/tesla> , online available 12/26/20
- [30] Audi AG, "Kontakt Cyber Security Audi - Vulnerability Reporting Policy", <https://www.audi.de/de/brand/de/kontakt-und-haendler/kontakt-audi-cyber-security.html> , online available 12/26/20
- [31] Toyota AG, "Vulnerability Disclosure Program – Policy", <https://hackerone.com/toyota/hackitivity?type=team> , online available 01/03/21
- [32] BMW AG, "Bug Bounty Program – Hackitivity", <https://hackerone.com/bmwgroup/hackitivity?type=team> , online available 01/03/21
- [33] Robert Bosch GmbH, "Bosch PSIRT", <https://psirt.bosch.com/about-us/> , online available 12/26/20
- [34] Robert Bosch GmbH, "Bosch PSIRT- report a vulnerability", <https://psirt.bosch.com/report-a-vulnerability/> , online available 12/26/20
- [35] Robert Bosch GmbH, "BOSCH Responsible Disclosure Program for Website Vulnerabilities", <https://bugcrowd.com/bosch-vdp?preview=fb8e0a17dfe4ebaf6ac9268fb57706c3> , online available 12/26/20
- [36] Robert Bosch GmbH, "Notification process for data protection requests", <https://www.bkms-system.net/bkwebanon/report/clientInfo?cin=18rbd519&c=-1&language=eng> , online available 12/26/20
- [37] Continental AG, "Continental Product Security Incident Response Management", <https://www.continental-automotive.com/en-gl/Passenger-Cars/Company/PSIRM-Website> , online available 12/26/20
- [38] ZF Friedrichshafen AG, Dr. B. Murray, "Den Hackern voraus sein", [https://www.zf.com/site/magazine/de/articles\\_18560.html](https://www.zf.com/site/magazine/de/articles_18560.html) , online available 12/22/20
- [39] BASF SE, "Responsible Disclosure Statement" , <https://www.basf.com/global/en/legal/responsible-disclosure-statement.html> , online available 12/22/20
- [40] Automotive Information Sharing and Analysis Center-AutoISAC, <https://automotiveisac.com/> , online available 11/20/21
- [41] National Institute of Standards and Technology , "Search: Intel", [https://nvd.nist.gov/vuln/search/results?form\\_type=Basic&results\\_type=overview&query=Intel&queryType=phrase&search\\_type=all&isCpeNameSearch=false](https://nvd.nist.gov/vuln/search/results?form_type=Basic&results_type=overview&query=Intel&queryType=phrase&search_type=all&isCpeNameSearch=false) , online available 02/10/21
- [42] National Institute of Standards and Technology , "Search: Apple" [https://nvd.nist.gov/vuln/search/results?form\\_type=Basic&results\\_type=overview&query=Apple&queryType=phrase&search\\_type=all&isCpeNameSearch=false](https://nvd.nist.gov/vuln/search/results?form_type=Basic&results_type=overview&query=Apple&queryType=phrase&search_type=all&isCpeNameSearch=false) , online available 02/10/21
- [43] National Institute of Standards and Technology , "Search: Microsoft", [https://nvd.nist.gov/vuln/search/results?form\\_type=Basic&results\\_type=overview&query=Microsoft&queryType=phrase&search\\_type=all&isCpeNameSearch=false](https://nvd.nist.gov/vuln/search/results?form_type=Basic&results_type=overview&query=Microsoft&queryType=phrase&search_type=all&isCpeNameSearch=false) , online available 02/10/21

- [44] VULDB, "Global TechStream bis 15.10.032 Pufferüberlauf", <https://vuldb.com/de/?id.159170>, online available 01/22/21
- [45] JPCERT/CC and IPA , "A vulnerability in TOYOTA MOTOR's DCU (Display Control Unit) ", <https://jvn.jp/en/vu/JVNVU99396686/index.html>; online available 01/22/21
- [46] Tencent Keen Security Lab, "Exploiting Wi-Fi Stack on Tesla Model S", <https://keenlab.tencent.com/en/2020/01/02/exploiting-wifi-stack-on-tesla-model-s/>, online available 01/22/21
- [47] VULDB, "NVIDIA Vibrante Linux 1.1/2.0/2.2 User Space Driver erweiterte Rechte", <https://vuldb.com/de/?id.135962>, online available 01/22/21
- [48] VULDB, "Audi A7 2014 MMI Multiplayer Format String", <https://vuldb.com/de/?id.164741>, online available 01/22/21
- [49] HackerOne, "Subdomain takeover on usclsapipma.cv.ford.com", <https://hackerone.com/reports/484420>, online available 01/22/21
- [50] Bugcrowd, "Disclosed ServiceNow credentials leading to data leak", <https://bugcrowd.com/disclosures/41991d9c-3866-40e7-97c2-f33b812565d7/disclosed-servicenow-credentials-leading-to-data-leak>, online available 01/22/21
- [51] CERT/CC Computer Emergency Response Team/Coordination Center, "What is Vulnerability Coordination? ", <https://vuls.cert.org/confluence/pages/viewpage.action?pageId=4718642>, online available 05/28/2020
- [52] M. Gurman, "Apple Accelerates Work on Car Project, Aiming for Fully Autonomous Vehicle"
- [53] Waymo LLC, <https://waymo.com>, online available 11/20/21
- [54] National Institute of Standards and Technology , "CVEs and the NVD Process", <https://nvd.nist.gov/general/cve-process>, online available 11/23/2021

# Detecting Hardware Modifications on Vehicles

Faruk Altun

*Faculty of Mechanical Engineering  
and Mechatronics  
Hochschule Karlsruhe  
Karlsruhe, Germany  
alfa1016@h-ka.de*

Felix Müller

*Institute of  
Energy Efficient Mobility  
Hochschule Karlsruhe  
Karlsruhe, Germany  
mufe0001@h-ka.de*

## Abstract

Modern vehicles consist of a number of electronic control units (ECUs) dedicated to specific tasks. The behavior of the car can be modified by introducing additional, third-party ECUs. This can be a result of users wanting to add or modify functionality, or a result of an attack on the vehicle. Especially the modification of functions that impact the vehicles behavior in regard to driving functions or performance figures might result in the vehicle no longer being road legal. The scope of this project report is to showcase ideas, on how to identify additional ECUs.

## Index Terms

ECU, Third-Party, Power Measurement

## I. Introduction

The number of electronic components in vehicles is constantly increasing. Today, a modern luxury-class vehicle contains up to 100 electronic control units (ECUs) [1], most of which interact with each other via complex networks to exchange data and use shared resources such as sensors connected to these networks. Accordingly, changes to vehicles and their networks can also have a serious impact on road safety.

For the majority of future vehicle applications it is important to have a high level of communication security. Among other things, this includes ensuring that critical information is only transmitted in encrypted and authenticated form to prevent possible manipulation. Modern security mechanisms are based on cryptographic algorithms and protocols. These enable secrecy, tamper-resistance and authentication and are consequently able to ensure the most essential challenges to automotive communication security. [2] Commonly used communication technologies in a car today do not support this throughout, so a modification can not be excluded in its entirety. [3]

The goal of the semester project titled "Detecting Hardware Modifications on Vehicles" was to develop approaches to detect third-party hardware on

vehicles. Therefore, the work has gone through a research phase and a conceptualization phase. In the research phase, the relevant and affected bus systems were explained in more detail, including their use, advantages and disadvantages.

The basic knowledge about bus systems in vehicles is important, because they are affected by the attachment of third-party hardware. [4], [5] This is followed by researching the variety of possible modifications to vehicles, as well as their detection and impact on the vehicle. Afterwards, potential intrusion points, such as various interfaces, at which attackers can tap, were pointed out. The main part of the work is the conceptual design phase. In this phase, three concrete solution approaches were developed, which should enable the detection of foreign hardware on vehicles.

## II. Possible Solutions

The following approaches deal with various ways of foreign hardware detection. Certainly, it is difficult to detect foreign hardware with the naked eye unless one has a trained eye and has already studied such methods, or they are readily visible. [6] In addition, third-party hardware can be designed to not be noticeable for the car itself, and can be hidden in various compartments and behind cladding. As a consequence, one would not necessarily be able to detect foreign hardware. Without a suspicion of foreign hardware, one would also not get the idea to start looking for it. In the following we will take a closer look at various approaches.

### A. *Scanning the vehicles*

Proposed solution 1 is based on the use of mobile scanners. Mobile scanners can already be used to detect contraband and hidden objects. This technology is already used for trucks. However, this involves objects that do not interact with the vehicle or have any influence on the vehicle. X-ray technology also finds its application at the airport. The suitcases of the numerous passengers are scanned and if dangerous and illicit goods are suspected, they are unpacked and disposed of. The technology is also used in health care for people. Here, for example, existing fractures can be detected, as the bones then do not correspond to normal findings, but look pathological.

Exactly this technology, can be used for the recognition of foreign hardware. In order to be able to use the technology here, mobile X-ray scanners for cars of any kind would be needed in the first place. This scanner would then x-ray the cars and show the x-ray image on a connected monitor. For comparison, an reference X-ray image of the car is needed, which shows how the vehicle looks originally or freshly after delivery from the production.

The numerous images of the various cars according to their original form are stored in a database. The software is then able to compare the current image with the original file to detect extraneous hardware, such as on



the ECU. In order to establish the method, one would need a functioning software, which can fulfill the requirements. Furthermore, one has to rely on the trustworthiness of the source for the reference X-ray images.

A currently available example can be seen in a giant tube for X-raying cars at the Fraunhofer-Entwicklungszentrum Röntgentechnik EZRT in Fürth. The gigantic turntable with a diameter of three meters of the XXL CT rotates slowly and drives into an X-ray source, in which a four-meter-long X-ray detector is located. The computer tomograph of a different dimension can scan entire cars, airplane wings and even entire shipping containers [7].

Furthermore, there are already fully automated X-ray scanners that can find damage to cars in the shortest possible time [8] and the so called ProofStation that utilizes images for the inspection of cars for damages. A total of 25 cameras take 600 images, providing a 360-degree analysis of the car. [9] Merging these concepts could help in the current quest.

Now the question also arises, what are the advantages and disadvantages of this approach? Among the advantages is the duration of the process. In a very short time one has the detection if foreign hardware is attached and the result is very reliable because the hidden hardware can be detected by the X-ray image. The disadvantages include the great effort until you have the reliable X-ray images and can build the database. Likewise, the rays could be dangerous for the people who are near the device. Additionally, objects cannot be distinguished 100%. For example, a pen in the back seat could be seen as a hazard, or third-party devices hidden in the cladding or in metal folds of the cars structure can be hard to identify. Especially, if the inspected cars are not cleaned of additional stuff transported in the vehicle before inspection. The device would also be very impractical due to its size.

#### *B. PIN query when adding a third-party hardware*

Another approach is the prevention of third-party hardware in advance. Vehicle Software is becoming increasingly standardized. [10] A software monitoring network activity could be implemented. The idea is to actuate a two-way confirmation, such as a PIN request, when adding third-party hardware. With this, the car owner is warned in case foreign hardware wants to act on the car. Nowadays, there are various apps from car manufacturers, which can be connected to smartphones. In case of an attack, installation of the foreign hardware is carried out without the presence of the driver, the linking of the software to be developed with the mutual confirmation also with the app is of important significance. This of course would fail as a possible measure, if the third party hardware would be installed by the owner of the vehicle.

Now we come to the advantages and disadvantages here as well. Among the advantages is the ease of use, since the third-party hardware can be controlled by the smartphone. For the user, this is just an additional function

without much effort, since he does not have to be on site as, for example, with the mobile scanner. In case of non-authorization, the third-party hardware does not bring any benefit. This solution also serves for control purposes and one can follow conscious or even unconscious changes as a user of the vehicle. Disadvantages would arise if, for example, the smartphone is stolen and one has not secured the app well. Additionally this could be implemented into the fleet management software. [11]

### *C. Foreign hardware detection through power consumption*

Modern cars have many comfort and safety functions. All kinds of devices and assistants consume electricity and put a strain on the battery. The 3rd solution refers to this train of thought, as third-party systems installed in a vehicle have to be powered as well.

The solution would be to install a consumption meter or consumption analyzer on the vehicle. This device would record measured values and warn the driver via the on-board computer if the difference from the standard consumption is too great. The warning would indicate that third-party hardware is installed and is causing the increased consumption. This proposal also has advantages and disadvantages. The advantage is that the user will be alerted when excess consumption occurs and will check his vehicle. The attachment of such a device is not visible from the outside and does not damage the appearance, for example. A disadvantage is that the excess consumption cannot be localized.

## III. Experimental evaluation

An experiment was set up to test the detection of third-party hardware via its power consumption. The test setup consisted of a vehicle network from a German executive class vehicle. Powered by a power supply from the wall-outlet. The Vehicle network was in a state that is somewhat equivalent to that of a parked car with the ignition set to power certain accessories, although not all devices that would consume power in this scenario were present and installed in the vehicle network. To simulate a Third-Party-Device that is connected to the vehicle, an additional electronic control unit was attached to the system. Even though the power was fluctuating, the added device could clearly be measured.

To get the necessary measurements, a MSO07034B Infini Vision Oscilloscope was used in conjunction with a multi-meter of the type 34405A, both made by Agilent. To measure the power draw, a resistor with 0.823 Ohm was connected in series with the test bench to measure the voltage across the resistor and calculate the power draw.

As shown in Figure 1, while the power was fluctuating as expected, the added power consumption of the third-party-hardware can be seen quite clearly. In order to get a better understanding of the current draw, the

TABLE I

Comparison between the power consumption of the system under test, without the added hardware and with the third-party-device, mean values.

Measurement	Reference	with Third-Party-Hardware
Mean over 100s	7,047 A	8,210 A
Min (Mean)	6,990 A	8,146 A
Max (Mean)	7,102 A	8,425 A
Std. Deviation (Mean)	25,027 mA	60,476 mA

mean average was calculated for the different scenarios over a period of 100s each and can be seen in Table I.

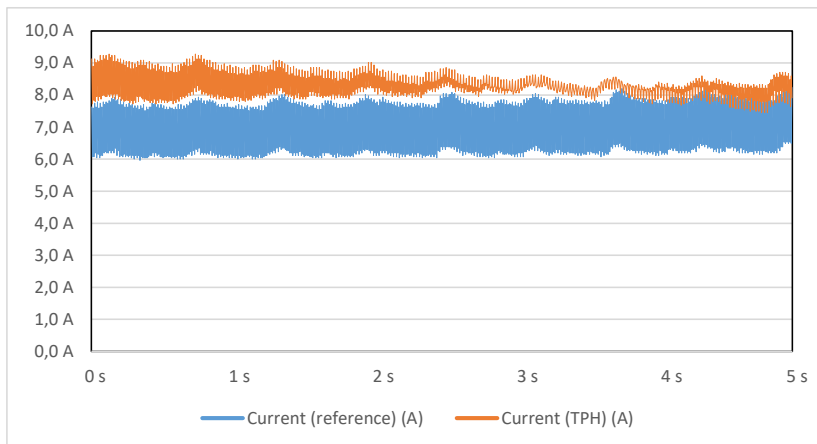


Fig. 1. Current readings of the test setup over a period of 5 seconds. Displayed is the reference power draw in comparison to the power draw (shown in blue) with the added third-party-hardware (marked as TPH and shown in orange).

#### IV. Conclusion

As technology is always being advanced, safety and security should be a focus and solutions regarding safety issues should be put in place. Based on the research that is presented here, three solutions have emerged in this semester project. These were scanning the vehicle for modifications, requiring access to vehicle functionality via PIN query to ensure the owner or driver is aware of the modification and monitoring the power consumption of the vehicles electronic components. All three solutions could be very helpful for detecting third-party hardware. While the evaluation of the power consumption seems promising, in detecting variations in the

power draw caused by additional hardware it could also only detect devices, that were powered through the car. A more in depth look in how such a monitoring system could be incorporated would be the next step to explore this approach.

### Acknowledgment

This report is the short version of a student research project from the curriculum of the bachelor's degree program in automotive technology, allowing the participation of students in ongoing research projects. The goal was to evaluate new ideas for the detection of third-party-hardware in vehicles. The project has been completed by Faruk Altun. Felix Müller contributed with the translation of this report and with the setup of the experiment.

### References

- [1] J. Mathes, "Revolution durch die Hintertüre", <https://www.hanser-automotive.de/a/fachartikel/revolution-durch-die-hintertuere-326780>, last accessed 21. December 2021.
- [2] W. Rudschies, and T. Kroher, "Autonomes Fahren: So fahren wir in Zukunft", <https://www.adac.de/rund-ums-fahrzeug/ausstattung-technik-zubehoer/autonomes-fahren/technik-vernetzung/aktuelle-technik/>, last accessed 21. December 2021.
- [3] A. Dausenau, "Autonomes Fahren und Angriffssicherheit", <https://www.uni-koblenz-landau.de/de/koblenz/fb4/ist/AGZoebel/Lehre/sommer2016/SeminarASidA/A4>, last accessed 21. December 2021.
- [4] Kunbus, "KFZ Bussysteme, Protokolle und Standards 2", <https://www.kunbus.de/kfz-bussysteme-protokolle-und-standards-2.html>, last accessed 21. December 2021.
- [5] softing, "Bussysteme", <https://automotive.softing.com/de/standards/bus-systeme.html>, last accessed 21. December 2021.
- [6] A. Bender-Napp, "So wird Ihr Auto zum Smart-Car", <https://www.auto-motor-und-sport.de/tech-zukunft/mobilitaetsservices/vernetzung-per-obd/>, last accessed 21. December 2021.
- [7] Fraunhofer IIS, "Hochenergie-Computertomographie", <https://www.iis.fraunhofer.de/de/ff/zfp/tech/hochenergie-computertomographie.html>, last accessed 21. December 2021.
- [8] D. Stoller, "Riesenröhre durchleuchtet ganze Autos", <https://www.ingenieur.de/technik/fachbereiche/messtechnik/riesenroehre-durchleuchtet-ganze-autos/>, last accessed 21. December 2021.
- [9] Proovstation, "KI gestützte Fahrzeug Inspektion Automatisierte Schadenserkennung", <https://de.proovstation.com/>, last accessed 21. December 2021.
- [10] Vector Informatik, "AUTOSAR Classic Plattform Der serienerprobte Standard für Fahrzeugsteuergerte", <https://www.vector.com/de/de/know-how/autosar/autosar-classic/>, last accessed 21. December 2021.
- [11] Kommsoft, "Software für Fuhrparkmanagement", <https://www.kommsoft.de/fuhrparkmanagement>, last accessed 21. December 2021.

# A Benchmark Study of Chemical Reaction Mechanisms for Ignition Delay Calculation of Natural Gas/Hydrogen Mixtures under Internal Combustion Engine Conditions

1<sup>st</sup> Jörn A. Judith  
2<sup>nd</sup> Maurice Kettner  
*Karlsruhe University of  
Applied Sciences*  
Karlsruhe, Germany  
joern.judith@h-ka.de  
maurice.kettner@h-ka.de

3<sup>rd</sup> Thomas Koch  
*Karlsruhe Institute of Technology*  
Karlsruhe, Germany  
thomas.a.koch@kit.edu

## Abstract

The usage of chemical reaction mechanisms in numerical simulations offers the possibility to describe combustion phenomena in a high level of detail. However, different degrees of complexity and varying limits of mechanism validation complicate the commitment to a specific reaction mechanism in practical application. This work presents a benchmark study using natural gas reaction mechanisms well-established in literature for calculation of ignition delay under internal combustion engine conditions. Simulation results of the chemical reaction mechanisms GRI 3.0, USC II, San Diego, CRECK 2003 and AramcoMech 1.3 were tested against ignition delay data from shock tube and rapid compression machine experiments reported in literature. The experimental reference data were selected in order to cover the operating ranges of both spark and compression ignition engines, thus including lean and stoichiometric mixtures ( $\lambda = 1.0$ -3.3), pressures between 10-160 bar and temperatures in the range of 850-1500 K. Significant amounts of hydrogen admixed to natural gas up to 80 %-mol were considered in order to address the rising significance of hydrogen. While the absolute differences between experiments and simulations strongly depend on the specific boundary conditions, it is found that there is overall excellent agreement of AramcoMech 1.3 and CRECK 2003 with measurement results. The less complex mechanisms GRI 3.0, USC II and San Diego reasonably predict ignition delay in the high temperature regime above 1100 K as well, yet a tendency to underestimate ignition delay exists. Concerning elevated hydrogen fractions, the accuracy

of AramcoMech 1.3 and CRECK 2003 found for pure hydrocarbons holds, USC II tends to overpredict reactivity and the prediction accuracy of GRI 3.0 and San Diego depends on the actual hydrogen fraction studied.

### **Index Terms**

Chemical reaction mechanism, Natural gas, Hydrogen, Ignition delay, Simulation

## **I. Introduction**

The possibility to buffer surplus electrical, regenerative energy in form of pure hydrogen and hydrogen-based gaseous or liquid fuels, also known as Power-to-X technologies, emphasizes the significant role of hydrogen within the imminent transformation process from fossil to regenerative energy sources. Considering the energy sector, the existing European natural gas pipeline infrastructure already allows for significant amounts of ca. 10 %-mol hydrogen admixed to natural gas [1] [2]. Though the natural gas network allows even higher hydrogen enrichment, maximum hydrogen content is limited by the capability of domestic appliances and energy facilities to process those hydrogen enriched fuel mixtures without failure [2] [3]. Owing to its high reactivity, increasing hydrogen fraction notably changes combustion characteristics compared to pure natural gas with respect to the flammability limits, flame speed, ignition delay and peak temperatures. For the design of future combustion systems fueled with natural gas, e.g. combined heat and power plants or gas turbines, thus the challenge arises to ensure long term operation and high efficiency while complying with emission legislation using natural gas as well as natural gas/hydrogen mixtures.

The usage of chemical reaction mechanisms represents an advanced method to describe combustion phenomena in numerical simulations, yet at high computational costs. Owing to the gradual decomposition of larger fuel species to smaller molecules, i.e. to smaller, stable fuel species, the oxidation schemes of  $H_2/O_2$  and  $C_1-C_4$  form the core mechanisms of smaller as well as larger hydrocarbon fuels [4]. Natural gas mainly comprises  $C_1-C_3$  alkanes as well as minor fractions of  $C_4-C_6$  alkanes, which amounts significantly vary depending on the region the natural gas is extracted from. In comparison, liquid fuels incorporate larger fuel species, exhibiting complex chemistry and thousands of reaction pathways to be considered during combustion. The lower complexity of natural gas fuel species allows for much lower computational effort at still high level of detail compared to liquid fuels, leading to a high applicability of natural gas mechanisms within numerical simulations.

In recent years several groups of researchers made great effort in developing chemical reactions mechanisms with different level of detail to model combustion of main natural gas constituents. Table I presents an overview of mechanisms frequently used for simulation of natural gas combustion

and details their count of species and reactions and the year of release. Owing to the large number of mechanisms available in literature, the overview shown represents a sample based on a literature review, which does not seek completeness. The selection of reaction mechanisms herein strives for significant archival mechanisms as well as latest mechanisms accessing most recent literature data.

The Gas Research Institute (GRI) 3.0 [5], San Diego [6] and University of Southern California (USC) II [7] mechanisms represent prominent examples of mechanisms comprising a concise set of species and reactions allowing for a sound calculation of fundamental kinetics. GRI 3.0 includes rate parameters fitted to match a comprehensive experimental database including alkanes up to  $C_3$  and is thought to be one of the first freely available natural gas mechanisms. Ignition delay validation targets cover a pressure range from 0.25-83.7 bar, high temperatures 1350-1850 K and a wide range of air-fuel ratios ( $\lambda$ ) within  $\lambda = 0.2$ -3.7.

USC II includes  $C_4$  alkane chemistry and its rate parameters definition bases on an extensive literature review. The mechanism is validated against ignition delay data of hydrogen, alkanes as well as alkenes under high temperatures and high pressure conditions up to 2200 K and 87 bar, while covering lean and rich mixtures in the range of  $\lambda = 0.5$ -2.0.

The San Diego mechanism is built on the approach of keeping the number of species and reactions to a minimum, thus involving only species up to  $C_4$ , which are expected to dominate the ignition process. Though the mechanism does not come along with direct comparison against measurement data as provided for other mechanisms, its authors state that it covers a wide range of high temperature combustion phenomena.

The most recent mechanisms of the National University of Ireland (NUI) in Galway, termed as AramcoMech, represent comprehensive mechanisms based on an extensive review of available mechanisms as well as of theoretical and experimental studies on rate parameters and thermochemical species properties. AramcoMech was developed in a hierarchical manner for  $C_1$ - $C_4$  hydrocarbon and oxygenated fuels and subsequently validated for  $C_1$ - $C_2$  (version 1.3 [8]),  $C_3$  (version 2.0 [9]) and  $C_4$  species (version 3.0 [10]). An exhaustive set of validation data using flames, shock tubes, rapid compression machines (RCM), jet-stirred and flow reactors accompanies the mechanisms. Ignition delay data of shock tubes used for validation cover air-fuel ratios in the range of  $\lambda = 0.17$ -16.7, low as well as high temperature oxidation within 800-2600 K and a large pressure range of 0.65-260.0 bar.

The latest version of the Chemical Reaction Engineering and Chemical Kinetics Laboratory (CRECK) modeling group mechanism including  $C_1$ - $C_3$  alkanes (version 2003 [11]) is based upon a coupling of former work of Ranzi et al. [12] [13] and AramcoMech 2.0 and includes several modifications for the sake of performance outlined in [11]. In addition to the AramcoMech validation data, the validation of CRECK 2003 includes a set

of methane MILD and OXY-fuel combustion data indicated by high dilution rates in  $N_2$ ,  $H_2O$  and  $CO_2$  at high temperatures within 1330-2000 K.

Heavier fuel species decompose to smaller fuel species during oxidation. This hierarchical relation enables simulation of natural gas combustion also using mechanisms focusing on larger C species. Two prominent mechanisms capturing n-heptane combustion are the Lawrence Livermore National Laboratory (LLNL) mechanism 3.1 [14] and the California Institute of Technology (Caltech) mechanism 2.3 [15]. Low temperature ignition was validated for both mechanisms in a temperature between 650 and 1250 K, pressures up to 50 bar and air-fuel ratios in between  $\lambda = 0.5$ -3.3 (LLNL) and  $\lambda = 0.5$ -4.0 (Caltech), respectively. It can be noted that additional data supporting the validation of the Caltech mechanism emphasize its performance with respect to smaller hydrocarbons as well, though at significantly lower pressures. Both mechanisms feature the advantage of capturing the negative temperature coefficient behaviour of larger hydrocarbon fuels, which becomes more important as the fuel mixture includes notable amounts of heavier species, e.g. in dual fuel engines.

TABLE I  
Reaction mechanisms for simulation of natural gas and natural gas/hydrogen combustion

No	Mechanism	Species	Reactions	Year	Ref
1	GRI 3.0	53	325	1999	[5]
2	USC II	111	784	2007	[7]
3	San Diego	58	270	2016	[6]
4	AramcoMech 1.3	253	1542	2013	[8]
5	AramcoMech 2.0	493	2716	2015	[9]
6	AramcoMech 3.0	581	3037	2018	[10]
7	LLNL n-hept 3.1	654	2827	2012	[14]
8	CRECK 2003 C3 HT	114	1999	2020	[11]
9	Caltech 2.3	192	1156	2015	[15]

### *Scope of this Work*

From the brief review of natural gas mechanisms presented, two main challenges evolve when it comes to the selection of a reaction mechanism for a specific application. On the one hand, different level of detail exist, accompanied with severe differences in calculation costs. On the other hand, the validation ranges outlined for each mechanism are not full factorial, i.e. not every species is validated over the complete temperature



and pressure ranges mentioned. Thus, the usage of a reaction mechanism requires evaluation of its performance under the specific operating conditions of the target combustion system. The aim of this work is the evaluation of multiple reaction mechanisms described in the previous section with respect to the quality of their ignition delay predictions under internal combustion engine conditions. Analyses explicitly focus on the deviation between measurement data and mechanisms predictions with respect to the general ignition tendencies for varying mixture quality, ignition pressure and ignition temperature. Examination of effects promoting or inhibiting ignition, e.g. higher temperatures or larger amounts of higher hydrocarbons, lies not within the scope of this work. In the subsequent section a brief description of the methodology employed is presented, followed by the results section including the comparison between measurement and simulation data. Eventually, the main findings of the benchmark analysis are summarized.

