VDI

Automotive and
Traffic Systems Technology

VDI-Berichte 2357

19th International Congress

ELIV 2019

October 16-17, 2019, Bonn

VDI-BERICHTE
Herausgeber:
VDI Wissensforum GmbH

Bibliographische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter www.dnb.de abrufbar.

Bibliographic information published by the Deutsche Nationalbibliothek (German National Library)
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at www.dnb.de.

Der VDI-Bericht, der die Vorträge der Tagung enthält, erscheint als nichtredigierter Manuskriptdruck.
Die einzelnen Beiträge geben die auf persönlichen Erkenntnissen beruhenden Ansichten und Erfahrungen der jeweiligen Vortragenden bzw. Autoren wieder. Printed in Germany.

ISSN 0083-5560
ISBN 978-3-18-092357-4

Content

Foreword 1

ADAS

Seeing With Sound – Next-level 3D ultrasonic sensors based on echolocation 5
N. Knappstein, Toposens, Munich

Ensuring the reliability, availability and safety of fully automated and autonomous transport systems through modern system architectures 11
J. Heinrich, A. Braasch, Institut für Qualitäts- und Zuverlässigkeitsmanagement GmbH, Wuppertal;
F. Plinke, Institut für Qualitäts- und Zuverlässigkeitsmanagement GmbH, Hamburg

ADAS/AD Systems: Efficient Testing & Validation – From data acquisition to data analytics . . . 21
M. Kremer, M. Kreutz, M. Luxen, S. Christiaens, FEV Europe GmbH, Aachen

Problems and solution spaces for driver-initiated handover from automatic to manual driving mode 31
J. Klesing, Nexteer Automotive, Auburn Hills, USA;
S. Safour, Nexteer Automotive, Paris, France

UX

User-centred development of a display concept for fully automated driving – A methodical approach 45
L. Gauer, I. Totzke, Audi Electronics Venture GmbH, Gaimersheim

UX in the Automotive Industry – How to make it comparable? 57
R. Ludwig, P3 automotive GmbH, Stuttgart

Development of the Cockpit-UI/UX of the Taycan in an Agile Way – Less is More 67
L. Krauß, E. Kögler, M. Bayer, S. Wiechmann, M. Worch, M. Mohamad,
Dr. Ing. h. c. F. Porsche AG, Weissach

3D-Displays with Lightfield Technology for a natural look and feel – User experience between attention-guiding and brand emotion 79
K. Hohmann, F. Rabe, C. Menzenbach, Continental Automotive GmbH, Babenhausen

E-Vehicles

Future e-mobility and the change in system requirements – The interplay between battery and thermal management for different mobility concepts 87
L. Schindele, D. Schütz, F. Heber, P. Sailer, G. Le Hen, N. Müller, Robert Bosch GmbH, Stuttgart

Modeling and identification of electrochemical energy storage for drive train development – Review and evaluation 101
P. Gesner, F. Kirschbaum, F. Landenberger, J. Scheiffele, Daimler AG, Stuttgart;
L. Morawietz, B. Bäker, Technische Universität Dresden, Institut für Automobiltechnik (IAD), Dresden

Condition monitoring for failure monitoring of power electronic assemblies 115
S. Wagner, F. Wüst, S. Trampert, F. Sehr, A. Middendorf, O. Wittler, Fraunhofer Institut für Zuverlässigkeit und Mikrointegration (IZM), Berlin;
M. Schneider-Ramelow, Technische Universität Berlin

Holistic Energy Management of 48V Mild Hybrid Vehicles 127
P. Griefnow, J. Andert, M. Engels, RWTH Aachen University, Aachen;
J. Richenhagen, D. Jolovic, FEV Europe GmbH, Aachen

Easy Integration of 48V Mild Hybridization by Dual Voltage Battery Management – Realizing CO_2 saving potentials by low implementation efforts. 139
B. Fähnrich, A. Körner, HELLA GmbH & Co. KGaA, Lippstadt

48 Volt High Power: Electric Drive for Excellent CO_2 Emissions & Electric Driving Features . . . 153
F. Graf, S. Baensch, T. Knorr, D. Ellmer, C. Marechal, Division Powertrain, Continental Automotive GmbH, Regensburg, Continental France SAS, Toulouse, France

Efficiency Advantages of SiC in Electric Drive Train Applications 169
T. Grasshoff, O. Tamm, SEMIKRON International GmbH, Nürnberg

The Transition of EV Applications from Silicon to Silicon Carbide – Helping the power electronics design community overcome reliability challenges for EV applications that use silicon carbide 177
A. Kashyap, A. Gendron-Hansen, D. Sdrulla, B. Odekirk, Microchip Technology Inc., Bend, Oregon, USA

Modular DC-DC converter for high-performance fuel-cell systems in trucks and buses 189
T. Bürger, Kunal Goray, AVL SFR, Regensburg;
F. Berg, W. Resende, AVL List GmbH, Graz, Austria

End-2-End

Future electric/electronic architecture – Sustainable design of a digital in-vehicle backend infrastructure 201
M. Traub, H.-U. Michel, BMW Group, München

Function- and Service-Orientation – a Game Changer for the E/E-Architecture of Tomorrow . . 213
R. Roppel, M. Görber, Dr. Ing. h.c. F. Porsche AG, Weissach

Vehicular RF Architectures – Managing integration of next generation automotive wireless systems 227
T. Zipper, Continental Automotive GmbH, Regensburg;
R. Gee, Continental Automotive Japan KK, Yokohama, Japan

Going from an Electronic Unit Centric Development to Application Software Centric Requires a Different Architecture Mindset in Automotive 239
A. Magnuson, Volvo Group Truck Technology, Gothenburg, Sweden

Using Cloud-Based Electronic Horizons to Enable Distributed Driving Functions 255
P. Engel, A. Geraldy, J. Wolter, Robert Bosch GmbH, Hildesheim

Use of open source software in automotive safety projects – A decision tree for the usage of open source software components in safety projects 269
R. Grave, Elektrobit Automotive GmbH, Erlangen

Mission D

Trucks as the drivers of connectivity-based innovation – What the passenger car sector can learn from the experience already gained in trucks today 275
G. Mabire, Continental Automotive GmbH, Frankfurt am Main

Functions on demand – Enabler for digital business with car functions – Challenges of implementation of a high complex security mechanism 283
J.-K. Landgraf, A. Fabri, AUDI AG, Ingolstadt

System of systems structured data for mobility services 293
Y. Chazal, Renault, Paris, France;
A. M. Hein, Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay, Paris, France;
S. Boutin, Knowledge Inside, Versailles, France

Data Management

Building a Standardized Data Pipeline from the Cloud to All In-Vehicle ECUs and Sensors – A New Opportunity for the Connected Car 307
S. Acharya, Excelfore, Fremont, California, USA;
M. Gardner, Molex, Lisle, Illinois, USA;
S. Herz, Hella GmbH, Lippstadt;
C. Hosner, Alpine Electronics, Auburn Hills, Michigan, USA;
F. Lesbroussart, ZF Friedrichshafen AG, Friedrichshafen

Data Structures and Interfaces for High-resolution Maps in Rapid Prototyping Applications of Highly Automated Driving 319
M. Giertzsch, Opel Automobile GmbH, Rüsselsheim

Multilateralism at its best: A blockchain-based platform enabling data sharing, monetization and service differentiation in the automotive industry 333
K. Bader, V. Knaup, S. Schneider, Continental Secure Data Germany GmbH, Aschheim

Mission D

AI and the Evolution of Model-Based Design 347
J. Tung, MathWorks, Natick, Massachusetts, USA

On modern automotive software development – Forever stuck in the middle? 353
R. Schmidt-Clausen, U. Reder, R. Lange, e.solutions GmbH, Ingolstadt

The Future of Digital Car Access – Service Potentials and Ecosystem Challenges 359
K. L. Barbehön, O. Müller, D. Knobloch, BMW AG, München

ADAS KI

Potential of Training Neural Networks Using Virtual Environments 365
R. Pfeffer, N. Ahn, IPG Automotive GmbH, Karlsruhe

Mission AI in Automotive – Collaboration Models and Functional Safety 375
U. Bodenhausen, Vector Consulting Services GmbH,
U. Bodenhausen AI Coaching, Stuttgart

ADAS

Engineering and Hardening of Functional Fail-Operational Architectures for Highly Automated Driving – Identifying and shaping the operational design domain 383
R. Adler, D. Schneider, Fraunhofer IESE, Kaiserslautern;
T. Fukuda, Hitachi Automotive System Europe GmbH

Safety for Automated Driving with High Performance ECUs 395
M. Oertel, J. Wolf, Vector Informatik GmbH, Stuttgart

Impact of Cybersecurity and Safety Standards on ADAS Software Development Practices . . . 407
O. Ur-Rehman, G. Wallraf, B. Holderbaum, M. Jentges, FEV Europe GmbH, Aachen

Security

Are you Security Compliant? – Current Automotive Security Legislations, Potential Impacts to Automotive OEMs & Suppliers, and First Action Proposals 419
M. Minzlaff, Marko Wolf, ESCRYPT GmbH, Munich

Integration of Cybersecurity into Development Processes – A Case Study 423
F. Stahl, AVL Software and Functions GmbH, Regensburg

The transition to HPC-based vehicle architectures – Cyber Security Implications 431
A. Shomer, Argus Cyber Security, Tel-Aviv, Israel

Enhancing In-Vehicle Communication by Authentication and Security – An incremental approach with an example for CAN message authentication 443
A. Hahn, Automotive Security Group, Microchip Technology Munich GmbH, Heilbronn

Hardware matters: how one chip can impact the security of a connected vehicle 455
M. Brunner, H. Adlkofer, Infineon Technologies AG, Neubiberg

Embedded Intrusion Detection based on AI – A Data-Driven Approach 469
A. Weichslgartner, Audi Electronics Venture GmbH, Gaimersheim

Continuous Security Testing for the Automotive Domain 479
S. Greiner, H. Löhr, P. Duplys, Robert Bosch GmbH, Renningen

Architectures + Software

AUTOSAR Adaptive Platform – A standardized SW platform for intelligent vehicles with functional safety and data integrity 493
G. Reichart, M. Niklas, AUTOSAR partnership, Aschheim near Munich

Service-Oriented HPC Communication Standard for Vehicle Lifecycle Management 503
A. Schleicher, DSA Daten- und Systemtechnik GmbH, Aachen

How to Improve Automotive Testing in an Agile Development Process – A Review of Popular Testing Methods and Overview of Advanced Automated User Interface Testing 519
D. Robinson, Altia Europe GmbH, Nuremberg

Mission D – Charging

800V Fast Charging is Reality – From the Vision in 2015 to Reality in 2019 529
O. Bitsche, Dr. Ing. h.c. F. Porsche AG, Weissach

On-Board 2.0

Addressing the challenges in designing fail-operational architectures for autonomous driving platforms – Tailoring fail-operational systems based on production experience in the aerospace industry for the automotive use cases 537
S. Poledna, TTTech Auto AG, Vienna, Austria

Boost Safety & Styling for vehicle lighting – Individualization and new Functionalities 551
M. Kleinkes, W. Pohlmann, C. Wilks, HELLA GmbH & Co. KGaA, Lippstadt

CAN FD Light – A novel communication bus supporting digitalization and customization of automotive lighting for the broad market 567
F. Rennig, J. Barthel, M. Sanza, D. Tagliavia, STMicroelectronics Application GmbH, Aschheim-Dornach near Munich

Digital Light – Function & Design on Demand utilized for Car2X Communication. 581
M. Kruppa, W. Thomas, AUDI AG, Ingolstadt

Foreword

Mission: Transformation

At the time we were ourselves not very confident about our idea of relocating ELIV 2017 from Baden Baden to Bonn. So we were all the more pleased with the overwhelming feedback from participants, speakers, exhibitors and the press regarding this step. Above all, the professionalism and internationalization of that conference as well as the significant increase in the number of participants far exceeded all our expectations.

Two years later, we are facing an elaborate transformation process into climate-friendly, automated and networked vehicle concepts that at minimum goes beyond the predictions of the boldest amongst us. General conditions have become tougher due to legislation, climate change, trade wars and public opinion. And our world continues to change dramatically.

Software is becoming the game-changer:
Artificial intelligence, blockchain, cyber security, big data and autonomous systems in the vehicle as well as new business models are changing the industry permanently. Disruptive approaches in technology, organisation and processes are being developed to ensure future competitiveness. Suppliers and OEMs of the automotive industry are adapting their structures to this software orientation. After a first phase of this orientation, OEMs have researched and developed new concepts and products. Now it is time to deliver. There is the best opportunity at ELIV 2019 to sum up and discuss technology and strategies in the mix of management lectures and technical depth, doing so provocatively and in a way typical of ELIV. Under the motto “Mission: Transformation”, we seek to provide a platform for the “future of mobility”, the “next level of highly automated driving” and the further development of “total networking”.

Over the last 25 years **E**LIV has been constantly evolving. The **e** as in electronics has long been synonymous with the **e**volution of technology – from hardware to software, the **e**volution of the market, of communication and of society.

The program committee is constantly working in close cooperation with the VDI to further develop ELIV into an innovative platform that not only throws light on the trends of tomorrow, but also sets standards as the most important driving force in the industry.

With this in mind, experience what ELIV 2019 has to offer in Bonn this year. I for one am looking forward to it.

Uwe Michael
Chairman of the Program Committee

Program Committee

Dr. Klaus Büttner, *AUDI AG, Ingolstadt*
Dipl.-Ing. Harald Deiss, *ZF Friedrichshafen AG, Auerbach*
Stefan Juraschek, *BMW Group, Munich*
Dipl.-Ing. Christof Kellerwessel, *Ford-Werke GmbH, Cologne*
Ralf Lenninger, *Continental AG, Regensburg*
Dipl.-Ing. (FH) Helmut Matschi, *Continental AG, Regensburg*
Dipl.-Ing. Uwe Michael, *Dr. Ing. h. c. F. Porsche AG, Weissach (Chairperson)*
Dr. Burkhard Milke, *Adam Opel AG, Rüsselsheim*
Dipl.-Ing. Bernd Münsterweg, *Hella KGaA Hueck & Co., Lippstadt*
Dr.-Ing. Dieter Rödder, *Robert Bosch GmbH, Stuttgart*
Dr. Jutta Schneider, *Daimler AG, Sindelfingen*
Dipl.-Ing. Stefan Teuchert, *MAN Truck & Bus SE, Munich*
Dr. Rolf Zöller, *Volkswagen AG, Wolfsburg*

Sponsors

We would like to thank our sponsors for their support

Gold Sponsors

www.iav.com

www.kpit.com

www.maximintegrated.com

www.plm.automation.siemens.com

www.synopsys.com/software-integrity

Silver Sponsors

www.brose.com

www.nxp.com/connected-car

www.qualcomm.de

Bronze Sponsors

www.mathworks.com

Sponsor of the evening event

www.esg.de

Seeing With Sound

Next-level 3D ultrasonic sensors based on echolocation

Nick Knappstein, Toposens, Munich

Abstract

Even though ultrasound has been studied by scientists for over 200 years, its capabilities in practical applications are yet to be fully harnessed. Even today, ultrasound has primarily been used for one-dimensional applications.

Up until now bats have been natures prototypes for sound-based navigation. They rely on echolocation to detect obstacles in flight, forage for food and to see in dark caves. The Munich-based startup Toposens has managed to mimic the bat's technique of navigating and uses it to detect people and objects in 3D in real-time. The company has thus developed the first commercialized 3D ultrasonic solution that provides echolocation, i.e. the ability to "see with sound" for intelligent automobiles and robotics.

Close-range perception for intelligent automobiles

How can it be possible that a staggering 1 out of every 5 motor vehicle accidents takes place in a parking lot? Even at the slow speeds in parking scenarios, the driver cannot perceive his environment flawlessly, despite parking assistance functions like cameras and PDC's. Tight parking spots in car parks or crowded parking areas in front of shopping malls display everyday driving challenges. Come to think of it can be quite reckless to not support your car with the latest sensor technologies and leave multiple tons of moving metal without eyes.

Up until now, sensors helped drivers steer more safely and guide them into a parking spot via sound and light signals. In order to advance in areas like autonomous driving, mapping and collision avoidance, sensors now need to detect more complex environmental scenarios, like steering a car through a construction site or reliably detecting people in a crowded parking area. As every ride starts and ends in a parking position, involving a parking maneuver, it is crucial to reliably perceive a car's immediate environment!

While existing sensor technologies mostly focus on covering long distances, the immediate environment around a car (0 – 5 m) is often left out of the discussion. That is where Toposens 3D sensors come into play. Toposens sensors provide reliable, rich three-dimensional data for the close-range environment around a vehicle. The sensors are therefore well-suited for applications in the automotive field and add yet another level of safety and redundancy to conventional radar, lidar and camera technologies. “Because our ‘Bat Vision’ sensors are compact, affordable and integration-ready,” explains Tobias Bahnemann, Managing Director of Toposens, “engineers can easily add them to their perception stacks to replace or complement their existing optical sensing systems, providing both redundancy and an improved level of accuracy compared to standard ultrasonic sensors in various autonomous navigation applications.”

The sensor data can further be used for additional comfort features, e.g. gesture control to open doors and trunks, positioning the vehicle for automated charging (for EVs), and collision avoidance for automatically opening doors.

Fig. 1: Ultrasonic perception for automobiles (animation)

Passenger monitoring in the interior of the car

With further improvements in the autonomous driving space, the behaviour of the driver is also likely to change. While drivers today must be completely focused on the road -ready to react at a moment’s notice- this is likely to change once cars are able to drive and steer fully automatically. Drivers would then be able to lean back and relax, work on their computers, turn to their children in the back seat, or temporarily enjoy an expanded infotainment program.

Such an eventuality puts new demands on assistance systems. Just like the numerous sensors available for analysing a car's external environment, similar knowledge is needed for the interieur in order to realize a more secure and intuitive interaction experience. In this context,

the use of 3D ultrasound again provides interesting advantages. Data gained from the 3D ultrasound sensor can be used to identify the number of people sitting in the car, their size and their posture. Based on the information regarding where people are sitting and their physical characteristics, airbags could be adjusted to individual body sizes and further improve safety.

Toposens technology does not collect any personal data since ultrasound cannot evaluate visual input, and only records anonymous pointcloud data. This is an especially important consideration in terms of privacy and data protection. Furthermore, gesture simulation in the interior of the car can be used for information and entertainment purposes, like controlling your car's infotainment systems with simple pre-configured actions.

There is no doubt that autonomous vehicles of the future need more assistance from sensors to operate safely in populated places. Whether you are living in a big city where detecting people and accurate parking plays a major role or on the countryside where automated charging is indispensable, sensors from Toposens will have several benefits to the everyday life of future drivers.

Sensing principle of (3D) ultrasonic sensors

In general, ultrasonic sensors are devices that make use of high-frequency sound waves for a range of applications such as distance measurement, non-destructive testing and medical imaging. For distance measurement, a typical ultrasonic sensor uses a transducer to periodically send out ultrasonic pulses in the air. These pulses get reflected from objects in the detection area of the sensor and are received back by the sensor. By measuring the time, it takes an ultrasonic pulse to travel to the object and get captured by the sensor, the distance to the object can be calculated. This principle is called time-of-flight measurement.

Conventional ultrasonic sensors used for parking assistance only record one-dimensional data, which is the distance to the closest object. Azimuth and elevation angles of objects are not calculated with this method, and vertical opening angles are severely limited. Thus, many objects – such as the curb and low-lying obstacles – are not picked up by 1-D sensors.

In addition to measuring the distance to objects, 3D ultrasonic sensors apply a wide opening angle of up to 180° to also calculate their horizontal and vertical positions relative to the sensor. This localization of objects in three-dimensional space also allows a 3D ultrasonic sensor to

detect and **distinguish between multiple objects** in a single scan. In this sense, the principle of 3D ultrasonic sensors is similar to echolocation methods used by bats and dolphins

The sensor sends out ultrasonic signals, evaluates the echoes that come back and precisely detects where in a given space both static as well as dynamic objects are located in real time. It can detect multiple objects and people to generate a 3D pointcloud of its ambient environment.

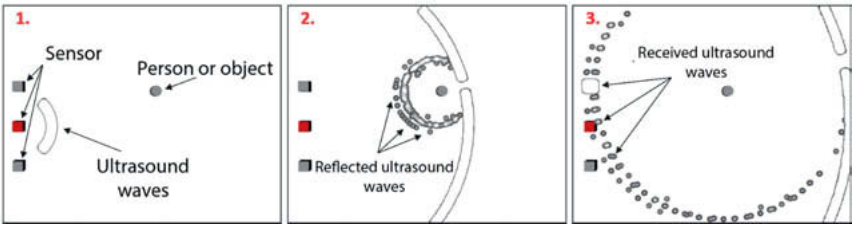


Fig. 2: Principle of Echolocation

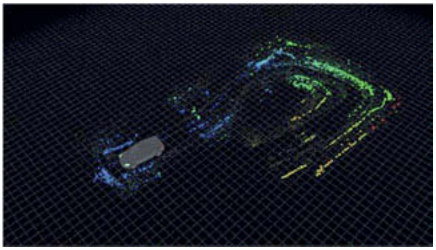


Fig. 3: Point Cloud Example recorded by Automotive Kit (real data recorded by sensor)

Toposens flagship product, the TS3, combines carefully selected hardware components with proprietary signal processing algorithms built on their proprietary Toposens' SoundVision1™ chip that easily adapts to a variety of product designs. This makes the TS3 the perfect technology platform for developing next-level mass market applications like automated parking and ADAS components.

TS3 comes in a small form factor (71.4 x 27.4 x 11.6 mm), consumes minimal energy (200 mA nominal) and is light weight (below 20g). Its technical specifications include a detection range of up to 5 meters and a scan rate of 28 Hz. The sensor returns up to 200 points per second

with each 3D point containing info about the cartesian coordinates and an additional intensity measurement of the echo returned by the detected object. A unique feature of this device is its ability to immediately evaluate the received raw signals on an onboard microprocessor, making an external control unit obsolete and keeping latency extremely low. Further clustering and processing operations can be added by the user after receiving the data via UART.

Toposens current sensor generations

Toposens' current sensor versions, the **TS3** and the **Automotive Kit**, are robust, against dust, dirt and lighting conditions and can seamlessly detect transparent and reflecting surfaces. Along with the sensors, Toposens offers a feature rich toolkit for Robotics Operating System (ROS), making the sensors easy to integrate into preexisting systems and enabling fast prototyping for high-level navigation capabilities. The team is currently working advanced decision-making algorithms to complement the software stack.

Fig. 4: TS3 – robotics sensor

Fig. 5: TS Bravo – automotive sensor

About Toposens

Toposens develops sensor products that accelerate the advancements in machine perception. Toposens was founded in 2015 by Alexander Rudoy, Rinaldo Persichini and Tobias Bahnemann. It is headquartered in Munich, Germany with an additional office in Sunnyvale, CA. The team currently consists of 22 people. Toposens is working together with well-known companies from the automotive and robotics industry as well as with established research institutes.