## II. Methodology

### *Selection of Reaction Mechanisms*

For the sake of conciseness, the benchmark study includes a reduced set of mechanisms presented in Table I. The work of Schuh et al. [16] compares the performance of the three versions of AramcoMech at 100 bar, compression temperature below 1000 K and four different methane/propane mixtures comprising propane fractions up to 30 %. They found good agreement of AramcoMech 1.3 and 3.0, matching the experimental data very well. An additional comparison at higher temperatures 1000-1300 K and lower pressures 24-28 bar within their work revealed minor advantages of version 3.0 over version 1.3 at high propane fractions  $\geq 30\%$ . Taking into account that a) version 3.0 results in much higher computational effort and b) the agreement of version 3.0 and 1.3 is good even at low temperature and high propane fractions, AramcoMech 1.3 was chosen for the benchmark study of this work. The mechanisms of LLNL and Caltech address higher carbon number fuels and yet exhibit moderate counts of reactions of 2827 and 1156, respectively. Compared to the detailed  $C_1$ - $C_4$  mechanisms comprising similar reaction counts it can be assumed that LLNL and Caltech capture the influence of intermediate species to a lesser degree, yet they benefit when the importance of higher hydrocarbon low temperature oxidation rises. For the reasons presented, CRECK 2003 and AramcoMech 1.3 were selected as representative mechanisms possessing a high level of detail. In addition, GRI 3.0, San Diego and USC II were examined representing compact mechanisms. In the sense of clarity the mechanisms are termed as CRECK, Aramco, GRI, USC and San Diego in the further course of this work.

### *Selection of Validation Data*

The accuracy of ignition delay predictions is examined by comparing simulation results with experimental data of shock tubes and rapid compression machines taken from literature. Experiments cited herein are chosen based on their experimental boundary conditions to fit the scope of this work, rather than on their consideration during the development of a specific mechanism. That means that the reference data might be part of the set of validation data for some mechanisms, yet they do not explicitly have to. The majority of the experimental data selected for validation addresses the operating range of compression ignition (CI) gas engines, i.e. high dilution and high pressure. An additional set of measurement data including stoichiometric mixture completes the benchmark study, accounting for the prediction of ignition delay in spark ignited (SI) gas engines. For both engine types, i.e. CI and SI, results include significant amounts of hydrogen admixed to natural gas. Though natural gas comprises minor fractions of  $C_4$ - $C_6$  alkanes, solely  $C_1$ - $C_3$  alkanes are considered to ensure a just comparison among the mechanisms.

### *Numerical Methods*

The reference data accessed in this work originate from both shock tube (ST) and rapid compression machine (RCM) experiments. Facility-specific nonidealities such as heat losses and premature radical formation in rapid compression machines as well as pressure gradients within the mixture in shock tubes induced by boundary layer effects may cause discrepancy between measurement and simulation data. Such nonidealities predominantly result in a decrease/increase in the measured pressure trace before the actual ignition event. A way to consider the change in pressure before ignition within the simulation framework is to apply a change of the specific volume according to the measured pressure trace.

Owing to its mechanical limitations, RCM experiments encompass larger time-scales and thus, more pronounced facility effects compared to shock tubes. Therefore, the use of a time-dependant change of the reactor volume is vital for the calculation of RCM experiments. Since there is inconsistent access to the detailed pressure history along the ignition delay data of RCM measurements referenced herein, the RCM experiments are not directly simulated in this work. Instead, the RCM measurements are compared with simulations using the ideal shock tube assumptions and the RCM end of compression conditions as the initial state. This simplification perturbs the comparison of experiment and simulation, though results calculated still should presume comparable ignition tendencies.

Regarding the simulation of shock tube data, Chaos et al. [17] point out that the consideration of nonideal effects becomes crucial in experiments encompassing mild ignition predominantly occurring below  $\approx 1000$  K and for

larger hydrocarbons. In such mild ignitions spatial inhomogeneities initiate local deflagration which result in an additional compression prior to the main ignition event shortening ignition delay time. This low temperature behaviour falls below the conditions investigated within the shock tube experiments cited herein. Hence, simulations performed in this work neglect pressure effects prior to ignition.

All calculations shown model shock tube experiments and were performed using the open source code Cantera [18]. Following the convention of frequent modeling work, simulations use the pressure and temperature behind the reflected shock wave as initial conditions within a homogeneous, adiabatic constant volume reactor. Maximum pressure rise rate defines ignition delay time. Though this criteria differs from [19] and [20] when CH and OH emission was used to define onset of ignition, Healy et al. [21] reported ignition delays of the CH and the pressure signal in their work to vary within 1 %.

### III. Results

#### A. Lean Mixture

Maximum pressure rise rate and abnormal combustion, such as knocking, limit the load range of compression ignition gas engines to a minimum  $\lambda$  often beyond the lean limit in SI application. The pressure at the timing of ignition commonly ranges between 50 and 100 bar and depends on the ignition trigger, e.g. high intake temperature and high compression ratio in homogeneous charge compression ignition (HCCI) engines or pilot injection in dual fuel engines. Table II depicts literature data used in this work for the validation of reaction mechanisms within the operating range of compression ignition gas engines, including a large variety of mixture quality as well as a large pressure range from 30 to 160 bar.

TABLE II  
Literature data referenced for reaction mechanism validation under lean conditions

ST - Shock Tube, RCM - Rapid Compression Machine

No	Fuel	Pressure [bar]	$\lambda$	Facility	Ref
1	C <sub>1</sub> /C <sub>3</sub>	30	2.0	ST	[19]
2	C <sub>1</sub>	44–160	2.5	ST	[22]
3	C <sub>1</sub> -C <sub>3</sub>	30	2.0	ST	[21]
4	C <sub>1</sub> /C <sub>3</sub>	80–120	1.9	RCM	[23]
5	C <sub>1</sub> /H <sub>2</sub>	20–80	2.0	RCM	[24]
6	C <sub>1</sub> -C <sub>3</sub> /H <sub>2</sub>	30	2.0-3.3	ST	[20]

The experimenters listed in Table II examined multiple gas compositions comprising the main natural gas species methane, ethane and propane. In addition, investigations of [24] and [20] include significant amounts of hydrogen. Table III details the fuel compositions used in the associated work. It has to be noted, that mixture 6.2 and 6.4 originally included small percentages of butane and pentane (total  $\leq 7.5$  %-mol within the fuel). As GRI and CRECK mechanisms studied herein solely capture alkanes up to  $C_3$  the fractions of  $C_1$ - $C_3$  were increased to compensate the molar fraction of the higher C species while keeping the relative ratio of  $C_1/C_2$  and  $C_1/C_3$  constant within the fuel. Mixture and diluent for each simulation were set up according to the information given in the corresponding paper.

TABLE III  
Detailed fuel composition data for lean mixture validation

Species fractions in %-mol, numbers mark the work in Table II the fuel refers to

No	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>3</sub> H <sub>8</sub>	H <sub>2</sub>	$\lambda$
1	90.0	0.0	10.0	0.0	2.0
2	100.0	0.0	0.0	0.0	2.5
3	90.0	6.6	3.3	0.0	2.0
4.1	100.0	0.0	0.0	0.0	1.9
4.2	95.0	0.0	5.0	0.0	1.9
4.3	70.0	0.0	30.0	0.0	1.9
5.1	100.0	0.0	0.0	0.0	2.0
5.2	70.0	0.0	0.0	30.0	2.0
6.1	40.0	0.0	0.0	60.0	2.0
6.2	59.0	7.3	3.6	30.0	2.0
6.3	16.9	2.1	1.0	80.0	3.3
6.4	26.9	8.6	4.3	60.0	3.3

Fig. 1 shows simulation results for methane/propane and methane/ethane/propane mixtures at 30 bar in comparison to experimental data of Petersen et al. [19] and Healy et al. [21]. San Diego and GRI slightly overpredict mixture reactivity and thus underestimate ignition delay. Aramco, CRECK and USC show excellent agreement with respect to magnitude and slope of measurement data for both mixtures.

The trends shown persist in case of pure methane at high pressures and  $\lambda = 2.5$  (s. Fig. 2). Aramco, CRECK and USC match the measurements very well, though USC predicts inhibited reactivity at 160 bar and lower temperatures. GRI and San Diego capture the influence of higher pressures as well, yet they recurrently underestimate ignition delay times.

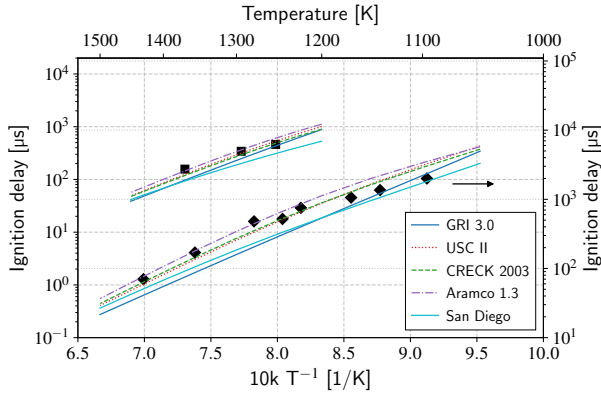


Fig. 1. Ignition delay times of mixture 1 and mixture 3 ( $\lambda = 2.0$ ) at 30 bar in air. Lines refer to simulation data, symbols refer to experimental data from [19] [21]. ■ mixture 1, ◆ mixture 3.

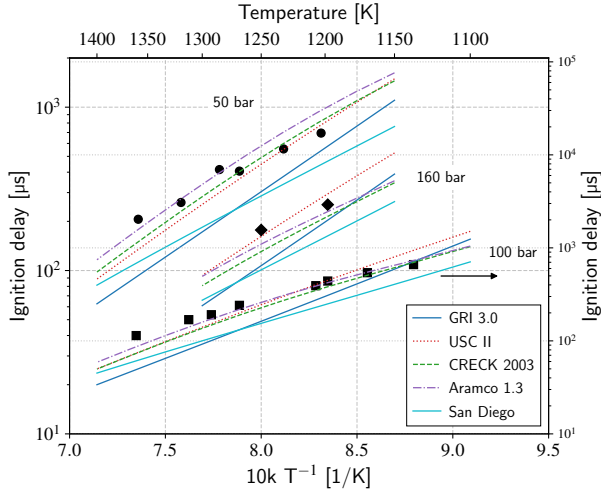


Fig. 2. Ignition delay times of 100 %  $\text{CH}_4$  (mixture 2,  $\lambda = 2.5$ ) at ● 50, ■ 100 and ◆ 160 bar in Ar. Lines refer to simulation data, symbols refer to experimental data from [22].

In contrast to the shock tube data presented so far, Pachler et al. [23] investigated a temperature range below 1000 K in a RCM. Measurements include pressures of 80, 100 and 120 bar and three binary methane/propane mixtures up to propane concentrations of 30 %. Due to the shock tube assumptions within the simulation framework (s. section II), an a priori deviation between the experimental and simulation data is expected. Fig. 3 compares the set of experimental data and the associated simulation results of this work. To illustrate the discrepancy resulting from the shock tube assumption, the original simulation data provided in [23] using Aramco 1.3 and taking facility effects into account is added (s. Fig. 3 Aramco 1.3 RCM). The notable difference in results of Aramco 1.3 and Aramco 1.3 RCM reveals the influence of model assumptions on simulation results. However, since Aramco 1.3 RCM predicts RCM measurements very well under all conditions, Aramco is assumed to reproduce dominant physical effects even at low temperature and high propane fractions. Thus, for the comparison of reaction mechanisms using the ideal shock tube assumptions Aramco is used as the reference.

In case of pure methane, results of CRECK and Aramco nearly overlap. Compared to Aramco, USC overestimates ignition delay, whereas San Diego calculates earlier ignition. Results of GRI significantly differ from Aramco in slope and magnitude, indicating a notable difference in activation energy at lower temperature. It is readily seen that predicted ignition trends among the mechanisms start drifting apart as a) temperature falls below 1000 K and b) propane content rises. Though this temperature regime is less relevant for engine application, it is still considered for completeness. Pachler et al. found propane low temperature chain branching pathways to start dominating ignition at higher propane fractions and lower temperatures. Albeit Aramco was not extensively validated for propane, it captures the influence of the existing low temperature chemistry very well. The poor performance of the remaining mechanisms still is reasonable, as the low temperature conditions mostly exceed their validation range. Pure methane exhibits no negative temperature coefficient behavior, hence the difference in ignition tendency among the mechanisms decreases for lower propane fraction.

In order to emphasize this assertion, a more comprehensive version of the CRECK mechanism including low temperature reactions (LT) and alkanes up to C5 is included for the 70/30 mixture for comparative reasons. Clearly, the low temperature ignition is in much higher agreement compared to the results of Aramco, though the C5-LT versions still predicts lower reactivity. An additional investigation not explicitly shown here contrasts the performance of CRECK and Aramco at higher temperature using the 70:30 blend. As the difference in calculated ignition delay approaches to levels shown in Fig. 1-2 for temperatures above 1050 K, results clearly point out the low temperature range to cause the discrepancy among the

mechanisms. Eventually, the discussion above highlights the importance of using reaction mechanisms only within their validation range and exemplarily illustrates the deviation to be expected for low temperature ignitions and at elevated fractions of propane. Further, results point out the exceptional capability of Aramco to accurately predict ignition under these conditions.

The next analysis focuses on the capability of predicting the ignition delay of mixtures including significant amounts of hydrogen. Fig. 4 depicts calculated ignition delay times of 100/0 and 70/30 methane/hydrogen at 50 bar and temperatures above 950 K. In addition, a sweep of compression pressure at constant compression temperature of 1010 K is presented in Fig. 5. Though no direct quantitative comparison is reasonable due to the modeling assumptions, experimental data of RCM experiments taken from Gersen et al. [24] at similar end of compression conditions are added to emphasize the general ignition tendencies. All mechanisms capture the change of ignition delay for varying pressure and temperature. CRECK and Aramco access the same baseline  $H_2/O_2$  and  $C_1$ - $C_2$  core mechanisms, hence they hardly reveal any difference in methane/hydrogen results. GRI, USC and San Diego significantly differ from CRECK and Aramco in both magnitude and slope. San Diego evinces a distinct sensitivity of reactivity to hydrogen fraction, as ignition delay rises for higher hydrogen contents compared to the remaining mechanisms. When temperature approaches 950 K, the discrepancies known from Fig. 3 become apparent. GRI and USC predict almost identical ignition delays for varying pressure at 1010 K. Compared to Aramco and CRECK, mechanisms GRI, USC and San Diego estimate lower mixture reactivity for the hydrogen blend at higher pressure.

The evaluation of multiple  $C_1$ - $C_3/H_2$  blends at 10 and 30 bar and  $\lambda = 2.0$ -3.3 in Fig. 6 completes the benchmark study of lean mixture (s. Table III, mixtures 6.1-6.4). For validation reasons the measurement data of shock tube experiments taken from Donohoe et al. [20] are considered as a reference. As mentioned earlier, the molar fractions of butane and pentane within the original mixture were compensated by adequately increased fractions of  $C_1$ - $C_3$  fuel species. Though this change in fuel composition slightly biases the comparison with the experimental data, the general ignition tendencies are considered to persist as the fractions of  $C_4$  and  $C_5$  are modest (total  $\leq 7.5$  %-mol within the fuel). In all cases USC calculates fastest ignition and overpredicts reactivity compared to the measurement results. At 10 bar, a discrepancy with respect to the predictions of GRI exists. For mixture 6.1 (60 % hydrogen) ignition delay is underpredicted whereas increasing hydrogen fraction to 80 % (mixture 6.3) results in the underprediction of reactivity. Donohoe et al. reported a discrepancy in this context as well, as they included chemistry of higher C species into the GRI mechanism and compared it to their data. They ascribe the inconsistent ignition trend to the tendency of the GRI mechanism to overpredict reactivity in case of hydrocarbons and to underestimate reactivity of hydrogen

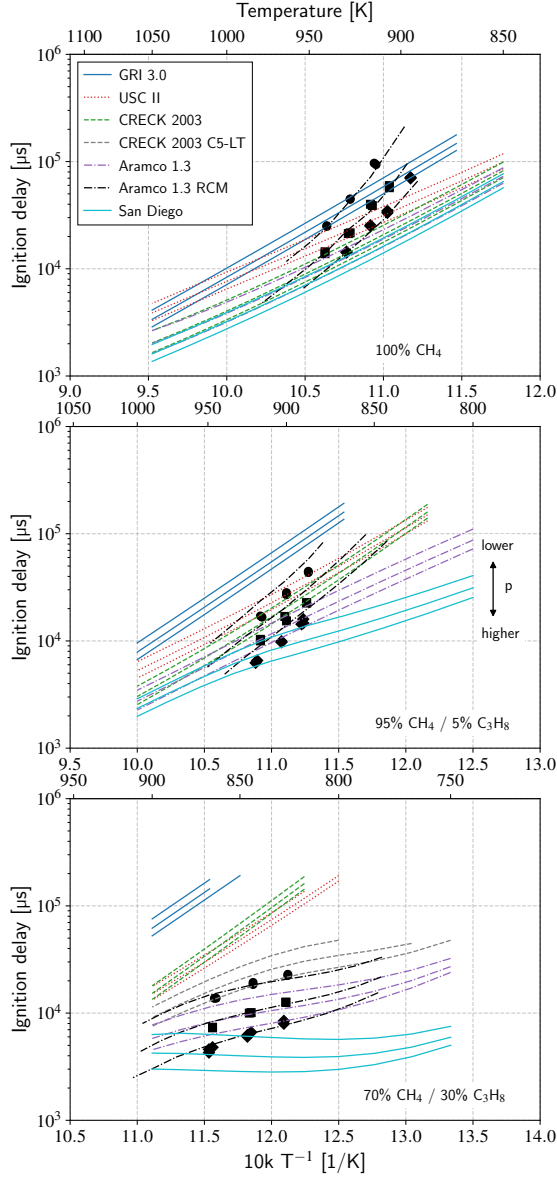


Fig. 3. Ignition delay times of mixtures 4.1-4.3 ( $\lambda = 1.9$ ) at  $\bullet$  80,  $\blacksquare$  100 and  $\blacklozenge$  120 bar. Lines refer to constant volume simulation data in air, symbols refer to rapid compression machine experimental data in  $\text{N}_2$  / Ar from [23].



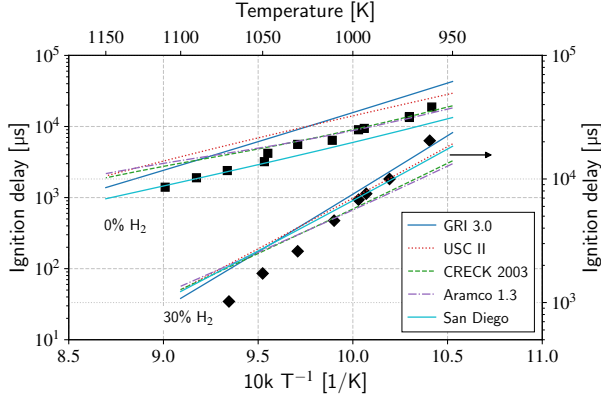


Fig. 4. Ignition delay times of mixtures 5.1-5.2 ( $\lambda = 2.0$ ) at 50 bar in  $N_2 / Ar$ . Lines refer to constant volume simulation data, symbols refer to rapid compression machine experimental data from [24]. ■ mixture 5.1, ◆ mixture 5.2.

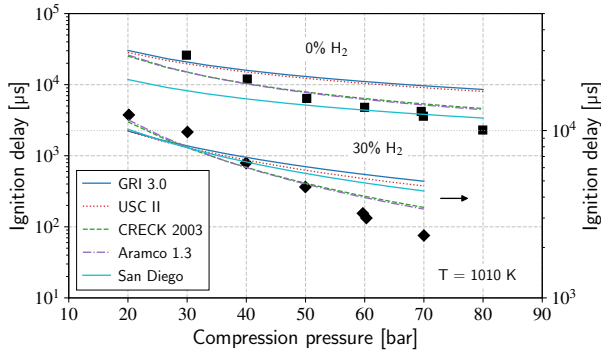


Fig. 5. Ignition delay times of mixtures 5.1-5.2 ( $\lambda = 2.0$ ) at 1010 K and pressures 20 - 80 bar in  $N_2 / Ar$ . Lines refer to constant volume simulation data, symbols refer to rapid compression machine experimental data from [24]. ■ mixture 5.1, ◆ mixture 5.2.

enriched mixtures. The results shown herein of pure hydrocarbon fuels (mixtures 1-5.2) support this assertion. Results of San Diego match the data of GRI in most cases. Aramco and CRECK perform very well as they accurately predict slope and magnitude of the experimental data for varying mixture composition,  $\lambda$  and pressure.

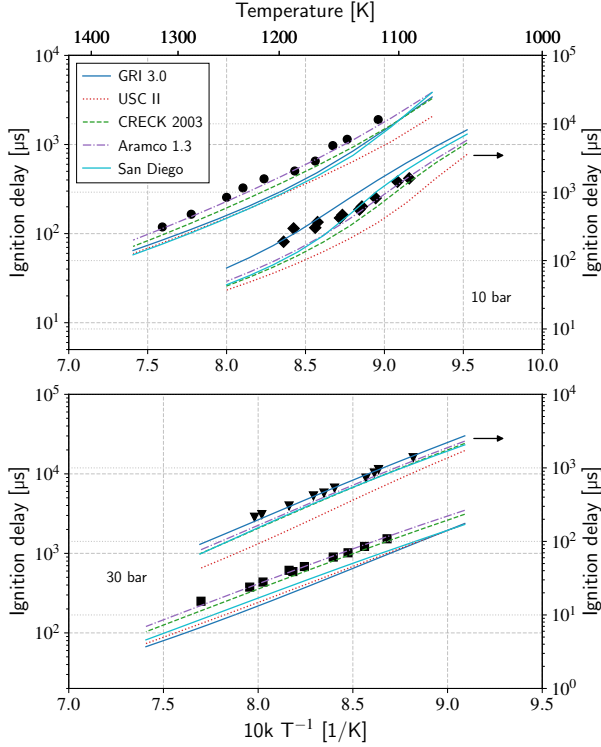


Fig. 6. Ignition delay times of mixtures 6.1-6.4 at 10 and 30 bar in Ar. Lines refer to simulation data, symbols refer to experimental data from [20].  $\lambda = 2.0$ : ● mixture 6.1, ■ mixture 6.2,  $\lambda = 3.3$ : ◆ mixture 6.3, ▼ mixture 6.4.

### B. Stoichiometric Mixture

This section provides a brief evaluation of the mechanisms performance under stoichiometric conditions. Table IV summarizes the literature data used as a reference. Since the data was generated as a part of the studies accessed for lean mixture validation, fuel compositions also cover a large range of  $C_1$ - $C_3$ / $H_2$  blends (s. Table V).

Fig. 7 presents ignition delay data at 17 and 24 bar for mixture 7 compared to simulation results. The slope is captured very well by all mechanisms besides GRI, which notably overpredicts ignition delay at temperatures below 1200 K. San Diego predicts ignition slightly faster compared to experimental results. Aramco, CRECK und USC tend to underestimate reactivity, yet results lie within a reasonable range of the measured ignition

TABLE IV  
Literature data referenced for reaction mechanism validation under stoichiometric conditions

ST - Shock Tube, RCM - Rapid Compression Machine

No	Fuel	Pressure [bar]	Facility	Ref
7	C <sub>1</sub> /C <sub>3</sub>	10-25	ST	[19]
8	C <sub>1</sub> -C <sub>3</sub>	30-55	ST	[21]
9	C <sub>1</sub> /H <sub>2</sub>	40	RCM	[24]
10	C <sub>1</sub> -C <sub>3</sub> /H <sub>2</sub>	10-30	ST	[20]

TABLE V  
Detailed fuel composition data for stoichiometric mixture validation

Species fractions in %-mol, numbers mark the work in Table IV the fuel refers to

No	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>3</sub> H <sub>8</sub>	H <sub>2</sub>	$\lambda$
7	90.0	0.0	10.0	0.0	1.0
8	90.0	6.6	3.3	0.0	
9.1	100.0	0.0	0.0	0.0	
9.2	70.0	0.0	0.0	30.0	
10.1	20.0	0.0	0.0	80.0	
10.2	47.3	15.1	7.6	30.0	

delay times.

In contrast to mixture 7, the tendency of GRI to overpredict ignition delay vanishes in case of mixture 8 (s. Fig. 8). GRI and San Diego overestimate reactivity within the pressure range between 29 and 53 bar. USC, Aramco and CRECK are in excellent agreement for all pressure and temperature conditions shown in Fig. 8.

Fig. 9 compares ignition delay times gathered by constant volume simulations and RCM measurements of pure methane and a 70/30 methane/hydrogen blend at temperatures below 1050 K and 40 bar. The RCM data illustrate the ignition tendencies. In case of pure methane the predictions of Aramco, CRECK and San Diego fairly match. Instead, GRI and USC calculate a significant decrease in reactivity. In case of the hydrogen enriched mixture the predictions of Aramco and CRECK still agree very well. GRI and USC recurrently calculate lowest reactivity and results of San Diego match GRI and USC to a much higher level compared to Aramco and CRECK. The notable change in simulation results indicates a distinct sensitivity of the

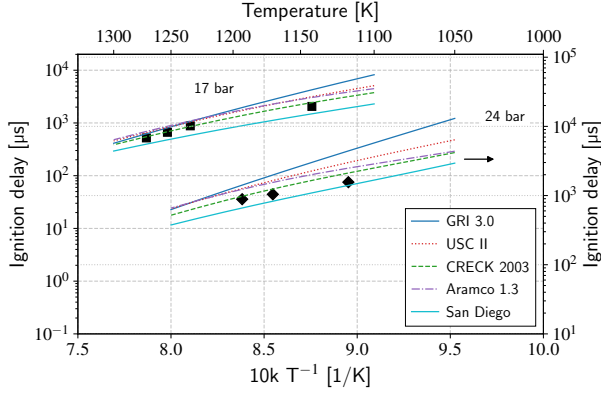


Fig. 7. Ignition delay times of mixture 7 ( $\lambda = 1.0$ ) at 17 and 24 bar in air. Lines refer to simulation data, symbols refer to experimental data from [19]. ■ 17 bar, ◆ 24 bar.

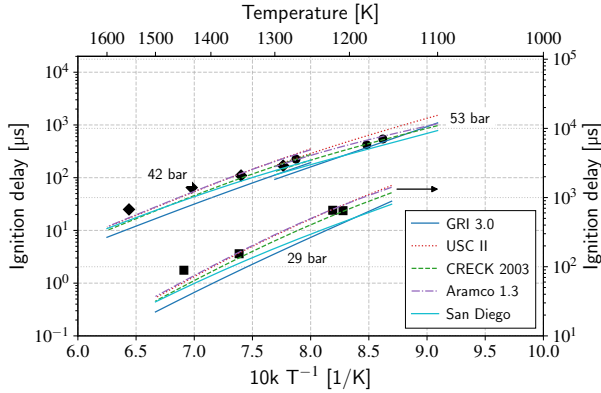


Fig. 8. Ignition delay times of mixture 8 ( $\lambda = 1.0$ ) at 29, 42 and 53 bar in air. Lines refer to simulation data, symbols refer to experimental data from [21]. ■ 29 bar, ◆ 42 bar, ● 53 bar.

ignition delay predictions with respect to hydrogen content in case of the San Diego mechanism under the particular conditions of Fig. 9.

The analysis of mixtures 10.1-10.2 at 10 and 30 bar in Fig. 10 extends the validation of mixture quality towards higher hydrogen contents. Similar to the simulations under lean conditions referring to the experiments of Donohoe et al. [20] the original fractions of  $C_4$ - $C_5$  were compensated by higher amounts of  $C_1$ - $C_3$  in this work. Results underline the discrepancy of

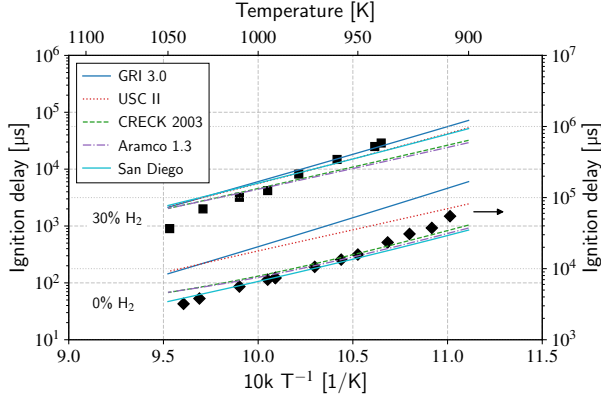


Fig. 9. Ignition delay times of mixture 9.1-9.2 ( $\lambda = 1.0$ ) at 40 bar in  $N_2 / Ar$ . Lines refer to constant volume simulation data, symbols refer to rapid compression machine experimental data from [24]. ■ mixture 9.2, ◆ mixture 9.1.

GRI to predict varying ignition tendencies with increasing hydrogen content already found for lean mixture. San Diego and USC exhibit a pronounced sensitivity to hydrogen fraction as well, since the agreement of predicted and measured ignition delay notably varies for different hydrogen contents.

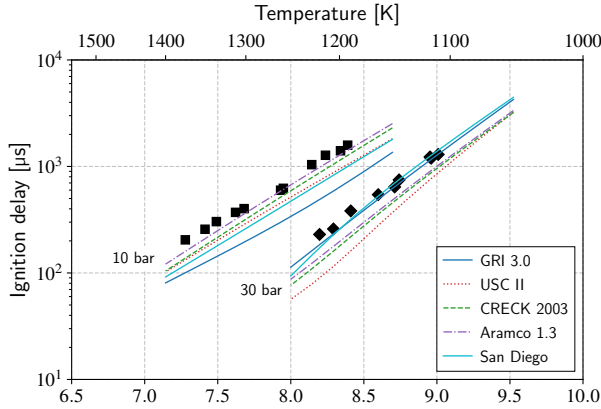


Fig. 10. Ignition delay times of mixture 10.1-10.2 ( $\lambda = 1.0$ ) at 10 and 30 bar in  $Ar$ . Lines refer to simulation data, symbols refer to experimental data from [20]. ■ mixture 10.2 at 10 bar, ◆ mixture 10.1 at 30 bar.

#### IV. Summary

In this work the performance of archival and recent chemical reaction mechanisms was evaluated to predict ignition delay under conditions relevant for internal combustion engines fueled with natural gas and hydrogen enriched natural gas mixtures. Based on a review of literature data, the benchmark study examines the mechanisms GRI 3.0, USC II, San Diego, CRECK 2003 and AramcoMech 1.3. The simulation results were compared against measured ignition delay times from shock tube and RCM experiments. Relative air-fuel ratios of  $\lambda = 1.9$ -3.3, ignition pressures of 30-160 bar and a wide range of  $C_1$ - $C_3/H_2$  blends address the operating range of compression ignition engines. Further, the analysis of multiple stoichiometric  $C_1$ - $C_3/H_2$  blends at pressures up to 55 bar encompass a segment of operating conditions relevant for spark ignition engines. The ratios of methane/ethane/propane within the mixtures examined strive to represent natural gas surrogates, exhibiting methane as the main fuel component. With respect to the hydrocarbon/hydrogen blends, maximum hydrogen enrichment reaches 80 %-mol. The general trends of high temperature ignition beyond  $\approx 1100$  are predicted within a reasonable range by all mechanisms. Results of a more detailed analysis can be summarized as follows:

- Since validation of mechanisms GRI, USC, San Diego and CRECK focuses on high temperature oxidation, significant deviation among the mechanisms ignition delay predictions arise for temperatures falling below  $\approx 1000$ -1100 K.
- The evaluation of low temperature RCM data emphasizes Aramco to capture the dominating effect on ignition delay of low temperature chain branching pathways induced by elevated propane fractions of 30 %-mol reported in [23].
- Results of USC are mostly on par with predictions of CRECK and Aramco for pure hydrocarbon fuels.
- USC tends to overpredict reactivity when notable amounts of hydrogen beyond 30 %-mol are admixed.
- GRI and San Diego exhibit contrary trends with respect to hydrocarbon/hydrogen blends: Both mechanisms tend to underestimate ignition delay of hydrocarbon blends, while underestimated reactivity of pure hydrogen is likely to enhance the agreement with measurement data of hydrocarbon/hydrogen blends.
- Keeping the low temperature limitations of CRECK in mind, the overall agreement of Aramco and CRECK with measurement results under the conditions studied herein is found excellent.

#### Acknowledgment

This work was supported by the Baden-Wuerttemberg Ministry of Science, Research and Arts. The authors also acknowledge the support from the German Federal Ministry of Education and Research.

## References

- [1] K. Albrecht and D. Pinchbeck, "Admissible hydrogen concentrations in natural gas systems," 2013.
- [2] Naturalhy Project Consortium, "Using the existing natural gas system for hydrogen: preparing for the hydrogen economy by using the existing natural gas system as a catalyst," Project contract no.: Ses6/ct/2004/502661.
- [3] O. Florisson, "Naturalhy: assessing the potential of the existing natural gas network for hydrogen delivery," 04.06.2010.
- [4] C. K. Westbrook and F. L. Dryer, "Chemical kinetic modeling of hydrocarbon combustion," *Prog. Energy Combust. Sci.*, vol. 10, pp. 1–57, 1984.
- [5] G. P. Smith, D. M. Golden, M. Frenklach et al., "Gri-mech 3.0," 1999.
- [6] Mechanical and Aerospace Engineering, "Chemical-kinetic mechanisms for combustion applications: San Diego mechanism," 2016.
- [7] H. Wang, X. You, A. V. Joshi et al., "High-temperature combustion reaction model of H<sub>2</sub>/CO/C<sub>1</sub>-C<sub>4</sub> compounds: USC Mech version II," 2007.
- [8] W. K. Metcalfe, S. M. Burke, S. S. Ahmed et al., "A hierarchical and comparative kinetic modeling study of C<sub>1</sub>-C<sub>2</sub> hydrocarbon and oxygenated fuels," *International Journal of Chemical Kinetics*, vol. 45(10), pp. 638–675, 2013.
- [9] S. M. Burke, U. Burke, R. Mc Donagh et al., "An experimental and modeling study of propene oxidation. Part 2: ignition delay time and flame speed measurements," *Combustion and Flame*, vol. 162(2), pp. 296–314, 2015.
- [10] C.-W. Zhou, Y. LI, U. Burke et al., "An experimental and chemical kinetic modeling study of 1,3-butadiene combustion: ignition delay time and laminar flame speed measurements," *Combustion and Flame*, vol. 197, pp. 423–438, 2018.
- [11] G. Bagheri, E. Ranzi, M. Pelucchi et al., "Comprehensive kinetic study of combustion technologies for low environmental impact: mild and oxy-fuel combustion of methane," *Combustion and Flame*, vol. 212, pp. 142–155, 2020.
- [12] E. Ranzi, A. Frassoldati, R. Grana et al., "Hierarchical and comparative kinetic modeling of laminar flame speeds of hydrocarbon and oxygenated fuels," *Progress in Energy and Combustion Science*, vol. 38(4), pp. 468–501, 2012.
- [13] E. Ranzi, A. Frassoldati, A. Stagni et al., "Reduced kinetic schemes of complex reaction systems: fossil and biomass-derived transportation fuels," *International Journal of Chemical Kinetics*, vol. 46(9), pp. 512–542, 2014.
- [14] M. Mehl, W. J. Pitz, C. K. Westbrook et al., "Kinetic modeling of gasoline surrogate components and mixtures under engine conditions," *Proceedings of the Combustion Institute*, vol. 33(1), pp. 193–200, 2011.
- [15] K. Narayanaswamy, G. Blanquart, and H. Pitsch, "A consistent chemical mechanism for oxidation of substituted aromatic species," *Combustion and Flame*, vol. 157(10), pp. 1879–1898, 2010.
- [16] S. Schuh, A. K. Ramalingam, H. Minwegen et al., "Experimental investigation and benchmark study of oxidation of methane–propane–n-heptane mixtures at pressures up to 100 bar," *Energies*, vol. 12, doi:10.3390/en12183410, 2019.
- [17] M. Chaos and F. L. Dryer, "Chemical-kinetic modeling of ignition delay: considerations in interpreting shock tube data," *International Journal of Chemical Kinetics*, vol. 42(3), pp. 143–150, 2010.
- [18] D. G. Goodwin, R. L. Speth, H. K. Moffat et al., "Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes: version 2.4.0," doi:10.5281/zenodo.1174508, 2018.
- [19] E. L. Petersen, D. M. Kalitan, S. Simmons et al., "Methane/propane oxidation at high pressures: experimental and detailed chemical kinetic modeling," *Proceedings of the Combustion Institute*, vol. 31(1), pp. 447–454, 2007.

- [20] N. Donohoe, A. Heufer, W. K. Metcalfe et al., "Ignition delay times, laminar flame speeds, and mechanism validation for natural gas/hydrogen blends at elevated pressures," *Combustion and Flame*, vol. 161(6), pp. 1432–1443, 2014.
- [21] D. Healy, H. J. Curran, J. M. Simmie et al., "Methane/ethane/propane mixture oxidation at high pressures and at high, intermediate and low temperatures," *Combustion and Flame*, vol. 155(3), pp. 441–448, 2008.
- [22] E. L. Petersen, D. F. Davidson, and R. K. Hanson, "Ignition delay times of ram accelerator CH/O/diluent mixtures," *Journal of Propulsion and Power*, vol. 15(1), pp. 82–91, 1999.
- [23] R. F. Pachler, A. K. Ramalingam, K. A. Heufer et al., "Reduction and validation of a chemical kinetic mechanism including necessity analysis and investigation of CH<sub>4</sub>/C<sub>3</sub>H<sub>8</sub> oxidation at pressures up to 120 bar using a rapid compression machine," *Fuel*, vol. 172, pp. 139–145, 2016.
- [24] S. Gersen, H. Darmeveil, and H. Levinsky, "The effects of CO addition on the autoignition of H<sub>2</sub>, CH<sub>4</sub> and CH<sub>4</sub>/H<sub>2</sub> fuels at high pressure in an RCM," *Combustion and Flame*, vol. 159(12), pp. 3472–3475, 2012.



# Real Route Generation for Simulation Based Development

1<sup>st</sup> Tuyen Nguyen

*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
tuyen.nguyen@h-ka.de

2<sup>nd</sup> Yannick Rauch

*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
yannick.rauch@h-ka.de

## Abstract

As a gap between the simulation of defined driving cycles and the execution of test drives, virtual test drives on real infrastructure can simulate real traffic situations. However, this requires the planning of the necessary test routes and the generation of all relevant information for the test drive. This functionality is provided by a corresponding tool, which generates the route and other relevant data based on route planning. Various services are used via *APIs* to provide the necessary information. The generated route data can then be fed to a simulation environment via corresponding interfaces. An initial implementation of the route data generation is carried out with a *MATLAB* tool, which makes it possible to plan and generate test routes. The test route data is then made available to the *CarMaker* simulation tool via a corresponding interface. This makes it possible to quickly and easily map virtual test drives on real routes in *CarMaker*.

## Index Terms

simulation, route data generation, driving cycle, energy demand

## I. Introduction

A simulation-driven development of functions or systems with an influence on the dynamics and energy requirements of a vehicle requires corresponding driving cycles as a basis for the description of the vehicle dynamics [1]. Such a driving cycle consists of a speed and altitude profile and thus represents certain scenarios relevant for the respective application. Synthetic driving cycles such as *WLTP* (*Worldwide Harmonised Light Vehicles Test Procedure*), *NEDC* (*New European Driving Cycle*), etc. are used, for example, which are intended to represent the typical operation of a vehicle for homologation purposes. However, the real use of vehicles cannot always be comprehensively represented in this way, so that substitute driving cycles are increasingly applied. These are based on real test drives from which a corresponding speed profile is obtained. Thus, the replacement driving cycles represent an average of the test drives and cannot conceptually represent all possible situations. Therefore test drives

are necessary in addition to the simulation of driving cycles, as is the case with Real Driving Emissions.

Virtual test drives, which represent the drive on real infrastructure, can act as a gap between the simulation of drive cycles and the execution of real test drives. This requires a simulation environment that can represent the journey based on data using route guidance, traffic signs, etc. If necessary, the influence of traffic or other random influences on the vehicle dynamics can be represented.

For this type of representation, it is necessary to transfer data from real infrastructure into the corresponding simulation environment. This can be done by manually reproducing the route to be driven, which, however, requires huge effort for longer routes. For certain simulation tools or simulation environments, there are corresponding add-ons that can obtain the data for the required route from map services. An example of this is the *Here Map Interface* for the tool *CarMaker* from *IPG Automotive GmbH*, which enables the transfer of road networks [2].

However, only the specific simulation tool and the specified services can be used. Accordingly, a tool is needed which provides variable data formats and interfaces to import them into different simulation tools or environments. The generation of data should be as variable and expandable as possible so that different information services can be used.

## II. Basics on the influence of energy consumption

The 3F-method (in German: Fahrer, Fahrzeug und Fahrumgebung), which includes the influencing variables driver, vehicle and driving environment, offers a well-known method for subdividing the range influence groups. This is used to determine representative load spectra for electric machines, transmissions, parallel hybrid powertrains and the chassis [4]. The approach presented here uses a leaning subdivision. According to [5], the driving environment is divided into route and environment. This subdivision into 4 categories (1.Driver, 2.Vehicle, 3.Route, 4.Environment) has the background that the environmental influences have far-reaching effects on the energy consumption of an electric vehicle compared to an internal combustion vehicle. The driving resistances to be overcome, efficiency losses, recuperation energy, the power of the electrical consumers and their useful life together form the total energy consumption of the vehicle [3].

Driver: As an interface between the other influencing groups, the driver group forms a special role among the energy consumption influences. This group reacts, for example, to the existing environment and route influences (acceleration, steering angle and speed of the vehicle). Driving behavior, which is influenced by time of day, physical condition and emotional state, also determines vehicle speed and acceleration phases [6], [7]. Also, the driver is responsible for managing the controllable electrical components in the vehicle (electrical auxiliaries, e.g. climate system) [3].

Vehicle: The values of this influence group are already fixed at the beginning of the journey (a change of the properties during the journey, such as the warming of the tires, are neglected in this work due to the relatively low influence). This includes the combination of different vehicle components such as body, powertrain and chassis. For example, the drag coefficient and the frontal area of the vehicle result from the given geometry of the body. The engine used has an engine characteristic which has different efficiencies in different load ranges [3].

Route: As a framework condition for the motion profile and the drive energy required with it, the selected route plays a decisive role. The characteristics of the route are divided into permanent and temporary factors. Permanent factors include, for example, speed limits, route type (highway, interurban or urban) and elevation profile and curve. On the other hand, there are the temporary route characteristics, such as road works, detours, and route utilization (no traffic, gridlocked traffic, or full closure) [3].

Environment: Any external influence that occurs during the journey and cannot be assigned to either the traffic volume or the route characteristics is grouped in the environment group. Examples are temperature, humidity, air pressure and air density, wind speed and direction, sun intensity, precipitation as well as day and season [3].

In the following work, the focus is on the generation of route data and the associated environmental influences, as this has a significant impact on the driver influence group. Thus, except for the vehicle influence group, the majority of the identified energy consumption influences can be considered.

### III. Concept

#### A. Structure

For the generation of route data a Route Generating Algorithm [8] is used and extended by functions to develop an application (called *RouteGenerator* in this work). Figure 1 shows the schematic structure of this application.

The two core elements of the application are the Route Generating Algorithm and the individual Application Programming Interfaces (*API*) (see chapter III-C ). The *APIs* serves as information sources and returns the input data for the Route Generating Algorithm [8]. The *APIs* are divided into three categories: Data for the visual display of the route (Map Display), data for the compilation of the route (Route Data) and additional data like slope, weather or traffic data of the route (Additional Data). The Route Generating Algorithm collects this data and in combination with the user's input data (start and destination of the route, display type or weather/traffic influence) generates the computed route. In addition, the route is displayed graphically in the *RouteGenerator* application. The generated data can then be transferred to external programs via an implemented interface. In this concept the simulation program *CarMaker* of the company *IPG Automotive*

*GmbH* is used for the generation of the energy consumption data. Therefore the *RouteGenerator* contains the function to convert the created route data into a *CarMaker* compatible format and to transfer it directly to *CarMaker* by using the embedded interface. The advantage, which results from the modular structure of the *RouteGenerator*, is that any formats and interfaces to external applications can be implemented.

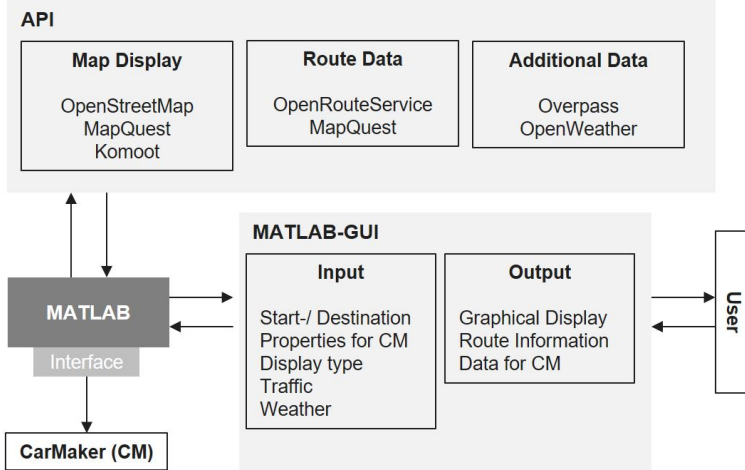


Fig. 1. Structure of the application *RouteGenerator* with two core elements, a *MATLAB-GUI* for the users input and visualisation and the *API* to gain the needed data from web services.

### B. Route Algorithm

As a basis for the generation of route data, a *Route Generating Algorithm* [8] is used. This takes as input parameters the start and destination coordinates, possibly individual waypoint between start and destination and date of the simulation. Using these input parameters, the route geometry, speed limits, traffic and weather data are generated and combined into a route.

To obtain the data for the route generating algorithm, various *APIs* are utilized. The *APIs* serve as information sources and return the input data for the Route Generating Algorithm [8]. The *APIs* are divided into three categories: Data for the visual display of the route (map display), data for the compilation of the route (route data) and additional data like slope, weather or traffic data of the route (additional data). The Route Generating Algorithm collects this data and in combination with the user's input data (start and destination of the route, display type and weather/traffic influence) generates the computed route.

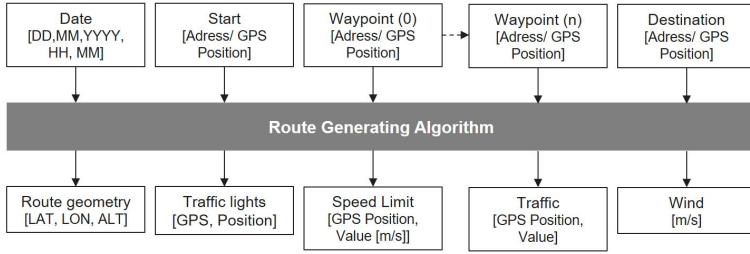


Fig. 2. Blackbox schematic with input and output data for the route generating algorithm [8]

### C. Application Programming Interfaces

APIs are interfaces of a program or a website that provide other programs with a connection to the system. In this paper, APIs of online services are used. The API call is made via a web URL standardized for the respective API [16]. In this concept, the map related online services *OpenStreetMap* [9], *MapQuest* [10] and *Komoot* [11] are used for the visual output of the route. All three map services are performant, free to use and in the case of *OpenStreetMap* it is an open source project with constant updates [9]. There is the possibility, to add new services to the *RouteGenerator* at any time by using APIs. The advantage of *OpenStreetMap* and *Komoot* is that they provide a detailed data base for creating the route. *MapQuest* was added for longer routes as it is more performant than the other two services. This is because the downloaded map sections from *MapQuest* can be up to 3840x3840 pixels in size. This shortens the route planning procedure in comparison to the other two providers presented, as with them the typical image size is 256x256 pixels. For instance, for a map section in the *RouteGenerator* of the size of 600x600 several map images have to be downloaded and compiled, which accordingly takes longer than downloading only one map element. In addition, *MapQuest* offers the possibility to download a hybrid map data (including satellite pictures) and then display them in the application. For the route data generation *OpenRouteService* (which gets its input data directly from *OpenStreetMap*) and *MapQuest* due to the available degree of detail of the data. In this case, the route data and height data are queried using the entered start, stopover and destination points. Since the map service has only stored a maximum of 500 coordinate pairs per request for the retrieval of height data, an additional algorithm was implemented for routes with more than 500 coordinate pairs. For a more realistic representation of the route, this concept uses the *OverPass API* [13] service to retrieve additional data such as local speed limits, traffic lights, intersections, and any traffic signs within a specified radius along the

route. Additionally, real weather data is retrieved using the *OpenWeather* [14] service, as this can also have an impact on energy consumption of the vehicle [15]. Since the *OpenWeather API* does not provide information about the air density, this must be determined from the given data for each request point.

#### IV. Conclusion

With the presented concepts and *MATLAB*-tool for route generation, it gets possible to simulate virtual test drives based on real infrastructure. The route to be completed can be planned, while the generation of the required information and data is automated. The generated data set can then be transformed into the appropriate data formats required by the simulation tools/environments. Not all information on the track and environment necessarily has to be generated or transferred to the simulation environment. This enables a flexible use of *APIs* and the simulation representation. This concept is currently implemented by a *MATLAB* tool for data generation including an interface to the simulation tool *CarMaker*.

The independent implementation of the route data generation makes it possible to achieve a connection to other simulation tools by means of corresponding interfaces. In addition, the structure of the data generation is designed in such a way that it can be expanded to include the collection of additional or redundant data.

#### References

- [1] R. Fechert, *Fahrzyklen*, TU Dresden, Professur für Fahrzeugmechatronik, 2018, Available online: <https://tu-dresden.de/bu/verkehr/iad/fm/forschung/forschungsfelder/ebf/fahrzyklen>, 07.01.2022
- [2] IPG Automotive GmbH, *CarMaker Add-Ons*, Available online <https://ipg-automotive.com/de/produkte-loesungen/software/add-ons/>
- [3] K. Kruppok, R. Kriesten, "Analyse der Energieeinsparpotenziale zur bedarfsgerechten Reichweitenerhöhung von Elektrofahrzeugen", Doctor of Philosophy. Karlsruhe Institut für Technologie, Karlsruhe. Elektrotechnik und Informationstechnik, Germany, 2020.
- [4] F. Küçükay, T. Kassel "Anforderungsoptimierung für Getriebe und Komponenten", *ATZ - Automobiltechnische Zeitschrift* volume 109, pages 812–819, 2007
- [5] K. Kruppok, R. Kriesten, C. Sackmann, P. Sautter, "Elektrisches Energiemanagement als Assistenzfunktion für zuverlässige Reichweitemaussagen - Simulation zur Abschätzung des Energieeinsparpotenzials", 18. Kongress SIMVEC Simulation und Erprobung in der Fahrzeugentwicklung, Band 18, 2016
- [6] S. Al-Sultan, A. H. Al-Bayatti, H. Zedan, "Context-Aware Driver Behavior Detection System in Intelligent Transportation Systems", *IEEE Transactions on Vehicular Technology*, Volume: 62, Issue: 9, pages 4264 - 4275, 2013
- [7] G. Castignani, T. Derrmann, R. Frank, T. Engel, "Driver Behavior Profiling Using Smartphones: A Low-Cost Platform for Driver Monitoring", *IEEE Intelligent Transportation Systems Magazine*, Volume: 7, Issue: 1, pages 91 - 102, 2015
- [8] C. Gutenkunst, D. Chrenko, R. Kriesten, P. Neugebauer, B. Jaeger, "Route Generating Algorithm Based on OpenSource Data to Predict the Energy Consumption of Different Vehicles", *IEEE IEEE Vehicle Power and Propulsion Conference (VPPC)*, 2015

- [9] OpenStreetMap. Available online: [https://wiki.openstreetmap.org/wiki/Main\\_Page](https://wiki.openstreetmap.org/wiki/Main_Page), 01.08.2021
- [10] MapQuest. Available online: <https://developer.mapquest.com/documentation/>, 01.08.2021
- [11] Komoot. Available online: <https://static.komoot.de/doc/external-api/v007/index.html>, 01.08.2021
- [12] MapQuest. Available online: <https://developer.mapquest.com/documentation/>, 01.08.2021
- [13] Overpass. Available online: <https://osmlab.github.io/learnoverpass/en/docs/>, 01.08.2021
- [14] OpenWeather. Available online: <https://openweathermap.org/api>, 01.08.2021
- [15] A. Donkers, D. Yang, M. Viktorovic, "Influence of driving style, infrastructure, weather and traffic on electric vehicle performance", Elsevier Transportation Research Part D: Transport and Environment, page 88, 2020
- [16] H. Schulte, P. Litzinger, M. Gerhardt, R. Kriesten, T. Nguyen, Y. Rauch, "Entwicklung eines Funktionsdemonstrators für die Simulation von Energieströmen von E-Fahrzeugen", R+D project, University of Applied Sciences Karlsruhe - IEEM, Karlsruhe, Germany, 2021

# Methods of developing Model Predictive Control for Air Conditioning systems (A Literature review)

1<sup>st</sup> Arsema Derby Chekol

*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
Arsema.Derbie2@h-ka.de

2<sup>nd</sup> Maurice Kettner

*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
Maurice.Kettner@h-ka.de

## Abstract

Physical systems in controlled environments ought to involve limited and determined variables. Notwithstanding this fact, real-world instances are found to be influenced by tremendous parameters. An aggressively escalated interest in control attempts to capture the dynamics of such systems and manipulate them to achieve optimal working constraints. The focus of this review is on model predictive control (MPC) which is by far the most efficient technique, particularly for process control. Its trending implementation for thermodynamic systems especially air-conditioning applications using neural networks is discussed. As various thermal conditioning researches have been carried on heating, ventilation, and air conditioning (HVAC) of buildings, most of the research papers are dominated by such a special application. The review states MPC has many advantages over conventional controllers. It presents the two important steps in the development of MPC namely system Identification and Controller design. Under these topics, neural network models and methods of optimizing such models are thoroughly explained.

## Index Terms

Model predictive control (MPC), air conditioning, System identification, Optimization

## I. Introduction

From the very first time of its introduction as an efficient Control for dynamic processes in the early 1970s, MPC (Model Predictive Control) also known as RHC (Receding/Moving Horizon Control) has proved to be an advanced form of a control strategy for process control which has underlying principles for satisfying the constraints [1]. Due to its simple structure and dynamic performance, it is implemented in various application areas including power transmission, automatic vehicles, chemical processes, thermal processes



like air conditioning, oil drilling, and more. MPC has gained popularity in the field of heating, ventilation and air conditioning (HVAC) since the classic control techniques like state control, bang-bang control, P (Proportional), PI (Proportional-Integral), and PID (Proportional- Integral-Derivative) controls disregard the knowledge of the process, possess constant parameters and solely rely on the measurements from sensors [2]. As shown in [3], [4] poor control performance was provided for noisy and nonlinear processes having large time delays. Their control performance is enhanced by cascading multiple PID controllers or linking feedback and feed-forward controllers but improper gain selection makes the entire controlled systems unstable. The demand for optimal, predictive, and adaptive techniques emanated from such challenges [5]. In contrast to these classical control techniques, MPC is a promising candidate for efficient control, higher energy savings, better stability, etc. [6].

MPC provides worthy prospects for the pursuance of energy-efficient control, system dynamics, and stiffness along with time delays. It can deal with constraints on input/output signals in multi-input and multi-output (MIMO) systems. This is mainly significant for thermal (conditioning) control because it allows the maximum potential of the commands in order to impose temperature bounds [5].

This review explores literature focused on definitions, working principles, important elements of MPC, especially from a neural networks' integration perspective. It is structured as follows. Basic terminologies and working principles of MPC are discussed in Section 2, while Section 3 describes system identification. Neural networks as system models are treated under Section 4 proceeded by controller design with the focus of optimization in Section 5. Finally, the conclusion is stated in section 6.

## II. Basic terminologies and working principles of MPC

Being majorly constituted of the classical control principles [7], MPC comes with an intuitive parameterization through adjusting a process model at the cost of a higher computational effort. It predicts future system behavior based on a system model considering it in the optimization that determines the optimal trajectory of manipulated variables[8]. A typical MPC consists mainly of a system model, an objective function, and a control law. The last two comprise the controller design. Figure 1 shows a simplified schematic diagram of MPC-based control loop.