Further information: https://toposens.com/ | info@toposens.com | +49 89 2375 1540 (Germany) | +1 (669) 206 2139 (US)

Ensuring the reliability, availability and safety of fully automated and autonomous transport systems through modern system architectures

Johannes Heinrich, M.Sc., Dr.-Ing. **Andreas Braasch**,
Institut für Qualitäts- und Zuverlässigkeitsmanagement GmbH, Wuppertal;
Dr.-Ing. **Fabian Plinke**,
Institut für Qualitäts- und Zuverlässigkeitsmanagement GmbH, Hamburg

Abstract

Fully automated and autonomous vehicles place new demands on reliability, availability and safety. Eliminating the driver as a fallback path in the event of technical breakdowns or failures, forces a self-driving vehicle to operate safely even in the event of a fault, in order to reach a risk-minimum state. Similarly, minor disturbances, such as software crashes or failures, must be compensated for safety terms in real time.

These requirements are realized through redundancy-based, fail-operational on-board architectures. In the event of an error, a fault-tolerant subsystem can be switched on or off in order to obtain a (possibly degraded) operability.

In aviation, such system designs have been standardized and steadily developed for the use of complex, software-based flight control systems.

The article gives an overview of the technical requirements of the aviation and automotive industries, as well as the presentation of aviation methods and principles and their transfer to the automotive development.

This includes methods for fault detection, fault tolerance, strategies for continued operation, live repair, degradation and the safe shutdown of the system into a state of minimal risk. This procedure has not been implemented in the current automotive development because the responsibility for driving the vehicle lies always with the driver.

A quantitative assessment of reliability, availability and safety – considering the above system properties – can be performed using multi-stage simulation models, which are also presented in this paper. The goal is the statistical validation of an economic system design while complying with the safety requirements as established by common standards and rules in the automotive sector (for example ISO 26262, SOTIF via ISO / PAS 21448, etc.).

Introduction

The development of automated and autonomous systems is currently being heavily promoted across many industries, producing new results at regular intervals. A large share of this lies within the automotive industry, but also industries such as the railway industry or the process engineering research in these areas and try to use new technologies for themselves.

One aspect which, in addition to the pure technical implementation of the systems, represents a major challenge is the safe design of the systems and the safety proof to be provided. As a result of the elimination of humans as a control level and the assumption of control by software, the demands placed on the systems regarding safety and availability increase.

In particular, the handling of failures and the associated fault-tolerant design of the system have a major role. The person who, in the case of an error, e.g. in the car, has to intervene, is no longer tangible, so that this fallback path must be replaced.

In this paper, chapter 2 addresses the safety-relevant challenges regarding the safety-relevant and fault tolerant system design and explains the topics redundancy and Monitor-Control-Principle. The safety and reliability proof are discussed in Chapter 3, first new characteristics of automated and autonomous systems that lead to problems using known safety and reliability models are described and afterwards a state-based simulation model for the safety and reliability assessment is shortly presented. A methodical outlook is given in Chapter 4 in which the FDIR-Principle to handle different kinds of faults is explained. Chapter 5 summarized the paper and gives an outlook on further activities.

Fault tolerance and safe architectures

The safety and reliability concept of future automated and autonomous systems must be renewed due to the omission of a human person. Today's systems are often designed by using the fail-safe-principle, where a person can immediately take over the control of the system in case of a failure that lead to a function shutdown. Besides that, the person acts as a controller in terms of automated functions and must intervene in the event of a wrongdoing. For example, a driver has to monitor an automated car up to SAE Level 2 "Partial Automation" [1] and is responsible for the action of the system. Therefore, there is no need for fallback modes in terms of the system design to maintain or supervise the function. Automated driving functions of Level 3 "Conditional Automation" do not need a driver in the loop, but the driver has to be available as a fallback, e.g. in case of failure, after a takeover time. The system therefore has to remain capable of manoeuvring in the event of a fault within a defined time interval. In case of Level 4 "High Automation" and Level 5 "Full Automation" the driver does not need to be in

the loop and be available as a fallback, so there are special fault tolerance requirements for the system architecture due to the fail-operational-principle.
Besides the fault tolerance the system has to supervise itself in terms of safety-relevant actions, no accidents, especially with human damage, due to a malfunction can be tolerated.
Because of the described challenges in automated and autonomous systems suitable strategies must be used in order to guarantee the maximum of the safe state probability (availability of safety). A possible approach is to use the aerospace principles for Monitor-Control and redundancies.

A self-resolving system must be monitored in real-time. Pilots are monitoring the system in an aircraft. However, autonomous systems must be using redundancy to implement a Monitor-Control-Principle (Fig. 1).

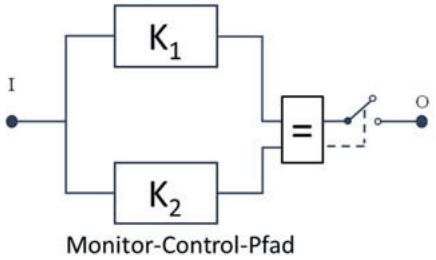


Fig. 1: General Monitor-Control architecture

A difference in the reliability and safety redundancy need to be considered. A simple parallel system increases the reliability but in order to use a redundancy with a Monitor-Control-Principle a mvn, nvn-systems or serial system with a voter is required.
The implementation of redundancy is more complicated and leads to an increase of redundant and partly dissimilar systems. In case of failures, where the system needs to access a safe failure state, the availability of base function needs to be ensured (probably without requirements to the redundancy, etc.). From the view of reliability engineering different paradigms are considered: reliability, availability, safety and the parameters are assessed for deviations. So far, the classical quality management often only determines reliability parameters and assigns them to safety critical scenarios.

Safety and reliability proof

Another challenge of automated and autonomous systems are the safety and reliability proof. The classic reliability system structures for the conventional mechanical and electrical dominated systems cannot describe the whole picture of mainly software-based infrastructures. The main reason for that is the dynamic of the software application and their failures and repair. After a software failure the components can be switched off, restarted or the whole application can be moved to another processor, resulting in a new repair and failure behaviour that must be methodically defined. A new start of an application, a flash of an image or external repair processes via WLAN or mobile networks are examples that must be considered for the analysis. As a result of these boundary conditions, the parameter of availability of dynamic repairable systems is becoming increasingly important.

In quantified methods for reliability analysis the possibility for considering repair rates and repair behaviour are already possible. However, describing the repair behaviour for components realistically cannot be considered by classical methods, yet. The implementation of strategies, e.g. Markov-Processes or Petri-networks is mathematically challenging or even in some cases impossible.

Through the increasing realization of functions by means of software and the existence of redundant structures a temporary or permanent shift of a whole function implemented in software to a different hardware is possible and necessary [2] [3]. The result is a change of the system structure during operation, which is adapting its tasks and modes according to fixed rules. A reliability optimization problem is created, which must consider the technical and economic efficiency of the system's design. Furthermore, safety requirements can depend on different use cases. Different requirements for the computing power of an automated or autonomous function exist depending on the size of existing influencing factors and the resulting complexity of the function. The system must be technically flexible to adapt to different requirement and environments. A solution is that dynamic system structures are introduced that can adapt their structure automatically according to the use case. For the reliability engineering multi-layered calculation on different levels must be done and summarized.

This point poses another challenge to the reliability and safety analyses, since such dynamics combined with the repair behaviour of the software can hardly be represented and calculated analytically.

This section presents an approach to handle the above-mentioned challenges in terms of safety and reliability proof. A possible solution to represent the boundary conditions is to use

the Monte Carlo Simulation (MCS) [5] [6] [7] [8] whereby the systems can be modelled very flexible.

The simulation is based on state diagrams, shown in Fig. 2, in which the system is mapped with regard to active and failed hardware and software. The system corresponds to a duo-duplex-system with four active software elements on two hardware components.

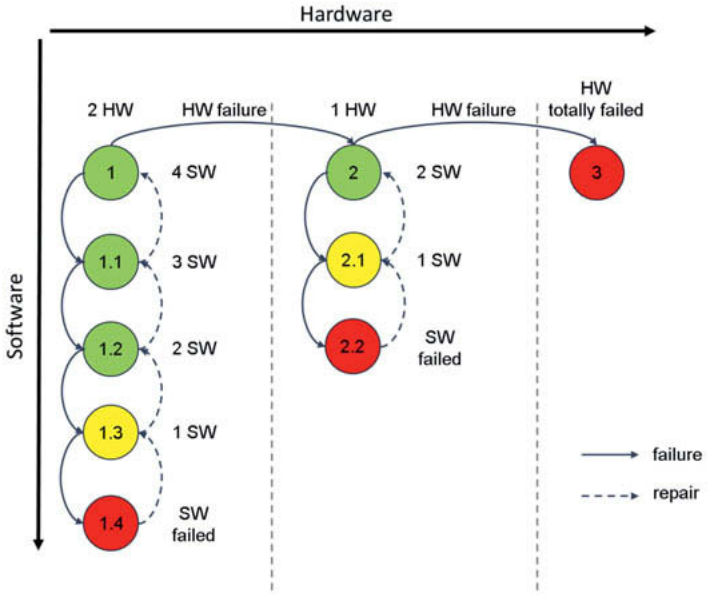


Fig. 2: State diagram of a hard- and software architecture

The state-diagram differs between irreparable hardware failures (horizontal direction) and reparable software failures and repairs (vertical direction). This means that should the system be in state 1.4 (system failed) due to four software errors, it may work again through restarting a software. If the system is in state 3 due to two hardware failures, the system has failed completely and must be fixed, e.g. in a workshop.

For every transition from one state to another special rules can be implemented, for example different kinds of parallel error and repair modes, depending transitions or superordinate rules. Moreover, it is possible to adjust the transition rates so that even a dynamic system can be mapped.

By the means of the state diagrams, it is possible to assign key performance indicators (KPI) to the different states, for example *safety* (green states), *availability* (green and yellow states) and *reliability* (green and yellow states, until the system reaches a red state, after that the reliability equals zero).

In this paper the KPI's are defined with regards to the explained challenges for the software-based safety-critical systems. Safe states are characterized by the fulfilment of fault tolerance and the Monitor-Control-Principle, i.e. there is at least one redundant component that can compare the result of the own calculation with another component, or which could transfer the system in a safe condition. Available states are all states where the system is functional, even after it has failed, and has been repaired due to a software reboot. Reliable states are very similar to the available states, but the reliability of the system cannot be restored after a system failure, as the system has initially failed.

The repeated simulation generates a large number of status profiles. Fig. 3 shows two simulated exemplary status profiles depending on the available redundancy.

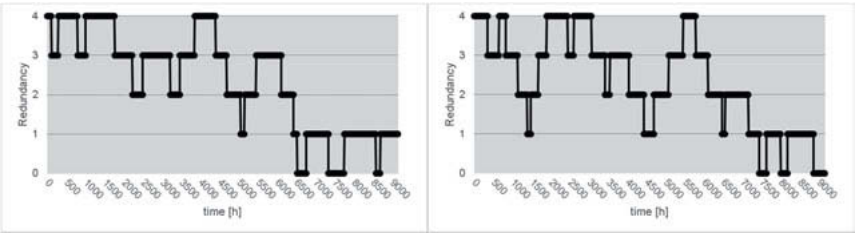


Fig. 3: Status profiles of two iterations

Based on the MCS though the huge amount of status profiles a probability can be calculated that the system is in state X at time t. Combining the probability with the defined state KPI's temporal profiles for safety, availability and reliability can be derived. If different system architectures are examined by means of the MCS, quantitative statements about these systems can be affected. Fig. 4 shows an example of the course of the three KPI's over a time of 12 months.

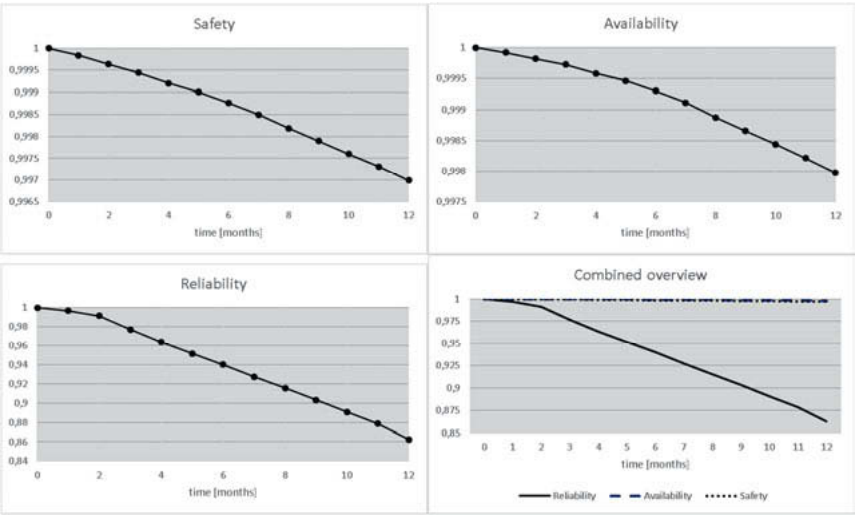


Fig. 4: Diagrams of safety, availability and reliability

As can be seen, the graph of the safety and availability keep a very high level of 99,7% of safety and 99,8% of availability after 12 months. Neglecting a repair after a system failure, the reliability decreases more and is after a period of 12 month at a value of about 86%. The course of security and availability are very similar. This is because both parameters are based on the state probabilities of states 1, 1.1, 1.2 and 2. In the case of availability, however, the state probabilities of 1.3 and 2.1 are added.

Failure management

An important issue for automated and autonomous systems is to distinguish between systems whose control in the case of an error can be taken back by a person after an acceptance period, e.g. by a driver in the vehicle of Level 3 or via a remote connection by a person in a control center, and systems in which a takeover by a human is not easily possible and the system must continue to operate autonomously in the event of a fault, e.g. space systems that can't be easily controlled remotely due to space communication issues.

In the case of the first-mentioned systems, it is sufficient to ensure a temporary functioning until someone else takes control. In systems in which the control can't be transferred to a person in a simple way, continuous functioning in the event of a failure must be ensured with

regard to a previously defined fault tolerance. Therefore, an intelligent fault management system is necessary that can identify different types of failures and decide how to handle the failures, either by an independent processing or by a processing via remote.

For space applications, where immediately no person is part of the control loop, the Fault Detection, Isolation and Recovery (FDIR) principle exists [4] [5]. Based on this it is possible to classify faults and to take appropriate measures. As shown in Fig. 5 the classification is split up in levels 0 to 4, in which the criticality of a failure increases from level 0 to level 4.

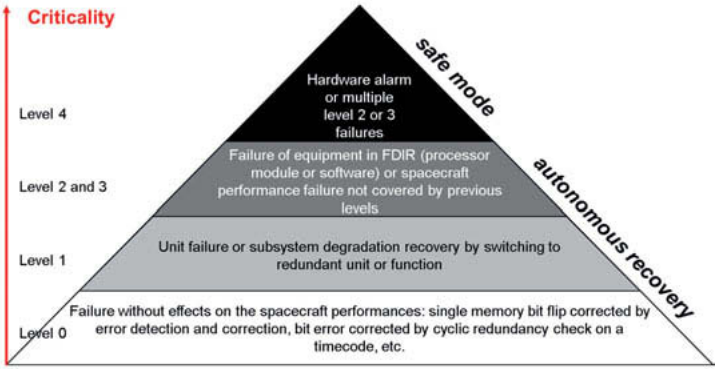


Fig. 5: Fault diagnosis and management architecture for satellite/spacecraft [5]

If failures of levels 0 to 3 occur, the system can recover itself autonomously. Depending on the severity of the failure, it will be detected and fixed at different levels of the system within variable reaction times. Failures of Level 0 are detected and fixed in each unit of the system by local correction, independent of other units. Level 1 failures must be detected outside the unit and fixed due to a subsystem, which consists out of several units, by switching a unit to its redundant one. If a failure from level 0 or 1 cannot be fixed the failure is categorized in level 2 and has to be fixed more globally at subsystem or platform level. Failures of level 3 are usually software or processor module failures that can be recovered by switching to a redundant processor module. In case of level 4, there are hardware failures or multiple level 2 or 3 failures, the system can no longer handle it autonomously. In case of a satellite it is transferred into a safe mode and the ground station has to intervene and control the system.

This method shows a way how different kinds of failures can lead to different behaviour and that the system can handle most of the failures autonomously by itself, only in very critical scenarios (level 4) a person is needed to deal with the failures and to control the system.

Conclusion

This paper presents on the one hand approaches for a fault tolerant and safely acting architecture, Monitor-Control-Principle and redundancies, and an intelligent failure management, called FDIR, based on concepts from aerospace industry that enable the transition from fail-safe to fail-operational architectures, on the other hand a state-based simulation model to calculate different KPI's such as safety, availability and reliability for software-oriented automated and autonomous systems with a realistic modelling and without tight restrictions. In addition to considering software as an individual component within the system structure, it also extends system characteristics, e.g. Monitor-Control-Principle and dynamic switching of operational modes have been considered for the assessment methodology. Furthermore, the simulation shows that some characteristics like different failure and repair modes for the transition between two states, that could not have been modelled using e.g. Markov-processes, can be now applied with the presented model. A disadvantage of the simulation is the time-consuming effort for the implementation and calculation of the systems because the whole modelling has to be implemented in a computer algebra system. The more accurate the modelling is performed with dynamic operating modes and different failure and repair rates, the more complex is the implementation of the code. For future activities, it is planned to transfer the presented aerospace models to specific application areas. Depending on the industry, the application of the concepts presents new challenges that did not exist before, e.g. due to space and financial reasons, it is not easy to implement multiple redundancies in the automotive industry. Suitable alternative solutions have to be found. Furthermore, it is planned to implement the FDIR principle in the simulation model. In this way it will be possible to see the effect on different failure-level on the system and to model in detail the dealing with the various kinds of failures.

References

[1] SAE J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE International, 15.06.2018.

[2] Raksch, C.: Eine Methode zur optimalen Redundanzallokation im Vorentwurf fehlertoleranter Flugzeugsysteme, Dissertation, Technische Universität Hamburg-Harburg, 2013.

[3] Rehage, D.: Zustandsmodellierung und Zuverlässigkeitsanalyse fehlertoleranter Systemarchitekturen auf Basis von Integrierter Modularer Avionik. Dissertation, Technische Universität Hamburg-Harburg, Schriftenreihe Flugzeug-Systemtechnik Band 1/2009, Shaker Verlag, Aachen ,2009, ISBN: 978-3-8322-8650-7.

[4] Jalilian, S.; Salar Kaleji, F.; Kazimov, T.: Fault Detection, Isolation and Recovery (FDIR) in Satellite Onboard Software. https://ict.az/uploads/konfrans/soft_eng/87.pdf, invoked: 19.07.2019, Azerbaijan, 2017.

[5] Zolghadri, A., Henry, D.; Cieslak, J.; Efimov, D; Goupil, P.: Fault Diagnosis and Fault-Tolerant Control and Guidance for Aerospace Vehicles – From Theory to Application. Springer-Verlag London, 2014, ISBN: 978-1-4471-5312-2.

[6] Heinrich, J.; Horeis, T.; Plinke, F.: Zustandsbasierte Verfügbarkeitsanalyse von Hard- und Softwarearchitekturen mittels Monte-Carlo-Simulation. Tagung Technische Zuverlässigkeit 2019, Nürtingen, VDI-Berichte 2345, VDI-Verlag GmbH, Düsseldorf, 2019, ISBN: 978-3-18-092345-1.

[7] Plinke, F.: Beitrag zur Weiterentwicklung der zuverlässigkeitstechnischen Sensitivitäts- und Ausfallanalyse mittels Monte-Carlo-Simulation. Dissertation, Bergische Universität Wuppertal, 2015.

[8] Zio, E.: The Monte Carlo Simulation Method for System Reliability and Risk Analysis. Springer Verlag London, 2013, ISBN: 978-1-4471-4587-5.

[9] Meyna, A.; Pauli, B.: Taschenbuch der Zuverlässigkeitstechnik: Quantitative Bewertungsverfahren. 2nd edition, Carl Hanser Verlag, Munich, 2010, ISBN: 978-3-446-41966-7.

ADAS/AD Systems: Efficient Testing & Validation

From data acquisition to data analytics

Dipl.Ing.(FH) **M. Kremer**, Dr. rer. nat. **M. Kreutz**,
Dipl. Ing. (FH) **M. Luxen**, Dipl. Ing. **S. Christiaens**,
FEV Europe GmbH, Aachen

Abstract

Highly and fully automated driver assistance systems place new demands on test and validation solutions. In particular, the huge amount of data to be collected and analysed constantly increase the requirements both on the data loggers themselves as well as on the data analytics framework. In the context of the European-Union funded L3Pilot-Project, where FEV Europe GmbH is involved, a focus is placed on data collection, storage and the processing framework.

The presentation first describes the selection criteria of a data logger solution for the acquisition and processing of the required vehicle and sensor data by means of an exemplary data logger evaluation. A focus is made on the importance of having time-synchronous and highly precise log data from different types of vehicle sensors such as ultrasound, radar, LiDAR, video and vehicle bus communications like CAN(-FD), Flexray or Ethernet in real driving situations.

The second part of the presentation is about introducing an efficient way to transmit and handle these data in order to optimize the data transmission as well as the storage and analysis. Some key elements of the framework, such as data generation, private or open cloud backend, data format conversion and in-depth analysis are addressed with practical examples from real world testing.

Background & Motivation

Most challenging topics regarding ADAS/AD and verification as well as validation of autonomous vehicle functions are:

1) How can these functionalities be validated and verified?
2) What is needed for verification and validation?
3) Which data needs to be collected and how can an overview of the collected data be kept?
4) How to efficiently handle the big amount of data?
5) How not to lose the overview about the already verified and the still to be verified functionalities?
6) How to keep the costs for storage under control?

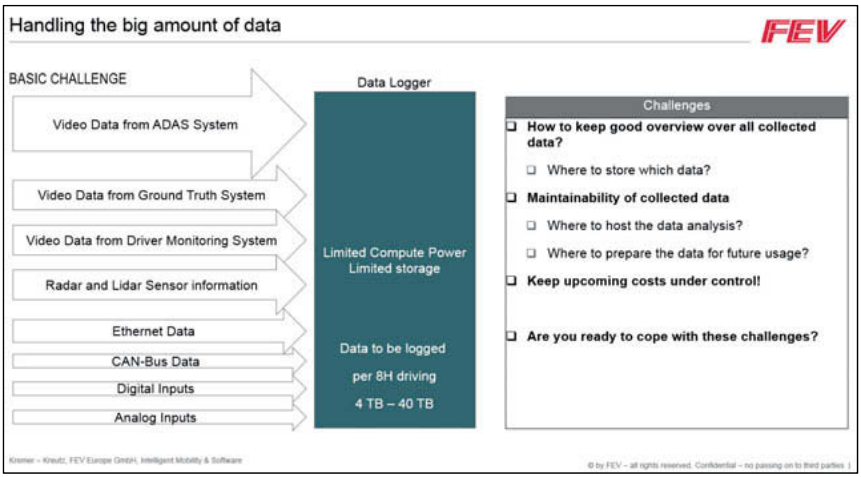


Fig. 1: Challenges for data acquisition and handling for ADAS / AD vehicle developments

Intelligent data collection for autonomous vehicles

System Overview

To cope with all boundary conditions of autonomous driving, first of all, the creation of an overview of the relevant data to be collected to verify and validate the systems, is essential. In this context, Fig. 2 depicts an exemplary system overview of an autonomous vehicle.

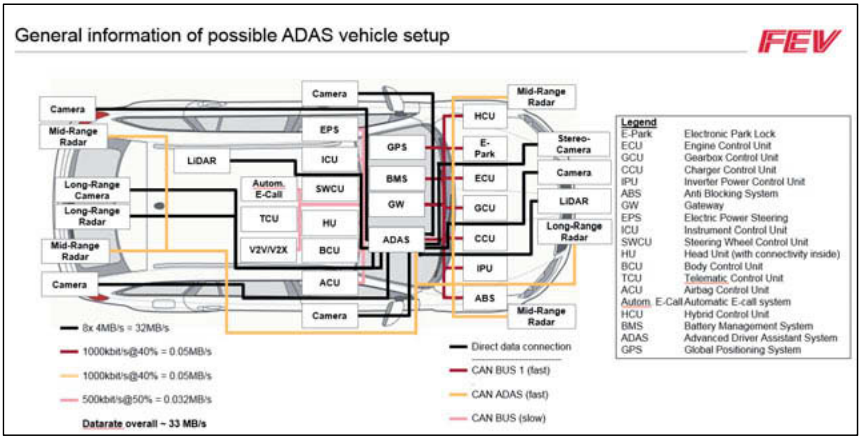


Fig. 2: Possible Vehicle setup with ADAS functions

Even if not all of the relevant control units to be found within a state-of-the-art premium class vehicle are considered in this Fig., it provides a rough impression of the system complexity. Considering that not only one camera for front view is needed, but at least at second one for redundancy, the amount of video data is already increasing by factor 2 for each view.

The direct data connection between the ADAS control unit and Radar or LIDAR sensors is also a communication line with a high data throughput. Here, also redundant setups of sensors for long range and short range object detection must be assumed.

Depending on the setup of the vehicle, the overall amount of data transfer, which must be handled inside of such a complex system, is between 4 and 40 TByte for 8 hours of driving.

This data volume is needed to be reused for e.g. simulating situations in HiL, MiL or SiL environments.

Especially for ViL test-benches, the complete data set is required to set up a surrounding environment for the vehicle under test.

Additionally to the in car information, it is also necessary to collect map data, environment information, weather data, test driver behavior and collect information about the driver's sensations, when being chauffeured by an autonomous vehicle e.g. for the first time.

Inside of L3Pilot project, all of these topics are being taken into account to get enough data for evaluating especially the driver's acceptance of autonomous vehicles.

That means that after having found a proper data logger, which can take care of these needs, a proper solution must be found for data handling of the vehicles which are performing daily test drives.
Right now the only possibility to handle 4 to 40 TBytes from each car of a fleet and not to lose relevant information is to take the hard disk drives from the data logger and upload the data after the vehicle has returned from test drive.

Just an example for the complexity of uploading such data volumes to the cloud or storage:
Data connection with glass fibers at local homes is limited to maximum 1 GBit/s right now.
That means for each car that 40 TBytes == 320000 GBit need to be uploaded to the data storage. This would result in, if one normal glass fiber connection is being used for only one car, 320000 seconds equal approximately 90 hours (3,75 days) to upload the data to the storage space from an 8 hours test drive of one car.

Because of this temporal constraint, it was decided to not only install a data logger for the big amount of data inside of the car, but also a second connected one (Fig. 4), that is capable of logging 4 CAN-Buses with all the relevant data inside. This enables to evaluate onboard the relevant situations or scenarios, which occur during the test drives. The logger logs the 4 CAN buses and GPS information and transfers the data directly through a normal mobile network to a cloud system, via a secure data connection as can be seen in Fig. 3, flow number 2.

System components in detail

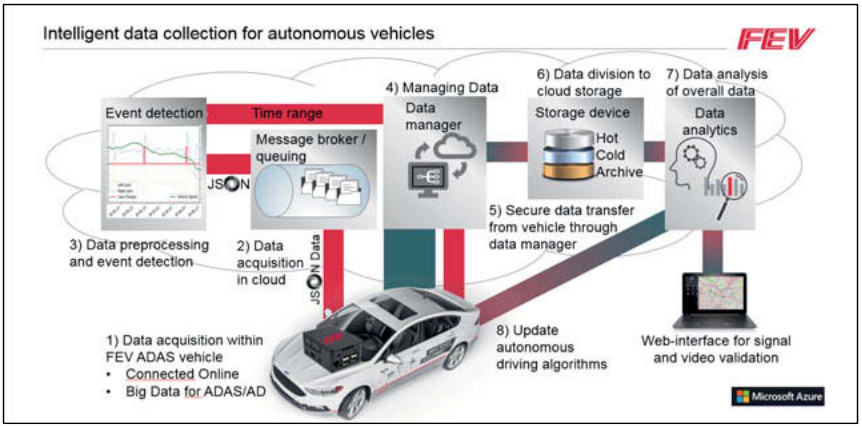


Fig. 3: Eco-system for data collection from autonomous vehicles with connected logger additionally installed to ADAS/AD Logger system (ADAS/AD Logger is not shown in the picture, but installed inside of the car)

The data from the vehicle is transferred in JSON format to a message broker towards the cloud-system.

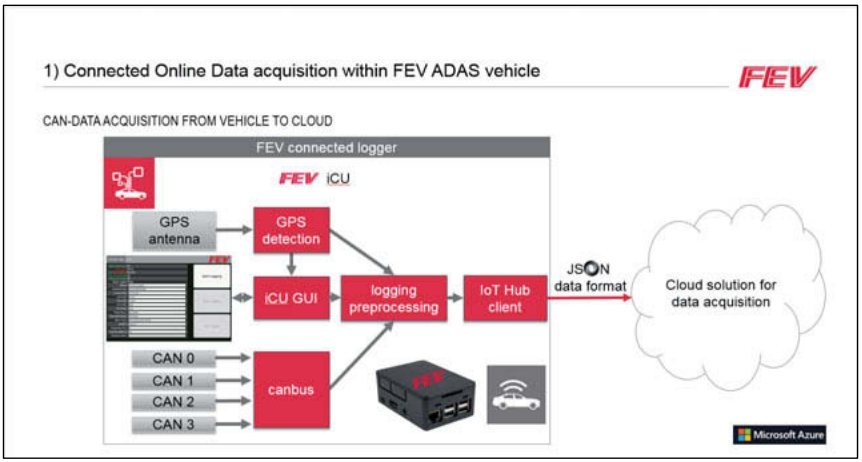


Fig. 4: Principle of the connected data logger

Once the data is available inside of the cloud, which refers to “only” about 200 relevant signals from the CAN-Buses inside of the car, the data conversion and analysis can start.
To be able to react on peak loads and since not only one car might be running inside of a fleet, this data preparation and analysis is running on cluster machines inside of the cloud, which can be scaled according to the needs of the upcoming data traffic.
The data is then resampled and events, like e.g. lane change detection, harsh braking, near contacts with other objects and so on, are detected and the time interval, e.g. 2 mins before and after the event, are marked.

The whole data preparation and analysis is already running, when the car is still on the test drive. After the car is coming to the workshop for retrieving the data from the ADAS/AD data logger, first results and information of the test drive are already available. It is even an option to already give feedback to the driver during the test drive what kind of scenarios are already tested and which are still missing.
This helps to manage the already mentioned 40TByte data-handling, which normally would take 90 hours to upload.
Since the most relevant events and scenarios are already known, the data manager in Fig. 5 can then takes care of uploading the data from the hard disk drives to the relevant storage inside the cloud. Very special or focused scenarios can be stored in hot storage. Scenarios which need to be investigated at a later point of time, can be stored in cold storage and data from the car, where nothing relevant happened, can be stored in archive storage.
This allows the user of this environment to keep the costs for the data storage under control and use the available storage very efficiently. Furthermore, it facilitates data pre-processing and data preparation activities for deeper analytics.

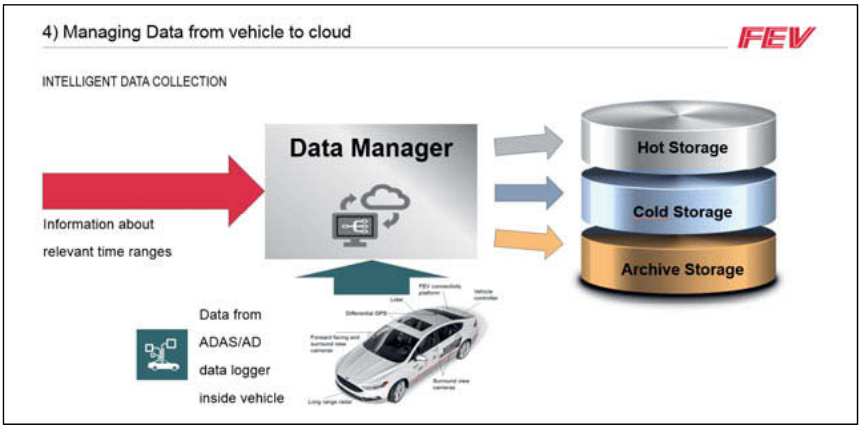


Fig. 5: Principle of Data Manager and data handling from vehicle to cloud

After the upload of the data, the next part of the data preparation can begin. The data can be supplemented with weather data, additional map data, and annotation or labeling of video data can be performed.

Fig. 6 shows the video annotation tool developed by FEV Polska, which is able to annotate automatically for example vehicles, lanes, traffic lights, traffic signs, etc.

After this supplement, the data is ready to be used for simulation purposes, AI training or validation of new software for AD/ADAS systems.

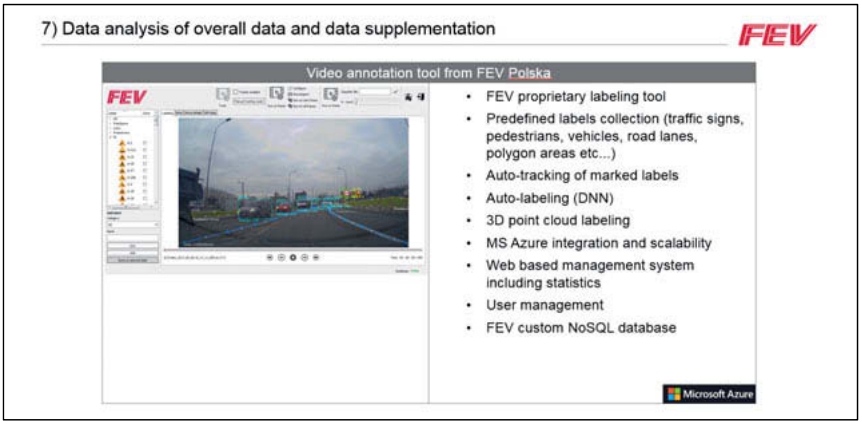


Fig. 6: Video annotation tool from FEV Poland, which is also running in Azure Cloud as a cloud service

Summary

Considering the list of questions initially formulated regarding ADAS/AD function verification and validation, appropriate solutions based on an efficient toolchain have been developed. The following applies:

1) How can these functionalities be validated and verified?

   With the help of the prepared logging and cloud solution, the data measured from autonomous vehicles can be supplemented with the needed additional information and used for simulation purposes for future usage.

2) What is needed for verification and validation?

   Two data-loggers are needed, one connected logger and one logger for ADAS/AD data logging connected to a cloud system. Also required are algorithms to find relevant scenarios and supplement the collected data with additional information (e.g. video labelling, weather data, …).

3) Which data needs to be collected and how can an overview of the collected data be kept?

   A data-manager is needed, which can move the collected data to relevant storage locations based on scenario detection or relevance evaluation. Whenever the data is categorized and located, the reuse of the data can start.

4) How to efficiently handle the big amount of data?

   With an efficient setup of the cloud environment, e.g. using cluster machines for data preparation and enrichment, locating the relevant data in easy accessible storage devices and preparing a proper solution for mobile data transfer, efficient preparation and execution of the next validation and verification process steps for the ADAS/AD functions will be made possible.

5) How not to lose the overview about the already verified and still to verify functionalities?

By categorizing the detected scenarios during the test drives, it can be assured that all previously defined relevant scenarios for validation purposes are tested and verified.

6) How to keep the costs for storage under control?

   By categorizing the data transmitted to the cloud into most relevant, semi relevant and non-relevant, the costs can be reduced for further usage of the data and the analytics can be optimized adequately.

   In Fig. 7 you can find some buzzwords for the efficient data handling within the data-management solution from FEV Europe GmbH which helps you to keep costs under control.

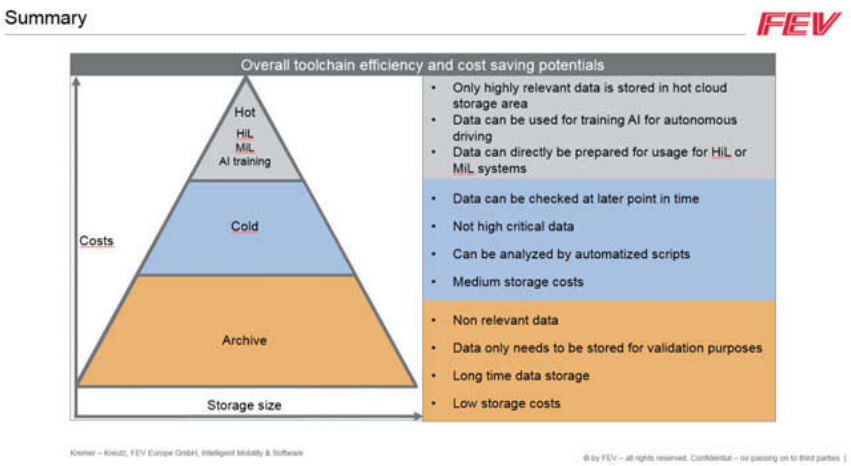


Fig. 7: Overview about the cost efficient storing inside of the Cloud Ecosystem

Problems and solution spaces for driver-initiated handover from automatic to manual driving mode

Dipl.-Ing. **Joe Klesing**, Nexteer Automotive, Auburn Hills, USA;
Dr.-Ing. **Salaheddine Safour**, Nexteer Automotive, Paris, France

Abstract

This paper discusses use cases for automated driving in conjunction with the driver-initiated handover from automated to manual driving. Potential problem areas that may arise are identified, as well as ways to overcome them.

These potential problems include:

- Loss of situational awareness
- Degradation of driver skills and impaired performance
- Mode confusion regarding who is responsible for the control of the vehicle.

By referencing current research, we discuss the following solution spaces:

- Shared haptic control as a means of bi-directional communication between the automated system and the driver
- Fusion of driver gaze data with steering behavior to evaluate driver readiness and ability to assume manual control
- Seamless integration enabled by steer-by-wire

We present the results of user group studies and evaluate the proposed approaches for haptic shared control and sensor fusion. Finally, we give an outlook regarding future work on options for shared haptic control.

Handover Process from Automated to Manual Driving

The Handover or Transition from Automated to Manual Driving is defined as the process and period of transferring responsibility of, and control over, some or all aspects of a driving task, between a human driver and an automated system [1].

Figure 1 shows the framework for a handover process from a Level 4 Automated Driving situation (definition according SAE J3016), to a Level 2 Assisted Driving.

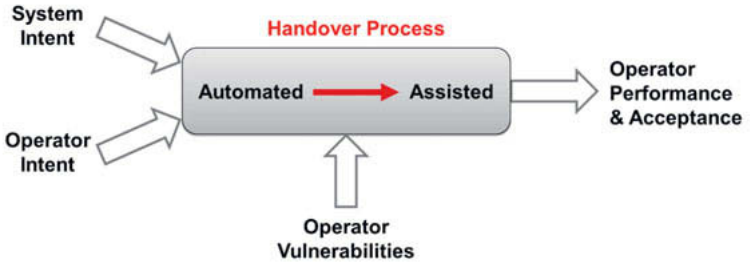


Fig. 1: Handover Process from Automated Driving to Assisted Driving

The inputs comprise the system intent and operator intent. The role of the operator changes from being a passenger during the automated driving mode to being the driver in the assisted/manual driving mode. The operator can, for example, be alert and aware of the surrounding traffic or disconnected and Out-Of-The-Loop (OOTL). If the operator wants to assume control of the vehicle from the automated system and change its trajectory (see Figure 2), she is subject to vulnerabilities. The driver performance after handover, and thus driver safety, as well as the driver acceptance of the handover process complete the framework.

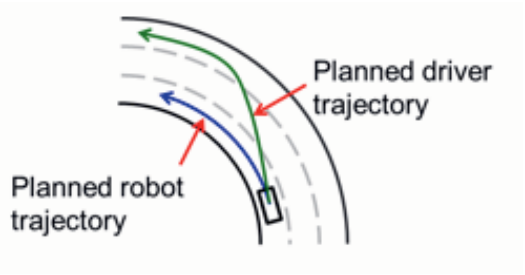


Fig. 2: Operator seeks to assume control to change trajectory of automated vehicle

The operator is subject to the following vulnerabilities:

- **Loss of Situation Awareness** [2], i.e. the ability to perceive elements in the environment (e.g. a nearby vehicle), comprehend the current situation (e.g. nearby vehicle indicates a lane change) and project the future status (e.g. nearby vehicle will pull out in front). Accordingly, drivers with lower situation awareness take additional time to anticipate latent hazards [3].

- **Mode confusion** [4], i.e. confusion over “who” has authority over “what” controls at any given time. The operator may not fully understand the mode of automation and its limitations.
- **Skills degradation and impaired performance** [5], i.e. the operator gradually loses the skills of performing the driving task. This decay of skills may result in exaggerated steering corrections and a wider variation of lateral vehicle positions in the driving lane.
- **Workload fluctuation** [6], i.e. during automated driving the workload is low whereas during and after the handover it may be high, causing an overload condition that negatively impacts the performance of the driver. The workload level can be assessed by measuring eye movements of the driver, such as gaze dispersion, blink frequency and duration.

The outputs of the proposed handover process are the driver’s performance during and after the handover as well as her acceptance of the handover process. A non-comprehensive list of metrics to assess the driver’s performance includes driver reaction times, steering wheel reversal rate, lateral lane position, time-to-contact e.g. to a leading vehicle, time-to-lane crossing, steering effort, mode awareness and situation awareness [7].

Considered Use-Case for Dual Mode vehicles

The driver re-engagement from a Level 4 Automated Driving System (ADS) can occur in two ways (see Figure 3): Either the vehicle is approaching the exit of an Operational Design Domain (ODD), such as a geo-fenced highway, and the operator is prompted to change her role from a passenger to a driver. Alternatively, the operator voluntarily requests the control of the vehicle and thus subsequently assumes the role of the driver. The following discussions are based on the latter use case.

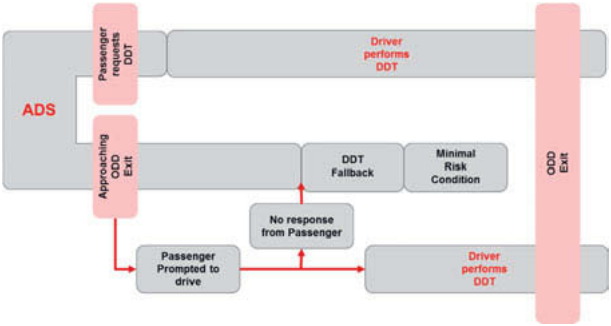


Fig. 3: Example Use-Cases for Driver Re-engagement from L4 Automated Driving System

Experimental Configurations

The main body of work in this paper was performed with two vehicle configurations:

- **Steer-by-Wire**: The vehicle has a Steer-by-Wire system implemented, i.e. there is no mechanical connection between the road wheels and the steering wheel. The automated driving mode is emulated via a GPS-guided controller that steers the vehicle on a fixed course on a proving ground (Figure 4).

Fig. 4: Proving ground setup

- **Wizard-of-Oz (WiOz)**: This vehicle features two sets of steering wheels and two sets of pedals. The automated driving mode is emulated via a second driver (Figure 5).

Fig. 5: Wizard-of-Oz experimental vehicle

The Steer-by-Wire–based vehicle offers a more realistic automatic driving experience; however, it is limited to a fixed driving course. The WiOz offers a wider range of driving manoeuvres.

The experimental design was based on previous foundational work in a laboratory environment [8] and in a simulator with six degrees of freedom in movement [1]. Apart from the experiments in the laboratory, which took place in a static environment, all other experiments occurred in a dynamic setup, i.e. the test objects were driving in a vehicle or sitting in a moving simulator. All the above experimental setups were based on a similar approach (Figure 6):

1. Test subjects drive in a highly automated mode while being induced into an OOTL-state. The tasks vary from reading and comprehending a text, solving a demanding quiz or working on work emails and texting on a mobile phone.
2. Control is handed over from automated mode to manual driving mode.
3. Test subjects have to perform a driving maneuver immediately after handover while their performance is evaluated.
4. Driver and vehicle are stabilized and repositioned.
5. Driver’s subjective perception is recorded with a questionnaire.

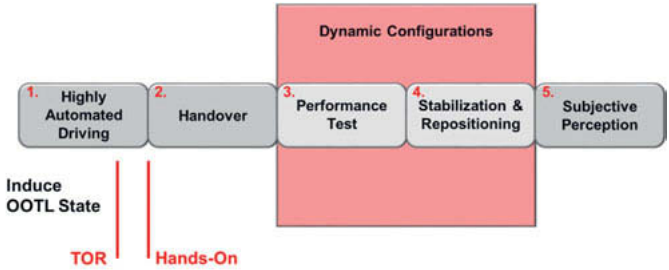


Fig. 6: Experimental approach

Findings from foundational work

The work by Louw [1] concludes that drivers who focussed their gaze early and more consistently towards the point of a potential hazard – the road centre (Figure 7) – were more likely to avoid a crash than those who were late to fixate on the potential hazard. By avoiding indecisive eye scanning, the test subjects could make better decisions and execute a safer driving manoeuvre. The study also concludes that the further drivers were taken OOTL, the worse was their ability to recognize and respond to road-related hazards.

Fig. 7: Setup from [1]

The study [8] concludes that an intuitive interaction of the driver with the steering wheel helps her to focus on the driving task faster. Generally, the study participants wanted to assume steering control as soon as their hands grabbed the steering wheel. This went hand-in-hand with the desire for a fast deployment of a steering wheel that was stowed away in the instrument panel during the automated driving mode. Taking the eyes off the road to find a switch was not accepted. Also, there is a need to offer a mobile phone friendly environment because most participants refused to put their mobile phone to the side during handover.