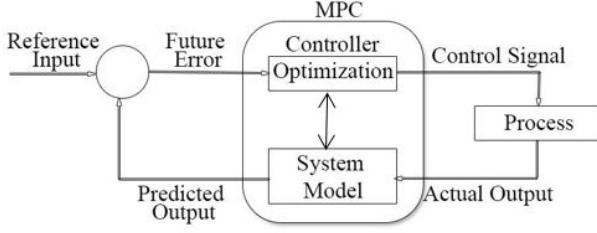


Fig. 1. Simplified schematic diagram of MPC-based control loop

For a discrete-time linear time-invariant system that evolves in time, the system can be explained by equation 1.

$$X_{t+1} = AX_t + BU_t \quad (1a)$$

$$y_t = CX_t + DU_t \quad (1b)$$

where  $x \in \mathbb{R}^n$  denotes the state vector and  $u \in \mathbb{R}^m$  denotes the manipulated or control vector, respectively; A, B, C, and D denote the system matrices. The MPC for an infinite time horizon entails solving the constrained optimization problem shown in equation 2:

$$\min_n \sum_{t=0}^{\infty} y_t^T Q x_t + u_k^T R u_k \quad (2a)$$

Subject to:

$$X_{t+1} = AX_t + BU_t \quad (2b)$$

$$y_t = CX_t + DU_t \quad (2c)$$

$$x_0 = \bar{x} \quad (2d)$$

$$y_{min} \leq y_t \leq y_{max} \quad (2e)$$

$$x_{min} \leq x_t \leq x_{max} \quad (2f)$$

where Q and R are the positive semi-definite weight matrices [9][10]. Equation 2a defines the cost function (J) which is optimized to its minima. The objective, in this case, is to achieve a desired  $y_{t+1}$  with minimal control actions. The derivation of optimization functions is based on the objectives

and behavior of the controlled system. These functions are generally defined and designed by researchers and energy designers. Studies have generally defined these functions using economic goals (e.g. power consumption and energy cost), thermal comfort goal (for the applications of HVAC), and control goal (tracking error and set-point) [11]. To explicitly define the working principle of MPC, the foundational elements of the control are shown in Figure 2. The basic terminologies in MPC are:

- Control time step ( $c$ )/ Sampling time: the time between control updates and iterative receding horizon optimizations.
- Prediction horizon ( $H$ ): the number of control time-steps the controller looks ahead in the future to optimize the cost function under constraints
- Control Horizon/execution horizon/Manipulated input horizon ( $C$ ): the number of possible different values the manipulated variables can take in the future

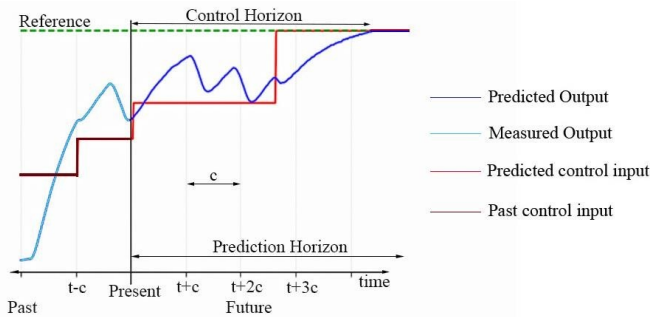


Fig. 2. Schematic diagram of MPC control

Prediction horizon is an important term that greatly attributes the performance of the controller. It must be designed long enough to capture the transient dynamics of the system. Control horizons mostly have smaller values than prediction horizons mostly limited between 1 and the prediction horizon. Control variables remain unaltered sampling time. Lowering this time unit improves the rejection of unknown disturbances but adversely increases the computational effort, therefore, should be optimized. Proper implementation of MPC requires well-defined numerical values of these factors since they directly affect computational efforts, the stability of controller, rejection capability of disturbances factors, and setting time of MPC controller [11].

MPC is characterized by the features of handling multi-input multi-output systems [12] [13] [14], the possibility of minimization of objective cost function within well-defined constraints [15], and easy to implement control law[16].

MPC has different variants depending on its purpose. The most common ones include robust model predictive control (RMPC), stochastic model predictive control (SMPC), distributed model predictive control (DMPC), adaptive model predictive control (AMPC), and hybrid model predictive control (HMP). Robustness and scalability are compromised due to uncertainties that arise from model development, external disturbances, and system complication. RMPC protects system states from violating the constraints when uncertainty is bounded [17][18]. As an alternative to the conservative performance of RMPC, SMPC exploits stochastic models of uncertainty and typically enforces constraints up to a certain probability [19][20]. For large-scale applications like distribution systems, a central network form where calculations of states are optimized online at once can be tedious and cumbersome. A control scheme namely distributed model predictive control (DMPC) has been put forward that can split the large-scale applications into smaller and simpler sets of sub-problems regulated locally [11]. Adaptive and Hybrid model predictive controls are used when the system models ought to be updated in accordance with uncertainties and when the models suffer from poor accuracy respectively.

### III. System Identification

MPC incorporates knowledge about the process through a system model and solves a dynamic optimization problem at each time step to yield an optimal control sequence, applying the first control action in the sequence to the system before proceeding to resolve the problem at the next time step [21]. System identification is therefore a foundational task where adequately representative and manipulable system models are identified. The requirements for an MPC dynamic system model are sufficient accuracy in describing system dynamics for a given set of design variables and computational tractability allowing real-time control and optimization [22].

The accuracy and effectiveness of an MPC are highly dependent on the identified model. Modeling and identification are the most difficult and time-consuming parts of automation processes. The basic conditions that each model intended for MPC usage should satisfy are reasonable simplicity, well estimated system dynamics and steady-state properties as well as satisfactory prediction properties [23].

System modeling techniques can be broadly classified into three: physical modeling (or white box, mathematical, forward), data-driven (or black-box/empirical/inverse), and gray box (or hybrid). System models can be dynamic or static (steady) depending on the variability of parameters with time. Depending on the linearity of the system, the models can be linear or nonlinear. Most physics-based models fall under inductive types of system models while data-based models are mostly deductive. Other classifications include explicit and implicit, discrete or continuous, deterministic or probabilistic/stochastic models [24].

#### *A. Physics-based models*

Physics-based models are a set of mathematical equations based on fundamental physical and chemical laws. Commonly used laws for deriving mathematical equations of thermodynamic processes include heat transfer law, mass transfer, momentum equation, heat balance method, etc. Incorporating specific assumptions and with no necessary demands of system prototype, such modeling techniques are proficient tools for initial design [11].

Physics-based dynamic HVAC system models are commonly developed for the slow-moving temperature and humidity processes (e.g., zone temperature dynamics, zone humidity dynamics, heating/cooling coil dynamics, etc.), while static models are implemented for the fast-moving dynamics of the system [e.g. mixed air temperature and carbon dioxide (CO<sub>2</sub>) concentration in mixing box, and flow rate of air and water through damper and valve respectively] and energy consumption (fan or pump energy consumption) [25].

Mathematical equations are interpretable and distinctive. The general model of a system is mostly a series of equations representing the elements and subsystems that constitute the entire process.

#### *B. Data-based modeling*

The concept of the black-box model is to fit a transfer function model to the input/output real model data to yield coefficient polynomials that can be factored to provide resonance frequencies and characterization of damping coefficients without knowledge of the internal working [26]. From a pool of mathematical relationships, the best fit for the input and output variables of the system are discovered. The black box modeling is very useful for systems where the system processes are very complex and they are time-consuming [27].

Black-box models have multiple advantages, including lower engineering cost because they follow a data-in-data-out approach; less domain knowledge because it is based on the mapping of input and output data, and greater adaptability because the model will evolve itself with new data. Black-box models also have some drawbacks, including high demand for data quality: missing, wrong, or biased data lead to low-quality models [28].

A complete classification of data-driven models is summarized by Abdul Afram et al [25] and shown in Figure 3.

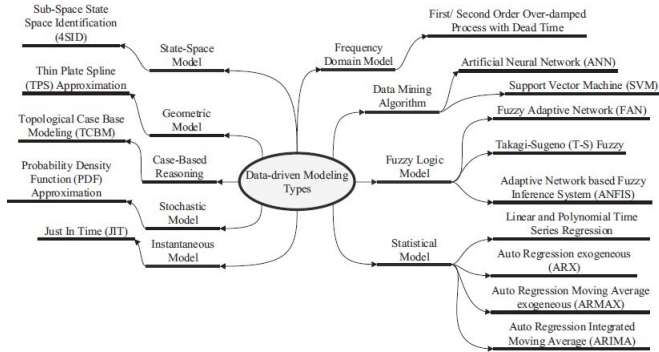


Fig. 3. Inverse modeling types (adapted from [29])

Out of all these inverse modeling methods, ANN is the most popular method due to its high accuracy to model nonlinear systems compared to other methods [29].

### C. Grey-box modeling

Sometimes called reduced-order models or simplified models, grey-box models are blends of both white-box and black-box models incorporating good generalization from the former and better accuracy of the latter. The basic structure of the model is formed from physical or thermodynamic principles while the model parameters are determined using the parameter estimation algorithms on measured data of systems (from a catalog, commissioning, or operating data). The grey-box model is more interpretable than the black-box model. It is more computationally efficient and simpler than the white-box model [28][11][24]. Thermal process dynamics are mostly captured by the electric analogy of the resistance-capacitance (R-C) equivalence networks, also called thermal networks, in grey-box modeling [30].

Thermodynamic processes are nonlinear and cumbersome which leads to an increase in computational effort. The detailed physics-based model requires many iteration processes that may result in problems instability of the system and increased computational time. Such features are highly undesirable for online computation especially for real-time applications [11]. The modeling process of most thermal systems leads to dynamic, nonlinear, and very high-order models because of the physical properties such as high-thermal-inertia, real lag time, uncertain disturbance factors, etc. of the system [26]. From possible retrofitting and inevitable adjustments during installation, acquiring precise and thorough information is seldom readily available. In addition, the complexity of a physical model as systems become large-scale poses limitations for the use of white-box models [31]. The need for a thorough understanding of physical systems for such white-box model developments also limits its usage.

For typical physical systems, black-box models provide better accuracy without a comprehensive knowledge of the operations. Nevertheless, as data beyond training data is not extrapolated, the prediction accuracy is no more maintained. Data acquisition plays a key role by including instances of all working conditions. Table 1 summarizes the points of comparison between white box and black box models.

MPC requires a reasonably accurate model of physical systems. Air conditioning systems are very complex systems to model. Thus, the traditional physics-based modeling approaches like the white-box and the grey-box techniques, although detailed, are cost and time-prohibitive [32].

#### IV. Artificial Neural Network-based Model Predictive Control

The trend of using Artificial Neural networks (ANNs) for the application of MPC in air conditioning applications has escalated in the past two decades. Inspired from the biological nervous system, ANN is a network of units with information processing capacity. The small units of both networks are called neurons which receive, transfer and send signals either in response to a stimulus or an action command.

ANNs are mostly used for universal function approximation in numerical paradigms because of their excellent properties of self-learning, adaptivity, fault tolerance, nonlinearity, and advancement in input to output mapping [33]. Perception provides neural networks with the capability to capture relationships between variables, recognize anomalies, predict outcomes, and so forth for a wide spectrum of applications.

Artificial intelligence algorithms have long been used for modeling decision-making systems as they provide automated knowledge extraction and high inference accuracy [34]. Appropriately trained ANN can approximate any nonlinear process to a high degree of accuracy [35]. ANNs for MPC can be trained with labeled or unlabeled data giving rise to both supervised and unsupervised learning [36].

TABLE I  
Comparison between black box and white box modeling

Comparison parameters	White box	Black box
Generalization Capability	High, based on natural laws	Low, work best for specific systems
Thorough understanding of physical systems	Highly required	Not necessarily every detail
Handling system complication	Low, iterations complicate with system complication	High, capable of capturing complicated relationships
Accuracy for specific systems	Low, downgrades with complication and unknowns	High, based on datasets from unique systems
Handling Missing/distorted data	Unknowns are assumed (scientific guess)	Models don't extrapolate out of the dataset range
Interpretability	High, easy to grasp relationships	Low, meaningful relationships are not explicitly explained

Another significant advantage of using neural network system models is their explicit nature. Once a neural net of suitable properties is constructed, it can be evaluated on moderate hardware, and hence no optimization solver is needed. Another main advantage is that the neural net is in no way limited by the length of the prediction horizon or by the size of the system, as it is in explicit model predictive control [10].

ANNs are adaptive machines that are made from the interconnection of artificial neurons which act as simple processors. This network, through a learning process, acquires knowledge from the environment, and such knowledge is stored in the synaptic weights of the network [37]. The continuous update of weights in the learning process is referred to as training. Neural networks are ensembles of layers comprising perceptrons that contain monotonic functions named activation functions. These functions define neuron outputs from the weighted input signals. Most common activation functions include sigmoid  $\frac{1}{1+e^{-ax}}$ , hyperbolic tangent  $\tanh x$  and rectified linear units ( $\max(0, x)$ ). The effectiveness of neural networks is highly dependent on the optimal selection of hyperparameters like the



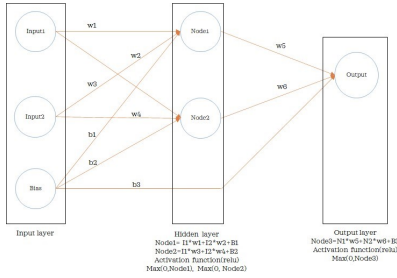


Fig. 4. MLP elements and working principle

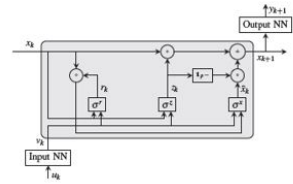


Fig. 5. Internal dynamics of RNN [39]

number of hidden layers, number of nodes, activation functions, and so forth. The efficiency of training is verified by standard performance index indicators like Mean Standard Error (MSE) or Mean Average Error (MAE).

Neural network architectures define how two or more neurons are interconnected to form the networks. Static or feedforward architectures can approximate any continuous function by performing static mapping. They can have single-layer or multilayer architectures. In the dynamic or feed-back/ recurrent architecture, dynamic mapping is achieved using feedback loops and delays. The presence of a recurrent structure has a profound impact on the learning and representation capacity of the neural network [38]. Figures 4 and 5 depict the major elements and working principles in a multi-layer perceptron (MLP) which is a common feedforward architecture in MPC implementations of air conditioning systems and recurrent neural networks which are also implemented in MPC for process industries [39].

Application of neural network models for MPC purposes has been undertaken by several researchers. The studies focus on hyperparameters optimization, best architecture selections, and optimization techniques based on the neural network models [40][41][42].

## V. Controller design

Optimization is the very essence of MPC next to the system model. Since optimal control actions are taken at regular steps, further manipulation of the model is compulsory. This becomes way easier when dealing with linear systems. That explains why linearization is a crucial step in the controller design of MPC. Local/global minimum and maximum values a linear optimization problem can be efficiently found by using optimization approaches such as the active set method or interior-point method [43] which are the

most commonly used optimizers in optimization tools like CasADi [44]. In addition to iterative methods, recursive and direct methods are linear programming techniques for optimization in MPC applications. Tail-call method is a typical instance under the recursive method while Gauss Jordan and Simplex fall under direct linear optimization techniques [11].

Shi Li et al. presented a model predictive control based on RBF neural network model for a CSTR process. After developing an inverse model with excellent nonlinear approximation performance, two RBF-MPC algorithms one with nonlinear optimization and another with local linearization were derived. The neural network model is linearized at the current sampling time, obtaining a linear model with time-varying parameters, assuring convex optimization that means the global optimum solution and the online calculation can be guaranteed. The performance evaluation by simulation experiments verified that MPC with local linearization presents higher accuracy and much less computation time than MPC with nonlinear optimization [45]. Comparatively low computational burden was achieved by the application of linear control for nonlinear systems by successive linearization also by Pramuditha Mendis et al [46].

Nonlinear programming constitutes a variety of techniques to determine local and global extremums of nonlinear systems. Direct and Derivative Based techniques make up the nonlinear optimization techniques that are used to find local extreme points. Grid Search, Random Search, Univariate Method are a few direct methods while Levenberg- Marquardt, Steepest Descent, Conjugate Gradient, and Penalty Function approach are Derivative Based Methods. Deterministic and Stochastic methods together with Heuristic and Metaheuristic methods constitute nonlinear techniques that determine global extremes of nonlinear systems. Under the Deterministic method, Band B method, Intervals Method, and Cutting plane methods are found. Stochastic tunneling is an example of Stochastic method while Particle Swarm Optimization (PSO) and Evolutionary Algorithms like Genetic Algorithm (GA) and Evolutionary Strategy are commonly used Heuristic and Metaheuristic global minima identification methods [11].

In most papers, as trained ANNs are black-box models, they require iterative, heuristic approaches to find solutions. For the algorithms, solutions often fall into numerous local optima. Moreover, the scheduling problems need to be solved iteratively to select the best local optimal solution among those obtained so far, which ultimately increases the computational time [47]. Even though the global optimality of the solutions is not guaranteed, [48] and [49] replicated trained ANNs using a set of nonlinear equations, which enabled the application of non-iterative, deterministic non-

linear solvers Nonlinear optimization problems may be non-convex and it may have many local minima. Thus, Young-Jin Kim proposed explicit ANN replication using a set of piecewise linear equations. The ANNs were then explicitly integrated into the optimal scheduling problem. Optimal solutions were then obtained for electricity prices and building thermal conditions [36].

GA and PSO stand out to be the most common nonlinear optimization methods among ANN-MPC applications. GA is empirically proven to provide robust searches for optimal solutions in complex spaces. Objective functions are minimized without calculating their derivatives and it is not restricted to the estimation of uncorrelated parameters. It may also produce more than one solution. The drawback of GA is that it is slow to converge for complex problems [50]. PSO can enhance and adapt to global and local exploration problems. PSO is excellent in solving single-objective optimization problems with fast convergence [51]. They are both heuristic approaches that do not necessarily need to be differentiated.

#### *A. Input convex neural networks (ICNN)*

Deep neural networks have proven to be successful in many identification tasks, however, from a model-based control perspective, these networks are difficult to work with because they are typically nonlinear and nonconvex. To bridge the gap between model accuracy and control tractability faced by neural networks, networks that are convex concerning their inputs are constructed explicitly [52]. Brandon Amos et al. [53] proposed the architecture of input convex neural networks which are scalar-valued neural networks  $f(x; y; \theta)$  where  $x$  and  $y$  denote inputs to the function and  $\theta$  denotes the parameters, built in such a way that the network is convex in (a subset of) inputs  $y$ . The fundamental benefit to these ICNNs is that optimization over the convex inputs to the network given some fixed value for other inputs is possible globally and efficiently.

The construction of the networks is highly dependent on the selection of convex and non-decreasing activation functions and non-negative weights. The proof of convexity in these networks is given by the fact that non-negative sums of convex functions are also convex and that the composition of a convex and convex non-decreasing function is also convex [54]. When joint convexity is maintained over the entire input to the function, fully input convex neural networks (FICNN) architecture is achieved. When convexity over only some inputs to the network is maintained partially input convex neural network (PICNN) architecture is achieved. The network architectures are shown in Figures 6 and 7.

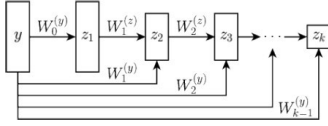


Fig. 6. A fully input convex neural network architecture [53]

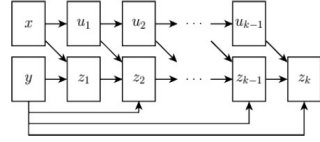


Fig. 7. A partially input convex neural network architecture [53]

A nonlinear optimization problem may be non-convex and it may have many local minima. To leverage the problem of convexity special neural networks namely input convex neural networks (ICNNs) were implemented. Yize Chen et al. [52] showed that input convex networks can be trained to obtain accurate models of complex physical systems. They designed input convex recurrent neural networks to capture the temporal behavior of dynamical systems and achieved optimal controllers via solving a convex model predictive control problem. They demonstrated the good potential of the proposed input convex neural network-based approach in a variety of control applications. Similarly, Felix Buenning et al. [55] adapted Input Convex Neural Networks that are only convex for one-step predictions for use in HVAC MPC. They introduced additional constraints to their structure and weights to achieve a convex input- output relationship for multistep ahead predictions, assessed the consequences of the additional constraints for the model accuracy, and finally tested the models in a real-life MPC experiment. Although these networks showed a decreased model accuracy compared to one-step ahead ICNN, the controller in the experiments kept the room temperatures within comfort constraints, while exploiting periods with relaxed constraints to save cooling energy.

Rebecka Winqvist et al. [9] presented a framework for off-line training and evaluation of neural network approaches for MPC. The basic idea is to approximate the MPC mapping from state to control input with constrained ReLU based neural networks including a projection layer. They considered the study of neural network architectures in PyTorch [56] with the explicit MPC constraints implemented as a differentiable optimization layer using CVXPY which is a state-of-the-art technique for modeling convex problems amenable to mathematical operations. CVXPY-layers realize the projection by solving a constrained optimization problem. The neural network used in conjunction with a projection layer is Projection NN (PNN). PNN structure is shown in Figure 8. The outcomes of their research show the trade-off between the number of training data and the approximation properties of the resulting controller.

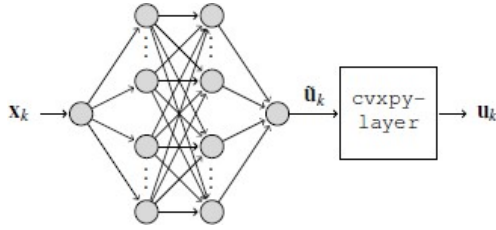


Fig. 8. Projection neural network structure [9]

## VI. Conclusion

In this review, the various advantages of using MPC over conventional controllers are thoroughly discussed. After the principal types, parameters, and steps in the process of developing an MPC are described, the two major steps namely system identification and development of control scheme were presented. Regarding the system identification, the three major approaches White-box, Black-box, and Grey-box modeling techniques were introduced and further exploration of neural network models takes place. The advantages and challenges of using neural networks for MPC applications are high representation capability and intractability respectively. Multi-layer Perceptrons (MLP), recurrent neural networks (RNN), and input convex neural networks (ICNN) are the three major neural network architectures discussed for MPC applications. The control scheme development mainly focuses on the exploitation of system models especially for optimization since MPC provides optimal control steps in line with the principles and constraints entailed in a system model. Nonlinearity and non-convexity are typical characteristics of Neural network system models which pose challenges for optimization. To tackle the optimization hindrance, general heuristic approaches, linearization, and convex networks are researched solutions. Genetic Algorithm and Particle Swarm Optimization are the two most commonly used heuristic approaches with slow response time. Linearized models give out guaranteed optimal solutions but are not representative of nonlinear systems. Finally, convex networks and layers with modifications for continuity were stated. Points in the summary of this review are:

- MPC outperforms conventional controllers.
- Neural networks have high system representation capabilities.
- Neural network models are as good as the quality of collected data and identified hyper parameters.
- The nonlinearity and non-convexity of neural network system models can be addressed resulting in high MPC performances.

## References

- [1] C. B. A. Eduardo F. Camacho, *Model Predictive Control*. Springer science & business media, 2013.
- [2] M. Tesfay, F. Alsaleem, P. Arunasalam, and A. Rao, "Adaptive-model predictive control of electronic expansion valves with adjustable setpoint for evaporator superheat minimization," *Build. Environ.*, vol. 133, no. December 2017, pp. 151–160, 2018, doi: 10.1016/j.buildenv.2018.02.015.
- [3] N. Tan and D. P. Atherton, "Improved cascade control structure for enhanced performance \_," vol. 17, pp. 3–16, 2007, doi: 10.1016/j.jprocont.2006.08.008.
- [4] B. Thomas, M. Soleimani-mohseni, and P. Fahle, "Feed-forward in temperature control of buildings," vol. 37, pp. 755–761, 2005, doi: 10.1016/j.enbuild.2004.10.002.
- [5] P. Hameed, N. Bin, M. Nor, P. Nallagownden, I. Elamvazuthi, and T. Ibrahim, "A review on optimized control systems for building energy and comfort management of smart sustainable buildings," *Renew. Sustain. Energy Rev.*, vol. 34, pp. 409–429, 2014, doi: 10.1016/j.rser.2014.03.027.
- [6] D. S. F. Belic, Z. Hocenski, "HVAC control methods-a review," in *19th International Conference on System Theory, Control and Computing, ICSTCC 2015 - Joint Conference SINTES 19, SACCS 15, SIMSIS 19*, 2015, pp. 679–686.
- [7] J. C. Hung, "Internal model control," *Control Mechatronics*, pp. 8.1-8.3, 1981, doi: 10.1201/9781315218403-8.
- [8] M. Schwenzer, M. Ay, T. Bergs, and D. Abel, "Review on model predictive control: an engineering perspective," *Int. J. Adv. Manuf. Technol.*, vol. 117, no. 5–6, pp. 1327–1349, 2021, doi: 10.1007/s00170- 021-07682-3.
- [9] R. Winqvist, A. Venkitaraman, and B. Wahlberg, "On Training and Evaluation of Neural Network Approaches for Model Predictive Control," 2020, [Online]. Available: <http://arxiv.org/abs/2005.04112>.
- [10] K. Kiš and M. Klaučo, "Neural network based explicit MPC for chemical reactor control," *Acta Chim. Slovaca*, vol. 12, no. 2, pp. 218–223, 2019, doi: 10.2478/acs-2019-0030.
- [11] Y. Yao and D. K. Shekhar, "State of the art review on model predictive control (MPC) in Heating Ventilation and Air-conditioning (HVAC) field," *Build. Environ.*, vol. 200, no. May, p. 107952, 2021, doi: 10.1016/j.buildenv.2021.107952.
- [12] M. B. Kane, J. P. Lynch, and J. Scruggs, "Run-Time Efficiency of Bilinear Model Predictive Control Using Variational Methods, with Applications to Hydronic Cooling," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 2, pp. 718–728, 2019, doi: 10.1109/TMECH.2019.2896020.
- [13] Y. I. Alamin, M. Del Mar Castilla, J. D. Álvarez, and A. Ruano, "An economic model-based predictive control to manage the users' thermal comfort in a building," *Energies*, vol. 10, no. 3, pp. 1–18, 2017, doi: 10.3390/en10030321.
- [14] L. Cavanini, G. Cimini, and G. Ippoliti, "Computationally efficient model predictive control for a class of linear parameter-varying systems," *IET Control Theory Appl.*, vol. 12, no. 10, 2018, doi: 10.1049/iet-cta.2017.1096.
- [15] T. Zakula, P. R. Armstrong, and L. Norford, "Advanced cooling technology with thermally activated building surfaces and model predictive control," *Energy Build.*, vol. 86, pp. 640–650, 2015, doi: 10.1016/j.enbuild.2014.10.054.
- [16] X. Chen, Q. Wang, and J. Srebric, "Model predictive control for indoor thermal comfort and energy optimization using occupant feedback," *Energy Build.*, vol. 102, pp. 357–369, 2015, doi: 10.1016/j.enbuild.2015.06.002.
- [17] W. H. Chen et al., "Data-driven robust model predictive control framework for stem water potential regulation and irrigation in water management," *Control Eng. Pract.*, vol. 113, no. May, p. 104841, 2021, doi: 10.1016/j.conengprac.2021.104841.