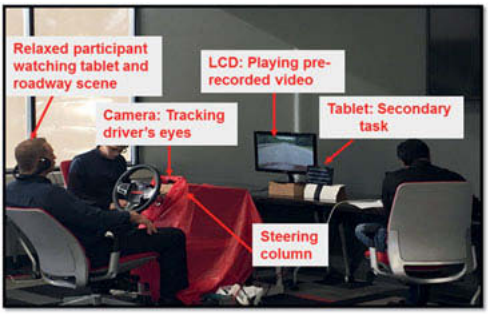


Fig. 8: Setup from [8]

From this foundational work and a theoretical analysis [7] it was concluded that the operator performance during handover and her acceptance of the process may vary with the applied handover strategy. From [9] and [10], the following handover procedures can be identified:

1. Immediate handover:
   Shifts the control from system to driver from one instant to the next. The foundational work from above suggests that the operator prefers this procedure; however, her performance and thus safety can be compromised without any form of additional support.
2. Stepwise handover:
   Longitudinal and lateral control occur at different instances. The evaluation is essentially the same as in 1.
3. Driver-monitored handover:
   e.g. countdown to transfer control after the driver initiated the handover. This mode also does not take any additional help from an automated system into consideration.
4. System-monitored handover:
   System monitors the driver and interferes, if needed. In this case, the driver does not have the freedom to override the system e.g. to compensate for system limitations.
5. Shared haptic control:
   Enables immediate control of driver, but with support from the system. Driver has ability to override the system if needed.

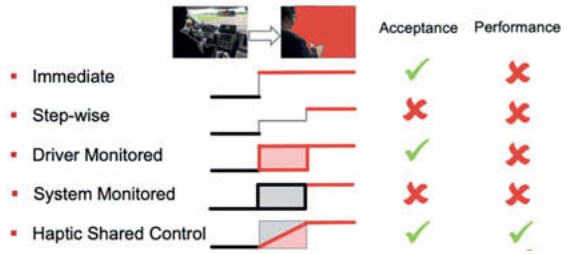


Fig. 9: Evaluation of potential handover procedures

On the basis of this evaluation, shared haptic control was chosen to be the focus of the main study.

Shared haptic control and Steer-by-Wire

Shared haptic control is a framework whereby human and automation cooperate to achieve the required control action together [10]. This enables communication and mediation between the operator and the automated system (Figure 10).

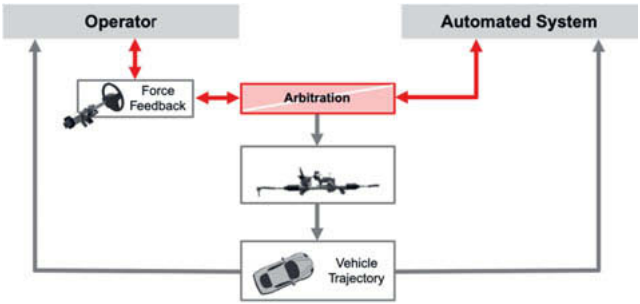


Fig. 10: Shared haptic control

To maximize the degrees of freedom in providing the driver haptic feedback while keeping the vehicle stable, a Steer-by-Wire architecture was chosen (Figure 11). This system consists of two parts:

1. Force Feedback Actuator:
   Provides haptic guidance to the operator. The torque felt by the driver emulates the same steering feel as a conventional steering system with an intermediate shaft that connects to the front wheels. Additional torque signals can be overlaid without impacting the steering angle and thus the trajectory of the vehicle. Depending on the situation, the steering ratio, i.e. the ratio of the steering wheel angle and the angle of the road wheels, can be modified within an instant. This helps the operator to keep control of the vehicle in any driving situation.
2. Road Wheel Actuator:
   Executes the arbitrated steering angle command generated by the operator via the force feedback unit and the commands from the automated system.

Fig. 11: Steer-by-Wire arrangement

Results gained from Steer-by-Wire test vehicle

- Gaze and steering performance are viable metrics for driver readiness to assume control:
  The test results confirm the findings from [1] that gaze and steering performance constitute viable metrics for situation awareness and driver performance. The gaze of 200 participants of different ages, genders, ethnicities, eye colors and types and strengths of glasses were tested under varying light conditions in the vehicle. The results show that gaze alone is not reliable enough to determine whether the driver is ready, willing and capable to safely assume control of the vehicle (Figure 12).

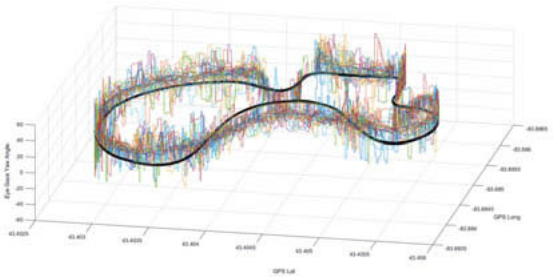


Fig. 12: Gaze dispersion of participants during test track drive

Whereas the steering-related measurements are very reliable and accurately represent the drivers' control over the vehicle (Figure 13).

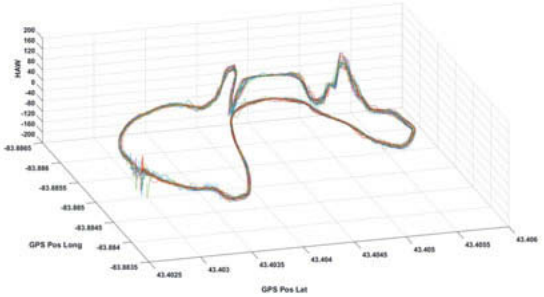


Fig. 13: Steering angle distribution of participants during test track drive

- Once driver wants to take over control, she needs perception of control within three seconds or faster

  Any delays beyond 3 seconds were not accepted by any participant. All of the 30 participants understood that the position of the steering wheel (either stowed or un-stowed) was the indicator of who was in charge of controlling the vehicle. A moving steering column, i.e. the process of un-stowing the steering wheel, was also accepted, as long as the driver had control while turning the steering wheel.

- The state machine needs to be robust to a number of corner Use Cases

In particular the following Use Cases need to be taken into account in the state machine design:

1. Intended handover by driver who is aware of the surrounding situation.
2. A distracted driver requests a handover of control.
3. The driver initiates the handover procedure unintentionally.
4. The driver initiates the handover procedure but changes her mind during the process.
5. The driver wants to override the automated system, exceeding a steering torque threshold.

- Visual signals (Figure 14) can supplement communication to driver during handover but are not as effective as haptic feedback

All test subjects were familiar with the driving course prior to the experiments. No unexpected driving situations occurred. However, the handover process led to a high work load for the driver. As a result, 13 out of 40 drivers did not notice changing light patterns in front of them during the procedure and 30 out of 40 drivers did not register audio signals to mark changing driving modes.

Fig. 14: Visual Signals during handover procedure

Conclusion: Steer-by-Wire–enabled architecture for handover

The state machine (Figure 15) comprises the states “manual”, “autopilot” and “shared” modes. The signals comprise “activate autopilot / hands-on”, “hands-off”, “torque override” and the “confidence level” that the driver is able and willing to assume control switch between the states.

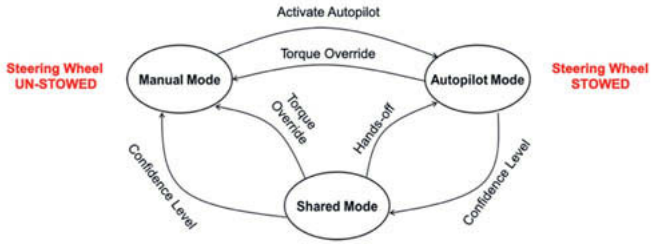


Fig. 15: State Machine

Steer-by-Wire enables independent shared torque and steering angle controls (Figure 16).

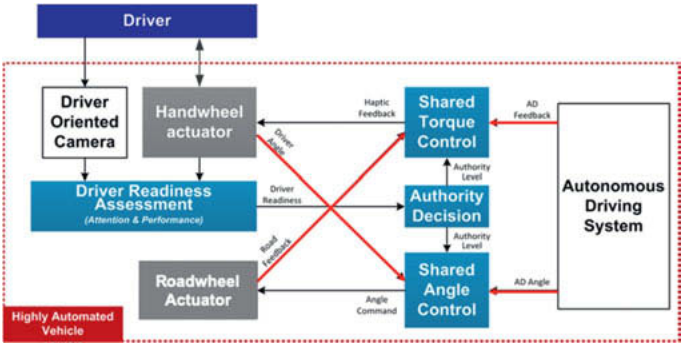


Fig. 16: Functional architecture for shared torque and steering angle control

The driver submits her intent via the steering wheel / handwheel actuator. Based on the calculated confidence in the driver's intent and ability, the shared angle control sends steering angle commands to the roadwheel actuator to control the trajectory of the vehicle. The driver receives feedback via force feedback, which is generated by the handwheel actuator, to assume full control of the vehicle faster and safer.

The timing of the handover and degree of driver control at any time is determined by the software component "driver readiness assessment". The driver maintains the ultimate control by having the ability to override the automatic system by exerting an override torque.

Summary

- The handover from automatic to manual (assisted) driving mode is NOT a point in time but a period of time.
- The Steering Wheel becomes a bi-directional HMI Element (handwheel actuator) to facilitate Driver / System Communication and mediates between the Automated Driving System and the Driver.
- The handover procedure requires routines designed to assess whether and to what degree the operator is performing the role specified for her.
- Steer-by-Wire offers an additional degree of freedom to interpret the driver's intentions and her readiness as well as the means to compensate for driver errors.

[1] Louw T. (2017), The human factors of transitions in highly automated driving, PHD thesis, Institute for Transport Studies, Leeds

[2] Endsley, Mica; Jones, Debra (2016-04-19). Designing for Situation Awareness (Second ed.). CRC Press

[3] Samuel, Borowsky, Zilberstein & Fischer (2015), Minimum time to Situation Awareness in scenarios involving transfer of control from an automated driving suite, transportation Research Record Journal of the Transportation Research Board 2602(2602):115-120

[4] Victoria A. Banks, Neville A. Stanton (2015), Discovering driver-vehicle coordination problems in future automated control systems: Evidence from verbal commentaries, 6th International Conference on Applied Human Factors and Ergonomics

[5] Merat N. et. al. (2014), Transition to manual: Driver behavior when resuming control from a highly automated vehicle, Transportation Research Part F, 274 – 282

[6] de Waard (1996), The measurement of drivers' mental workload, PHD thesis, The Traffic Research Centre VSC, University of Groningen

[7] Maggi D. (2018), Seamless and safe transfer of control authority exploring haptic shared control during handovers, PHD transfer report, Institute for Transport Studies, Leeds

[8] Klesing, Zuraski, Rezaeian (2018), Steering on Demand for dual-mode vehicles, 9th International Munich Chassis Symposium 2018

[9] Marcel Walch, Kristin Lange, Martin Baumann, and Michael Weber. Autonomous driving: Investigating the feasibility of car-driver handover assistance. In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '15, pages 11–18, New York, NY, USA, 2015.

[10] David A. Abbink, Mark Mulder, and Erwin R. Boer. Haptic shared control: smoothly shifting control authority? Cognition, Technology & Work, 14(1):19–28, Mar 2012. ISSN 1435-5566.

User-centred development of a display concept for fully automated driving

A methodical approach

M. Sc. **Leonie Gauer**, Dr**. Ingo Totzke**,
Audi Electronics Venture GmbH, Gaimersheim

Abstract

The role of the driver is changing with increasing vehicle automation from driving actively to being a passive passenger after having handed over the driving task to the vehicle entirely. This implies both fears and opportunities**.** Fearing to lose control due to the elimination on interaction options is accompanied by the comfort of not having to focus on the driving task. In order to introduce the driver to his new role as a passenger, the presentation of driving related information has to be adapted to the drivers´ needs. Therefore, it is necessary to understand the drivers´ needs during the initial contact with fully automated driving (SAE Level4/5) and to consider them in the development process of HMI-concepts.

To make sure the driver is introduced to the new role adequately, the User Centered Design Approach was chosen. In this approach quantitative and qualitative methods, (i.e., focus group discussions and Wizard of Oz driving studies as part of the context analysis), were used. Based on these results, user needs and requirements were deducted and implemented in a display concept for fully automated driving. This will be validated with the user´s involvement.

In this paper, the User Centered Design Approach to the development of an information concept will be presented and supplemented with selected results.

1. Introduction

Increasing vehicle automation indicates remarkable changes in the driver vehicle interaction [1]. The role of the driver changes from active driving over supervising the system to being a passenger [2] in fully automated driving (SAE Level 4 & 5 [3]). This implies both fears and opportunities for the driver**.** Fearing to lose control due to the elimination on interaction options is accompanied by the comfort of not having to focus on the driving task [4]. With regard to these changes, developing system trust and comfort are major challenges in the field of human factors [5]. These challenges have not been solved yet. That becomes apparent when people are asked which driving mode they would prefer. They indicate that manual driving is the most

enjoyable mode, whereas fully automated driving would be the least enjoyable one [6]. This implies that skepticism and a low level of trust predominates the common attitudes and expectations of prospective users towards fully automated driving. Particularly for the initial contact with fully automated driving, it is important to adapt the human-machine-interaction to the new conditions.

In consideration of these premises, it is crucial to allocate feedback on automation states and behaviors to the driver [7] in order to gain understanding of the automation and avoid mistrust or even distrust [8] [9] [2] [10] [11] [12] [13] [14]. By presenting feedback the driver gets the chance to develop an adequate mental model towards automated driving [15]. This feedback should be provided in a simple and uncomplicated manner [16].

However, deciding how much as well as what kind of feedback and advice is required is an essential challenge for HMI design [17].

Facing this challenge, it is essential to include the drivers´ needs and requirements into the design process of complex human machine systems to identify how much feedback and advice is required. By involving the driver in the development process of an information concept for the initial contact with fully automated driving, a successful and long-term interaction with the system could be ensured [18]. With this background it is reasonable to refer to the so called User Centered Design Process (UCDP). So the driver becomes a user. The UCDP is “an approach for a system’s design and development, which aims at making interactive system’s more usable by focusing on the use of the interactive system and by applying knowledge and methods from the areas of occupational science / ergonomics and usability” ([19] p. 10). The UCDP is characterized by four general phases, in which the user is involved: The first phase includes the understanding of the user by analyzing the user’s context. In the following phase users´ needs and requirements are derived. Based on these results, design solutions and prototypical concepts are generated. The validation of the created solutions marks the last step in this process. This process can be repeated iteratively until the developed solution fulfills the user´s needs and requirements. Fig. 1 shows the schematic overview about the UCDP [19].

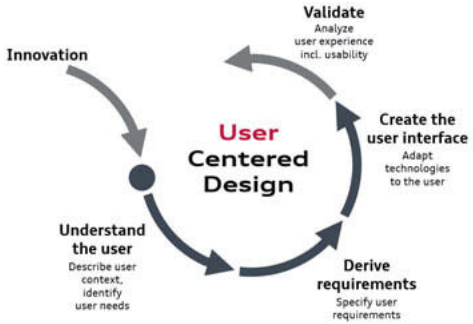


Fig. 1: User Centered Design Process (own representation, adapted from [19])

By this approach an information concept for the initial contact with fully automated driving is developed. To understand and define the users' context a literature research was accompanied by a mixed-method design. The design combined quantitative data from real driving studies in a Wizard of Oz setting [20, 21] with qualitative data derived from focus group discussions. Selected results of this context analysis will be presented in this contribution.

2. Focus group discussion

Aim of the focus group discussion was to gather a deeper understanding of information needs of potential users during the initial contact with fully automated driving. Focus groups were chosen in order to get an in-depth understanding of the users perceptions and feelings about issues or products in collaborative group atmosphere [22].

Procedure: After welcoming the participants, they were made familiar with the concept of fully automated driving by a video with fully automated driving situations. In the main part of the discussion four specific scenarios during a fully automated ride were presented by videos:

1. **Motorway:** Ego vehicle drives onto the highway
2. **Overtaking:** Ego vehicle overtakes a bus on the highway
3. **Construction zone:** Ego vehicle drives into construction zone with track narrowing
4. **City:** Ego vehicle drives through a city with traffic light, pedestrians and oncoming traffic

The scenario were discussed consecutively with regard to the expected information needs of potential users, aka the participants, during the first ride in an autonomous vehicle. Finally the participants prioritized the information needs by importance for a trustworthy and comfortable ride.

Each session took approximately two hours. The sessions were recorded via audio and video

equipment and were transcribed afterwards. After the process of material reduction (paraphrasing, selecting and bundling of the material) the transcripts were analyzed using qualitative content analysis [23]. 373 statements have been merged into six information categories by two independent raters. The inductive development of categories were chosen in order to develop the aspects as near as possible to the material.

Participants: *N* = 28 took part in 6 separate focus group discussions (*n* = 8 male, *n* = 20 female) aged between 24 to 53 years (mean age 31.29 years, *SD* = 7.63). *N* = 27 participants hold a driving license for on average 13.81 years (*SD* = 7.69), n = 1 participant hold no license. *N* = 15 Participants (53 %) stated to possibly use fully automated driving because of relaxation reasons, safety issues or time profit. Participants who would not use fully automated driving (n = 13, 47 %) indicated fear to lose driving and control over the vehicle as well as being mis- or even distrustful.

Results: With respect to the everyday scenarios (e.g., motorway and overtaking on a motorway) the participants requested most strongly for the following information:

- Detection of other vehicle – “Does my car detect everything important?”
- Current speed – “How fast is my car driving?”
- Upcoming maneuver – “What will my car do?”

In specific driving situations (e.g., construction zones), participants ask for relevant information which are specific for this situation. Examples were:

- Detection of the changed road management
- Width of roadway

However, participants prefer to get information about which dynamic and static elements of the driving situation were detected especially in the city scenario:

- Detection of pedestrians
- Detection of obstacles
- Detection and observing of traffic signs

To sum up, Fig. 2 gives an overview which information categories are requested by the participants. Most information needs relates to maneuver information (37.53 %) and environmental information (27.73 %). Navigational (14.75 %) as well as driving relevant information (12.86 %) are requested less often. Needs regarding system functionality are mentioned by 7.51 %. Options to influence the systems behavior are only mentioned by a minority of participants (1.61 %). This indicates that the participants understand their role as passenger in a fully automated vehicle.

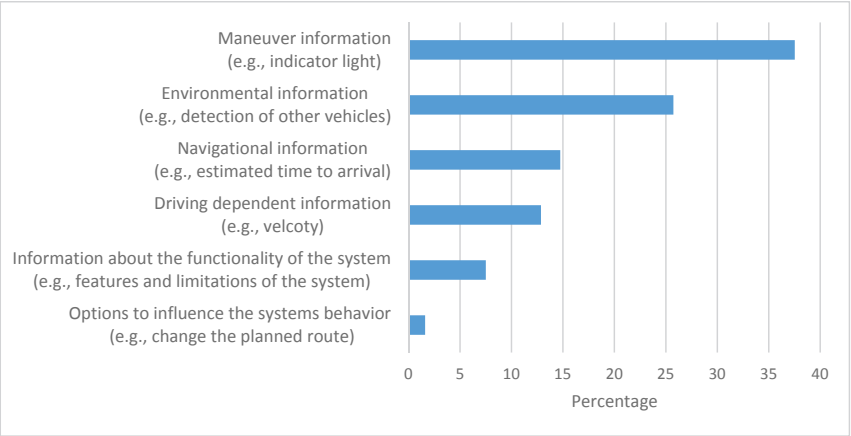


Fig. 2: Percentage of naming information categories in focus groups (based on 373 statements of *N* = 28 participants)

In general, several arguments in the focus group discussion refer to overall evaluation of fully automated driving. These arguments can be merged into two major user needs: Transparency and predictability of the system to gain trust and perceived comfort.

Transparency should be provided by displaying the current situation and how the system handle the situation. This includes displaying environmental information and the upcoming maneuver. Displaying environmental information means the detection of driving-dependent factors in real time. The upcoming maneuver should be comprehensible to the own driving abilities and explained (e.g. braking because of pedestrians on the crosswalk).

Predictability should be provided by displaying a preview of the next planed maneuver. An early announcement if driving behavior changes, were considered especially for maneuvers with strong dynamics, i.e. emergency braking. In addition, navigational information about routing and road management should be provided as well. The participants stated, they would adapt their non-driving related activities based on these information. Navigational information like points of interest could also be used to learn about the environment. In addition, participants would also use the navigation information to intervene the guidance for having a brake or to refuel.

Over the period of the discussion the participants assume, their information needs will change in dependence of the experience with the system. Information needs will decrease with greater experience and trust in the system. For this purpose participants suggest to make the information display customizable.

Not only growing experience has an influence on the need of information, but also interindividual differences influence these needs. If the amount of information needs (measured by the participant with the greatest need for information) is related to the attitude regarding fully automated driving expressed by the respondents and rated by two independent raters, three groups of approx. equal size are identified (see Fig. 3):

- The technically interested
- The safety-oriented
- The ingenious

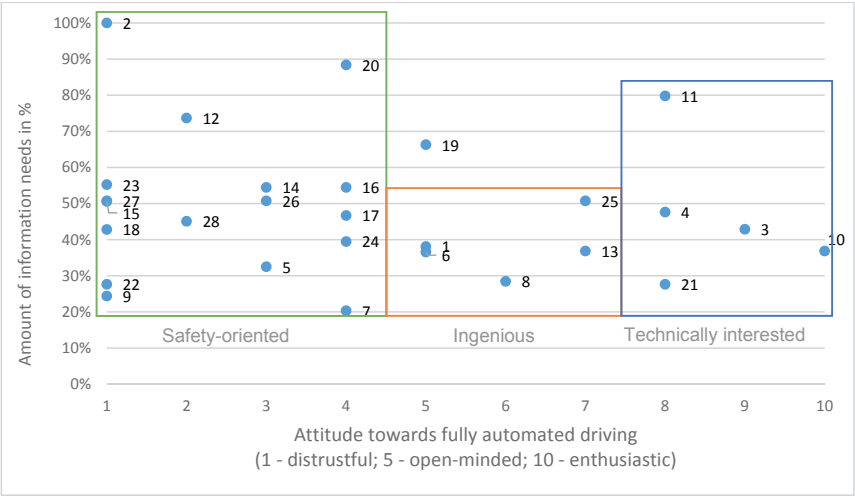


Fig. 3: Amount of information needs depending on the attitude towards fully automated driving

The technically interested participants are fascinated by driving fully automated. These participants like to see the sensor activity and the function potentiality in detail. The safety-oriented participants need a large amount of information. They want to make sure that everything important is detected especially in prospective dangerous situation. The ingenious are sure that the system will manage every situation by its own therefore these participants generally need less information to feel comfortable.

Based on this findings it is important to provide differentiated information.

3. Real driving Wizard of Oz studies

Aims of the real driving studies were to understand more about the real experience of the initial contact with fully automated driving. In order to enable participants to experience fully automated driving, a Wizard of Oz setting were chosen. Three consecutive studies were conducted with different focuses. Beside others, the following questions were considered:

- Which information needs do users express during fully automated driving?
- How does the experience influence perceived comfort and trust in the system?
- How can the gaze behavior be described?

Experiment vehicle and route: A Audi Q7 was prepared as a Wizard of Oz vehicle. An in-vehicle experimenter (seated in the rear seat or the passenger seat) drove the vehicle by using a joystick-system. This joystick were covered by a construction so the participant got the impression that he/she was seated in al fully automated vehicle. A cover story were used to explain the presence of the wizard: The wizard were presented as intern student or safety driver [20].