- [18] W. H. Chen and F. You, "Smart greenhouse control under harsh climate conditions based on data-driven robust model predictive control with principal component analysis and kernel density estimation," *J. Process Control* , vol. 107, pp. 103–113, 2021, doi: 10.1016/j.jprocont.2021.10.004.
- [19] A. Mesbah, I. V. Kolmanovsky, and S. Di Cairano, "Stochastic model predictive control," in *Handbook of model predictive control* , Springer, 2019, pp. 75–97.
- [20] X. Li, N. Li, I. Kolmanovsky, and B. I. Epureanu, "Stochastic model predictive control for remanufacturing system management," *J. Manuf. Syst.* , vol. 59, no. August 2020, pp. 355–366, 2021, doi: 10.1016/j.jmsy.2021.02.002.
- [21] E. Chee and X. Wang, "Generalized system identification for nonlinear MPC of highly nonlinear MIMO systems," *IFAC- PapersOnLine* , vol. 54, no. 3, pp. 366–371, 2021, doi: 10.1016/j.ifacol.2021.08.269.
- [22] M. Trčka and J. L. M. Hensen, "Overview of HVAC system simulation," *Autom. Constr.* , vol. 19, no. 2, pp. 93–99, 2010, doi: 10.1016/j.autcon.2009.11.019.
- [23] F. Oldewurtel, C. Sagerschnig, and Z. Eva, "Building modeling as a crucial part for building predictive control," vol. 56, pp. 8–22, 2013, doi: 10.1016/j.enbuild.2012.10.024.
- [24] Z. Afroz, G. M. Shafiullah, T. Urmee, and G. Higgins, "Modeling techniques used in building HVAC control systems: A review," *Renew. Sustain. Energy Rev.* , vol. 83, no. December 2017, pp. 64–84, 2018, doi: 10.1016/j.rser.2017.10.044.
- [25] A. Afram and F. Janabi-Sharifi, "Review of modeling methods for HVAC systems," *Appl. Therm. Eng.* , vol. 67, no. 1–2, pp. 507–519, 2014, doi: 10.1016/j.applthermaleng.2014.03.055.
- [26] R. Z. Homod, "Review on the HVAC System Modeling Types and the Shortcomings of Their Application," *J. Energy* , vol. 2013, pp. 1–10, 2013, doi: 10.1155/2013/768632.
- [27] S. CHEN, S. A. BILLINGS, and P. M. GRANT, "Nonlinear system identification using neural networks," *Int. J. Control* , pp. 1191–1214, 2007, [Online]. Available: <https://doi.org/10.1080/00207179008934126>.
- [28] Y. Li, Z. O'Neill, L. Zhang, J. Chen, P. Im, and J. DeGraw, "Grey-box modeling and application for building energy simulations - A critical review," *Renew. Sustain. Energy Rev.* , vol. 146, no. May, p. 111174, 2021, doi: 10.1016/j.rser.2021.111174.
- [29] A. Afram, F. Janabi-Sharifi, A. S. Fung, and K. Raahemifar, "Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system," *Energy Build.* , vol. 141, pp. 96– 113, 2017, doi: 10.1016/j.enbuild.2017.02.012.
- [30] A. Li, Y. Sun, and X. Xu, "Development of a simplified resistance and capacitance (RC)-network model for pipe-embedded concrete radiant floors," *Energy Build.* , vol. 150, pp. 353–375, 2017, doi: 10.1016/j.enbuild.2017.06.011.
- [31] C. Cui, X. Zhang, and W. Cai, "An energy-saving oriented air balancing method for demand controlled ventilation systems with branch and black-box model," *Appl. Energy* , vol. 264, no. February, p. 114734, 2020, doi: 10.1016/j.apenergy.2020.114734.
- [32] A. Jain, F. Smarra, E. Reticcioli, A. D'Innocenzo, and M. Morari, "NeuroOpt: Neural network based optimization for building energy management and climate control," vol. 120, no. 2016, pp. 1–10, 2020, [Online]. Available: <http://arxiv.org/abs/2001.07831>.
- [33] D. Wang, "for a Nonlinear Overhead Crane System," *IEEE Trans. Ind. Informatics* , vol. 14, no. 7, pp. 2932– 2940, 2018.
- [34] Z. R. Yang and Z. Yang, *Artificial Neural Networks* , vol. 6. Elsevier B.V., 2014.
- [35] A. Afram, F. Janabi-Sharifi, A. S. Fung, and K. Raahemifar, "Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC sys-

- tems: A state of the art review and case study of a residential HVAC system,” *Energy Build.* , vol. 141, pp. 96– 113, 2017, doi: 10.1016/j.enbuild.2017.02.012.
- [36] Y. J. Kim, “A supervised-learning-based strategy for optimal demand response of an HVAC system,” *arXiv* , 2019.
  - [37] S. Haykin, *Neural Networks and Learning Machines* , 2nd ed. Pearson Education, 2009.
  - [38] J. D. Rios, A. Y. Alanis, N. Arana-Daniel, and Carlos Lopez-Franco, *Neural Networks Modeling and Control, Applications for Unknown Nonlinear Delayed Systems in Discrete Time* . Academic Press, 2020.
  - [39] N. Lanzetti, Y. Z. Lian, A. Cortinovis, L. Dominguez, and M. Mercang, “Recurrent Neural Network based MPC for Process Industries,” pp. 1005–1010, 2019.
  - [40] D. Kreuzer, M. Munz, and S. Schlüter, “Short-term temperature forecasts using a convolutional neural network — An application to different weather stations in Germany,” *Mach. Learn. with Appl.* , vol. 2, no. October, p. 100007, 2020, doi: 10.1016/j.mlwa.2020.100007.
  - [41] J. Jin, S. Shu, and F. Lin, “Prediction of indoor air temperature based on deep learning,” *Sensors Mater.* , vol. 31, no. 6, pp. 2029– 2042, 2019, doi: 10.18494/SAM.2019.2290.
  - [42] B. M. Åkesson and H. T. Toivonen, “A neural network model predictive controller,” *J. Process Control* , vol. 16, no. 9, pp. 937– 946, 2006, doi: 10.1016/j.jprocont.2006.06.001.
  - [43] A. N. N. Approach, *Computationally Efficient Model Predictive Control Algorithms* . .
  - [44] J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, “CasADi a software framework for nonlinear optimization and optimal control,” *Math. Program. Comput.* , vol. 11, no. 1, pp. 1–36, 2019, doi: 10.1007/s12532-018-0139-4.
  - [45] S. Li, P. Jiang, and K. Han, “RBF Neural Network based Model Predictive Control Algorithm and its Application to a CSTR Process,” pp. 2948–2952, 2019.
  - [46] P. Mendis, “Adaptive Model Predictive Control with Successive Linearization for Distillate Composition Control in Batch Distillation,” pp. 366–369, 2019.
  - [47] E. Mocanu et al. , “On-Line Building Energy Optimization Using Deep Reinforcement Learning,” vol. 10, no. 4, pp. 3698–3708, 2019.
  - [48] S. Silva and A. Ruano, “IMBPC HVAC system Wireless Sensors and IoT Platform The HVAC system Wireless Sensors and IoT Platform The IMBPC HVAC system Wireless Sensors and IoT Platform,” *IFAC- PapersOnLine* , vol. 51, no. 10, pp. 1–8, doi: 10.1016/j.ifacol.2018.06.227.
  - [49] Y. M. Lee, R. Horeish, and L. Liberti, “Optimal HVAC control as demand response with on-site energy storage and generation system,” *Energy Procedia* , vol. 78, pp. 2106–2111, 2015, doi: 10.1016/j.egypro.2015.11.253.
  - [50] A. Garnier, J. Eynard, M. Caussanel, and S. Grieu, “Predictive control of multizone heating , ventilation and air-conditioning systems in non- residential buildings,” *Appl. Soft Comput. J.* , vol. 37, pp. 847–862, 2015, doi: 10.1016/j.asoc.2015.09.022.
  - [51] A. Kusiak, M. Li, and F. Tang, “Modeling and optimization of HVAC energy consumption,” *Appl. Energy* , vol. 87, no. 10, pp. 3092–3102, 2010, doi: 10.1016/j.apenergy.2010.04.008.
  - [52] Y. Chen, Y. Shi, and B. Zhang, “Optimal control via neural networks A convex approach,” *7th Int. Conf. Learn. Represent. ICLR 2019* , pp. 1–25, 2019.
  - [53] B. Amos, L. Xu, and J. Z. Kolter, “Input Convex Neural Networks,” 2017.
  - [54] S. Boyd and L. Vandenberghe, *Convex Optimization* . Cambridge University Press, 2009.



- [55] F. Bünning, A. Schalbetter, A. Aboudonia, M. H. de Badyn, P. Heer, and J. Lygeros, "Input Convex Neural Networks for Building MPC," vol. xxx, pp. 1–11, 2020, [Online]. Available <http://arxiv.org/abs/2011.13227>.
- [56] A. Paszke et al., "PyTorch: An Imperative Style , High-Performance Deep Learning Library," no. NeurIPS, 2019.

# Approaches of a Human-Machine-Interface for Augmented Reality in Automotive Systems

1<sup>st</sup> Weber Michael

*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
michael.weber@h-ka.de

2<sup>nd</sup> Weiß Tobias

*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
tobias.weiss@h-ka.de

3<sup>rd</sup> Nötzold Marlene

*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
noma1027@h-ka.de

## **Abstract**

To use Augmented Reality (AR) in an automotive vehicle for testing Advanced Driver Assistance Systems (ADAS) a new development approach with high computing power is needed. Reasons for this are a high vehicle speed as well as fewer possible orientation points on an urban test track compared to using AR applications inside a building. The objective of this paper is the conceptual design of an Human-Machine-Interface (HMI) for AR-ADAS test systems. For this purpose, a requirements profile for the use of HMI in real automobiles is first developed in order to identify and subsequently evaluate different HMI approaches based on this profile. Based on the AR term specified by Azuma [1] and the associated display classes, as well as the HMI requirements profile developed, three see-through (ST) variants of the head-up display (HUD) type and two ST variants of the head-mounted display (HMD) type are identified as possible HMI approaches. By deriving the relevant criteria from the requirements profile, the utility analysis conducted yields significant findings with regard to technical feasibility.

## **Index Terms**

Augmented Reality, Advanced Driver Assistance Systems, Human-Machine-Interface

## I. Introduction

Advanced Driver Assistance Systems (ADAS) such as the active lane departure warning system and traffic sign recognition support the driver, offer comfort, and take responsibility for increasing road safety. These complex systems go through an extensive testing phase, which results in optimization potential regarding quality, reproducibility, and costs. ADAS in the future will support ever-larger proportions of driving situations in increasingly complex scenarios. Due to the increasing complexity of vehicle communication and the rising demands for functional safety in a complex environment, support for the driver and increased road safety, the test scenarios for ADAS are constantly further developed and adapted to higher requirements. European New Car Assessment Programme (Euro NCAP) has introduced a series of new safety tests for ADAS into its program and created a road map until the year 2025 [2] [3]. Today's test methods can be separated into two categories. On the one hand, the testing of the ADAS with the help of virtual worlds (Simulation) and on the other hand, real world testing on the test track using vehicles and other objects in real life. The central idea of the virtual test procedure is to transfer vehicle behavior to virtual test drives as realistically as possible. The approach for virtual tests is aimed to benefit from the advantages of simulation in terms of reproducibility, flexibility, and reduction of effort. In this way, specifications and solutions derived from this virtual tests should be able to be evaluated at an early stage of the development process. The use of suitable simulation methods enables the efficient design, development, and application of vehicles and vehicle components. However, virtual development methods cannot yet replace real world testing on the test track in all respects. Due to the complex physical conditions in which a vehicle is exposed when testing ADAS, real world testing on the test tracks are still necessary to the current status. For example, the weather and the surface texture of the road take a decisive role in the evaluation process of ADAS test drives [4] [5]. The presented research background of this approach combines the advantages of ADAS-tests in a virtual simulation and these of ADAS-tests in a real environment. Particularly the camera images of the vehicle are augmented with additional virtual information. The augmentation of virtual road lanes allows, for example, the testing of a lane departure warning system independent of the test track. Scenarios such as the appearance of temporary lanemarkings or the absence of sections can be tested on the same test area. Narrowing and widening of lane markings can be represented as well as international differences between road markings. For testing traffic jam assistance systems, vehicles driving ahead can be augmented with camera images. In the first phase of testing, second vehicles including drivers can thus be dispensed with, reducing the costs of the tests and increasing the safety of the test engineers. Furthermore, ADAS-test cases with traffic signs

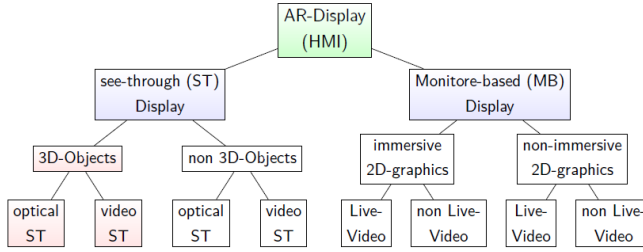


Fig. 1. AR display classes inspired by Milgram [15]

as well as pedestrians and cyclists can be augmented situationally and quickly. Furthermore, by using Augmented Reality (AR) for testing ADAS, new possibilities for testing complex, critical, and even forbidden test cases arise. For example, for testing the lane departure warning system, the traffic lane can be inserted into the image in any given width, regarding the lane and the white stripe itself. Therefore it is possible to test the system to its limits, a feature not possible by testing in reality on the test track.

Based on this, the aim is to develop a real-time AR-ADAS electronic control unit (ECU) with a suitable human-machine interface (HMI) for use in a real vehicle. The HMI serves as an interface between the driver and the AR-ADAS-ECU. This supports the test driver in the execution of the test cases by visualising the extensions of the test track that do not exist in reality for the driver by means of AR. This paper deals with the conceptual design of a suitable HMI for the AR-ADAS test system as well as the identification and evaluation of HMI approaches with regard to their suitability for the overall project.

## II. Definition of HMI Approaches for AR

The main task of the HMI is to make the AR - which is also transmitted to the vehicle - perceptible to the test driver in real time in such a way that the ADAS functions of the test vehicle can be evaluated, including human interaction. The acceptance of the different HMI solutions, as an interface to the virtual experience, plays an important role here. This depends, among other things, on the quality of the visualisation, the interaction, but also the haptics [6]. For the presented approach, the selection of a suitable HMI concept focuses exclusively on visualisation and interaction. In order to visualise AR, the use of a suitable HMI is absolutely necessary. As it can be seen in figure 1, the various display approaches can be divided into characteristic classes based on their properties.

Displays that allow a medium-direct view through to the real environment in 3D belong to the class of see-through (ST) displays. Monitor-based (MB)

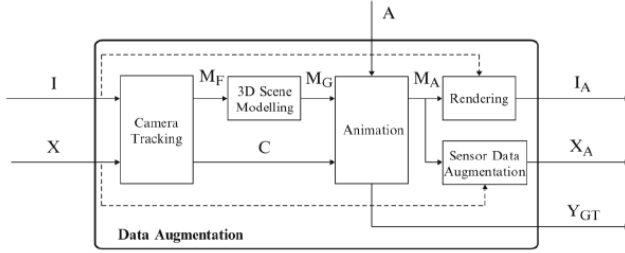


Fig. 2. 3D Scene Modelling [8]

displays, on the other hand, only allow an indirect view of the real environment. This is done with the help of live or stored videos (2D). Indirect displays (3D objects: video ST), which visualise AR three-dimensionally using video, also belong to the group of ST displays. The 3D concept is crucial here. The processing of the 2D camera data used for the real environment, through 3D scene modelling, makes it possible to integrate the virtual objects in the correct perspective (Fig. 2).

If the recording and playback of this same AR on an indirect display take place almost simultaneously, they are called video-based ST displays (video ST). Displays are called Optical ST displays (3D-Objects:optical ST) when the reproduction of the virtual objects in combination with the direct view of the real environment are correctly integrated. The visualisation of AR according to Azuma [1] limits the AR-capable displays to those that are able to display virtual three-dimensional objects correctly oriented in perspective (groups marked in red, Fig.1). For the identification of suitable HMI approaches for testing camera-based ADAS, only these ST displays fulfil the necessary criteria [15].

### III. Requirements profile for AR-HMI

The requirements profile describes the entire set of the functional requirements for the HMI. The requirements are formulated as generally as possible and as restrictively as necessary in order to be able to make a suitable selection for the AR-ADAS test system within the conceptual design. The conception of the HMI approach elaborated in this work has the objective of obtaining a comprehensive picture of all for the application in the test vehicle applicable HMI approaches. For this purpose, the function-based requirements profile offers a high degree of flexibility in implementation. The basic requirements (No. 1-2, Table I) for the AR-ADAS-ECU with HMI as an overall system must be fulfilled by the HMI and are therefore a prerequisite. Based on these, the requirements for the HMI are derived (No. 3-11, Table I).

Nr.	Category	Description
1.	Basic requirement	AR-ADAS-ECU with HMI simulates and visualises test track using AR to test ADAS functions
2.	Basic requirement	Adaptive system - is compatible with most common vehicle types and OEMs
3.	Requirement HMI	Relevant virtual three-dimensional objects must be fully visible to test drivers
4.	Requirement HMI	HMI visualises AR in real time (min. 25fps, delay not visible for test drivers)
5.	Requirement HMI	HMI visualises AR correctly at all times (geometric registration of the virtual three-dimensional objects)
6.	Requirement HMI	High acceptance of the HMI by the test driver
7.	Requirement HMI	The HMI input function must not distract the driver while driving
8.	Requirement HMI	Interface between HMI and AR-ADAS-ECU
9.	Requirement HMI	Operating scope HMI: ON/OFF function
10.	Requirement HMI	Interaction between HMI and test driver possible
11.	Requirement HMI	Minimal field of vision impairment due to HMI

TABLE I  
Basic requirements and special requirements for HMI

The basic requirements for the overall system identify suitable HMI approaches. The requirements profile for the HMI serves as the basis for a utility analysis, which is used to evaluate the identified HMI approaches. Based on the results, suitable HMI models of the favoured HMI approach are identified, evaluated and applied. For their evaluation, the HMI requirements have to be specified and completed according to the conceptual design of the overall system. With regard to interaction between the HMI and the test driver, these are, for example, the following specifications:

- Operating scope HMI: Selection of vehicle type and test case
- HMI calls for action: reference reality OK/NOK for test
- HMI informs: Test OK / NOK / optional NOK Deviation from OK

#### IV. Identification of the HMI approaches

The identified HMI approaches result from the Azuma [1] definition of AR and the associated display classes as well as from the criteria from the requirements profile for the HMI for the test case of camera-based ADAS functions in the vehicle. The HMI approaches described belong to the class of ST displays (optical or video-based), which can visualise AR generated with virtual three-dimensional objects. They fulfil the basic requirements 1. and 2. of the requirements profile and have an interface to the AR-ADAS-ECU. HMI approaches in which stationary displays are mechanically fixed to the vehicle for the duration of the test belong to the Head-Up Display (HUD) group. Those in which the display is attached to the head belong to the Head-Mounted Display (HMD) group [8]. In both HMD and HUD, HMI approaches of optical and video-based ST displays are identified, which are described in more detail below.

##### A. HMD - Head-Mounted Display

HMDs are displays that are worn directly in front of the eyes. This can be realised in the form of glasses or a helmet. HMDs usually have tracking systems (determination of position and orientation) with which the change in the user's viewing direction and position can be continuously readjusted accordingly. In addition, the glasses can be further divided into those with an open design (AR glasses) and those with a closed design (Virtual Reality (VR) glasses) [8].

- HMD ST optical: Head-Mounted Display (see-through) medium-direct
- HMD ST video: Head-Mounted Display (see-through) via Video

Both AR and VR glasses are suitable for visualising AR for testing ADAS functions in the vehicle. They are referred to below as 'HMD ST optical' for AR glasses and 'HMD ST video' for VR glasses.

1) *HMD ST optical*: The HMI approach is based on the use of AR glasses, in which the real environment and the virtual three-dimensional objects are optically overlaid for the user's view. The vision of the real environment is medium-direct and means the view through a medium (e.g. glass) onto the real environment. By superimposing this view around the virtual three-dimensional objects, the three-dimensional AR is created for the user. For the perspective-correct integration of the virtual three-dimensional objects, the real environment is tracked via the camera of the glasses. The open design allows the user to perceive the environment outside the field of vision provided by the glasses [8]. Typical representatives of this approach are, for example: Microsoft HoloLens, Magic Leap One and Moverio BT-300. The well-known Google Glass, on the other hand, is not suitable for three-dimensional AR due to its monocular principle. Typical representatives of



Fig. 3. Typical representatives of AR glasses - left: Microsoft HoloLens [9], right: Epson Moverio BT-300 [10]



Fig. 4. Typical representative of VR glasses - Samsung Gear VR [11]

AR glasses are shown in Figure 3.

2) *HMD ST video*: With the HMD ST video, the real environment is also captured via the camera of the glasses, but - in contrast to the HMD ST optical - "overlaid with the virtual content as part of the rendering" [8]. Subsequently, the three-dimensional AR created in this way is visualised as a video. This is an indirect view of the real environment, as this is viewed through the playback of a video. The closed design of VR glasses restricts the user's field of vision to the field of vision provided by the glasses, and the environment outside the field of vision cannot be perceived accordingly [8]. Typical representatives of HMD ST video are, for example: HTC Vive and Samsung Gear VR (Figure 4).

#### B. HUD - Head-Up Display

HUDs are displays that are stationary and mechanically fixed to the vehicle for the duration of the test. The ratio of the field of view to the driver's





Fig. 5. Typical representatives of HUD ST optical - Hudway Drive [13]

field of vision depends on the display size and the driver's distance from the display and has a major influence on the perception and acceptance of the AR visualised in this way [8]. Stationary displays also allow the user to perceive the environment outside the display's field of view. Depending on whether the display supports a medium-direct (HUD ST optical) or an indirect view (HUD ST video) of the real environment and whether the HMI's own or the vehicle's internal camera (car) is used for tracking and rendering, there are three different HUD-based HMI approaches:

- HUD ST optical: Head-Up Display (see-through) medium-direct
- HUD ST video: Head-Up Display (see-through) via Video
- HUD ST video (car): Head-Up Display (see-through) indirectly with camera tracking vehicle

All three approaches are also suitable for the above application and are presented below.

1) *HUD ST optical*: This HMI approach is based on the principle of AR glasses, in which the real environment and the virtual three-dimensional objects are optically overlaid for the user's view. The view of the real environment is medium-direct. For the correct perspective integration of the virtual three-dimensional objects, the real environment is tracked by the user's own camera. A portable representative of a suitable HUD ST optical approach is for example the Hudway drive in Figure 5.

2) *HUD ST video*: This HMI approach is based on the principle of VR glasses, in which the video recordings of the real environment are overlaid



Fig. 6. Typical representatives of HUD ST video - Apple iPad with Holder [13]

with the virtual three-dimensional objects as part of the rendering. Via the HUD ST video, the three-dimensional AR is visualised as a video. This is also an indirect view of the real environment without the disadvantage of a closed design. Typical representatives of the HUD ST video approach are mobile devices such as Microsoft Surface, Apple iPad, tablets with suitable hardware and software (Figure 6).

3) *HUD ST video (car)*: This HMI approach is based on the principle of the HUD ST video approach. In contrast to this approach, the HUD ST video (car) approach uses the image material of the vehicle's internal camera instead of its own camera. Depending on the type of vehicle, tracking takes place at the level of the bumper or the rear-view mirror. The AR generated in this way is visualised in an offset manner from the driver's point of view (angle of view of the vehicle camera). By means of a relation matrix, the tracked position can be adapted to the driver and, with this approach, enables a significantly reduced computational effort for the AR-ADAS ECU, since only one image is used to generate the AR instead of two different camera images. Typical representatives of the HUD ST video (car) approach are also mobile devices with suitable hardware and software (see Figure 6).

In the following chapter, the HMI approaches described here are compared and evaluated using a utility analysis.

## V. HMI Evaluation

In order to be able to carry out an overall evaluation of the identified HMI approaches, a utility analysis is carried out in this chapter based on all relevant criteria. In a first step, the relevant criteria are defined on the basis of the requirements profile of the HMI. In the second step, the utility analysis is carried out and finally the results of this analysis are evaluated.

### A. Criteria of the utility analysis

The analysis of the HMI approaches makes it necessary to derive the relevant criteria from the requirements profile and to specify them individually so that a meaningful evaluation is possible. The order of the criteria is chosen arbitrarily.

1) [Req. 1]: *100 % visibility of 3D objects*: Completely visible is considered to be when all virtual three-dimensional objects (v3D-Obj.) relevant for the test drive can be visualised and recognised by the test driver at any time during the test case. This means that, in the best case, the driver recognises every object that the driver would also recognise in the real scenario. In addition, a 1:1 representation of the AR is rated higher than a 1:k representation.

2) [Req. 2]: *Visualisation in real time*: The visualisation of AR in real time depends on data transmission, processing and the amount of data. When considering the approaches, the expected computing effort caused by the radius of movement is evaluated in terms of the amount of data and the amount of data expected for visualisation. The real-time capability can be evaluated by means of a quantitative estimate of the expected data volume from the HMI to the ECU and the associated processing time within the framework of the data augmentation process (Figure 2). The acceptance of AR visualised via HMI depends on various criteria. This is linked to the (a) quality of the AR visualisation, the convenience, the (b) realism and the (c) experience with the technology.

3) [Req. 3]: *Field of view ratio (a)*: The quality of the visualised AR depends on the programming, the video resolution difference between the real environment and the v3D objects as well as the expected field of view in relation to the field of view of the test driver. The size of the field of view depends on the display size of the HMI and the distance between the HMI and the driver. When considering the HMI approaches, only the ratio of the field of view to the field of vision is evaluated. The criterion is considered to be completely fulfilled at a ratio of 1:1 and fulfilled to a small extent at 1:k with k much greater than 1.

4) [Req. 4]: *Reality reference (medium-direct) (b)*: The reference to reality is considered to be fully fulfilled if the HMI approach allows a medium-direct view of the real environment. A medium-direct view is when the view of the real environment is through a medium (e.g. glass). In contrast to the indirect view (e.g. live video transmission), a much higher acceptance by the test driver can be expected with the medium-direct view.

5) [Req. 5]: *Technology experience (c)*: Experience with the technology is considered to be fulfilled if the approach is already widely used in other areas of application.

6) [Req. 6]: *Type of operation*: The extent to which interaction with the HMI distracts the test driver while driving depends on the one hand on the programming and on the other hand on the type of operation. Depending on the type of operation, the criterion is considered to be fully met if the expected distraction is low.

7) [Req. 7]: *Direct interaction HMI*: The type of interaction between the test driver and the overall AR-ADAS-ECU system, which takes place via the HMI, depends on the programming and the HMI model used. For the analysis of the HMI approaches, it is only evaluated whether the interaction with the HMI can take place directly or indirectly according to the current technical status. Here, the criterion is considered to be fully met if direct interaction is possible.

8) [Req. 8]: *Field of vision impairment*: From a safety point of view, the vehicle control by the test driver must be taken into account. This depends, among other things, on the test scenario, test driver, vehicle and the field of vision impairment caused by the HMI. Only the field of vision impairment that is to be expected from the HMI according to the current technical status is evaluated. The cause of the impairment can be: inactive, defective or faulty HMI. The criterion is considered to be fully met if the expected impairment is low and the vehicle control is high.

## B. Utility analysis

Based on the requirements 3-11 for the HMI (see Table I), the derived relevant criteria of the utility value analysis result. These are weighted in a pairwise comparison within the framework of the utility analysis. Subsequently, the identified HMI approaches are evaluated for each criterion on the basis of a defined scale from 0 to 3 (0 = criterion is not fulfilled, 3 = criterion is fully fulfilled). This scale results from the fact that only a quantitative assessment is possible within the framework of concept identification. In addition, an even number of the scale excludes the possibility of averaging. Multiplying the evaluation by the weighting results in the respective utility value of the HMI approach. The HMI approach with the highest utility value is considered the optimum. This makes it possible to reduce the complexity of the criteria with the help of a ranking and at the same time reveal the differences [13]. It should be noted that the evaluation is based on the authors' perception. The pairwise comparison of the criteria is shown in Figure 7.

	100% visibility of 3D objects	Visualisation in real time	Field of view ratio	Reality reference	Technology experience	Type of operation	Direct interaction HMI	Field of vision impairment	Sum	%
100% visibility of 3D objects		1	1	0	1	1	1	0	5	17,86%
Visualisation in real time	0		1	1	1	1	1	0	5	17,86%
Field of view ratio	0	0		0	1	0	1	0	2	7,14%
Reality reference	1	0	1		1	1	1	0	5	17,86%
Technology experience	0	0	0	0		0	1	0	1	3,57%
Type of operation	0	0	1	0	1		1	0	3	10,71%
Direct interaction HMI	0	0	0	0	0	0		0	0	0,00%
Field of vision impairment	1	1	1	1	1	1	1		7	25,00%

Fig. 7. Utility analysis - pairwise comparison

Nr.	Criteria	% of 100
1.	[Req. 8]: Field of vision impairment	25
2.	[Req. 1]: 100% visibility of 3D objects	17,86
3.	[Req. 4]: Reality reference (medium-direct)	17,86
4.	[Req. 2]: Visualisation in real time	17,86
5.	[Req. 6]: Type of operation	10,71
6.	[Req. 3]: Field of view ratio	7,14
7.	[Req. 5]: Technology experience	3,57
8.	[Req. 7]: Direct interaction	0,00

TABLE II  
Utility analysis - Criteria Ranking

- 0 = Criteria (horizontal) is less important than criteria (vertical)
- 1 = Criteria (horizontal) is more important than criteria (vertical)

Based on the weighting of the criteria, the following ranking results in Table II.

The safety-relevant criterion '[Req. 8]: Field of vision impairment', with a weighting of 25 percent, has the greatest influence on the utility value

Requirements	Weights	HUD ST optical		HUD ST video		HUD ST video (car)		HMD ST optical		HMD ST video	
		Evaluation	Value	Evaluation	Value	Evaluation	Value	Evaluation	Value	Evaluation	Value
100% visibility of 3D objects	17,86%	0	-	2	0,36	2	0,36	3	0,54	3	0,54
Visualisation in real time	17,86%	2	0,36	2	0,36	3	0,54	1	0,18	1	0,18
Field of view ratio	7,14%	1	0,07	2	0,14	2	0,14	3	0,21	3	0,21
Reality reference	17,86%	3	0,54	2	0,36	2	0,36	3	0,54	0	-
Technology experience	3,57%	1	0,04	2	0,07	3	0,11	1	0,04	2	0,07
Type of operation	10,71%	0	-	1	0,11	1	0,11	2	0,21	1	0,11
Direct interaction	0,00%	0	-	3	-	3	-	2	-	1	-
Field of vision impairment	25,00%	3	0,75	1	0,25	1	0,25	3	0,75	0	-
Summe			1,75		1,64		1,86		2,46		1,11

Fig. 8. Utility analysis - Evaluation

Nr.	HMI-Approach	Utility Value	%
1.	HMD ST optical	2,46	82,0
2.	HUD ST video (car)	1,86	62,0
3.	HUD ST optical	1,75	58,3
4.	HUD ST video	1,64	54,7
5.	HMD ST video	1,11	37,0

TABLE III  
Utility analysis - Overall Ranking

of the HMI approaches. The evaluation of the approaches based on the criterion '[Req. 7]: Direct interaction' has no weighting and thus no influence on the utility value. In addition to the criterion of minimum field of vision impairment, '[Req. 1]: 100% visibility of 3D objects', '[Req. 4]: Reality reference (medium-direct)' and '[Req. 2]: Visualisation in real time' have a high influence on the utility value of the HMI approaches. Based on the weighted criteria, the next step is to evaluate the HMI approaches (Figure 8).