In order to create a nearly information-free environment the cluster element were covered as well as any mirrors (side mirror and interior mirror) were adjusted.

The driving studies took place at the AUDI test track in Neuburg an der Donau (Germany). The routes consisted of 10 to 12 driving scenarios each. These scenarios were derived from lists about relevant inner-city driving situation (e.g., overtaking a vehicle, pedestrians crossing) and were validated in a previous study [24]. Each drive lasted 5-7 minutes. See Fig. 4 for an example of a test track.

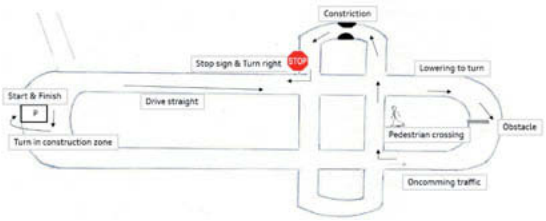


Fig. 4: Test track (example)

Procedure: The study procedure were divided into three parts: Firstly, after arriving and reception, the participants were informed about the experiment´s procedure. The participants were instructed, that they will experience their first autonomous rides and the experimenter is interested in their experiences.

Secondly, the rides took place. Each participant completed three consecutive rides in the experiment vehicle. After each ride, participants answered questionnaires regarding trust [25] and eeriness [26], amongst others. After the last ride, the participants should additionally rate the presentation and the absence of specific driving related information which could be displayed during a fully-automated ride (KANO-method) [27].
Finally, the participants were interviewed about their experience. They were enlightened about the Wizard of Oz setting and remunerated for their effort.

Participants: Overall, the sample includes *N* = 78 participants (Study 1: *n* = 31, study 2: *n* = 32, study 3 *n* = 15). *N* = 45 men and *n* = 33 women are on average 36 years old (*SD* = 14.76) and hold a driving license on average 19 years (*SD* = 14.76). *N* = 59 participants (75.6 %) stated they would use fully automated driving.

Results: With regard to the ranking of several driving related information with the KANO-method [27], some information are robust basic information at all, i.e. fuel level, own route, current speed and system activity. The estimated time to arrival is identified as performance attribute. Information about the environment (e.g. coming rest areas) and about sights along the route are assigned to excitement factor. See Fig. 5 for the visualized results of the KANO-method. [27]

Fig. 5 Selected results of the information-ranking [27]

Beside the explicit information needs, the development of trust and perceived eeriness within three rides is investigated. The results of a one-way repeated-measures ANOVA overall studies indicates, the more experience the more trust (F (2, 154) = 7.86, p = .001 η^2 = .093) and

less eeriness (F (2, 148) = 4.722, p = .010 η^2 = .060) is expressed by the participants in the questionnaires (see Fig. 6).

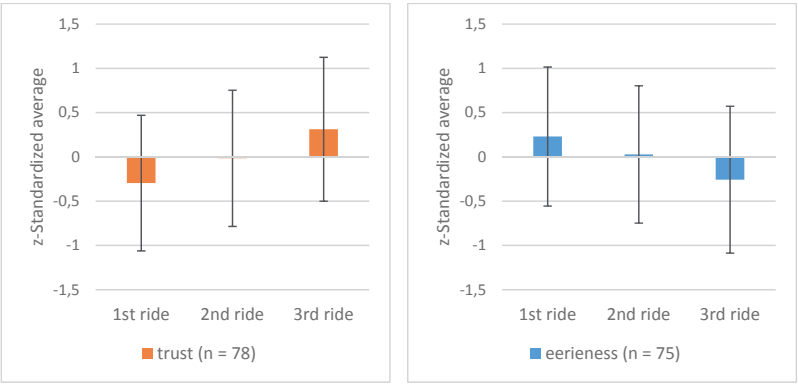


Fig. 6: Trust and perceived eeriness over three measuring time points

4. Discussion and conclusion

The current study examined the experience with fully automated driving in a multi-methodical approach. The studies were aimed to identify the context of use during the initial contact with fully automated driving. Therefore the following questions were considered: Are information needed at all during fully automated driving? What information needs express users? How do users experience fully automated driving? How experience changes the information needs? The results of the studies allows first indications for the users´ context of the initial contact with fully automated driving:

User needs several information during the initial contact with fully automated driving

Regarding the expressed information needs, the results of focus groups and the information ranking indicates that user needs information at all [14]. They require information regarding transparency and predictability [2, 5] in order to know the fully automated vehicle acts as responsible and reliable as the user itself. The user wants to be able to check, if the vehicles intention meets the exceptions of the user. Therefore, users request displaying of information concerning the driving dependent environment and the driving behavior of the autonomous system. It is not surprising that the user requires basic information, i.e. current speed and current fuel level. The users´ expectation to receive information about the functionality suggests that they need adequate knowledge about the system to create a mental model and gain trust in the automated system [9, 10, 14].

Information needs change with experience

The development of trust could be demonstrated by the Wizard of Oz studies. With growing experience the users individual trust level increases [2, 8]. It could be shown that skepticism and the perceived eeriness regarding the autonomous system decreases [26]. The participants of the focus groups assumed this as well. With increasing trust the amount of expected information decreases, which means an information concept should be customizable.

Information needs differ between individuals

The difference between the users became obvious especially in the focus group analysis. The users differ in their information needs and their displaying expectations. Some users are fascinated by driving fully automated and technically interested. They would prefer an information concept filled up with technical information. Another user group need a lot of displayed information about the detected environment to make sure they are safe in the automated vehicle. A third user group do not want to be disturbed by an excess of information and prefer reduced information concepts.

Involve the user early in the HMI development process

The presented methodical approach combined quantitative and qualitative data collection and offers the possibility to develop a multifaceted understanding of the user's context during the initial contact with fully automated driving. It shows the importance to involve the user in the development of HMI concepts to make sure not to evade the users 'needs [18]. The involvement of the user even in early phases of the development process could enable the developer of HMI concepts to focus on the features and information content the user really needs to know.

References

[1] Bengler, K.: Driver and Driving Experience in Cars. In: Meixner, G. u. Müller, C. (Hrsg.): Automotive user interfaces. Creating interactive experiences in the car. Human–Computer Interaction Series. Springer International Publishing 2017

[2] Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J., Othersen, I. u. Petermann-Stock, I.: What would drivers like to know during automated driving? Information needs at different levels of automation. 7. Tagung Fahrerassistenz, München, 25.-26.11.2015. (2015)

[3] J3016 (TM) Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, SAE, 2016

[4] Müller, A., Stockinger, C., Walter, J., Heuser, T., Abendroth, B. u. Bruder, R.: Einflussfaktoren auf die Akzeptanz des automatisierten Fahrens aus der Sicht von Fahrerinnen und Fahrern. (Wie) Wollen wir automatisiert fahren? 2017, S. 1–22

[5] Diels, C. u. Thompson, S.: Information Expectations in Highly and Fully Automated Vehicles. Springer, Cham 2017. https://link.springer.com/content/pdf/10.1007%2F978-3-319-60441-1_71.pdf

[6] Kyriakidis, M., Happee, R. u. Winter, J.C.F. de: Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. Transportation Research Part F: Traffic Psychology and Behaviour 32 (2015), S. 127–140

[7] Wickens, C. D., Hollands, J. D., Banbury, S. u. Parasuraman, R.: Engineering Psychology and Human Performances. Boston: Pearson Education 2013

[8] Lee, J. D. u. See, K. A.: Trust in Automation: Designing for Appropriate Reliance. Human Factors (Human Factors: The Journal of the Human Factors and Ergonomics Society) 46 (2004) 1, S. 50–80

[9] Weißgerber, T., Damböck, D., Kienle, M. u. Bengler, K.: Erprobung einer kontaktanalogen Anzeige für Fahrerassistenzsysteme beim hochautomatisierten Fahren. "5. Tagung Fahrerassistenz". München 2012

[10] König, W.: Nutzergerechte Entwicklung der Mensch-Maschine-Interaktion von Fahrerassistenzsystemen. In: Winner, H. e. a. (Hrsg.): Handbuch Fahrerassistenzsysteme. Grundlagen, Komponenten und Systeme für aktiver Sicherheit und Komfort. 2015, S. 621–632

[11] Seppelt, B. D. u. Victor, T. W.: Potential Solutions to Human Factors Challenges in Road Vehicle Automation. In: Meyer, G. u. Beiker, S. (Hrsg.): Road Vehicle Automation 3. Switzerland: Springer International Publishing 2016, S. 131–148

[12] Walch, M., Mühl, K., Kraus, J., Stoll, T., Baumann, M. u. Weber, M.: From Car-Driver-Handovers to Cooperative Interfaces: Visions for Driver–Vehicle Interaction in Automated Driving. In: Meixner, G. u. Müller, C. (Hrsg.): Automotive user interfaces. Creating interactive experiences in the car. Human–Computer Interaction Series. Springer International Publishing 2017, S. 273–294

[13] Körber, M., Baseler, E. u. Bengler, K.: Introduction matters: Manipulating trust in automation and reliance in automated driving. Applied ergonomics 66 (2018), S. 18–31

[14] Wiegand, G., Schmidmaier, M., Weber, T., Liu, Y. u. Hussmann, H.: I Drive - You Trust. CHI 2019. New York: ACM op. 2019, S. 1–6

[15] Geisler, S.: Von Fahrerinformation über Fahrerassistenz zum autonomen Fahren. In: Reuter, C. (Hrsg.): Sicherheitskritische Mensch-Computer-Interaktion. Interaktive Technologien und soziale Medien im Krisen- und Sicherheitsmanagement. 2018, S. 357–376

[16] Guenes, E. B., Hottelart, K. u. Reilhac, P.: The Digital Driver of the Future—User Experience Research on Generation Z in Germany. In: Meyer, G. u. Beiker, S. (Hrsg.): Road Vehicle Automation 4. Cham, Switzerland: Springer International Publishing 2018, S. 57–68

[17] Simon, J. H.: Learning to drive with Advanced Driver Assistance Systems. Empirical studies of an online tutor and a personalised warning display on the effects of learnability and the acquisition of skill. Dissertation. 2005

[18] Grandt, M. u. Ley, D.: Unterstützung von Entscheidungsprozessen durch benutzerzentrierte Gestaltung von Führungssystemen. In: Schmidt, L., Schlick, C. M. u. Grosche, J. (Hrsg.): Ergonomie und Mensch-Maschine-Systeme. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg 2008, S. 79–102

[19] DIN EN ISO 9241-210:2010-10. *Ergonomie der Mensch-System-Interaktion. Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme*

[20] Gauer, L.: Erprobung der Methodik “Wizard of Oz”. 12. Doktorandenworkshop der Fachgruppe Verkehrspsychologie der DGPS. Aachen 2018

[21] Gauer, L., Totzke, I. u. Zehetleitner, M.: Fahrer oder Beifahrer – das ist hier die Frage. In: Vollrath, M. (Hrsg.): 3. Kongress der Fachgruppe Verkehrspsychologie. „Mehr Mensch im Verkehr?“. 2019, S. 10

[22] Krueger, R. A. u. Casey, M. A.: Focus groups. A practical guide for applied research. Singapore: SAGE Publications 2015

[23] Mayring, P.: Qualitative Inhaltsanalyse. Weinheim, Basel: Beltz Verlag 2015

[24] Gauer, L., Müller, J. u. Totzke, I.: Bewertung kritischer Fahrsituationen während einer vollautomatisierten Fahrt. In: Alexander Schütz, Anna Schubö, Dominik Endres, Harald Lachnit (Hrsg.): 60. TeaP 2018. Abstracts of the 60th Conference of Experimental Psychologists. Lengerich: Papst Science Publisher

[25] Jian, J.-Y., Bisantz, A. M. u. Drury, C. G.: Foundations for an Empirically Determined Scale of Trust in Automated Systems. International Journal of Cognitive Ergonomics 4 (2000) 1, S. 53–71

[26] Ho, C.-C. u. MacDorman, K. F.: Measuring the Uncanny Valley Effect. International Journal of Social Robotics 9 (2017) 1, S. 129–139

[27] Kano, N., Seraku, N. Takahasi, F. u. Tsuji, S.: Attractive quality and must-be quality. Journal of Japanese Society for Quality Control 14 (1984) 2, S. 147–156

UX in der Automobilwelt

Wie macht man sie vergleichbar?

Rico Ludwig, P3 automotive GmbH, Stuttgart

Kurzfassung

Die Digitalisierung der Automobilwelt hat dazu geführt, dass im zunehmend komplexen, vernetzten Fahrzeug neben harten Fakten wie Leistung oder Verbrauch auch weiche Faktoren Auswahlkriterien, etwa für eine Kaufentscheidung, sein können. Zu diesen weichen Faktoren zählt insbesondere das Thema User Experience (UX). User Experience wird heute maßgeblich durch die im Fahrzeug zur Verfügung stehenden Dienste beeinflusst. Fast jedes Neufahrzeug bietet vielfältige, oft internetfähige Dienste an, die das Erlebnis Autofahren noch besser machen und den Kunden zufriedener stimmen sollen.

Eine Herausforderung ist es, User Experience hinsichtlich der Erlebnisse und Erfahrungen im Fahrzeug messbar zu machen. In diesem Zusammenhang wurde untersucht, welche Einflussfaktoren und Messwerte die Beurteilung von UX in Fahrzeugen signifikant charakterisieren. Durch eine Kombination aus Relevanzbewertungen, Usability-Tests definierter Use Cases sowie Erfassung quantitativer Messgrößen ist schließlich eine quantitative Evaluierungskenngröße als Maß der User Experience erzeugt worden.

Das entwickelte Evaluierungsinstrument soll die Bildung eines ganzheitlichen Urteils über die jeweiligen Erlebnisse und Erfahrungen im Fahrzeug zulassen und somit eine Vergleichbarkeit der UX von gesamten Fahrzeugen aber auch von Entwicklungsständen (einzelner Dienste) ermöglichen. Dies unterstützt letztlich nicht nur Kunden bei ihrem Kaufentscheidungsprozess, sondern dient auch Herstellern bei der Entwicklung und Optimierung neuer sowie bestehender Lösungen.

Inzwischen wurde die Methode bereits bei mehreren Wettbewerbsvergleichen angewendet. Zurzeit wird eine angepasste Form für Connectivity-Tests eingesetzt, welche in Kooperation des Connect-Magazins und der Firma P3 durchgeführt werden.

Abstract

The digitization of the automotive industry has led to more complex and connected cars, that - besides hard factors like power or consumption - need soft factors being available as criteria

for purchasing decisions. One of these soft factors is user experience (UX). Today user experience is essentially influenced by services available inside the vehicle. Almost every new vehicle offers a variety of mostly Internet-enabled services to make the driving experience even better and to make customers happier.

One challenge is to make user experience measurable regarding experiences and knowledge inside the vehicle. In this context it has been researched what influencing factors significantly characterize the evaluation of UX inside vehicles. A combination of assessment of relevance, usability testing of pre-defined use cases as well as gathering of quantitative parameters helped create a characteristic variable as a measure of user experience.

The final evaluation tool is to allow the formation of a holistic opinion about experiences and knowledge within the vehicle and to enable a comparability of UX of complete vehicles as well as of developmental states (of single services). Ultimately this not only supports customers in their purchasing decision process, but also manufacturers developing and optimizing new and existing solutions.

Meanwhile, the method has been applied during several competitive comparisons. An adapted form is currently being used for connectivity tests, which are being carried out in cooperation with the Connect magazine and P3.

1. Einleitung zum Thema User Experience und Connected Car

User Experience oder kurz UX ist ein Schnittstellenthema aus den Bereichen Design, Technik und Psychologie. In den letzten 20-30 Jahren hat das Thema immer mehr an Bedeutung gewonnen. Dennoch gibt es bis heute keine einheitliche und vor allem allgemein anerkannte Definition. Eine der existierenden Definitionen ist beispielsweise in der ISO 9142-210 festgehalten. Sie lautet: “UX umfasst die Wahrnehmungen und Reaktionen einer Person, die aus der tatsächlichen und/oder der erwarteten Benutzung eines Produkts, eines Systems oder einer Dienstleistung resultieren“. Das bedeutet, um die User Experience, etwa eines Produktes oder Dienstes, zu erfassen, müssen Wahrnehmungen und Reaktionen von Nutzern evaluiert werden. Es ist somit nicht verwunderlich, dass dafür ebenso kein allgemeines Testverfahren existiert, welches am Ende einen konkreten Wert für die User Experience liefert.

Nun hat die Digitalisierung in der Automobilwelt dazu geführt, dass immer mehr digitale und vernetzte Dienste im Fahrzeug Einzug halten. Inzwischen gibt es Wetterdienste, Nachrichtendienste, Kalenderdienste, E-Mail-Dienste, Online-Radiodienste, Notruf- und Pannendienste, diverse Navigationsdienste, Musikstreaming-Dienste, Dienste zum Chatten, zum Spielen, zum Telefonieren, zum Fernsehen; Spracherkennung, Müdigkeitserkennung, man kann sein Fahrzeug fernsteuern oder den Schlüssel digital an Freunde senden, und natürlich gibt es einen

Browser zum Surfen im Internet direkt im Fahrzeug – im vernetzten Fahrzeug, dem Connected Car.
Waren Autokäufer in er Vergangenheit vorwiegend an PS, Hubraum, Drehzahl und Kofferraumvolumen interessiert, rücken heutzutage immer mehr die Gebrauchstauglichkeit und der Spaßfaktor bei der Benutzung eines potentiellen Neufahrzeugs in den Vordergrund. Und gleichwohl möchten Produktmanager und Entwickler heute wissen, wo sie mit ihren digitalen Innovationen im Wettbewerberfeld stehen. Diese Entwicklungen ziehen die Notwendigkeit einer Vergleichsgröße nach sich, um die User Experience ebendieser digitalen Dienste untereinander messbar zu machen.

2. Wie testet P3 die digitalen Dienste im Connected Car?
Bei der P3 automotive GmbH führen wir seit inzwischen mehr als fünf Jahren User Experience Benchmarks durch. 2014 haben wir erstmals mehrere Fahrzeuge bestellt und diese ausgiebig getestet. Die Ergebnisse unseres ersten Ausflugs in die Welt des User Experience Testings haben wir schließlich im connect Magazin veröffentlicht.

Nun sollte dieses Projekt nur der Startschuss in diesem Themenfeld sein, denn bereits ein Jahr später hatten wir unser Know-How ausgebaut und wurden schließlich beauftragt, zum ersten Mal einen User Experience Benchmark für einen unserer Kunden durchführen.
Für diesen Nutzertest führten wir ausführliche Untersuchungen durch und identifizierten für Nutzer relevante Dienste-Kategorien, wie Navigation oder Infotainment. Anschließend erarbeiteten wir spezielle Anwendungsfälle (Use Cases) und testeten diese live mit vorher akquirierten Testpersonen. Die Tester sollten sie sich zunächst in ein konkretes Szenario hineinversetzen, in welchem sie dann unterschiedliche Use Cases zu absolvieren hatten. Ein beispielhafter Use Case für die Kategorie *Infotainment* konnte hierbei sein: *Lassen Sie sich das Wetter an Ihrem Zielort anzeigen.* Die Tester wurden gebeten, diesen Use Case auszuführen und dabei ihre Gedanken und Gefühle sowie ihre positiven und negativen Erkenntnisse zu notieren.
Für jedes Fahrzeug mussten die Probanden verschiedene Use Cases an gleich drei sogenannten Touchpoints testen. Touchpoints sind als Berührungspunkte zwischen einem Benutzer und einem System (z.B. Marke) definiert. Unsere drei Touchpoints waren in erster Linie das Fahrzeug selbst aber auch die Smartphone-App des Herstellers sowie dessen Webportal, welches auf einem Laptop abgerufen wurde. Denn auch dort konnten und können Benutzer diverse Dienste einsehen und ausführen.
Nachdem nun von einem Tester nacheinander alle Use Cases für eine Dienste-Kategorie durchgespielt und alle Eindrücke notiert waren, wurde selbige Kategorie mit einer einfachen

7-Punkte-Likert Skala von gut bis schlecht bewertet. Dadurch ließ sich am Ende der Durchschnittswert für alle Probanden berechnen.
Dieses Prozedere wiederholte sich nun für alle Kategorien und Touchpoints. Zum Abschluss ließen wir die Tester den meCUE Fragebogen ausfüllen. Dieser basiert auf dem CUE-Modell von Manfred Thüring und Sascha Mahlke und unterscheidet zwischen der Wahrnehmung aufgabenbezogener und nicht-aufgabenbezogener Produktqualitäten, er erfasst zusätzlich aber auch Nutzeremotionen [1]. Dadurch wurde das getestete Fahrzeug noch einmal insgesamt detailliert bewertet.
Aus den Probandentests resultierten Richtwerte für die Qualität der UX aller Dienste-Kategorien. Zudem erhielten wir detailliertes, qualitatives Feedback zu den getesteten Diensten. Der Ergebnisse des meCUE Fragebogens waren letztlich ein weiterer Indikator für den Gesamtvergleich der Fahrzeuge.

Bild 1: UX Benchmark - P3 Connected Car Experience Days 2017

In den Folgejahren erhielten wir weitere Beauftragungen, um UX Benchmarks durchführen. Ein Höhepunkt waren die P3 Connected Car Experience Days (CCED) in Berlin, als wir die jeweils aktuellsten Fahrzeuge zehn verschiedener Hersteller getestet und die Ergebnisse zahlreichen Teilnehmern aus der Automobilbranche präsentiert und vor allem live in den Fahrzeugen vorgeführt. Bis zum heutigen Tag habe wir sehr viele Erfahrungen gesammelt und die

folgenden relevanten Fragestellungen identifiziert, welche im Vorfeld eines User Experience Testings für Fahrzeuge beantwortet werden müssen:

1. Was ist das Ziel des UX Benchmarks?
2. Wer sind die Tester und wenn ja, wie viele?
3. Welche Testmethodik wird angewendet?
4. Wie viele Fahrzeuge sollen getestet werden?

Was ist das Ziel des UX Benchmarks?

Die Frage, die sich grundsätzlich zu Beginn jedes Projekts stellt. Was soll mit dem Benchmark erreicht werden? Welche Erkenntnisse sollen gewonnen werden? Welche Ergebnistypen sollen generiert werden? Besonders wichtig ist es zu definieren, welche Dienste in welchem Umfang und an welchen Touchpoints getestet werden sollen.

Diese Frage ist für gewöhnlich abhängig vom Auftraggeber. Bei Marketing und Vertrieb wissen wir inzwischen, das andere Erkenntnisse wichtig sind als bei einem UX Benchmark für eine Entwicklungsabteilung. Während mitunter die UX von Verknüpfungsprozessen zwischen Fahrer und Fahrzeug im Mittelpunkt steht, ist bei anderen Aufträgen der detaillierte Blick auf Infotainment-Dienste wie beispielsweise Bluetooth-Audio-Streaming notwendig.

Dies gilt es vorab zu klären, denn auf diesen Entscheidungen bauen später die User Cases als Basis des gesamten Benchmarks auf.

Wer sind die Tester und wenn ja, wie viele?

Hierbei kann zunächst zwischen zwei Kategorien unterschieden werden. Entweder, der UX Benchmark wird von Experten durchgeführt oder es werden Probanden für die Tests akquiriert. Fällt die Entscheidung auf Experten, sind in der Regel zwei bis vier UX Experten ausreichend. UX Experten zeichnen sich grundsätzlich durch fundiertes Wissen in den Gebieten Usability und User Experience aus. Für UX Tests von Fahrzeugen sind weitreichende Kenntnisse im Bereich Automotive ebenfalls obligatorisch.

Anders sieht es bei Probanden aus, hier ist prinzipiell kein Vorwissen notwendig und teilweise auch unerwünscht. Jedoch bedarf es einer weitaus höheren Anzahl an Testpersonen. Hier kommt es wieder auf die Ziele des Benchmarks an: Sollen beispielsweise lediglich Usability Schwachstellen aufgedeckt werden, erzielt man bereits mit sechs bis acht Probanden gute Ergebnisse. Ein reliabler UX Test, der ein verlässliches Stimmungsbild zu einem Produkt oder Service liefern soll, erfordert jedoch wenigstens 20 Probanden.

Um ein konkreteres Bild zu erhalten ist es weiterhin sinnvoll, nur bestimmte Kundengruppen testen zu lassen und die Probanden vorher entsprechend auszuwählen.

Welche Testmethodik wird angewendet?

Die Testmethodik ist primär davon abhängig, ob es sich um einen Experten- oder einen Probandentest handelt. Bei erstgenanntem werden meist Heuristiken oder Checklisten eingesetzt. Für Probanden existieren hingegen von einfachen Fragebögen über Lautes Denken bis hin zum Eyetracking zahlreiche unterschiedliche Methoden.

Fragebögen sind die wohl am häufigsten verwendeten Mittel zur Bestimmung der Qualität von User Experience. Es gibt sie für sämtliche Aspekte, unter anderem zur Emotionserfassung, zur Erfassung subjektiver Usability und User Experience oder zur Erfassung der visuellen Ästhetik. Bekannte Beispiele sind das Affect Grid, das SAM, das AttrakDiff2 der UEQ oder der VisAWI. Vorteile sind die einfache Anwendung sowie Auswertung. Die Ergebnisse sind jedoch stets subjektiv und nicht während sondern erst nach einer Aufgabe erfassbar.

Bei der Methode Lautes Denken verbalisiert ein Proband während eines Tests seine eigenen Gedanken. Dies hilft, seine Denkprozesse bei der Bearbeitung von Aufgaben besser zu verstehen. Laut Jakob Nielsen könne Lautes Denken die Wertvollste Methode des Usability Engineerings sein [2]. Sie ist flexibel und mit geringem Aufwand einsetzbar. Allerdings kann Lautes Denken die kognitiven Prozesse beeinflussen und das Verhalten des Probanden ändern.

Eye Tracking eignet sich insbesondere für die Erfassung der visuellen Aufmerksamkeit, die ein Proband auf spezifische Bereiche richtet, z.B bei der Auswahl eines Radiosenders. Die Ergebnisse können anschließend zur Optimierung von Informationsdarstellungen dienen [3]. Zwar werden mittels Eye Tracking kleinste Details erfasst, ohne das Verhalten des Probanden zu stören. Dies ist jedoch mit hohen Anschaffungskosten verbunden.

Zusammenfassend ist zu sagen, dass jede Bewertungsmethode für die Bewertung der User Experience im Fahrzeug im Hinblick auf die zu erreichenden Ziele individuell gestaltet werden muss.

Wie viele Fahrzeuge sollen getestet werden?

Die Frage nach der Anzahl der zu testenden Fahrzeuge ist wichtig, denn von ihr hängen weitere Faktoren ab.

Zunächst ist die Fahrzeuganzahl entscheidend für die Dauer der Durchführung des Tests. Je mehr Fahrzeuge getestet werden sollen, desto mehr Zeit muss in der Regel insgesamt für den

Benchmark eingeplant werden. Zwar ließe sich dieses Problem durch Parallelisierung und bestimmte Studiendesigns gut lösen, jedoch steht selten oder nur mit erhöhtem Ressourcenaufwand die notwendige Anzahl an Probanden zur Verfügung.
Zusätzlich ist die Anzahl der Fahrzeuge wichtig für die Organisation, da im im Vorhinein eine ausreichend große und idealerweise überdachte Location bestimmt werden sollte.
Bei dieser Frage gehen die Meinungen zwischen unseren Kunden und uns nicht selten auseinander. Verständlicherweise soll zum eigenen Fahrzeug ein möglichst großer Vergleichspool erzeugt werden. Aus den oben genannten Gründen ist dies oft jedoch mit erheblichem Mehraufwand bzw. Mehrkosten verbunden. Aus unserer Erfahrung hat sich hier ein Testset von fünf Vergleichsfahrzeugen als optimal erwiesen.

3. Das UX Evaluierungstool für die CCED 2017

Die im folgenden Abschnitt vorgestellte Methode wurde bei den P3 Connected Car Experience Days 2017 angewendet. Die Konnektivitätsdienste der zehn aktuellsten, vernetzten Fahrzeuge verschiedener Hersteller wurden aus drei verschiedenen Perspektiven getestet:

1. Die nach Relevanz gewichtete Verfügbarkeit von Diensten
2. Die erlebte Nutzererfahrung der fünf wichtigsten Dienste
3. Die Erfüllung von Nutzerbedürfnissen

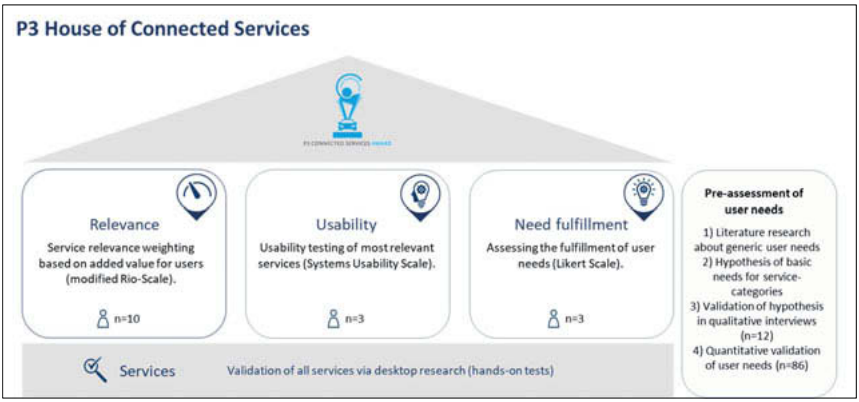


Bild 2: UX Testmethodik für die CCED 2017

Die Tests basierten auf einem globalen Dienste-Portfolio, welches alle zu der Zeit verfügbaren vernetzten Dienste enthielt. Das Portfolio wurde in neun verschiedene Kategorien unterteilt: Apple CarPlay, Kommunikation, End-to-End-Navigation, Unterhaltung, Innovation, Personalisierung, Remote-App und Sprachsteuerung. Die Relevanz eines jeden Dienstes aus dem Portfolio wurde von zehn Experten auf einer 5-Punkte-Likert-Skala mit „nicht relevant“ bis „extrem relevant“ bewertet. Die resultierenden Mittelwerte wurden auf einen Kategorierelevanzwert zwischen 0 und 1 normiert.
Danach wurde die Verfügbarkeit der vernetzen Dienste für jedes der Fahrzeuge getestet. Basierend auf dem globalen Dienste-Portfolio haben P3 Experten die Diensteverfügbarkeit in zwei Schritten ermittelt: Zunächst wurde die Verfügbarkeit in einer Online-Recherche überprüft, bevor eine weitere Validierung im Fahrzeug durchgeführt wurde. Weitere im Fahrzeug identifizierte Dienste wurden in das Dienste-Portfolio aufgenommen und auf ihre Relevanz hin bewertet, um eine gleichwertige Bewertungsgrundlage zu erreichen. Um die Dienste-Bewertung für jedes Fahrzeug in jeder Auszeichnungskategorie zu berechnen, wurde die Diensteverfügbarkeit (0…nicht verfügbar / 1…verfügbar) mit der Relevanz aus Schritt (0…1) innerhalb der zugewiesenen verbundenen Dienstekategorie multipliziert.

Im zweiten Schritt wurde die Nutzererfahrung der fünf relevantesten Dienste in jeder Kategorie von drei Experten anhand von Use Cases getestet. Zu den Anwendungsfällen gehörte zum Beispiel: „POI an das Fahrzeug senden“ für die Kategorie „End-to-End-Navigation“. Die Implementierung und Verwendbarkeit dieser Dienste wurde anhand des Systems Usability Scale (SUS), einen 10 Punkte Usability Fragebogen, evaluiert. Waren Use Cases in einem Fahrzeug nicht realisierbar, dann wurden die Ergebnisse in der Gesamtbewertung nicht berücksichtigt, um Fahrzeuge, die weniger verbundene Dienste anbieten, nicht zu benachteiligen.

Der dritte Schritt betraf die Erfüllung der Nutzerbedürfnisse im Bereich der vernetzten Fahrzeuge. Um diese Bedürfnisse zu definieren, wurde eine Vorabbewertung durchgeführt. Zwölf Testpersonen absolvierten dafür mehrere Use Cases in zwei verschiedenen Fahrzeugen (Tesla Model S 75D und VW eGolf). Ziel war es hierbei, die gesamte User Journey im Fahrzeug mithilfe mehrerer Use Cases abzubilden und die Grundbedürfnisse der Nutzer für die unterschiedlichen Dienste-Kategorien zu identifizieren. Während der Durchführung der Use Cases wurden die Testpersonen von einem Experten unterstützt und befragt. Basierend auf den Ergebnissen der Interviews wurde schließlich eine Online-Umfrage erstellt, um die ermittelten Nutzerbedürfnisse im Bereich vernetzter Fahrzeuge zu validieren. Das zusätzliche Ziel der Umfrage war es, eine Rangfolge der ermittelten Bedürfnisse zu erhalten. Daher wurden

die Bedürfnisse von den Teilnehmern (n = 86) von höchster bis geringste Wichtigkeit eingestuft.
Die Bedürfniserfüllung in den Testfahrzeugen wurde dann von drei Experten während einer Testdauer von zwei Tagen getestet. Um eine Vergleichbarkeit zu erreichen, wurden die gleichen Anwendungsfälle wie in der Vorbewertung verwendet. Die drei Experten erledigten die Aufgaben in allen Kategorien in allen Fahrzeugen, um sich einen vollständigen Eindruck des Fahrzeugs zu verschaffen. Nach Abschluss der Use Cases in einer Kategorie mussten die Teilnehmer die Erfüllung ihrer Bedürfnisse bewerten, indem sie die Aussagen anhand einer Likert-Skala von „Ich stimme überhaupt nicht zu“ bis „Ich stimme voll und ganz zu“ bewerteten. Der aus den Tests resultierende Durchschnittswert für jedes Bedürfnis in jeder Dienste-Kategorie wurde dann mit der durchschnittlichen Signifikanz jedes Bedarfs gewichtet, basierend auf den Ergebnissen der Online-Umfrage.

Die Tests ergaben drei Werte pro Fahrzeug und Kategorie: einen Dienste-Score, einen SUS-Score und einen Bedürfniserfüllungs-Score. Der endgültige P3-Konnektivitätswert für jedes vernetzten Fahrzeug in jeder Auszeichnungskategorie ergab sich dann aus dem Durchschnitt aller drei Bewertungen. Der Gesamtsieger, wie auch Sieger in mehreren Kategorien, war am Ende BMW. Mit einem sehr umfangreichen Angebot an Diensten konnte BMW in unserer Methodik viele Punkte erzielen.

Literaturverzeichnis

[1] Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology. International Journal of Psychology, pp. 253-264.

[2] Nielsen, J. (1993). Usability Engineering. London: Academic Press.

[3] Schiessl, M., Duda, S., Thölke, A., & Fischer, R. (2003). Eye tracking and its application in usability and media research. MMI-interaktiv Journal.

Development of the Cockpit-UI/UX of the Taycan in an Agile Way

Less is More

Dr.-Ing. **L. Krauß**, **E. Kögler**, **M. Bayer**, **S. Wiechmann**, **M. Worch**, **M. Mohamad**, Dr. Ing. h. c. F. Porsche AG, Weissach

Abstract

To develop an electric car with a sports car character from zero to zero is a challenge. To implement a new strongly reduced interaction concept is a challenge. To convert a business unit to Agile is a challenge. Each topic alone represents a great effort for all involved. In the Porsche Taycan project, all of these things were done simultaneously, and the goal was achieved; as demonstrated at the IAA 2019.

If you want to be competitive tomorrow, you have to face digital transformation today. The necessary and radical change from status-quo to a still unclear target state involves unprecedented challenges. This applies both to the content (new business models, organization, processes, technology/IT) and to the process. In this case, the end is also the means: an agile pilot project was created in order to create an agile organization. In workshops as well as their day-to-day work the team members involved in the "Cockpit and Infotainment System Experience" from the areas of Technology & Function and User Interface & Ergonomics were trained by agile coaches, Scrum and Design Thinking experts, finally forming independent product teams dedicated to smaller products (e.g. Navigation) within so-called clusters (e.g. Infotainment). Chapters were formed as a matrix to the product teams mimicking departments, with chapters such as "UI/UX Cross" tasked with ensuring a holistic user experience by enabling teams to produce products that appear to be cast from the same mold. In order to involve stakeholders and management more intensively and at an earlier stage, regular UI/UX reviews were conducted on tangible prototypes in the early phases, and stage presentations with Minimal Viable Products (MVPs) in later phases.

The Taycan is based on principles that have been part of the basic make-up of every Porsche since 1948: Openness, purism, clear architecture, driver orientation and suitability for everyday use. The all-electric drive concept allows a completely new interpretation of these principles: The Taycan's interaction concept, which is geared towards the driver as much as possible, shows the way into the future: Intuitive, fast, distraction-free - in other words; like a sports car.

The modernised Porsche interior is characterised by a continuous, clear dashboard and a powerful centre console with a large touch control panel. In front of the typical Porsche steering wheel, new digital instruments build on the classic 911 look, with elements extending beyond the steering wheel on the left and right.

The underlying maxim in system design "less is more" is a revolution in itself. For example, hardware operating elements are almost completely dispensed with, which allows interface to adapt to use-cases as well as adding comprehensive update capability. A single instrument cluster display, rather than a combination of small displays and hardware gauges, allows larger content such as a map to be brought into the drivers line of sight. During the conception of the interaction, the menu structures and the functions, the team asked itself again and again: *"Can this be made even easier? What else can we leave out? Does the driver really need that?"* Consequently a flatter menu structure and obvious interaction paths could be implemented, and the concept was streamlined to speed up typical use-cases in numerous agile evaluations, interviews and test drives with customers, the concepts were continuously refined and optimized.

Through a completely new form of cooperation and constant critical scrutiny of functions and UI concepts, we have succeeded in designing the next generation of cockpit UI/UX and further developing the driving experience.

While the UI is only one of the contributing factors to user experience, the focus of this manuscript has been reduced to the development of cockpit UI in order not to exceed the defined length for submission. At the conference MOBITAS 2020 the field will be extended to comprehensive touchpoint design as well as to the sphere and interplay of holistic transport and mobility services [MOB20].

Agile and UX are Siblings

At first, the title seems to be a buzz word battle: Agile and UX are common terms that are nowadays used in many variations. The Agile and User Experience Design worlds use a well-established realm of acronyms: HMI, UI, HCD, UX, IxD, IA, UCD, UXD, CX, agile UX, lean UX, guerrilla research, strategic UX, emotional design, etc. [TRE19]. We're swimming in a sea of strange words, but if you look back to the root of the meaning and intention of Agile and UX, then you will recognize that Agile and UI/UX are brothers and sisters at heart: Both are focused on the customer, on human factors and on the person who is interacting with a technical system.

In the Manifesto for Agile Software Development you can recognize the user centered mind set in the third value [BEC01]:

- Individuals and interactions over processes and tools
- **Working software** over comprehensive documentation
- **Customer collaboration** over contract negotiation
- **Responding to change** over following a plan

The four values are detailed in twelve guiding principles for the methodologies. They describe a culture in which change is welcome, and the customer is the focus of the work. The first principle is customer satisfaction through early and continuous software delivery [EBY16]. The first principle of the agile methodology at Porsche has also the focus on the customer: *"Our first priority is to serve the real needs of our customers".*

Even in the agile tooling, users' needs are a core part of Agile: The User Stories. User stories are one of the primary development artifacts for Agile [DOM19].

- A user story is short, specific and goal-oriented. It is a one-sentence statement that tends to have the following structure: "As a … , I want … so that … ".
- User stories are collaborative design tools. All project stakeholders are expected to participate in the definition and sorting of user stories.
- User stories focus the project on the perspective of those who will use it as "end-user".

User stories are – obviously – user-centered [DOM19]. Doesn't all this sounds like a user-centered philosophy and development process at its core? So let us explore the other topic: UX.

"User experience" (UX) encompasses all aspects of the end-user's interaction with the company, its services, and its products. It considers the periods before use, during use and after use. Furthermore UX is, like Agile, a process where UX teams create products that provide meaningful and relevant experiences to users [ISO9241]. This involves the design of the entire process of acquiring and integrating the product, including aspects of branding, design, usability and function.

In the UI/UX field it is obvious (and described in the ISO 9241) that the user, meaning the end consumer – and not any stakeholder in the development process – is the main part of a system. He or she is the person for whom a development process, a service or a product is made.

So, we should always begin with the users. Of course specification documents have their relevance, but first, we need to examine what the users' objective(s) are and what the business's objective(s) are. Going the other way, we would place the focus on a product and not on who will use it, if at all. Imagine building a car without thinking of who will drive it and for what it will be needed. Different vehicles match different user needs and crafting a digital product uses the same principals [NUN18].

At this point be aware that you should design experiences people need, not just what they say they want [UMB18]. You have to associate and to translate their expressed needs into their real needs and transform them into a product by means of concepts and specifications. That is the real job of an UI/UX developer no matter what education they have.
It follows that UX is not only research and evaluation, but also (even in the main part) the development, specification, implementation and realisation of a service or product - just as innovation is not only the idea, but also its realization.
“User Experience Design” is often used interchangeably with terms such as “User Interface Design” and “Usability”. However, while usability and user interface (UI) design are important aspects of UX design which take care of the look and feel of an interface, they are merely, albeit important subsets of it [FLO12]. Sometimes there is confusion that stems from people thinking that UX means interaction design, and assuming that UI only means visual design. However, the truth is that interaction design and visual design are complementary disciplines, both contributing to the user experience as well as the service design – the design of the processes that the user goes through in an overall effect chain [TOR19]. Therefore we have established an interdisciplinary co-creation system at Porsche in which visual designers and interaction designers with diverse educations as psychology, computer science, design, engineering, linguistics come together so they can leverage on each other’s strengths. This group only makes sense if they work as a proper team and are able to empathise with each-others disciplines. A good visual designer considers how the user will interact with what they’re designing, a good interaction designer that can layout the information in a way that will be easy to digest by the user.

Siblings may also have conflicts – but the conflicts are solvable

There are many articles about UX and Agile. Lots of them rant about how Agile is UX unfriendly, how these two approaches cannot work together, etc. Yes, it is difficult to work on software projects. Yes, it is challenging to work in collaboration with other disciplines [DOM19]. Even visual designers and interaction designers sometimes have disputes. And now we´re adding a third profession to the mix: Computer scientists.
Software is increasingly developed in cross-functional teams that combine the areas of concept, design and development. These teams use their knowledge from various areas and combine it with Agile and lean methods to lead their projects to success. The original concept of the agile approach was developed without a focus on user experience design. This may be why it is sometimes difficult for UX Designers to integrate into agile teams. The main reason lies in different ways of working (another topic results from different time schedules – a solution

is described in the paragraph “Integration of AGILE and UX at Porsche”) [STE19]: User experience designers usually work along scenarios to design the user interface. These describe a use case from the user's point of view and contain a completely different level of detail than user stories. User stories, on the other hand, do not depict a use case that makes sense from the user's point of view, but rather describe a series of individual requirements. However, a good user experience design cannot be developed from individual, detached user stories. The author Jeff Patton suggests combining the different ways of working in the so-called User Story Map [PAT15]. This map arranges user stories along usage scenarios and thus creates an important link between agile and user-centered procedures, because product increments are then planned taking into account related usage cases. This makes it easier for UX designers to get involved in the development process. While this is certainly a good compromise, it still seems like a stale specification of a product, rather than a method for user centricity.

Another possibility is to move the team to another way of thinking and to focus them on the end-user. A little imagination can work wonders in keeping mindful of the point that users are not like us. Just to emphasize the relevance for this way of thinking, among all stakeholders, a little scenario is explained [DOM19]:

“For instance, if you had to design an interface in the center console for car drivers you’d need to think about many types of car drivers. A picture pops into your head — from a situation you had this morning in your car. Quick, lose that thought, because you’re thinking “me” there. Instead, think of this:

- *“Me” could be any car driver. If you thought “me” was a frequent driver, go back and consider race track drivers, soccer mums, comfort drivers or sporty drivers.*
- *“Me” is also a car driver from any situation. That could commuting, holiday trips, country roads, completion trips, circuit discovery experience trips.”*

This way of thinking can be supported by personas. A persona is in user-centered design and marketing a fictional character created to represent a user type that might use a site, brand, or product in a similar way. In other words, do not simply place yourself in someone else´s shoes, but also think from their perspective. In the ideal case the personas are available, present and visible in the collocated workspaces for every team member. Maybe, sometimes, this is not based on real data, but it is a start. Maybe from this tiny empathy exercise, management and team members might understand the need for going out and finding out about the target users!

It is described above that AGILE and UX are not so far apart. There may be conflicts between the two processes but these can be resolved. Of course, cross-functional teams can still work together agilely. The basis for this is the common goal of creating an optimal product for the

user. How the integration of AGILE and UX took place at Porsche in the Taycan project is described in the following paragraphs.

Integration of AGILE and UX at Porsche for the Infotainment Development

The solution lies on different levels: First, on the level of the mindset, second, on the level of the team structure and third, on the process level.

With regard to the mindset and as described before, the common basis are the **needs of the user**. At Porsche, in the development of a function, a feature, a design or an operating sequence the primary consideration is always:

- *"What is the benefit for the customer, the user or the driver? What needs are met by this?"*

In addition, the aim of this product was **absolute simplicity**. The following questions were repeatedly asked in each team during the whole development process:

- *"Which functions and operating steps can we omit? Which interaction steps are ornamental or inherited from the past without thought?"*

For this mindset to become ubiquitous a team structure is required which supports the mindset and makes it implicit. Fig. 1.