The utility values of the individual HMI approaches determined on the basis of the utility value analysis result in the ranking (Table III). The HMI approach HUD ST optical represents the optimum with the highest utility value.

### *C. Evaluation of the utility analysis*

Based on the technical evaluation, the HMI approach HMD ST optical (AR glasses) is prioritised with a utility value of 2.46 (82 percent) of the maximum utility value 3 to be achieved. The HMI approach HMD ST video (VR glasses), on the other hand, has the lowest utility value in comparison (37 percent utility value HMD ST optical). Due to the closed design, the criterion '[Req. 8]: Field of vision impairment' is not fulfilled and endangers the safety of the test driver in case of a technical failure. This approach is therefore not considered for further investigations. The three HMI approaches HUD ST are comparable in terms of utility value and are below the utility value of the HMD ST optical. If only the HUD ST HMI approaches are compared with each other, it can be seen that HUD ST video (car) is to be prioritised. The HMI approach HMD ST optical scores well in the safety-relevant criterion due to the medium-direct view with minimal impairment of the field of vision. As a head-mounted display, the maximum ratio of field of view to field of vision is achieved. This automatically leads to 100 percent visibility of the v3D objects with a 1:1 ratio of the visualisation to the real environment. The medium-direct view also ensures a high degree of realism. Together with the criterion 'field of view: field of vision (1:1)', a high acceptance of the HMI by the test driver can be expected. Compared to the other two head-up displays, the HUD ST video (car) HMI approach scores highly in the criterion 'visualisation in real time', as the expected processing time is significantly reduced. Compared to the HUD ST optical, the higher field of view impairment of the HUD ST video as well as the HUD ST video (car) leads to reduced safety. In contrast to the HUD ST optical, however, the criteria 'field of view:field of vision(1:1)' and 100 percent visibility of the v3D object are rated higher for both HUD ST video with a 1:k ratio of the visualisation to the real environment. In conclusion, the two favoured HMI approaches of the technical utility analysis are as follows: 1. HMD ST optical and 2. HUD ST video (car).

### **VI. Conclusion**

The number and complexity of ADAS functions will continue to increase in the future [16]. The associated development and testing effort, especially for camera-based ADAS, can be significantly reduced by applying simulated test cases using AR in real vehicles. Azuma [1] specifies AR as the real environment augmented by virtual objects (e.g. road signs, lane markings or crossing people), whereby the virtual objects are integrated three-dimensionally and perspective-correctly into an interactive three-dimensional real-time environment. By applying this specified AR, the function of camera-based ADAS can be validly tested for a wide variety of test cases. The integration of the generated AR takes place on the one hand in the vehicle and on the other hand via an HMI suitable for visualisation, which

supports the test driver during the test drive. In order to enable a realistic behaviour of the test driver and at the same time ensure his safety, high technical requirements are placed on the HMI approach. The established requirements profile describes the totality of the functional requirements for the HMI. Using the pairwise comparison, eight criteria derived from the requirements profile for the HMI were weighted. Finally, the following two favoured HMI approaches result from the utility value analysis: 1. HMD ST optical and 2. HUD ST video (car).

## References

- [1] Azuma, R.: A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments. MIT Press. 6 (4): 355–385.
- [2] Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., Winner, H.: Three decades of driver assistance systems: Review and future perspectives. IEEE Intelligent Transportation Systems Magazine (2014) 6–22
- [3] Schuldt, F., Saust, F., Lichte, B., Maurer, M., Scholz, S.: Effiziente systematische Testgenerierung für Fahrerassistenzsysteme in virtuellen Umgebungen. (2013)
- [4] Kim, B.J., Lee, S.B.: A study on the evaluation method of autonomous emergency vehicle braking for pedestrians test using monocular cameras. Applied Sciences (2020)
- [5] Miquet, C., Schwab, S., Pfeffer, R., Zofka, M., Bar, T., Schamm, T., Zollner, J.: New test method for reproducible real-time tests of ADAS ECU: "vehicle-in-the-loop" connects realworld vehicle with the virtual world. (06 2014)
- [6] Jennifer Brade und Alexander Kögel. Presence in Virtual Reality - der Schlüssel zu Akzeptanz und Übertragbarkeit?!“ In: VAR hoch 2 2019 – Realität erweitern. Tagungsband TU Chemnitz. 2019.
- [7] Jonas Nilsson u. a. Using Augmentation Techniques for Performance Evaluation in Automotive Safety“. In: Handbook of Augmented Reality. 2011.
- [8] Ralf Dörner u. a. Virtual und Augmented Reality (VR/AR). Grundlagen und Methoden der Virtuellen und Augmentierten Realität. Berlin Heidelberg New York: Springer-Verlag, 2019. isbn: 978-3-662-58861-1.
- [9] Product image of Microsoft HoloLens. url: <https://www.microsoft.com/de-de/hololens> (besucht am 27. 02. 2021).
- [10] Product image of Moverio BT-300. url: <https://www.epson.de/products/see-throughmobile-viewer/moverio-bt-300> (besucht am 04. 03. 2021).
- [11] Product image of Samsung Gear VR. url: <https://www.samsung.com/de/mobile-accessories/gear-vr-r324-sm-r324nzaadb/> (besucht am 04. 03. 2021).
- [12] Product image of Hudway drive. url: <https://hudway.co/drive> (besucht am 02. 03. 2021).
- [13] Product image of Moverio BT-300. url: <https://www.epson.de/products/see-throughmobile-viewer/moverio-bt-300> (besucht am 04. 03. 2021).
- [14] Rudolf Fiedler. Controlling von Projekten - Mit konkreten Beispielen aus der Unternehmenspraxis – Alle Aspekte der Projektplanung, Projektsteuerung und Projektkontrolle. Berlin Heidelberg New York: Springer-Verlag, 2016. isbn: 978-3-658-11625-5.
- [15] Paul Milgram u. a. Augmented Reality: A class of displays on the reality-virtuality continuum“. In: SPIE 2351 (Jan. 1994), S. 282–292. doi: 10.1117/12.197321.
- [16] SAE International, (2021) Surface Vehicle Recommended Practice - Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, Online, pp. 30–32.



# How Significant are Business Models for Innovations and Technologies?

Rudolf Schnee  
*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
rudolf.schnee@h-ka.de

## Abstract

A major goal of many companies is to generate as many innovations and new technologies as possible. The aim is to increase competitiveness and ensure economic success. But not every innovation or new technology is a commercial winner. What are the reasons for this and what impact does business models have on this? Based on a short literature research, the terms innovation, technology and business model are defined concisely and the relationship between them are analyzed. And although innovations and technologies are the basis of new products, it is the business model that determines the demands on the development departments and creates the opportunities for successful commercialization.

## Index Terms

business models, business model innovation, innovations, technology

### I. Introduction

Technologies and business models are terms that are used inflationary in recent years [1]. In a constantly and ever more rapidly changing world, they are seen as a necessity and a solution to many emerging challenges facing humanity where the areas of usage are not fixed. Whether it is urbanization, the threat to the planet from pollution, changes in mobility or the silver society [2]. Innovations in particular are hyped and become a mantra for many companies. There are already metrics for innovation by which the

performance, success and position of a company is expressed and therefore a desired target in the strategies [3]. New innovation methods, like design thinking, are intended to break down existing and sometimes inflexible organizational structures in companies and get the workforce excited about innovations. This often leads to overheated and excessive expectations among the workforce [2]. But are innovations and technologies alone responsible for the success of companies and what role business models play in this? The relationship between innovations, technologies and business models is not clearly described in science and leads to the fact that their significance can be very different, depending on the point of view. In the following analysis, the terms will be defined compactly and their relationship to each other will be discussed in the following chapters.

## II. Invention, Innovation, Technology and Business Models

Inventions, innovations and technologies will be sometimes used in academic literature with the same meaning but there are significant distinctions. Fagerberger [4, p. 3] defines inventions as “the first occurrence of an idea for a new product or process” and for Freeman [5], inventions are “an iterative process initiated by the perception of a new market and/or new service opportunity for a technology-based invention which leads to development, production, and marketing tasks striving for the commercial success of the invention”. When inventions meet the market, it comes to innovations [4], [6], [7] (Fig. 1. Invention becomes an innovation through commercialization Base for successful innovations is additional value for the customers and can therefore create business value and the possibility for the company, to generate profit. It does not matter if innovations create a completely new product or if it is an extension of an existing one, which can be done by regular updates [1], [8, p. 283]. Even the increments must not be reduced to technical features. Changes in marketing, prices or in business models can contribute to the creation of a new product or service [9]–[11].

A special form of innovations are disruptive innovations. The theory, developed by Christensen [12] states that existing and successful innovations are replaced by new ones and displace it completely. Disruptive innovations can be found in new markets where start-ups can enter the market with limited investments. New features or enhancements of existing products satisfy the needs for the customer even if the power or the user-friendliness of the product is less than the original one.

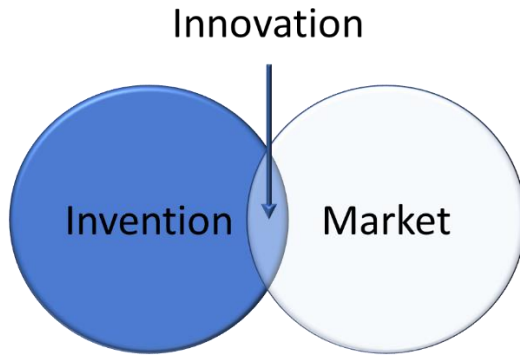


Fig. 1. Invention becomes an innovation through commercialization  
[13]

Technology, which has changed its meaning over the last 200 years, gained its current relevance with the second industrial revolution [14, pp. 190–192], [15]. Based on cause-and-effect statements, technology answer the questions, which resources and tools are necessary to solve a specific problem [16, p. 17], [17, p. 330]. It is “the use of scientific knowledge or processes in business, industry, manufacturing etc.” [18].

Drivers for innovations or new technologies are the R&D departments of companies, universities or technical institutes making improvements to existing products or develop new ones (innovation/technology push). Another reason can be consumer demands, which triggers innovations or new technologies (marketing pull). Furthermore, companies may be motivated by political measures such as laws or subsidies to develop innovations and new technologies and commercialize them on the market (regulatory push) (Figure2).

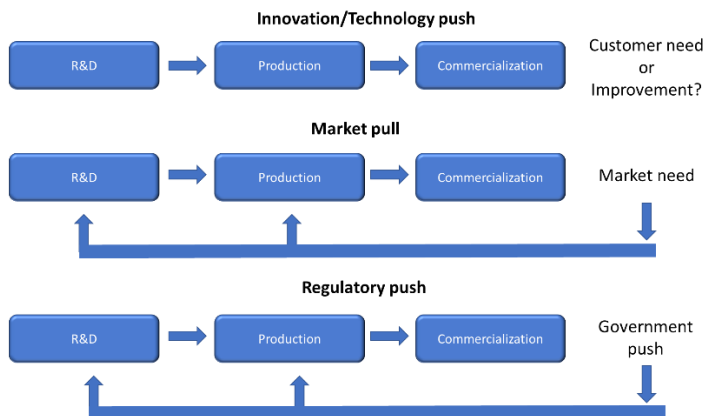


Fig. 2. Types of Motivations for Companies to Invest in Innovation and Technology [13]

The concept of business models was introduced in academic literature in 1957 [19] and publications increased dramatically at the beginning of the 1990's with the e-business hype when companies started to reduce transaction costs with the internet [20]. Nevertheless, science has not yet succeeded in creating a unified accepted definition. For Amit and Zott [20], business models describe the content, structure and control of transactions that are intended to create value by exploiting business opportunities. For Magretta [21], business models answer the questions: who is the customer? And how can we make money in this business? Teece [22] goes beyond the customer relationship and integrates the relationships between customer, suppliers and business partners. Also, business models should be well protected, otherwise they can be copied too easy by competitors. The different approaches have in common that business models consist of the elements value propositions, value creation and delivery and value capture [22] (Figure 3).

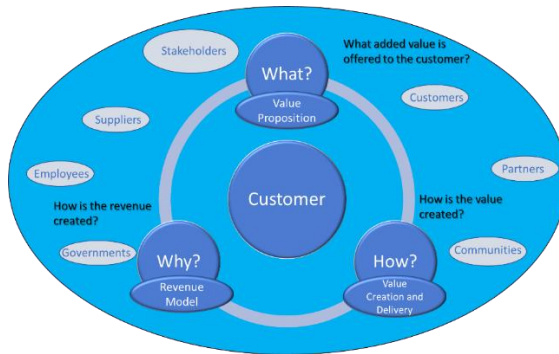


Fig. 3. Core elements of business models [13]

Value proposition means that the service to be provided in the form of products or services represents added value for the customer and that the customer is willing to pay a reasonable price for it and gives the company the opportunity to make a profit. A value is generated when it solves problems for the customer or provides convenience. The value creation and delivery describe the company's sources and capabilities by increasing the value of a product or services to generate new value. Value creation as the third element of business models describes the different forms of money collection from the customer. So, it can be the calculation of the price or an abo model [10], [21], [22].

Business models are subject for permanent adaptation due to external conditions and are further developed through business model innovations. One of the best-known examples of a successful business model innovation is probably the "power-by-the-hour" business model of the British aircraft turbine manufacturer Rolls-Royce. Before the new business model was introduced, building engines was a product for Rolls Royce: For a high one-off sum, the engine became the property of the aircraft manufacturer, who had to have it serviced and maintained. In order to reduce the high level of investment and the complex maintenance system and planning on customer's site, the new business model enables airlines to pay only for the hours the engines are in operation, rather than having to buy them. Rolls-Royce also takes care of the maintenance and repair schedules and expenses. The business model innovation has also the effect that lower capital expenditure enables aircraft companies to enter the market with lower capital resources.

Sustainable business models are becoming more and more crucial, due to the increasingly significant market demand for environmentally friendly

products that are produced under social and fair conditions. Therefore, the elements of environmental friendliness and social fairness are added to the conventional components of business models as equally ranking elements [23].

A business model is not the same as strategy and people sometimes use the terms interchangeably. While strategy is based on analyses and intends to ensure long-term market success, business models focus on short-term horizons and are based on assumptions and limited information. However, it is important that strategies and business models are aligned and do not pursue conflicting objectives.

### III. Relationship between Innovations, Technologies and Business Models

The relationship between innovations, technologies and business models is complex and not clearly defined in academic literature. Authors agree on the point that business models are necessary for value creation and delivery [24], [25]. Furthermore, they enable technologies to be monetized if they represent added value for the customer and generate demand on the market [26, p. 64]. The role of business models is to create the link between innovations and technologies and the realization of economic value [10]. Some economists are convinced to consider business models as the essential vector in the innovation process, since not every innovation or technology represents an economic success per se and they argue that companies should invest more in developing business models than in innovation or technology itself [27]. For example, at Xerox PARC, Palo Alto, many pioneering inventions were developed in the 1970s (the ethernet network technology, the graphical user interface with mouse navigation or the laser printer), but the successful commercialization was implemented by other companies [27, pp. 1–21]. Other scholars give greater emphasis to innovation or technology and state that the technology leaders are successful companies in their segments and gain significant advantages in the marketplace over their competitors. The first movers are characterized by modified and often innovative production processes over competitors within a market or market segment. Examples here are Apple and Google. Also, technology leaders are sometimes innovation leaders, but this is not a mandatory prerequisite and can even be a handicap since innovations and their marketing require different expertise.

Discussing the question if there is a sequence of innovations, technologies and business models, the innovation process has an important role to play. In a linear innovation process, innovations are developed to a product by the development department and marketing and sales are not involved until

shortly before the product is launched. Therefore, the benefits for the customer and the added value are already predominantly defined and there is only little room for improvements in the revenue model. This approach, which was very dominant in 1960 - 1990, can hardly consider of fast-changing customer needs and value chains and restricts the design of business models too much. For this reason, newer approaches pursue the early integration of customer requirements into the innovation process in order to avoid useless investments and cost-intensive product adaptations [26].

Whereas technological changes should not be brought to market until they are in a mature state in order to meet the quality demands of customers, in the case of business models it is possible and useful to make changes at short notice that provide a certain amount of testing on the market. Business model innovations can be started as a hypothesis, tested live at the market and adjusted at short-term if necessary. Other than innovations and technologies, long-term developments and preliminary considerations are not typical for business models and can also lead to problems when launching the business model, as customer requirements may have changed by the time the innovation or technology has reached the necessary market readiness [10].

#### IV. Discussion

Innovations, technologies and business models are closely related. In earlier years, innovations and technologies from R&D departments were the basis for the economic success or failure of companies. It was up to marketing and sales to generate the necessary revenues from the developed products on the market. Nevertheless, high investments in developing innovations and the resulting commercial success were difficult to estimate. There were many reasons for this: The products did not sufficiently satisfy the market's needs or the products were ahead of their time and the necessary infrastructure was not yet in place. In some cases, the companies had a high level of competence in developing products, but the marketing and sales capabilities were not sufficiently available for successful commercialization or the structures in the companies did not allow the marketing department to influence product development at an early stage. As a result of the e-business hype, the knowledge and relevance of business models received a considerable push and replaced innovations and technologies as the exclusive element for a commercially successful company. Companies and start-ups began to focus on value chains and the underlying business processes, which were automated by IT support.

Although the e-business itself lost momentum a short time later, business models became a main pillar in the development of products. In business models, the customer value represents one of the three elementary building blocks for a successful commercialization of products. It is the basis for the requirements for innovations and new technologies. It can significantly reduce the risk of misguided investments, since only functionalities are developed that potentially bring benefits to customers. The value chain as the second element of business models can influence the development process of companies. So, specifications for the value chain can be implemented in the design of production and can lead to reduced costs in production and distribution through efficient processes and faster logistics. The revenue model based on this is given greater freedom in its design on the market to collect money from the customer.

Different competencies are required for the development of products. The development of innovations and their implementation in cost-effective technologies requires the knowledge and experience of technically skilled people, while the development of a suitable business model requires knowledge of customer value proposition, value chains and pricing. Since all competencies are necessary for the successful commercialization of products, only the level of importance can be discussed and not the elimination of one of them. Innovations and technologies represent potentials for commercialization that can be examined and evaluated via the business model and also verify the impact they have on the company. If a customer demand is not recognizable on the market, it can be tried to stimulate it via marketing measures, but the success is not guaranteed. Depending if demand can be raised, expenses for the product are set against the price that can be achieved on the market, which must cover not only the operating costs, but also the expenses for development. If the marginal benefit for the customer exists, but is too low compared to existing products, it can be difficult to achieve the price required to cover all costs.

Innovations and technologies can have a fundamental impact on the existing product range and the organization of the company. New or modified products can lead to the elimination or cannibalization of present service offerings of the company. So, digitization can simplify maintenance work through updates via networks and technicians on site are no longer needed. As a result, employees are no longer required for these aftersales activities and the corresponding revenues are lost. These effects can be compensated for by a modified business model and can lead to new opportunities for the company.

Technology leadership can bring companies in a position to gain significant advantages over the competition in the market. However, companies must



constantly decide whether further development is necessary. The mobile phone manufacturer Nokia, for example, has permanently improved the performance of its batteries, but the product itself has been replaced by a completely new concept, the smartphone. Therefore, before investing in innovations and technologies, the necessities should be analyzed via the business model. And only if a corresponding value proposition is generated for the customer and the value chains can be expanded, these should be made.

It has also been considered that innovations and technologies can lead to a significant change in existing business models, like the value chain. However, if the changes are too fundamental, it can hamper the success of the innovation or technology. For example, the electric vehicle can hardly be successfully commercialized as a technology on its own, as important infrastructure measures are essential for this. These include nationwide charging stations and electricity providers that supply the charging stations with environmentally generated electricity. Integrating these components into the automotive companies' own value chain would be too cost intensive and time-consuming. Cooperations can enable and accelerate the build-up but are not sufficiently appropriate as a motivation for companies on their own, as high investments are required and success is unknown. The solution could be the development of an eco-system. There, the different companies work together and only through this joint eco-system, the success is achievable. This goes far beyond cooperations, as individual companies do not distinguish themselves at the expense of the other eco-system participants because otherwise, the entire eco-system and the success could be exposed. But to build a new eco-system is time consuming and in addition, further companies are to be motivated to invest in this eco-system where the result is uncertain.

The creation and selection of business models has a decisive influence on whether innovations and technologies are allowed to leave the boundaries of the company, as these are not successful per se. Depending on the evaluation by the suitable business model, commercialization in the market takes place. If no appropriate business model can be developed, there is a high risk that innovations and technologies will represent progress, but that this progress will be withheld from humanity because they cannot be commercialized.

## References

- [1] K. Fichter, "Grundlagen des Innovationsmanagements," Oldenburg, 2011.
- [2] Zukunftsinstitut GmbH, "Die Megatrends," *Zukunftsinstitut GmbH*, 2021.  
[Online]. Available:  
<https://www.zukunftsinstitut.de/dossier/megatrends/#12-megatrends>.  
[Accessed: 28-Dec-2021].
- [3] M. Schönfelder, D. Pathmaperuma, U. Reiner, W. Fichtner, H. Schmeck, and T. Leibfried, "Elektromobilität," *uwf UmweltWirtschaftsForum*, vol. 17, no. 4, pp. 373–380, 2009.
- [4] J. Fagerberg, "Innovation: A Guide to the Literature," Oslo, 20031012, 2003.
- [5] C. Freeman, "The nature of innovation and the evolution of the productive system," *Technol. Product. Chall. Econ. policy*, pp. 303–314, 1991.
- [6] N. Rosenbusch, J. Brinckmann, and A. Bausch, "Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs," *J. Bus. Ventur.*, 2011.
- [7] J. Schumpeter, *Business Cycles: A Theoretical, Historical, And Statistical Analysis of the Capitalist Process*. New York: Martino Fine Books, 1939.
- [8] S. Kline and N. Rosenberg, "An Overview on Innovation," in *The Positive Sum Strateg*, R. Landau and N. Rosenberg, Eds. Washington: Natioanl Academies Press, 1986, pp. 275–305.
- [9] O. Gassmann, E. Enkel, and H. Chesbrough, "The future of open innovation," *R and D Management*, vol. 40, no. 3, pp. 213–221, 2010.
- [10] H. Chesbrough and R. S. Rosenbloom, "The role of the business model in capturing value from innovation: Evidence from Xerox Corporation's technology spin-off companies," *Ind. Corp. Chang.*, vol. 11, no. 3, pp. 529–555, 2002.
- [11] D. J. Teece, "Reflections on 'Profiting from Innovation,'" *Res. Policy*, vol. 35, no. 8 SPEC. ISS., pp. 1131–1146, 2006.
- [12] C. M. Christensen, *The Innovator's Dilemma*. Cambridge: Harvard Business School Press, 1997.
- [13] R. Schnee, "Towards a Conceptual Framework of Sustainable Business Models for Electromobility," University of Borugogne Franche-Comte, 2021.
- [14] J. A. Stratton and L. H. Mannix, *Mind and Hand*. Cambridge, massachusetts: MIT Press, 2005.
- [15] E. Schatzberg, "Technik Comes to America: Changing Meanings of Technology before 1930," *Technol. Cult.*, vol. 47, no. 3, pp. 486–512, 2006.
- [16] T. J. Gerpott, *Strategisches Technologie- und Innovationsmanagement: Eine konzentrierte Einführung*. Stuttgart : Schäffer-Poeschel, 2005.
- [17] D. Specht and M. G. Möhrle, *Gabler Lexikon Technologie Management*. Wiesbaden: Gabler Verlag, 2002.
- [18] C. Cambridge Dictionary, "Cambridge Dictionary," *Relative Clause*, 2016.  
[Online]. Available:  
<https://dictionary.cambridge.org/de/worterbuch/englisch/technology>.
- [19] R. Bellman, C. E. Clark, D. G. Malcolm, C. J. Craft, and F. M. Ricciardi, "On the

- Construction of a Multi-Stage, Multi-Person Business Game,” *Oper. Res.*, vol. 5, no. 4, pp. 469–503, 1957.
- [20] R. Amit and C. Zott, “Value creation in e-business,” *Strateg. Manag. J.*, 2001.
  - [21] J. Magretta, “Why business models matter,” *Harvard Business Review*. pp. 86–92, 2002.
  - [22] J. Richardson, “The business model: an integrative framework for strategy execution,” *Strateg. Chang.*, vol. 17, no. 5–6, pp. 133–144, 2008.
  - [23] J. Elkington, “Partnerships from cannibals with forks: The triple bottom line of 21st-century business,” *Environ. Qual. Manag.*, vol. 8, no. 1, pp. 37–51, 1998.
  - [24] C. Baden-Fuller and S. Haefliger, “Business Models and Technological Innovation,” *Long Range Plann.*, vol. 46, no. 6, pp. 419–426, 2013.
  - [25] N. Kroichvili, K. Cabaret, and F. Picard, “New insights into innovation: the business model approach and Chesbrough’s seminal contribution to open innovation,” *J. Innov. Econ. Manag.*, vol. 15, no. 3, pp. 79–99, 2014.
  - [26] H. Chesbrough, *Open Innovation The new imperative for creating and profiting from technology*. Boston: Harvard Business school Publishing Corporation, 2003.
  - [27] J. Björkdahl, “Technology cross-fertilization and the business model: The case of integrating ICTs in mechanical engineering products,” *Res. Policy*, vol. 38, no. 9, pp. 1468–1477, 2009.

# IEEM-CMCNN model accuracy on different public datasets

1<sup>st</sup> Tanju Gofran

*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
Tanju\_Binte.Gofran@h-ka.de

2<sup>nd</sup> Maurice Kettner

*Institute of Energy Efficient Mobility*  
*University of Applied Sciences*  
Karlsruhe, Germany  
Maurice.Kettner@h-ka.de

## Abstract

In this work a deep convolutional neural network with large convolution filters is presented for condition monitoring of electric motor. The model is named as IEEM-CMCNN which is trained with newly created Vibration Dataset- IEEM CMDData to classify different bearing faults in the motor. One of the main significances of the model is that the network has 3-channels vibration input and trained for both single (1 - channel) and 3 - channel inputs so that the trained model can be tested on different test-bench dataset. The trained IEEM-CMCNN is then tested with different publicly available datasets (IMS, KAT and MFPT) where the accuracies are high.

## Index Terms

Convolutional Neural networks (CNNs), Feature Extraction, Test Data

## I. Introduction

In the ongoing fourth industrial revolution (industry 4.0) manufacturers and industries introducing large scale of machine to machine (M2M) and internet of things (IoT) for increasing automation, improving self-monitoring and smart production that can analyze and diagnose issues without human intervention. Machine learning (ML) based Condition monitoring (CM) of machineries has gained huge interest in both industrial applications and academic research because of huge improvement of the algorithms, faster computation power and involvement of numerous number of sensors with easy and fast data transfer process. Apart from requirement of automated fault diagnosis, fast, accurate and early diagnosis of rotating machines is essential for safely, uninterrupted production process and lower maintenance cost. Vibration based CM is very effective as it contains most of the information of the machine but the success of traditional ML depends on how good is the extracted characteristics or features from the measured vibration containing vibration of other rotating parts as well as background noise. The characteristics from the raw data or feature extraction requires the knowledge of the particular system and thus knowledge based feature extraction is domain dependent. Convolutional Neural networks (CNNs) have recently

become very popular in other field of sciences and engineerings other than in only computation vision problems. CNNs generally have stacks of convoulution layers which can extract features directly form data without prior knowlegde of the input's system and later fully connected nets do the classification. In this work we showed a CNNs model to classify or detect apperantly early stages of bearing fatigue faults from motor external vibration and the trained model is tested on different public dataset to validate model's accuracy. Recently there are many publications [6]; [14]; [7]; [12]; [13] showe CNNs based CM of bearings, which are very promising indeed but most of the time the trained models are tested with part of the data from same training sets. The goal of any data-driven classification or detection model is to achieve a model which should predict the result with good accuracy for any similar data as training data. In mechanical engineering problems it is difficult or not always possible because of variation in system domains. It is well known that if similar bearings have same fault, it should be detectable in the vibration pattern and theoritically it is pissible to determine the fault type by frequency analysis. So the objective of designing a fault detecting CNNs model should be that the deep layers are able to extract the features of similar faults from different test bench data than the trained data.

## II. Training Data: IEEM CMDData

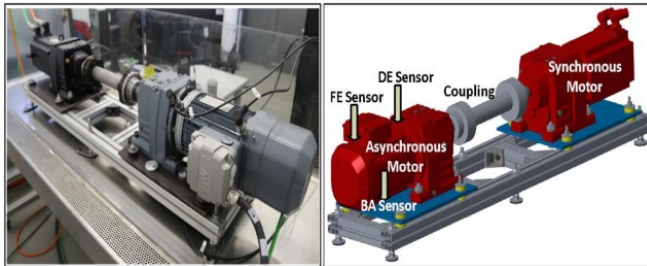


Fig. 1. IEEM-CMDData test bench

The challenge of creating a dataset for bearing faults is to seed the bearings with defctcs similar to real-life and achieving such faults in a test rig is complex and time consuming process. The mounted bearings on the motor of the test bench are deep groove ball bearing and generally it requires millions of cycles in normal conditions to attain degraded health. The idea applied here is to implement artificial fault to mimic fatigue type of defects in the bearing for creating the dataset. The bearing data generating test-bench is developed at the Institute of Energy Efficient Mobility (IEEM),

University of Applied Sciences Karlsruhe and supported by SEW-Eurodrive GmbH (SEW). In the Fig. 1 a view of the test-bench (left) and the CAD design (left) is shown. The dataset is called IEEM-CMData.

During measurements faulty bearings are installed either at Fan-End (FE) or at the Drive end (DE) of the asynchronous motor connected with a synchronous load motor coupled with highly flexible coupling. Three acceleration sensors are installed on motor cage near the FE, DE and under the base plate (BA) to measure the external vibrations. The measurements are recorded at different speed and load conditions.

The goal was to create number of different type of faults at different parts of the bearing such as at inner raceway, outer raceway, rolling element etc. Two methods are applied to implement artificial faults on the bearings:

#### **Create single spall at inner and outer ring to imitate surface fatigue:**

With an electric engraver small sized single dent are created either on inner ring surface or outer ring surface. To achieve different stages of fault, we used two sizes of graving tool and thus two sizes of spalls are made. Afterward for achieving clear or significant amount of vibration due to the fault, one third of the recommended lubrication is applied. Note that the implemented fault is very small and can be compared with very initial fatigue, so the vibration is minimal.

#### **Create roughness in the raceways to imitate rolling surface fatigue:**

In this method rough surface at both inner raceway and outer race was created to mimic rolling surface spall due to surface-initiated fatigue which can happen because of inadequate lubrication. A simple tool setup is made with to create the roughness – by putting some sand bearing outer ring was fixed in a wooden housing, the inner ring is rotated with help of a drill machine. Similarly, one third of the recommended lubrication is used afterwards.

IEEM-CMData contains the following data classes:

- Healthy: Data for both FE and DE bearings having No Fault
- Fault1: Data for both bearings having Inner ring spall (IRSpall) of 2mm
- Fault2: Data for both bearings having Inner ring spall (IRSpall) of 3.5mm
- Fault3: Data for both bearings having Outer ring spall (ORSpall) of 2mm

- Fault4: Data for both bearings having Outer ring spall (ORSpall) of 3.5mm
- Fault5: Data for both bearings having Rough rolling surface (RRSurface)

In this work we considered to classify types faults indifferent in fault progression; therefore the faults of both 2mm and 3.5 mm sizes are considered as same class. Finally the model is designed to classify four conditions of the monitoring bearing: Healthy, IRSpall, ORSpall and RRSurface.

### III. IEEM-CMCNN Architecture for Bearing Fault Classification

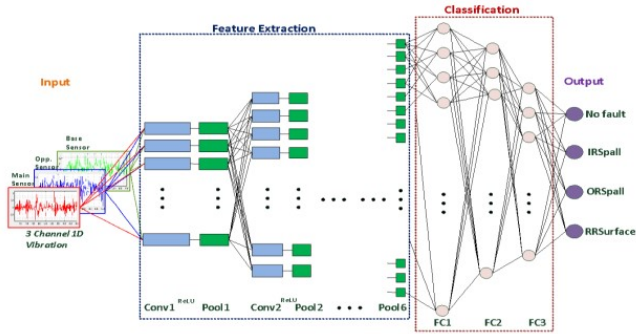


Fig. 2. IEEM-CMCNN architecture for four classes

The model is named as IEEM-CMCNN after the dataset and previously we modelled similar Condition monitoring Convolutional neural network (CM-CNN) [5] with CWRU dataset [1]. IEEM-CMCNN has input of three channels 1D data, six convolution layers, three Fully-connected layers and four output classes. The detail architecture of the IEEM-CMCNN is described in the Fig. 2.

The input is a three-channel 1D vibration data considering three sensors at three positions. The first channel contains the main-sensor data, second channel belongs to the opposite-sensor data and third channel for the base-sensor data. Main-sensor for the FE bearing is the sensor at FE and sensor at DE is the opposite-sensor; for DE bearing this is reversed accordingly. During training the input of IEEMCMCNN is a fixed-size:  $1 \times 1000 \times 3$  vibration data. The one dimensional vibration input length is considered as approximately one revolution of the motor shaft as described in previous paper [5]. No pre-processing is done on the training dataset.

The three-sensor vibration input is then passed through stack of six convolution (Conv) layers. At the first layer the filters have very large receptive field i.e.  $1 \times 50$  and gradually reduced sized filters towards higher level thus at final layer filter size become  $1 \times 2$ . The convolution stride is fixed to  $1 \times 1$  and padding varies from lower layers i.e.  $24 \times 24$  to higher layers i.e.  $2 \times 2$  (where the stride and padding size is to capture left/right center). The last Conv Layer padding is  $0 \times 0$ . The convolution stride and padding in layers are calculated in way to preserve the most of length of the input of each layer. After first Conv layer one batch normalization layer is kept. Each convolution layers are followed Rectified Linear unit Layer (ReLU) to remove the negative value, those followed by Max-Pooling layers (Pool) of window size  $1 \times 2$  with stride 2 and zero-padding.

The stack of Conv layers is then followed by three fully connected (FC) layers: first FC layer has 1024 channels with a ReLU layer, second FC layer has 1000 channels also with one ReLU layer and third has same number of channels as number of class. The final layer is soft-max layer. We compared different architecture of different number of filters after analyzing the filter activities at each Conv layer: in this work the developed architecture has similar number of filters as VGG16 [11]. The training was stopped when accuracy is not improving after 3 epochs. The IEEM-CMCNN design and how the feature is extracted in the deep layers are published in a previous work [4].

#### IV. Test Data: Different Public Dataset

Three different publicly available bearing fault datasets are tested with the trained model to validate the accuracy. Brief overview about the datasets used as test data are described in the following subsections and comparison among the datasets are shown at Table 1.

##### A. KAt Dataset:

The dataset is created by the chair of Design and Drive Technology (Konstruktions und Antriebstechnik- KAt), Paderborn University for condition monitoring of bearing fault [2]. The test-rig is designed to generate a Benchmark dataset for condition monitoring of bearing damage from motor current signal [9]. Two measuring signals are recorded in this testbench: the stator current of the drive motor and the vibration of the bearing housing. The authors found that the classification accuracy with current signal is lower than vibration signal [9]. Three types of artificial damage and some real-life damage data are created in this test setup. In this work we used



the artificial damage data to test model accuracy.

#### B. IMS Dataset:

For developing prognosis algorithms several universities and companies provided different dataset at the Dataset repository hosted by NASA where the IMS bearing dataset is provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati [8]. Four Rexnord ZA-2115 double row bearings were installed on a shaft which is rotated at constant speed of 2000rpm by an AC motor coupled to the shaft via rub belts. To accelerate damage, radial load of 6000 lbs is applied onto the shaft and bearing by a spring mechanism and failures occurred after more than 100 million revolutions [10].

#### C. MFPT Dataset:

The bearing fault dataset is provided by the Society for Machinery Failure Prevention Technology (MFPT) has baseline, inner-ring fault , outer ring fault data at different load condition at constant shaft speed [3]. The information of defect seeding process is not provided.

Differences among the datasets are compared in the following table:

TABLE I  
Comparison of different test-dataset with training dataset

Comparison parameters	IEEM Dataset	IMS Dataset	KAt Dataset	MFPT Dataset
<b>Bearing Type</b>	SKF 6304	Rexnord ZA 2115	SKF 6203	NICE
<b>Inner diameter</b>	20mm	40mm	17mm	19mm
<b>Sample rate</b>	12.8 kS/s	20 kS/s	64 kS/s	97.65 kS/s
<b>Shaft speed</b>	800 - 1440rpm	2000rpm	900 - 1500rpm	1500rpm
<b>Load</b>	5 - 30Nm	n/a	0.1 - 0.7Nm	n/a
<b>Radial force</b>	No	26.7kN	400 - 1kN	0 - 1.3kN

#### V. Data preparation for test the model

For testing any data for a trained model, the first step is to change the size of the data which can be pass through the trained model which is fixed for a certain input size. The test data are one dimensional (1D) data because of having one sensor monitoring the faults and the IEEM-CMCNN input is has three dimensions (3D) and also the test data are measured in different sample rate then IEEM-CMData. Therefore, data preparation require two

steps, first is to convert the sample rate of the test data and second step is to convert the 1D data to 3D input.

Sample rate conversion: As we mentioned in IEEM-CMCNN architecture the input should contain approximately one revolution of the bearing, for that simple sample rate conversion might lose the required signature of the fault pattern. In this work we applied two ways: 1) Method-1(M1): change the sample rate of the test data to same as training data and Method-2(M2): Change the sample rate of the test data according to shaft speed so the input has approximately one revolution (Eqn. 1 and Eqn. 2).

For given speed  $\omega_{test}$  and sample rate  $F_{stest}$  of test data for one revolution the test data length is:

$$l_{test} = round(F_{stest}/\omega_{test}) \quad (1)$$

And for given length of training input  $l_{train}$ , the new sample rate can be calculated as:

$$F_{stestnew} = \frac{(F_{stest} * l_{train})}{l_{test}} \quad (2)$$

The test dataset did not have three measuring sensors, so the converted 3D test input has the given sensor data at the first channel and other two channels have zero values.

## VI. Results: Test IEEM-CMCNN with Public Dataset

Three different publicly available bearing fault datasets are tested with the trained model to validate the accuracy. Brief overview about the test results are described in the following subsections and comparison among the datasets are shown at Table 1.

### A. Test KAt Dataset:

KAt dataset has Inner-ring Fault (IR) and Outer-ring (OR) with two fault seeding process: electric discharge machining(EMD) and electric engraver(EE). Accuracy showed IR fault detection is not satisfactory, but OR faults detection accuracy is very good. The result is shown in Fig. 3 where the x-axis belongs to 10 test inputs where 6 different color bars are different faults with two sample rate conversion methods (M1 and M2).

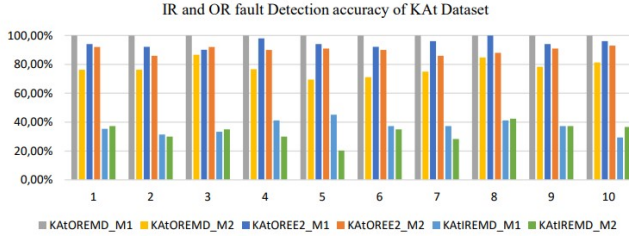


Fig. 3. KAt fault detection accuracy

### B. Test IMS Dataset:

IMS dataset is interesting because the test setup is completely different the training data and the rotating speed, load and bearing size are much larger too. In IMS test-bench four bearings were measured together. From the provided IMS dataset only the IR Fault of two bearings(IMSIR1 and IMSIR2) were relevant for testing. Similarly two data conversion methods are compared. The result is summarized in Table 2.

TABLE II  
IR Fault detection accuracy for IMS dataset

IMSIR1_M1	IMSIR1_M2	IMSIR2_M1	IMSIR2_M2
76,92 %	88,24 %	0,00 %	55,88 %

### C. Test MFPT Dataset:

MFPT dataset contains 7 load conditions data for IR and OR fault at 97.65 kS/s sample rate and 3 data for OR fault at 270 lbs load recorded at 48.82 kS/s sample rate. We tested the data for both sample rate conversion method (M1 and M2). The Fig. 4 shows in most cases fault detection accuracy is very high where x-axis presents the 7 loads and color bars are for different types of faults.

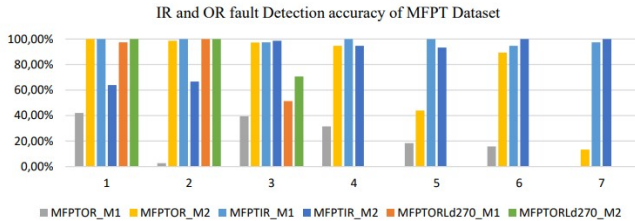


Fig. 4. MFPT fault detection accuracy

## VII. Conclusion

The significant contribution of the IEEM-CMCNN is that it can detect fault classes from other testbenches which certainly indicate the feature extraction at the deep layers is working. In the previous work [4] we found that deep layers are removing the noises from input from lower to higher layers and sharpening the characteristic vibration pattern for certain fault and the final classification accuracy is very high. Another important contribution of the model is that, fault detection is possible from raw data; no pre-processing like noise reduction, frequency conversion or normalization are required. The IEEM-CMCNN design idea also can be adapted for similar problems where multiple sensors are involved. The high accuracy of detecting fault classes from other test benches indicates the IEEM-CMCNN architecture is ideal for robust and accurate fault feature extractions from raw data.

## References

- [1] Case Western Reserve University. Jul. 27, 2021. Online. <https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-universitybearing-data-center-website>
- [2] Christian Lessmeier et al, KAT-DataCenter, Paderborn University. Jan. 10, 2022. Online. <http://mb.uni-paderborn.de/kat/datacenter>
- [3] Eric Bechhoefer, Society For Machinery Failure Prevention Technology. Jan. 10, 2022. Online. <https://www.mfpt.org/fault-data-sets/>
- [4] Gofran, Tanju; Kettner, Maurice; Schramm, Dieter. Condition Monitoring of Electric Motor with Convolutional Neural Network. Symposium on ARTIFICIAL INTELLIGENCE - APPLICATION IN LIFE SCIENCES AND BEYOND. 128–137
- [5] Gofran, Tanju; Neugebauer, Peter; Schramm, Dieter. Feature Extraction from Raw Vibration Signal and Classification of Bearing Faults Using Convolutional Neural Networks (2019). Artificial Intelligence From Research to Application. 15–21. 978-3-9820756-1-7
- [6] Guo, Xiaojie; Chen, Liang; Shen, Changqing (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement. 93. 490–502

- [7] Hoang, Duy-Tang; Kang, Hee-Jun (2019). Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cognitive Systems Research*. 53. 42–50
- [8] J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services. Jan. 10, 2022. Online. <http://ti.arc.nasa.gov/project/prognostic-data-repository>
- [9] Lessmeier, Christian; Kimotho, James Kuria; Zimmer, Detmar; Sextro, Walter. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors (2016). PHM Society European Conference
- [10] Qiu, Hai; Lee, Jay; Lin, Jing; Yu, Gang (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of sound and vibration*. 289. 1066–1090
- [11] Simonyan, Karen; Zisserman, Andrew (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv 1409.1556*
- [12] Xia, Min; Li, Teng; Xu, Lin; Liu, Lizhi; Silva, Clarence W. de (2017). Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks. *IEEE/ASME transactions on mechatronics*. 23. 101–110
- [13] Zhang, Wei; Peng, Gaoliang; Li, Chuanhao. Bearings fault diagnosis based on convolutional neural networks with 2-D representation of vibration signals as input (2017). *MATEC web of conferences*. 13001
- [14] Zhang, Wei; Peng, Gaoliang; Li, Chuanhao; Chen, Yuanhang; Zhang, Zhujun (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*. 17. 425

# A Piston-Bowl Design-Approach for a Gas Engine with Reactive and Partially Dry Exhaust Gas Recirculation

1<sup>st</sup> Youssef Beltaifa

2<sup>nd</sup> Maurice Kettner

*Karlsruhe University of  
Applied Sciences*

Karlsruhe, Germany

youssef.beltaifa@h-ka.de

maurice.kettner@h-ka.de

3<sup>rd</sup> Erik Kärcher

4<sup>th</sup> Marc Reutter

*Karlsruhe University of  
Applied Sciences*

Karlsruhe, Germany

erik.kaercher@h-ka.de

marc-reutter@arcor.de

5<sup>th</sup> Peter Eilts

*Technical University*

*of Braunschweig*

Braunschweig, Germany

p.eilts@tu-braunschweig.de

6<sup>th</sup> Bosse Ruchel

*WJ Power GmbH*

Kiel, Germany

ruchel@wj-power.de

## Abstract

The objective of this paper is to present an innovative piston bowl shape, which should enable a simultaneous reduction of all combustion-associated engine-efficiency losses (wall heat flow, imperfect combustion, and real combustion) for a stoichiometric operated gas engine with cooled, partially dry and reactive exhaust gas recirculation (EGR). The adopted strategy is to avoid efficiency losses associated with unnecessary turbulence-generating design-features, such as large squish areas, and to focus on wall heat losses reduction, taking advantage of the particular EGR properties regarding the burning delay, combustion stability and knock mitigation. Aside with lowering the charge temperature with the cooled EGR, reducing wall heat losses is also attempted through shrinking the surface area for heat transfer. To that purpose, the redesigned piston with a hemispherical bowl shape features a surface area decrease of about 28.8 % when compared to the baseline. To counteract the new piston bowl's negative influence on turbulence level near the firing top dead center (FTDC) in comparison to the baseline, swirl breakdown structures are implemented on the hemispherical piston bowl top. Motored 3D CFD simulations reveal that the swirl breakdown structures can somewhat accomplish the desired turbulence intensification.

## Index Terms

Piston Bowl Design, Gas Engines, Wall Heat Losses, Turbulence, 3D CFD

### I. Introduction

To comply with emissions standards, cogeneration-gas-engine manufacturers are converting their plants from the usually applied lean-burn process to the stoichiometric operation, in order to employ three-way catalyst. However, stoichiometric and undiluted gas engines suffer from diverse efficiency and component-load problems, associated with elevated combustion temperatures and high knock tendency. To cope with engine efficiency targets for the  $\lambda = 1$  combustion process, the mixture dilution with a cooled, reactive and partially dry EGR together with optimized combustion phasing and compression ratio are the followed measures in the scope of this research project. The compression ratio is to be improved in tandem with an upgrade in the piston bowl design to avoid further potential efficiency losses. The focus of this work is to select an appropriate piston bowl shape based on earlier investigations on SI natural gas engines with a flat cylinder head and to adapt it to the engine combustion process under development. To begin, the motivation and the procedure for selecting the new piston bowl shape are explained taking into account the outcomes of previous pertinent research, followed by a comparison between the baseline and new piston properties. Subsequently, an analysis of the impact of the new piston bowl shape on the charge motion inside the combustion chamber was conducted based on motored 3D CFD simulations, allowing optimization measures regarding the turbulence level near the FTDC to be identified. Finally, a summary of the main features of the optimized piston bowl shape and an outlook for future work are presented.

### II. Specifications of the Baseline Cogeneration Gas Engine

The engine used for investigations is a naturally aspirated lean-burn four-cylinder SI gas engine HMG 434 S 132A of the company WJ Power, which is used for cogeneration units of the power classes  $56 \text{ kW}_{\text{el}}$  ( $\lambda = 1$ ) and  $35 \text{ kW}_{\text{el}}$  ( $\lambda = 1.55$ ). The gas engine is derived from a diesel engine. Therefore, the cylinder head is flat, the intake duct exhibits a swirl-generating shape and the piston crown is adapted to the spark ignition (SI) operation, as shown in Figure 1 on the right. In order to accelerate the inflammation process and reduce the cyclic fluctuations of the lean-burn operation, ignition is initiated via a prechamber spark plug. The engine is operated with a compression ratio of 13.3:1 at a four-cylinder displacement of 4.9 liters. The technical data of the series engine used for the combustion process development are summarized in Figure 1 on the left.

Cylinders	4
Ports/Cyl.	2
Ignition	Prechamber
Stroke	134 mm
Bore	108 mm
Displacement	4900 cm <sup>3</sup>
CR	13.3:1
Speed	1500 rpm
Fuel	Natural Gas

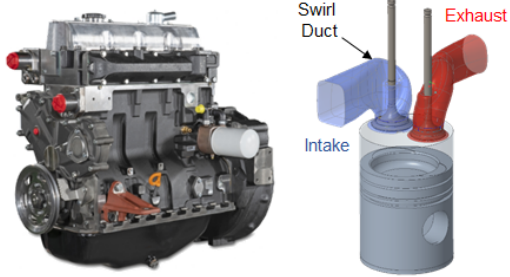


Fig. 1. WJ Power HMG 434 S 132 A: Engine technical data (left) and picture (middle [1]), CAD model of the gas path of one cylinder with the baseline piston (right)

### III. Motivation and Approach

For current state-of-the-art cogeneration gas engines, which are typically operated with the lean-burn mode, the piston bowl shape is designed in such a way to increase the turbulence levels in the combustion chamber to cope with lean-burn challenges as low dilution tolerance, combustion instability and high burning durations [2]. Increasing the turbulent kinetic energy by altering the flow conditions in the combustion chamber leads to acceleration of the turbulent flame propagation, resulting in faster combustion and reduced knock tendency, especially for large bore engines with long flame travel distances [3]. However, many works found in the literature confirm that piston turbulence-generating design-features such as large squish areas decrease the engine efficiency mainly because of the related large total combustion chamber surface area and thus large heat losses [3, 4, 5, 6]. Additionally, they are associated with long top land heights, leading to increased hydrocarbon emissions (HC) and consequently elevated imperfect combustion losses. The improvement of the piston bowl shape in this work is carried out as part of the development of the so-called EGRreact engine combustion process. The aim of this improvement is to promote the thermal efficiency of a stoichiometric operated gas engine, which expectantly features ultra-low emissions, due to the use of a three-way-catalyst. The  $\lambda = 1$  engine-efficiency issues to deal with are primarily the elevated wall heat losses, the limited compression ratio and restricted combustion phasing CA50 (due to knock tendency) and the relatively low isentropic exponent of the stoichiometric mixture in comparison to the lean one. A well-established and effective method for reducing wall heat losses for stoichiometric operated engines is the cooled exhaust gas recirculation (EGR). It reduces the temperature and reactivity of the unburned mixture in the combustion chamber. This lowers the knock tendency, allowing for an increase in compression ratio and an earlier CA50. As a result, the engine



thermal efficiency increases. However, EGR is associated with inhibition of the burning rate and a further lowering of the mixture isentropic exponent. To address these two drawbacks, the reactivity and isentropic exponent of the recycled exhaust gas should be increased. For this purpose, the dedicated EGR concept with in-cylinder thermal reforming and an appropriate EGR path, including a condensation heat exchanger, is proposed, which is depicted schematically in Figure 2.

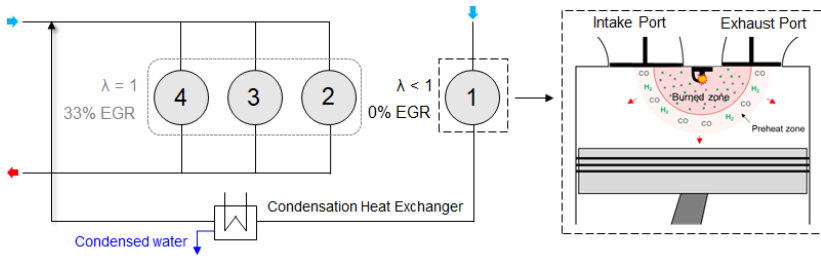


Fig. 2. Schematic description of the EGRreact working process (left) and an illustrative description of the in-cylinder thermal reforming in the dedicated cylinder (right)

The dedicated cylinder engine concept was first proposed by Southwest Research Institute in 2009. The fuel excess in the dedicated cylinder (cylinder 1 in Figure 2) does not only dilute the mixture and causes reduced wall heat losses; it is also reformed to  $H_2$  and CO during partial fuel oxidation, resulting in a high chemical energy, or reactivity, of the dedicated cylinder exhaust gas. Therefore, the dilution tolerance in the stoichiometric operated cylinders (cylinders 2 to 4 in Figure 2) can be significantly improved due to the increased EGR reactivity. This could be proven by the reference tests carried out so far with external hydrogen supply and exhaust gas recirculation. A further technique to increase the dilution tolerance in  $\lambda = 1$  cylinders is to extract condensed water from the recirculated exhaust gas by cooling it below its dew point. This also increases the ratio of specific heats  $c_p/c_v$  of the diluted stoichiometric mixture, which improves engine thermal efficiency.

Since the EGRreact working process envisages high dilution ratios with a reactive and partially dry EGR, the knock mitigation and dilution tolerance do not need to increase the turbulence intensity in costs of efficiency. Therefore, the focus of this piston-bowl shape improvement is to reduce wall heat losses, which represent especially in the high efficiency regime (early CA50) the first source for engine efficiency losses despite the significant mixture dilution ratios, as shown in Figure 3. Figure 3 depicts measurement results gathered on the HMG 434 S 132A engine test bench, introduced in section II. Figure 3 shows the dependency of the engine indicated efficiency and

losses on the combustion phasing CA50 for a diluted mixture ( $\lambda = 1.55$ ) at a constant engine speed of 1500 rpm. A significant reduction of the real combustion losses (from about 7 %-points to less than 2 %-points) can be reached by shifting the combustion phasings CA50 towards its optimum position, which results in turn in increased wall heat losses (about 7 %-points).

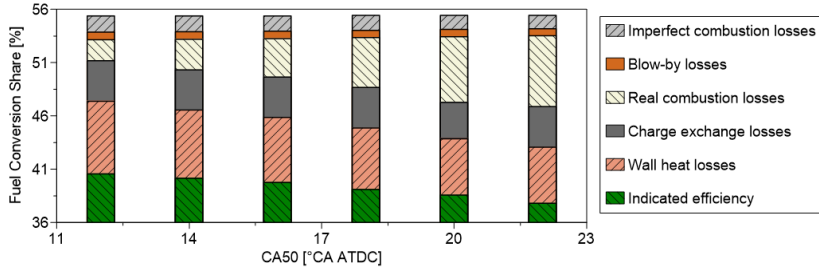


Fig. 3. Dependency of the efficiency and efficiency losses on CA50 for  $\lambda = 1.55$  and  $n = 1500$  rpm, measurement results of the engine HMG 434 S 132A test bench

Wall heat losses are the result of the transient gas-wall heat transfer in internal combustion engines, which is predominantly determined by forced convection. This depends on the charge motion and temperature gradients of the in-cylinder-mixture near the combustion chamber walls [7]. Here, the proportion of radiation is of negligible significance, especially in SI engines, due to the selective gas radiation that prevails there [7]. To describe the convective heat transfer, the Newtonian approach is usually used according to the equation.

$$Q_W(t) = A_{CC} \cdot \alpha_{CW}(t) \cdot [T_M(t) - T_W(t)] \quad (1)$$

with:

$Q_W(t)$  – Wall heat flow [W]

$A_{CC}$  – Total Surface of the combustion chamber walls [ $m^2$ ]

$\alpha_{CW}(t)$  – Heat transfer coefficient [ $W/m^2K$ ]

$T_M(t)$  – Mixture temperature [K]

$T_W(t)$  – Gas-side wall surface temperature [K]

From equation 1, it can be seen that wall heat losses are proportional to the total combustion chamber walls surface  $A_{CC}$ . Hence, the first aim of the piston bowl design is to minimize the surface area of the piston in order to reduce the wall heat losses. Furthermore, the heat transfer is also proportional to the heat transfer coefficient  $\alpha_{CW}(t)$ , which depends on the turbulence level in the combustion chamber. Therefore, the turbulence level

should be reduced. On the other hand, high turbulence levels are needed near FTDC in order to assist the turbulent flame propagation. Thus, a timed enhancement of the turbulence close to FTDC is the second target of this piston bowl design. Thereby, the baseline turbulence level close to FTDC shall be maintained.

In order to reduce hydrocarbon emissions (HC) and thus the imperfect combustion losses, the volume of the combustion chamber crevices must be minimized [8]. To that purpose, the reduction of the top land height and squish area can be effectively advantageous.