- **Product Team "X"** (PT): The core unit of this structure is the Product Team. Each team has the end to end responsibility for their product or service within the infotainment. Several teams can run in parallel, divided into (partial) products, to create a larger product.
- **Product Owner** (PO): The product owner is responsible for curating and planning the product team´s backlog, and for enabling the product team.
- **Function Owner** (FO): Within a product team certain members can focus on a subset of customer discernible functions within the team´s product. The function owners are representatives of the Chapter they belong to (e.g. FO-UI/UX is a member of the UI/UX Chapter, working in Product Team "X" on function "Y".
- **PO-External**: Service provider/supplier teams are integrated into the teams via a SPOC (Single Point of Contact)
- **Cluster Product Owner** (CPO): Manages the team of product owners in terms of content and planning. The CPO maintains the backlog for his cluster (eg infotainment), performs refinements of the backlogs with the POs, and forms the interface between the stakeholders / clients and the product teams.
- **Chapter Lead** (CL): Experts for specific topics come together in chapters to coordinate and align their work with one another. In the product teams, the experts are

regarded as chapter representatives and represent the values and decisions of the chapter.

- **MGMT** (management): Decisions regarding how to implement scopes of the backlog are made with the respective departmental management.
- The corresponding levels in this matrix structure are in constant exchange with one another, and it is welcome to see them attending each other's meetings.

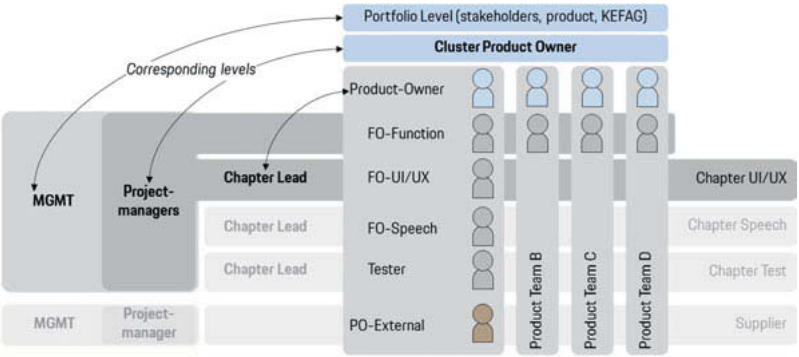


Fig. 1: Team Structure

In addition to the mindset and the team structure, another important building block is the process and the underlying timeline. In the schematic comparison shown in Fig. 2 of conventional product development with agile product development the difference becomes clear.
Traditional projects define the final goals first and then map the future in plans, which are then pursued in one piece or in abstract steps until the planned project object is created. Agile project plans are iterative, i.e. in small steps. Each step is in itself a user discernable product increment designed to fulfill a certain need. Each step is worked through and as this is happening the next step is planned until the entire project object has been created. The agile methodology is based on the assumption that the project life cycle cannot be predicted with sufficient accuracy at the beginning of the project and that its representation in plans is therefore not promising. This flexibility makes it easier to integrate new customer requirements and/or new functions into the product.

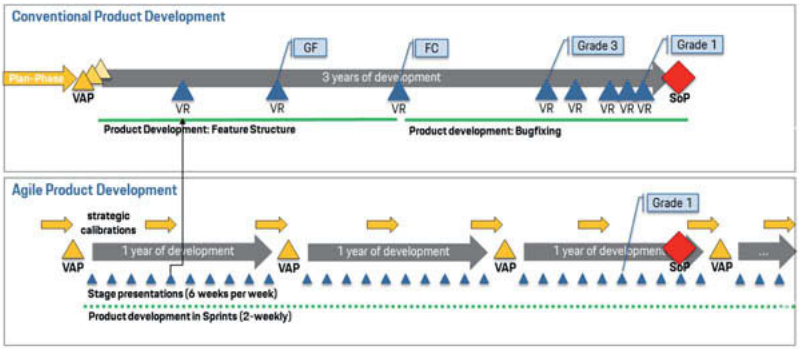


Fig. 3: Schematic Comparison of Conventional Product Development with Agile Product Development

One advantage is, among other things, that internal customers and stakeholders such as the members of the board (VAP) are more intensively involved in product development through regular presentations in stages. In Fig 4 a more detailed view is shown; the numbers in the brackets in the following explanation refer to the numbers in the image.

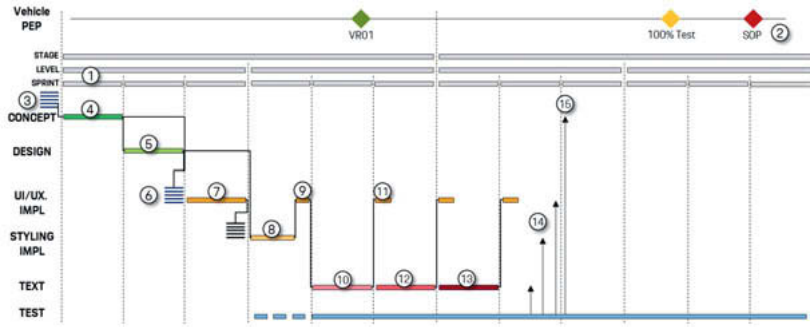


Fig. 3: Timeline of agile UI/UX development and implementation

For better clarity only one passage is displayed but the process overlaps continuously. This ensures that each chapter is constantly productive but also helps eliminate peaks and overloads due to the phase of the project. The increments of the development process are derived from the length of the stages, and specify the granularity of the items in the backlog (bigger than stage is epic, stage is use case, sprint is work package and smaller than sprint is task (1).

The events and increments of the agile development process are not necessarily synchronized with the product development processes of the vehicles. Agile development is a continuous development with vehicle specific releases (2). One agile backlog is used for all teams and the entire product (3). At step (4) the UI/UX concept, the corresponding technical concept and the resulting specification is created. It is followed by the GUI conceptual design if necessary – e.g. new widgets, new layout (5). Increments can return to the backlog and do not always have to follow the workflow in one go (6). This allows for buffering, enabling different chapter sizes and velocities. In step (7), the functional implementation takes place, the increment can be used. Step (8) is necessary because the HMI system is developed in HTML5. The CSS-styling, that is done here enhances the experience of the increment according to the design and enables further steps such as texting as GUI elements now have properties such as size. The step "Pull Request & Check-In" (9) is used to coordinate merging of code between several suppliers. After the text-IDs are available the translation process starts: (10) text translation I (developers text to German), (11) pull request & check-in, (12) text translation II (German to English, etc.), (13) text translation III (English to Chinese, etc.). Testing occurs as of approx. 90% experience capability. Backlog items are created for the responsible team (14).

But testing is not only performed to ensure the quality the implementation, but also of the concepts. In this way the product can be iteratively improved (15). For the continuous evaluation we implemented the "**Agile Testing**" or "**Agile Evaluation**". With this method the UI/UX-Designers and developers can test their products with test persons from other disciplines or real customers. The test takes place according to the principles of Nielsen, after the best results come from testing no more than 5 users and running as many small tests as you can afford [NIE00]. The tests take place with a low level or high level simulation once a month and each team is free to subscribe to a list and to test another variant of a feature, an operating sequence or a screen design. Every stakeholder is free to visit the test, but no report must be written – even if the UI/UX designers need it for themselves. With the described methodology of a combined AGILE and UX process was the digital cockpit of the new Porsche Taycan developed

Digital, clear, sustainable: the cockpit of the new Porsche Taycan

Porsche is entering a new era with the new Taycan, and the brand's first all-electric sports car is setting standards in cockpit design [TOB19].

Classic design features have been reinterpreted and brought into the digital age according to the goal "Less is more". The Taycan interior combines design elements typical for the brand with a type of user experience new to Porsche. The dashboard is designed for drivers and a sporty seating position. The cockpit signals the start of a new era with its clear structure and a completely new construction. It is clearly driver-focused. The instrument panel has a clean,

minimalist and ultra-modern design, and operating the controls is quick and free from distractions. The free-standing, curved instrument cluster forms the highest point on the dashboard is clearly focused towards the driver and ensures that everything that's needed for driving is in view.

Fig. 4: Cockpit of the Taycan

The instrument cluster consists of a curved 16.8-inch screen with the rounded look that's typical of Porsche. A cowl has been omitted, which ensures a slim and modern appearance in the style of high-quality smartphones and tablets.

Drivers can choose between four display modes for the instrument cluster:

- **Classic mode** (power meter) evokes the round instruments typical of Porsche. This display delivers information that's clearly arranged information, allowing for fast readability. A power meter replaces the rev counter in the middle instrument.
- **Map mode** replaces the central power meter with a map layout.
- **Full map mode** intentionally omits the round instruments in favour of a navigation map displayed across the full display.
- The **Pure mode** displays only essential driving information such as speed, traffic signs and navigation using a minimalist arrow.

There are also small, touch-control fields at the edges of the screen for operating the light and chassis functions. The instrument cluster is therefore wider than the steering wheel and reminiscent of the iconic original 911.

The upper and lower sections of the dashboard stretch across the entire width of the vehicle in the shape of wings. A central 10.9-inch infotainment display and an optional passenger display are combined to form an integrated glass band in a black-panel look, thereby blending in visually with the interior.

All user interfaces have been completely re-designed for the Taycan. The number of traditional hardware controls, such as switches and buttons, have been greatly reduced. Instead, control is intelligent and intuitive – via touch operation or a voice control function that responds to the command “Hey Porsche”. Porsche has teamed up with Apple Music to create the first fully

integrated music streaming experience. All vehicle configurations for the Taycan, such as Porsche Active Suspension Management (PASM), can easily be set up on the central screen via direct access. The driver can quickly access all apps via a clearly structured and customisable home screen. Apps include navigation, telephone, media, comfortand charging. With optimised voice control, drivers can access the required function even faster. For the first time, front passengers in the Taycan have the option of their own touch display, allowing them to easily alter settings without distracting the driver.

The elevated centre console intensifies the feeling of a low seating position, as you would expect from a Porsche. It features a large 8.4-inch touch panel with haptic feedback. This allows the air-conditioning settings to be altered directly. Integrated handwriting recognition also allows quick address inputs.

Every detail has been reduced to the essentials. Like the Porsche 918, the Taycan has a compact direction selector switch in the instrument panel instead of the classic selector lever. This gives the centre console a tidy look and creates storage space. In addition to the host of innovations, there is another detail that no Porsche should be without. Similar to the ignition lock on conventional Porsche models, the power button is located on the left behind the steering wheel.

[BEC01] Beck, Kent et al.: Manifesto for Agile Software Development (2001). 23.08.2019 https://agilemanifesto.org/

[BOL16] Bolognesi, Emanuele: Introducing Dual Track Agile - The Theory. 07.10.2016. https://medium.com/@emabolo/introducing-dual-track-agile-27a23d12268b

[DOM19] Domingo, Muriel: User Stories: As a [UX Designer] I want to [embrace Agile] so that [I can make my projects user-centered]. 10.08.2019. https://www.interaction-design.org/literature/article/user-stories-as-a-ux-designer-i-want-to-embrace-agile-so-that-i-can-make-my-projects-user-centered

[EBY16] Eby, Kate: Comprehensive Guide to the Agile Manifesto. 26.07.2016. https://www.smartsheet.com/comprehensive-guide-values-principles-agile-manifesto

[FLO12] Flowers, Erik: UX is not UI. 15.12.2012. http://www.uxisnotui.com/, http://www.helloerik.com/ux-is-not-ui

[ISO9241] ISO 9241-210:2019: Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems. 07/2019. https://www.iso.org/standard/77520.html

[MIC18] Milchling, Allison: A Framework for User-Centered Agile Development. 30.04.2018. https://uxplanet.org/a-framework-for-user-centered-agile-development-202f1d31948

[MOB20] MOBITAS: 2ND INTERNATIONAL CONFERENCE ON HCI IN MOBILITY, TRANSPORT AND AUTOMOTIVE SYSTEMS at HCI INTERNATIONAL 2020, 22nd International Conference on Human-Computer Interaction, Bella Center, Copenhagen, Denmark, 19-24 July 2020. http://2020.hci.international/MobiTAS.html

[NIE00] Nielsen, Jakob: Why You Only Need to Test with 5 Users. 18.03.2000. https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

[NOR19] Norman, Don; Nielsen, Jakob: The Definition of User Experience (UX). 23.08.2019. https://www.nngroup.com/articles/definition-user-experience/

[NUN18] Nunes, Michael: Focusing on the Users' needs. 06.09.2018 https://medium.com/@WeAreMonday/focusing-on-the-users-needs-a36071fd9010

[PAT15] Patton, Jeff: User Story Mapping- Nutzerbedürfnisse besser verstehen als Schlüssel für erfolgreiche Produkte. 2015. O'Reilly-Verlag

[STE19] Steimle, Toni: Agile, Lean und UX heute. 01.02.2019. https://www.ergosign.de/de/insights/2019/article-agile-lean-and-ux-today.html

[TOB19] Toberer, Nadine: Digital, clear, sustainable: The interior of the new Porsche Taycan, 23.08.2019. https://newsroom.porsche.com/en/2019/products/porsche-taycan-interior-digital-clear-sustainable-18432.html

[TOR19] Torre, José: UI vs UX — is NOT a thing. 03.01.2019. https://uxdesign.cc/ui-vs-ux-is-not-a-thing-28aef994fddc

[TRE19] Treder, Marcin: Lean UX vs. Agile UX – is there a difference? 08.07.2019. https://www.uxpin.com/studio/blog/lean-ux-vs-agile-ux-is-there-a-difference/

[UMB18] Umbach, Heath: Designing Experiences People Need, Not Just What They Say They Want. 16.02.2018. https://medium.com/fresh-tilled-soil/designing-experiences-people-need-not-just-what-they-they-say-they-want-d70c27e0feac

3D-Displays with Lightfield Technology for a natural look and feel

User experience between attention-guiding and brand emotion

Kai Hohmann, **Frank Rabe**, **Christof Menzenbach**,
Continental Automotive GmbH, Babenhausen

Kurzfassung

Gängige 2D-Displays gewinnen im Cockpit stetig an Bedeutung: Mehr Displays und größere Bildschirmdiagonalen werden eingesetzt, um dem Fahrer und den Beifahrern Informationen zu übermitteln bzw. sie zu unterhalten. Allerdings erweist sich die fehlende dritte Dimension dabei als Nachteil im Vergleich zu der echten dreidimensionalen Anmutung „physischer" mechanischer Instrumente mit 3D-Zeigern, Skalen, Skalenringen usw. Bereichert man ein Cockpit auf Basis von gängigen 2D-Displays um die dritte Dimension, so wird diese Einschränkung überwunden und das Nutzererlebnis im Cockpit wesentlich gesteigert. Mit einem 3D-Design lassen sich situativ relevante Informationen stark hervorheben, die Ergonomie wird verbessert, die Ablenkung kann durch Leiten der Aufmerksamkeit des Fahrers verringert werden, und die optische Attraktivität des Cockpits steigt deutlich. Die heute verfügbaren automobiltauglichen 3D-Technologien haben jedoch spezifische Nachteile, die Auswirkungen entweder auf die Qualität der 3D-Darstellung, die Komplexität, die Kosten – oder alle drei haben. Continental hat sich deshalb für Leia's Diffractive Lightfield Backlighting Technologie entschieden, um ein 3D-Display zu entwickeln, das die Vorteile anderer 3D-Technologien in sich vereint. Dieses 3D-Display kann natürlich anmutende und optisch attraktive 3D-Objekte zeigen, die für den Fahrer *und* für die Passagiere gleichermaßen sichtbar sind – und das ohne Blickverfolgung. Objekte stechen im wörtlichen Sinn vor der zweidimensionalen Oberfläche „hervor" und stellen ein zusätzliches Element der Aufmerksamkeitssteuerung und des visuellen Erlebnisses dar. Die Liste möglicher Anwendungen ist lang, und deren Umsetzung wird von einer sukzessive ausgebauten Zahl an Apps sowie einem Entwickler-Kit unterstützt.

Abstract

Mainstream 2D displays are gaining importance in the cockpit: More displays and bigger screen diagonals are employed to communicate information to the driver and passengers or to entertain. However the missing 3rd dimension is a downside in comparison to the genuine 3D look

and feel of “true” mechanical instruments with 3D pointers, dials and rings etc. By adding 3D to a display-based cockpit this 2D mainstream limitation is overcome and the user experience in the cockpit is strongly enriched. By utilizing 3D design instantaneously relevant information can be accentuated, ergonomics can be improved, distraction can be limited by guiding the driver’s attention, and the visual attraction of the cockpit can be greatly increased. Currently available automotive-suitable 3D technologies, however, have specific downsides which impact either the 3D quality, the complexity, or the cost – or all three. Continental has therefore chosen Leia’s Diffractive Lightfield Backlighting technology to develop a 3D display that combines the strong sides of other 3D technologies. This 3D display can show natural and visually attractive 3D objects which are visible to driver and passengers likewise – without eye-tracking. Objects can thus truly “pop up” out of the 2D surface and provide an additional element of attention guiding and visual experience. The list of potential use cases is very long and the implementation of them is supported by a gradually expanded number of apps and a development kit.

1. A new dimension – that is not so new after all

Looking back onto the history of driver information one quickly realizes that the speedometer – as the beginning of the modern vehicle cockpit – was a three-dimensional instrument from day one. Of course that is due to the fact that early speedometers were mechanical gauges consisting of many parts, integrated into a metal (or plastic) housing with a visible dial and pointer under a glass panel held in place by a decorative (metal/chromium) outer ring. With the advent of display technology this property of visual and tangible 3D was given up to utilize the overwhelming benefits of displays as today’s core part of the automotive human-machine interface. However, the two-dimensional limitation of traditional displays tends to equalize every content depicted on it. The freedom of showing many types of graphic content in full color (and perhaps with motion) is one way of selecting and prioritizing contents despite the limitation to one level of presentation. Still, the human eye is used to stereoscopic 3D vision. It is our natural way of perceiving the world. Distance or proximity to an object are a core part of the visual information flow we permanently process. Humans are “programmed” to react to approaching objects because approximation adds to the relevance of an object or person. A diamondback rattle snake ten yards away is one thing. A rattler near your foot is quite another.

By adding another dimension (3D depth to the rear *and* to the front) to the display contents, the automotive human-machine interface is given a new quality of perception. 3D information quite literally “stands out” and will thus be *perceived* more quickly, plus it will be *processed* more quickly because we intuitively understand it’s relevant to us. That is why automotive manufac-

turers have begun to look into 3D depiction of selected display contents. After all the main challenge of guiding the driver's attention to important information while generally keeping it on the road remains the main task of instrumentation. With the growing number of displays in the cockpit and the growing display surface this task is not getting any easier. Therefore 3D elements can provide an additional strategy to guide the driver's (and passengers') attention to instantaneously relevant information. At the same time 3D contents is an innovative way of enriching the user experience in the cockpit.

2. Possible 3D technologies for automotive use

For practical reasons automotive 3D technology cannot be based on 3D glasses (unlike 3D movies, for instance). It must be possible to perceive the 3D contents without putting on any eyewear. Also, there shall be the possibility to switch off the 3D effect to accommodate for a small percentage of the population who cannot perceive this 3D effect or who tend to develop motion sickness when they are exposed to it. Currently three mature technologies can meet these requirements. While they offer a varying level of 3D capability they also bring some limitations on the technical and economical levels.

- **Parallax barrier type 3D** employs either a barrier layer based on Liquid Crystal technology (LC) or a printed masking layer, which filter the pixel light emission (by absorbing light in specific areas) to create two separate views/images for the left and right eye. Inevitably this filtering causes brightness losses. The limitation to two views can give the 3D contents an "artificial" appearance. To ensure a good visibility of the 3D object, this kind of system requires a camera to adjust the 3D contents to the head position of the driver (head-tracking). Vehicle movement such as vibration and pothole impacts can cause disturbances of the views. By deactivating the LC barrier layer the 3D effect can be switched off to use full-resolution 2D. The 3D effect is only visible to the driver – not to anyone else in the car. The system cost for this solution is comparatively high because it includes the camera and two LC Displays (LCDs).
- **Lenticular 3D** is based on an additional 3D sheet above the display with a high number of microscopic lens structures (lentil-shaped, hence the name). These micro-lenses redirect the light emitted from the pixels and sub-pixels underneath the lens to different directions and thus create the different views for the 3D effect. If an additional switchable liquid crystal layer is added on top, the 3D effect can be switched off to utilize a full-resolution 2D mode. A camera to detect the head position of the driver is optional. This system can be combined with any standard display but the optical stack in front of the display can cause sunlight reflection. The angle of the

lenticular structure can also disturb depending on the viewing angle. Cross-talk which causes the 3D content to blur rapidly as it tries to “come out” of the screen is another potential issue.
- **Multilayer displays** create 3D by producing two images in two image cells/LC panels with an interstitial anti Moire film in between them. As the depth of the 3D objects depend on the distance between the two LC layers, 3D objects cannot truly “pop up”, they only have a certain depth, which limits the 3D effect. The lower transmission of two panels plus the anti Moire film require a strong backlighting. Among the benefits is a seamless 3D appearance across the horizontal axis. Also a full-resolution 2D mode is possible, however, at a comparatively high system cost. In addition the non-standard algorithms which partition the information for the two layers and resulting views are quite complex.

After analyzing the available 3D technologies, Continental decided early on to pursue a different technology path which offers a – currently – unique combination of benefits.

3. Multiview 3D Lightfield technology with directional backlighting

Continental is now adapting Leia’s technology for use in its automotive 3D display. Until recently, either parallax barriers or lenticular techniques were used to achieve a glasses-free 3D effect, each having its own issues for the automotive use (cf. above). Leia’s Lightfield display solution relies on different physics called “Diffractive Lightfield Backlighting” (DLB™) and combines the advantages of barrier and lenticular systems. This results in a simple and cost-effective system (one screen two backlights) creating 3D imagery for everybody in the car and their known eye boxes and viewing cones with no need for head-tracking. Lightfield technology exhibits low cross-talk, low sunlight reflections, and can be easily switched to a full resolution 2D mode that resolves super-fine details and text.

Fig. 1 shows the principle design of diffractive backlighting: An additional edge-lit backlighting plate receiving a planar light beam is installed underneath the standard LCD and above the 2D standard backlighting to create the 3D effect. The first 10" 3D demonstrator with Leia's DLB technology was presented at the CES show at Las Vegas in January, 2019.

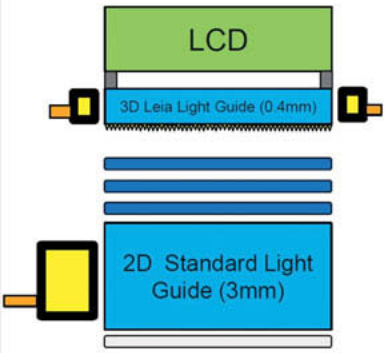


Fig. 1: Design of a Lightfield 3D solution

The 3D Leia light guide plate is very thin at only 0.4 millimeters. A nanoscale grating on the surface of the light propagation layer serves to bend and guide the lights rays which then pass through the LCD above and redirect the light emission from a certain group of pixels, changing the direction and angular speed of the pixel light (**Fig. 2**). By tailoring the printed grating to the display resolution and size it is possible to create multiple views (perspectives) according to customer requirement from directional pixels which gives a natural 3D effect.