A further important design aspect is to keep a short flame travel distance in the combustion chamber, which leads to reducing the burn duration, resulting in reduced real combustion losses. Poulos and Heywood [9] have shown that the combustion chamber geometry has a very strong effect on burn duration. Indeed, the changes in the combustion chamber, that lead to reduce the flame travel distance, tend to increase the thermal efficiency and reduce heat losses. However, added turbulence to achieve a comparable reduction in burn duration results in a rise in heat losses and a consequent drop in thermal efficiency [9]. To summarize, Figure 4 depicts an overview about the piston-bowl design approach and the targets followed within this work.

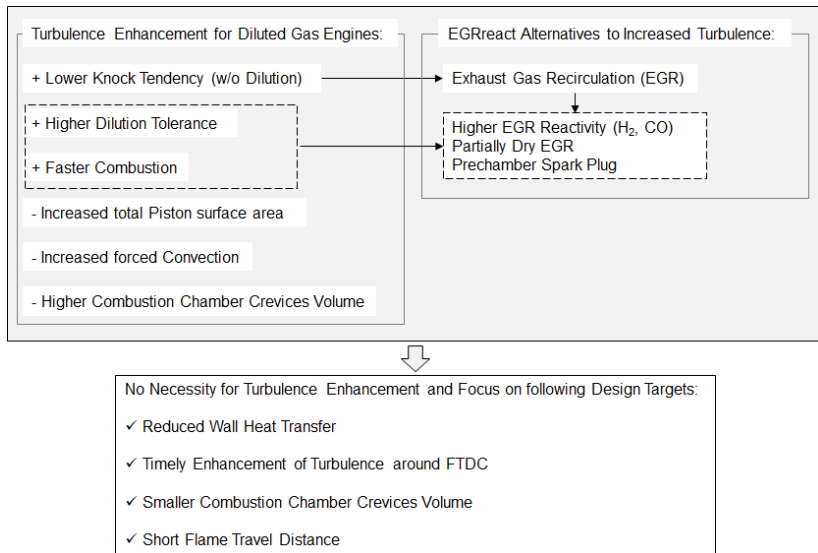


Fig. 4. Overview about the EGRreact Piston-Bowl Design-Approach

#### IV. Description of the Proposed Piston Bowl Design

The modification of the piston bowl shape is carried out while maintaining the geometric compression ratio constant at 13.3:1. The first aim of the new piston design is to reduce wall heat losses. Therefore, it is intended to minimize the surface area of the piston top (bowl and squish areas) and to reduce the in-cylinder turbulence level, when compared to the baseline (with exception for piston positions near the FTDC). To these purposes, the flat piston represents the ideal solution. Olsson and Johansson [5] performed engine measurements at maximum brake torque (MBT) with variations of the dilution ratio, both with excess air ( $\lambda = 1 - 1.8$ ) and with EGR (0 % - 30 % at  $\lambda = 1$ ) for ten different piston geometries including the flat one. The experimental investigations were conducted on a natural gas engine, derived from a diesel one, with a flat cylinder head and a swirling intake duct. The nominal compression ratio for all tested chambers was set to 12:1. The geometric features of their engine (bore and stroke) are very similar to those of the engine considered in this work, which motivates for adopting their findings. Regardless of the operating conditions, the engine with the flat piston exhibited by far the highest thermal efficiency despite the slowest combustion. This is due to its limited wall heat losses, caused by its minimum piston surface area and lowest turbulence intensity in comparison with other configurations. Considering that the new piston will be machined from the same piston blank as the baseline one, a flat piston with the same compression ratio as the baseline (13.3:1) is not attainable due to the minimum top ring land length between the piston top and the compression ring (about 5-6 mm).

In this case, the first alternative is the hemispherical piston bowl shape, which not only still provides the possibility to reduce significantly the piston-wall surface area in comparison with the baseline piston, but also features a better flame travel distance. Olsson and Johansson [5] have shown that the engine with the hemispherical piston is the second most appropriate option regarding the engine thermal efficiency. Indeed, it provides, similar to the flat one, a high thermal efficiency due to the limited piston surface area, moderate turbulence levels and thus low heat losses [5, 6, 10]. The low turbulence level here is due to the small squish area and minimum interference with the swirling flow [10]. Industrial SI gas engines, which are mostly evolved from diesel engines, are well known for their swirling flow, which does not regularly dissipate into turbulence [3]. Considering this fact, the interaction with the swirling flow can be used in order to provoke its dissipation into turbulent vortices and consequently maintain the baseline piston turbulence-level around FTDC. This should take place without incurring noticeable drawbacks in terms of the piston surface area. For this purpose, swirl breakdown structures (slats) with a maximum height of 2.5 mm are implanted on the hemispherical piston top. The baseline as

well as the hemispherical pistons with and without slats can be seen in Figure 5. Here, the hemispherical piston with the slats has a slightly higher compression ratio of approximately 13.46:1 due to the slats volume.

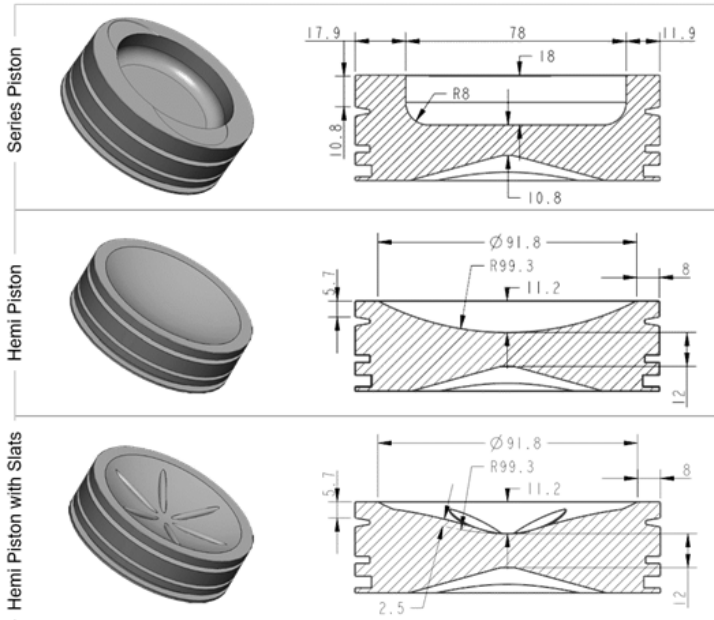


Fig. 5. Piston Bowl Shapes

The hemispherical (short: hemi) piston enables an important reduction of the piston top area of 28.8 % in comparison to the baseline one. This corresponds to a total combustion chamber surface area reduction of 15.72 %. The shape of the hemi piston was determined by maintaining a minimum squish length (8 mm), a minimum top ring land height (5.7 mm) and a constant compression ratio (13.3:1). Consequently, the radius of 99.3 mm results for the hemispherical bowl.

The squish gap height of the hemi pistons is almost four times larger than the one of the baseline. Additionally, the top land height is about 50 % shorter. Both changes are expected to evoke a significant reduction of the HC emissions and consequently of the imperfect combustion losses. A summary of the most important characteristics of the considered pistons can be found in Figure 6.

The swirl breakdown structures have a small height. As a result, they should affect the flow and increase the turbulence level only when the cylinder

Piston Design Parameters	Baseline	Hemi	Hemi with Slats
Geometric Comp. ratio [-]	13.30:1		13.46:1
Piston area [mm <sup>2</sup> ]	16110	11470	11790
Piston clearance [mm]	1.733	6.733	
Squish ratio [-]	0.427	0.273	

Fig. 6. Specifications of the baseline and hemispherical pistons

volume is small, namely when the piston is close to FTDC. To analyze the impact of the hemi pistons on the in-cylinder flow and to compare them to the baseline configuration, motored 3D CFD simulations were carried out using the commercial simulation software AVL FIRE. The simulation results are discussed in the following section.

#### V. Analysis of the Charge Motion using 3D CFD Simulations

For the mesh generation the Fame Engine Plus tool was used. It is a hexa-based automatic meshing tool, which automatically generates engine meshes for different crank angle positions according to the valve lift curves and the piston kinematics. During the movement, the volume mesh quality is constantly improved using smoothing algorithms.

Turbulence modeling is based on the unsteady Reynolds Averaged Navier-Stokes (RANS) equations, using the  $k$ - $\zeta$ - $f$  model. Together with a hybrid wall treatment, boundary layer cells are used to capture the near-wall flow, which allowed for low  $y^+$  values ranging between 1 and 70. The intake mixture was modelled as pure air since no combustion simulation was accounted for. The content of the cylinder was initialized as exhaust gas to be able to evaluate the distribution of the residual gas near the FTDC for different combustion chamber geometries. For a precise estimation of the boundary conditions, a tuned 1D model of the engine under investigation provided pressure and temperature traces over °CA for the boundaries of the computational domain. The simulations start during the charge exchange and end at 60 °CA AFTDC.

The change in combustion chamber geometry is associated with a corresponding alteration in the cylinder charge motion during the whole engine cycle. During the intake stroke, the flow into the cylinder interacts with the combustion chamber walls, resulting in intense turbulence development, which is important for the mixing and homogeneity of the in-cylinder mixture. Figure 7 on the left depicts the development of the mean turbulence kinetic energy in the combustion chamber from motored CFD simulations for the three piston geometries of this study. With the baseline piston, a significantly higher level of turbulence has evolved in the combustion

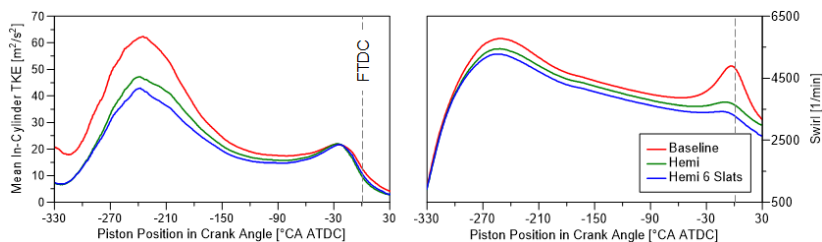


Fig. 7. Development of the mean turbulence kinetic energy (left) and of the swirling velocity (right) in the combustion chamber in dependence of crank angle

chamber since the beginning of the intake stroke. Digging deeper into this effect, Figure 8 illustrates the distribution of the turbulence kinetic energy in the combustion chamber at 330 °CA BTDC for the baseline (left) and Hemi (right) pistons. In the case of the Baseline piston, the deflection of the intake flow within the short squish clearance leads to a significant dissipation of the main flow, resulting in a considerable turbulence intensification in the intake valve area. In contrast, the Hemi piston with its considerably larger squish gap and its curved bowl shape causes an almost undisturbed inflow into the combustion chamber.

Figure 7 on the left demonstrates a difference in turbulence level also between both Hemi pistons during the intake and compression strokes between 290 °CA BTDC and 20 °CA BTDC. Therefore, the progression of the flow fields with both pistons during the intake process was observed. It was found that the differences in turbulence generation become more significant for crank angle positions starting from 265 °CA BTDC. Hence, the distributions of turbulence kinetic energy and flow velocity at this crank angle position are shown in Figure 9 for both pistons.

The differences in turbulence formation shown in Figure 9 are especially noticeable at the level of the plane A-A. There, the reduced swirl velocity in the case of the Hemi Piston with slats (Figure 9 bottom and Figure 7 on the right) especially in the middle of the combustion chamber results in reduced turbulence generation. Thus, less macroscopic kinetic energy has resulted in less microscopic turbulent kinetic energy.

As stated at the outset of this section, the high level of turbulence during the intake stroke is critical for the mixing and uniformity of the in-cylinder mixture. Because the turbulence level with the Hemi pistons is substantially lower than the turbulence level with the baseline one, the impact of the decreased turbulence, associated with the Hemi pistons, on mixture homogeneity, particularly near the FTDC, must be evaluated. Figure 10 depicts the residual gas concentration distribution in the combustion chambers with the baseline and Hemi pistons at a crank angle position of 10 °CA BTDC.

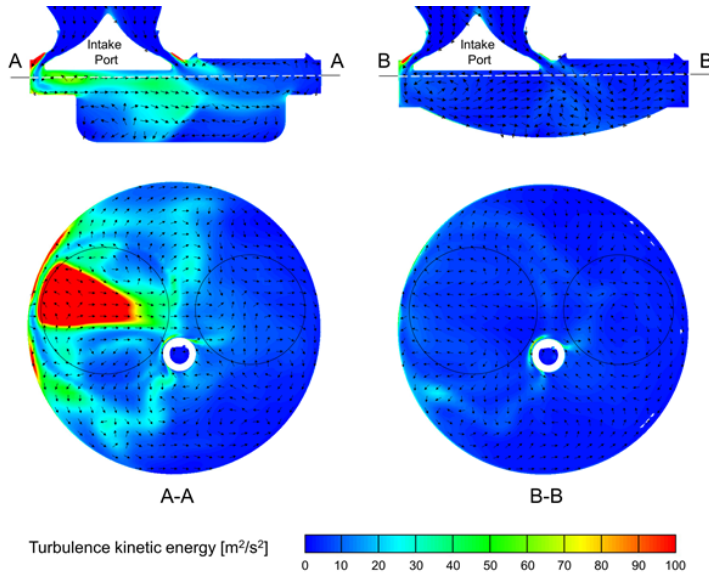


Fig. 8. Distribution of the turbulence kinetic energy at 330 °CA BTDC for the baseline (left) and Hemi (right) pistons, motored simulation results

Figure 10 demonstrates that the residual gas is fairly well dispersed in the main combustion chamber for all three cases, with local concentrations ranging between 2.5 % and 4.5 %. Despite the lower turbulence level, the Hemi pistons still have no drawbacks in terms of residual gas distribution, when compared to the baseline one.

Aside from cylinder charge homogeneity, the flow parameters in the combustion chamber during the flame propagation phase, beginning at a crank angle near to the FTDC, have a substantial influence on the combustion process and thermal efficiency. In this regard, Figure 11 depicts a comparison between the effects of the three simulated combustion chamber geometries on the in-cylinder flow characteristics close to the FTDC. Despite the alteration of piston design, the maximum turbulence kinetic energy and its crank angle position (near 22 °CA BTDC) remained unchanged. Thus, the Hemi piston offers no improvement compared to the baseline one regarding the crank angle position of the turbulence peak, which should best occur shortly after TDC. Afterward, across the whole crank angle range for turbulent flame propagation (approximately between 5 °CA BTDC and 20 °CA ATDC), the Hemi piston produces, as expected, a significantly lower turbulence level (about 28 %) compared to the baseline one.



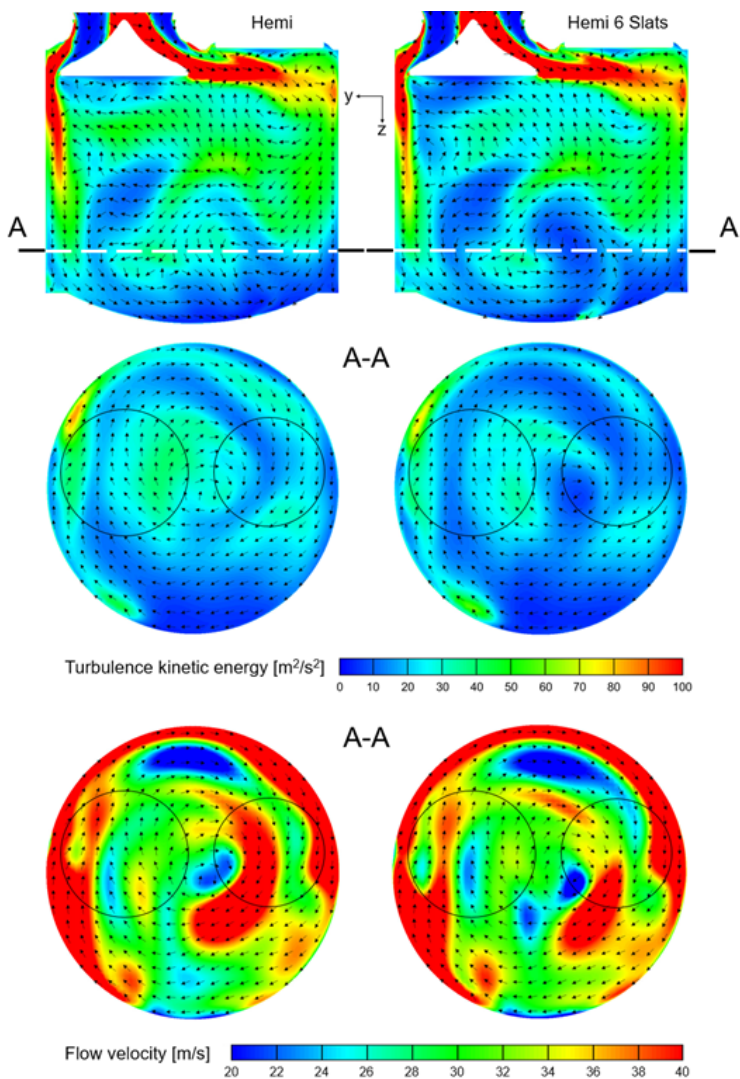


Fig. 9. Distributions of the turbulence kinetic energy (top and middle) and flow velocity (bottom) at 265 °CA BTDC for Hemi (left) and Hemi with Slats (right) pistons

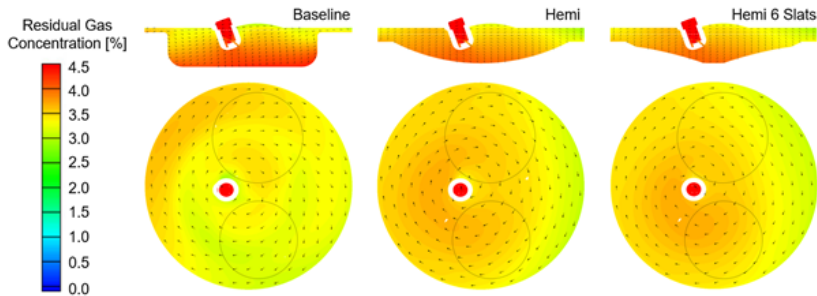


Fig. 10. Distribution of the residual gas concentration at 10 °CA BTDC for the baseline, Hemi, and Hemi-6-Slats pistons, motored simulation results

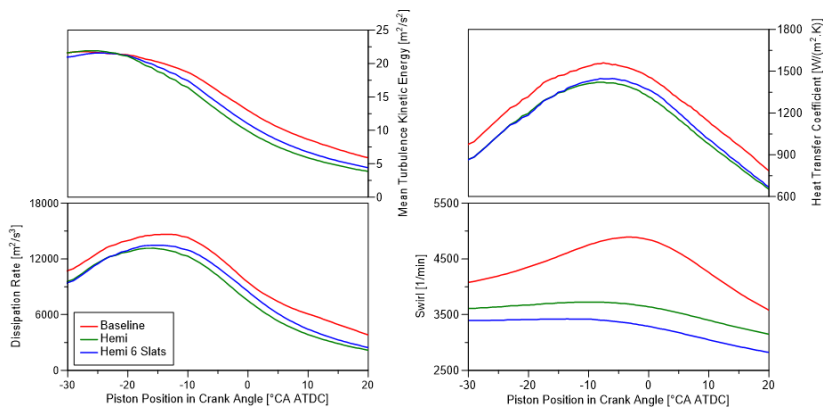


Fig. 11. Development of the mean turbulence kinetic energy (left, top), the turbulent dissipation rate (left, bottom), the heat transfer coefficient (right, top) and of the swirling velocity (right, bottom) in the combustion chamber in dependence of crank angle between 30 °CA BTDC and 20 °CA ATDC

Figure 12 demonstrates that the flow around, into (during compression), and out of (during expansion) the prechamber has a major impact in turbulence generation for all combustion chamber designs. Figure 12 further illustrates that turbulence development in the Hemi piston case is mostly generated by piston movement and flow deflection at the cylinder head. In the case of the baseline piston, the impact of the squish flow is additionally introduced, which causes the greater turbulence level in the case of the baseline piston.

With regard to the relatively weak turbulence generation with the Hemi piston geometry, the slats are an effective measure to intensify the tur-

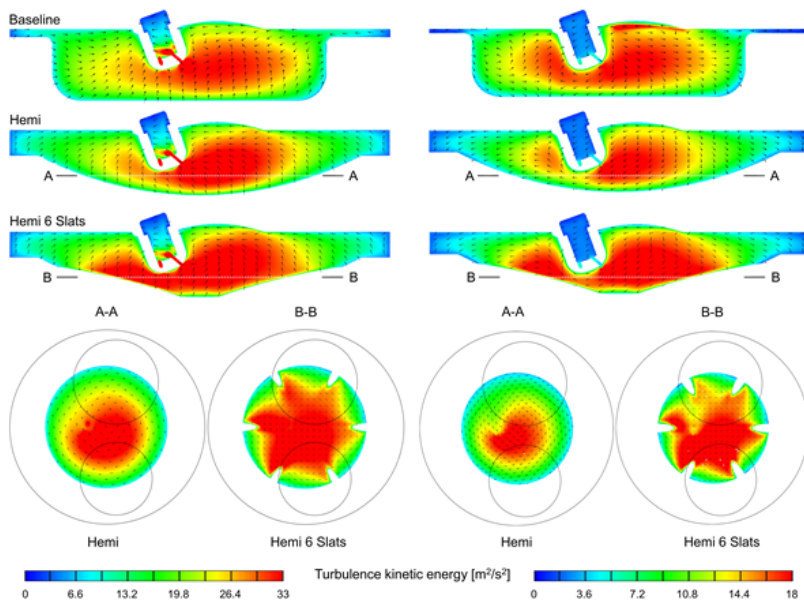


Fig. 12. Distribution of the turbulence kinetic energy at 10 °CA BTDC (left) and 5 °CA ATDC (right) for the baseline, Hemi and Hemi-6-Slats pistons, motored simulation results

bulence during the crank angle range for turbulent flame propagation, as shown in Figure 11 and in Figure 12. In the crank angle range -15...15 °CA ATDC, the impact of the slats on the dissipation rate and hence on turbulence as well as on the heat transfer coefficient becomes obvious. This corresponds to the intended anticipation that the slats benefit in terms of turbulence enhancement can only be realized with a limited combustion chamber volume for piston positions near to the FTDC. At 5 °CA ATDC, the slats cause a turbulence kinetic energy increase of approximately 13 % in comparison with the Hemi piston. Figure 11 depicts how the slats enhance the turbulence in their wake both at the end of compression and at the beginning of expansion strokes. However, the slats impact is inadequate to meet the baseline piston turbulence-level. Further optimization of the number and shape of the slats can assist attain this objective. In the case of the compression ratio increase by simply reducing the squish gap and without any changes in the slats and piston top geometry, an extra increase in turbulence owing to the resulting squish flow might be beneficial. The swirl flow velocity is another significant distinction between the Hemi pistons and the Baseline one. This is in the case of the Baseline piston sig-

nificantly higher due to the substantially smaller diameter of the combustion chamber for piston positions around FTDC. In the case of the Hemi piston with slats, the swirl speed is even lower due to the braking effect of the swirl-breakdown structures. The lowering of swirl velocity, along with the flattening of the piston bowl and the reduction of squish area, can increase the engine thermal efficiency. Adlercreutz et al. [12] shown that combining swirl levels in the mid swirling range with shorter squish lengths and a shallower piston bowl is the best practice in terms of reducing combustion time for a stoichiometric engine running with high amounts of EGR.

## VI. Conclusions and Outlook

Based on earlier research, this study proposes a pragmatic approach to design a piston bowl geometry for a stoichiometrically operated gas engine with reactive and partially dry exhaust gas recirculation. The purpose is to maximize the engine thermal efficiency by reducing wall heat losses and by the timely selective elevation of the turbulence level in the combustion chamber. The characteristics of the new piston bowl shape and its comparison with the baseline one can be summarized as follows:

- The new piston with its hemispherical bowl shape should help decreasing wall heat losses by reducing the piston surface area by about 29 %, the overall combustion chamber surface area by approximately 16 % (at FTDC) and the heat transfer coefficient by approximately 11 % (at FTDC), in comparison with the baseline piston.
- The slats on the top of the hemispherical bowl cause a negligible increase of the total combustion chamber surface area and allow for a significant reduction in turbulence across the entire engine cycle, with the exception of the crank angle range  $-15...15^{\circ}\text{CA ATDC}$ , when a higher turbulence level is required. Thus, by lowering turbulence when it is not essentially required, the slats contribute to lowering wall heat losses. At the same time, they cause a turbulence intensification, exactly when it is needed.
- The decreased turbulence level caused by the hemispherical piston and the slats, particularly during the intake stroke, has no negative effects on the homogeneity of the mixture at the ignition timing.
- The reduced swirl velocity in the combustion chamber with the hemispherical pistons can be advantageous regarding the flame detachment, the combustion duration and thus engine thermal efficiency.
- Considering that the new piston will be machined from the same piston blank as the baseline one, it provides a top land height reduction by approximately 50 %, which should very likely result in a reduction of the unburned hydrocarbon emissions and thus the imperfect combustion losses.

- Increasing the compression in the case of the Hemi-6-Slats piston by solely decreasing the squish gap without any change of the piston top shape can be highly beneficial for the engine thermal efficiency. This is not only due to the higher compression ratio, but also because of the further reduction in total combustion chamber surface area at FTDC (liner surface reduction) and owing to the slats stronger influence on turbulence intensification at FTDC by dint of the narrower squish gap.

In the next steps, numerical investigations with combustion simulation will be carried out to optimize the slats geometry, when the ignition timing and the compression ratio are varied. Based on the results to be obtained, a piston with the optimized compression ratio and slats geometry will then be manufactured and tested on the engine test bench.

### Acknowledgment

The German Federal Ministry of Education and Research supported this project through the grant program “Ingenieur Nachwuchs”. The authors would like to thank all persons involved. The authors also would like to thank the AVL List GmbH for providing the simulation software AVL FIRE within the framework of the University Partnership Program (UPP).

### References

- [1] *WJ Power GmbH: Gas Grundmotor HMG 434 S 132A.*
- [2] Wohlgemuth S., Roesler S., and Wachtmeister G., *Piston Design Optimization for a Two-Cylinder Lean-Burn Natural Gas Engine - 3D-CFD-Simulation and Test Bed Measurements*, SAE Technical Paper 2014-01-1326, 2014, doi:10.4271/2014-01-1326.
- [3] Mahendar S., Giramondi N., Venkataraman V., and Christiansen Erlandsson A., *Numerical Investigation of Increasing Turbulence through Piston Geometries on Knock Reduction in Heavy Duty Spark Ignition Engines*, SAE Technical Paper 2019-01-2302, 2019, <https://doi.org/10.4271/2019-01-2302>.
- [4] Krishnaiah R., Ekambaram P., and Jayapaul P., *Investigations on the effect of Piston Squish Area on Performance and Emission Characteristics of LPG fueled Lean Burn SI Engine*, SAE Technical Paper 2016-28-0123, 2016, doi:10.4271/2016-28-0123.
- [5] Olsson K. and Johansson B., *Combustion Chambers for Natural Gas SI Engines Part 2: Combustion and Emissions*, SAE Technical Paper 950517, 1995, <https://doi.org/10.4271/950517>.
- [6] Judith J., Neher D., Kettner M., Schwarz D. et al., *High Efficiency by Miller Valve Timing and Stoichiometric Combustion for a Naturally Aspirated Single Cylinder Gas Engine*, SAE Int. J. Adv. and Curr. Prac. in Mobility 2(2):1041-1057, 2020.
- [7] A. Wimmer, *Thermodynamik des Verbrennungsmotors*, Technische Universität Graz - Institut für Verbrennungskraftmaschinen und Thermodynamik, Graz, 2018.
- [8] Min K., Cheng W., and Heywood J., *The Effects of Crevices on the Engine-Out Hydrocarbon Emissions in SI Engines*, SAE Technical Paper 940306, 1994, <https://doi.org/10.4271/940306>.
- [9] S. G. Poulos, *The Effect of Combustion Chamber Geometry on S.I. Engine*, Massachusetts Institute of Technology, Massachusetts, 1982.

- [10] Johansson B. and Olsson K., *Combustion Chambers for Natural Gas SI Engines Part I: Fluid Flow and Combustion*, SAE Technical Paper 950469, 1995, <https://doi.org/10.4271/950469>.
- [11] Pischinger R., Klell M. and Sams T., *Thermodynamik der Verbrennungskraftmaschine*, 3. Auflage. In: List, H., *Der Fahrzeugantrieb*, Springer-Verlag, Wien, 2009.
- [12] Adlercreutz L., Cronhjort A., and Stenlaas O., *Variation in Squish Length and Swirl to Reach Higher Levels of EGR in a CNG Engine*, SAE Technical Paper 2019-01-0081, 2019, doi:10.4271/2019-01-0081.