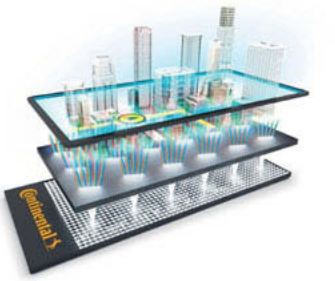


Fig. 2: Functional principle 3D backlighting

3D objects can pop out of the display with a height of up to 5 centimeters (this limit shall ensure that the driver is not disturbed by a 3D view hovering in mid-air in his view; so it is *not* a technical limitation), which allows new ways of depicting contents with a genuine 3D feeling. The initial automotive 3D solution currently under development will be based on a 12.5" 4K display with a

resolution of 3840x2160 pixels and an aspect ratio of 16:9. The DLB technology will create 8 views (8:1) with a 3D resolution of about 1357x764 pixels. Thanks to the high display resolution text is also depicted very sharp and crispy in 3D mode. As the additional 3D backlighting is integrated underneath the LCD, the Lightfield 3D solution is suitable for optical bonding and seamless integration. A switch to a 2D full-resolution view is possible. Readability under daylight conditions is very good, and DLB technology offers the lowest reflection level of the 3D technologies. The Continental 3D display achieves the same readability under bright ambient lighting conditions as an optimized 2D display.

4. Use cases for 3D effects

In a vehicle three-dimensional elements are not an end in themselves of course. In general 3D objects and images serve to interlink the digital world and the user environment more closely: The digital world materializes into “our” world. In the vehicle, 3D is a new element of human-machine interaction. It adds the quality of spatial proximity or distance (depth) to individual bits of information. Mainstream 2D displays depict everything on one level. This geometric uniformity provides no solution to highlighting individual design elements such as dials, pointers, decorative rings or others. By adding a natural 3D skin, the design concept can utilize the 3rd dimension to enrich the user experience (UX). As an example, 3D tutorials for diagnostics can show the tires and pressure reading of a tire pressure monitoring system in 3D.

One very clear and beneficial use case for 3D is the **accentuation** of an instantaneously important bit of information. An oncoming stop-sign, for instance, will be less likely overlooked if it pops out of the instrument cluster display. Similar use cases could include highlighting speed limits, red traffic lights, oncoming traffic, dangerous road situations like ice, or accidents ahead but also things like turn-by-turn navigation directions.

Navigation is also a good example of the benefits of 3D: Depicting maps in 2D is clearly useful but the driver is always challenged to “translate” and match the 2D view on his screen with his real-world 3D view of the actual scenery. However, if the 3D display contents match the real-world view by reproducing it in the correct spatial order and perspective, it will be easier for the driver to understand where his correct trajectory is in relation to buildings, turnoffs, landmarks etc. (**Fig. 3**).

Fig. 3: 3D offers a better display-to-world fit which can serve to make navigation easier

For **Automated Driving** (**AD**) 3D technology has a lot to offer. While AD is activated 3D could serve to provide the driver and passenger with selected bits of information about vehicle activities and "plans" to help build trust in the automation. During the back delegation of the driving task to the human driver, 3D objects can be a new element of the handover procedure which helps the driver to finds his/her way back to the driving task. Prioritizing information during this transition phase may help to improve the driver's situation and mode awareness. During an automated part of a journey the 3D effect can also be integrated into the entertainment options for the driver and passengers. For instance, during AD 3D-gaming would become an option, or watching movies in 3D. The graphic display of automated parking – such as the 360-degree bird view – would be a real eye-catcher in 3D.

3D contents will **enrich the UX** irrespective of the use case. Be it accentuation, navigation, user support during back delegation, or gaming, 3D adds a new quality to the way in which driver and passengers experience their trip. During certain phases of driving, such as setting off or arriving, more emotional contents such as an OEM logo can be accentuated through animation and depth. Complex light effects and sparkles can be used to add to the UX.

The enormous freedom of depicting information and entertainment on a display in combination with 3D will benefit from an ongoing change in the electronic architecture in the car: It is to be expected that a small number of high-performance computers (≈ servers) will provide the computing power for a great number of functions summarized within few vehicle domains. Looking at the cockpit this would span everything from driver information through driver assistance, entertainment/internet, navigation, and routines such as over-the-air-updates. The architecture of

this future type of high-performance computer will seamlessly integrate the digital world of cell-phone-based communication and services. 3D elements can be utilized within all of these different areas and will thus benefit from the trend towards a much more centralized electronic architecture as this potentially makes 3D available to every function once the display hardware is in place.

To support an automotive-optimized utilization of the Continental 3D display with Leia's technology the display hardware will be accompanied by a dedicated content-store that offers proven apps for individual use cases in combination with the Lightfield 3D display technology. An Automotive Software Development Kit (SDK) helps to create or convert content to the Lightfield format, with automatic settings to ensure visual comfort. Behind this content store is Continental's long-standing experience with developing elements of the human-machine interface to meet the highly specific ergonomic requirements of controlling a car with a maximum of safety, a minimum of distraction, and an optimum joy of use.

5. Summary and outlook

3D displays based on Diffractive Lightfield Backlighting (DLB™) offer an innovative and economical automotive solution which combines the strengths of other 3D physics while avoiding specific downsides. The 3D contents is visible within rather large eye boxes and viewing cones for all vehicle passengers. Objects or images can pop out of the display by up to 5 centimeters, which helps to make relevant information truly "stand out". By typically generating between 4 and 9 views, the 3D depiction is natural and realistic. As a part of a modern cockpit with its growing display number and surface 3D is a novel way of highlighting/prioritizing information and warnings. The visual quality of 3D catches the driver's attention and helps to guide it to selected information. Navigation can be supported with a "map" presentation that truly resembles the real-world scenery as the driver sees it to make orientation easier. From a very practical point of use, the 3D Lightfield display offers the lowest reflection and excellent readability even under sunlight conditions.

During Automated Driving immersive 3D-gaming, video calling, or watching films in 3D would be an addition to entertainment. In any case 3D enhances the human-machine interface by adding another dimension that can be utilized to improve the user experience. By providing a content store with apps supporting individual use cases, the 3D display hardware is complemented by tested tools to implement 3D in future vehicles.

An almost endless number of potential 3D use cases will help to differentiate the UX from mainstream 2D content by providing a realistic, valuable and natural look and feel to the car users.

Future e-mobility and the change in system requirements

The interplay between battery and thermal management for different mobility concepts

Dr. **Lothar Schindele**, Dr. **David Schütz**, **Frank Heber**, **Patrick Sailer**, Dr. **Gaël Le Hen**, Dr. **Norbert Müller**, Robert Bosch GmbH, Stuttgart

Kurzfassung

Die Wahl der thermischen Topologie sowie das Thermomanagement auf Gesamtfahrzeugebene sind ein wesentlicher Hebel für eine energieoptimale Betriebsweise eines Elektrofahrzeugs. In dieser Veröffentlichung werden für ausgewählte Fahrzeug-Nutzungsweisen Topologie- und Technologie-Vergleiche gezeigt. Anhand von Beispielen wird auf Ebene des Gesamtfahrzeugsystems das Temperaturmanagement der Hochvoltbatterie und die Wechselwirkung von Fahrzeugnutzung und thermischer Fahrzeugtopologie dargestellt.

Abstract

The choice of the thermal operating strategy and the thermal topology are decisive for the energy-optimal design of an electric vehicle. In this publication, topology and technology comparisons are shown for selected vehicle usages. Using examples, the temperature management of the high-voltage battery and the interaction of vehicle usage and thermal vehicle topology are presented on the level of the entire vehicle.

1. Motivation

Mobility of tomorrow is undergoing a radical change. Against the background of increasing electrification, automated driving, the connectivity of vehicles and ever stricter CO2 regulations, not only the powertrain technologies (e.g. BEV, FCEV, ...) but also its usage profiles (e.g. robo-taxi, car sharing, ...) and thus the requirements of the vehicle components will change considerably.

The segment of fleet operated cars will gain more importance in the mobility system of the future, and fleet operators will define vehicle requirements more and more. Their increased focus on total cost of ownership (TCO) for a given and specific use case requires flexible and

scalable vehicle architectures. In order to address diverging and specific use cases, new vehicle development and optimization tools have to be applied for optimized solutions on both, a vehicle architectural level, and on a component level.

Optimizing a vehicle for a specific use case, like for example parcel delivery, will typically lead to other vehicle architectures and component designs than optimizing a vehicle for an “average” usage [1]. An efficiency measure that is effective during a worldwide “average” vehicle usage, as it is replicated by a WLTP test condition, might show less benefit in a specific use case like parcel delivery, or vice versa. This represents a paradigm shift in vehicle design, with new requirements for design tools and a scalable modular component system.

Bosch has developed a holistic simulation tool chain, which allows to optimize an EV’s system topology for a specific vehicle usage. For this purpose, usage profiles (also called “use cases”) are considered in detail, which describe the usage of the respective vehicle over the entire year. All energy flows within the vehicle (e.g. mechanical, electrical and in particular thermal) are taken into account, not only during driving, but also during parking or charging phases, in order to determine annual energy consumption. This development tool chain allows to compare a wide variety of vehicle and powertrain topologies, to assess the effectiveness of efficiency measures for a specific use case, and it allows to derive component load profiles for all relevant components. [2]

2. BEV States, Use Cases and Assessment Methodology

In this chapter, the term “use case” is defined for the considered context, and exemplary use cases of todays and future vehicle usages are given.

2.1. Definition of use cases

In the context of system engineering, the term “use case” describes a list of event steps, defining the interactions between an actor (e.g. the driver or traveling person) and a system (e.g. the vehicle) to achieve a goal (e.g. drive from A to B). In the given context of mobility, a use case also describes a specific context of use (e.g. traffic and environmental conditions). Some exemplary use cases are shown in Table 1.

Table 1: Exemplary use cases and their characteristics [2]

Use case	Transport missions	Key parameters
Short Distance Commuter	Family care, shopping	Mileage: 9 tkm/year, Avg. speed: 20 km/h, Starts: 7/day
European Privately Owned Vehicle	Family care, shopping, commuting, business trips	Mileage: 15 tkm/year, Avg. speed: 46 km/h, Starts: 4/day
Taxi (human driven)	All (but restricted to a metropolitan area)	Mileage: 87 tkm/year, Avg. speed: 33 km/h, Starts: 7/day
Robo Taxi (fully autonomous)	All (but restricted to a metropolitan area)	Mileage: 87 tkm/year, Avg. speed: 39 km/h, Starts: 7/day

In order to determine a vehicle's energy consumption or the vehicle components' load profiles for such a use case, a detailed description of operation is required. This description does not only consist of vehicle driving conditions, but also of vehicle states like parking. Table 2 shows some of the relevant vehicle states. Accordingly, an analysis of energy flows in the vehicle needs to consider not only the energy consumption and energy flows during driving, but also during all other vehicle states, e.g. heat transfer to ambience during parking when the battery cools down, or heat generation during charging.

In addition to the drivetrain's energy consumption, the energy for heating, ventilation and air-conditioning (HVAC) of the passenger compartment is of major relevance for the vehicle's overall energy consumption and therefore for the TCO assessment of a BEV. The energy consumption for HVAC depends on both, climatic zones and seasons, which means that the vehicle location has to be considered. These states are used to describe a daily vehicle usage profile.

Table 2: Exemplary EV states [2]

EV States	Definition
Driving	The vehicle is driving according to a standardized or specific driving cycle in order to fulfill a transport mission.
Parking w/ infrastructure, on-board/AC charging	The vehicle is parked and connected to a charging station (e.g. 7 kW) and is able to charge the battery via its on-board charger.
Parking w/ infrastructure, Fast/DC charging	The vehicle is parked and connected to a DC charging station (e.g. 50 kW) and is able to fast charge the battery.
Parking w/o infrastructure	The vehicle is parked and not connected to a charging station.
Parking w/ infrastructure, Preconditioning	Infrastructure allows thermal preconditioning of cabin, battery and other components.

2.2. Today's versus future vehicle use cases

A detailed description of today's vehicle use cases is preferably based on measured vehicle and mobility data. However, a strategic component portfolio management has to consider future vehicle use cases as well, which might not be on the market until now, e.g. autonomous driving with shared vehicles. Therefore, in a first development step it is required to assess and describe in detail the usage of future vehicles or mobility concepts, i.e. of future use cases. Based on this, an appropriate development tool chain is to be used in a following step in order to derive new system and component requirements for such new use cases.

Table 3: Differences between the use cases Taxi (human driven) and Robo Taxi

	Taxi to Robo Taxi
max. acceleration [m/s^2], [3]	- 67 %
peak battery discharge power [kW]	- 73 %
driving time [h/day], [3]	- 15 %
idle time [h/day], [3]	+ 11 %
mean vehicle speed [km/h], [3]	+ 19 %
mean energy throughput [kWh]	+ 16 %

Based on the use case "human driven Taxi", a use case for a fully autonomous Robo Taxi has been derived in [3] by taking into account traffic simulation, lower acceleration rates and the interaction of the vehicle with other vehicles and the traffic infrastructure. This results in an

increase of the average speed and a reduction of the peak battery discharge power. Furthermore, the energy consumption of the automated driving kit (AD-Kit) has to be taken into account (Table 3).

2.3. Methodology and Tools

For a use case dependent technology assessment on a vehicle level, a 0/1D simulation model of an electric vehicle has been developed using the CAE system simulation software GT-SUITE®. It captures all relevant mechanical, electrical and thermal effects. Fig. 1 shows the applied optimization workflow. A more detailed description can be found in [2].

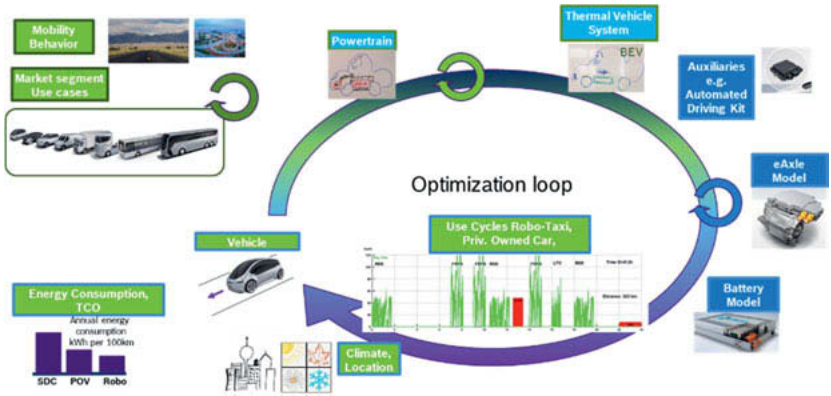


Fig. 1: System optimization workflow [2]

3. The interplay between battery and thermal management

As described in the first chapter, the way a specific type of vehicle is used (e.g. as Robo Taxi) can strongly influence the requirements of the vehicle architecture and its components' design. The traction battery has an optimal operating temperature range in which its performance (discharging and charging) is very high and the cell ageing is low. The aim of thermal control is to bring the battery to its optimum operating temperature range and keep it there.

As an example for this, this chapter describes the effects on traction battery performance for different thermal topologies, thermal operating strategies and use cases.

3.1. Battery recuperation capability at low temperatures

In this example, the influence of low ambient temperatures on the recuperation performance of the traction battery is shown. The aim is to demonstrate the influence of the thermal operating strategy on the operational range of the vehicle.

For this study, the following conditions are assumed:

- Cold start: ambient and vehicle component temperature < -10 °C
- Vehicle: passenger car (compact class)
- Traction battery type: 60 kWh energy cell
- Vehicle usage: the driving time is approximately 3.5 hours (moderate driving cycle, metropolitan area, mean velocity 24 km/h, total distance 84 km)

At low ambient temperatures, the sooner the battery can be brought to its optimum operating range, the more braking energy can be recuperated. For a water cooled battery, this can be done via an electric heater (PTC), integrated into the battery cooling circuit. Three thermal operating strategies are described and compared below (Table 4). Fig. 2 shows the applied thermal topology.

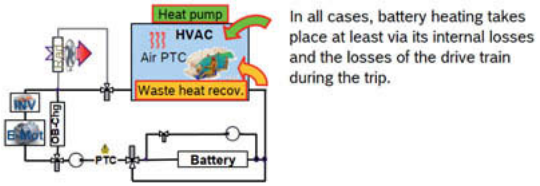


Fig. 2: Applied thermal topology

Table 4: thermal operating strategies

Operating strategy	Definition
w/o-heating	The traction battery is not heated actively.
w/-heating	The traction battery is actively heated during the trip.
w/-BatPreCond	In addition the traction battery is actively heated before the start of the trip. Energy supply via the power grid.

Fig. 3 shows the temperature profile of the traction battery for the three operating strategies considered (Table 4). In the case of "w/o-heating", the battery only reaches the temperature at which it can be recuperated at the end of the trip. In the case of active heating ("w/-heating")

via electrical heater while driving, the battery can be charged (i.e. recuperation is possible) after approx. 20 min. If the battery is preconditioned ("w/-BatPreCond") before the start of the trip, recuperation during braking phases is possible right from the start of the trip.

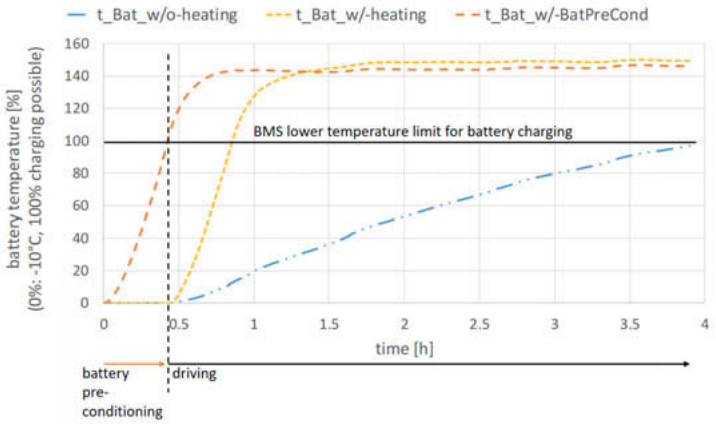


Fig. 3: Battery temperature profiles for the thermal operating strategies "w/o-heating", "w/-heating" and "w/-BatPreCond", with reference to the lower temperature limits (BMS) for battery charging

This example illustrates the influence of the thermal operating strategy on the operational range of the vehicle. In this case the preconditioning results in approx. 36 % more range compared to the case "w/o-heating" (see Fig. 4). In the case of “w/heating”, the thermal operating strategy (e.g. cabin heating) and the driving cycle have a large influence, so that the range advantage can be between approx. 0 and 28%.

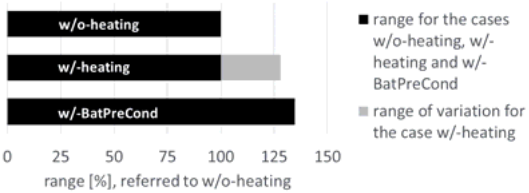


Fig. 4: Ranges for thermal operating strategies "w/o-heating", “w/-heating” and "w/-BatPreCond"

It becomes apparent that the trip properties (e.g. duration, average and peak power) as well as the ambient conditions have a decisive influence on the energy-optimal operating strategy of the vehicle. From an energy point of view, it would not be beneficial to heat up the battery for a short trip, since more energy would be used to heat up the battery than what could be recuperated during braking phases. Before a long distance trip, however, heating up the battery can boost driving range considerably.

3.2. Air-cooled versus water-cooled traction battery at high ambient temperatures

Traction batteries may only be operated within a certain temperature range. If it becomes too warm, for example at fast charging conditions, it must be cooled down in order to keep it within its operating limits. The following example compares a passively air-cooled and an actively water-cooled battery at vehicle level (Fig. 5 and Table 5). The influence of high ambient temperatures on the charging performance is described.

The following conditions are assumed:

- Warm start: ambient and vehicle component temperature > 35 °C
- Vehicle: passenger car (compact class)
- Traction battery type: 30 kWh performance cell
- Vehicle usage: first trip is approx. 45 minutes. Afterwards, the system is fast charged (DC) to 80 % SoC. The following second trip is approx. 45 min (with the battery depleted down to 20 % SoC). At the end it is fast charged (DC) to 80 % SoC again (dynamic driving cycle, highway, mean velocity (w/o charging time) 96 km/h, total distance = 170 km)

Fig. 5: Water-cooled versus air-cooled battery

Table 5: Thermal topologies for traction battery

thermal topology	Definition
air-cooled	The traction battery is not actively cooled. Cooling takes place via heat conduction and convection.
water-cooled	The traction battery is actively cooled by a coolant circuit. The heat is released from the coolant circuit via radiator to ambience or via chiller to the refrigerant circuit.

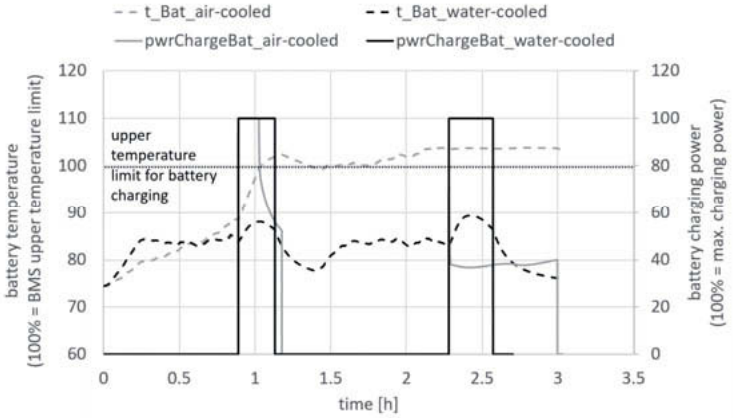


Fig. 6: Charging profiles and battery temperature profiles for the thermal topologies “air-cooled” and “water-cooled”, related to the upper temperature limit (BMS) for charging.

The air-cooled battery heats up to its operating limits during charging. In order to avoid overheating, the BMS limits the maximum allowed charge rate, i.e. it reduces the charging current. This reduces the charging losses of the battery so that it can be operated within the permitted temperature range. The water-cooled battery is kept within its permitted temperature range by a chiller integrated into the cooling circuit.

In this example, the charging time of the air-cooled battery is increased by approx. 95% compared to the water-cooled battery (see Fig. 7).

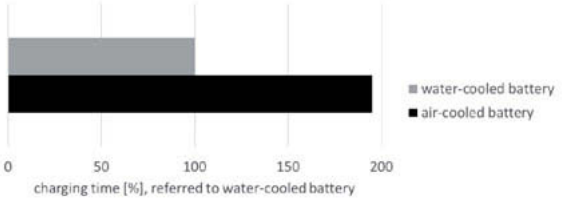


Fig. 7: Accumulated charging time for the entire trip

This example shows how vehicle design and vehicle requirements must be matched to each other. With a vehicle designed for urban applications (e.g. small battery and air-cooled), a change in vehicle usage (e.g. long distance and fast charging) can result in significant limitations for the vehicle user.

3.3. Influence of the thermal topology

The influence of different thermal topologies on vehicle energy consumption is shown in Fig. 8 for three different thermal topologies. The first topology ("Baseline topology") uses an electric heater (Air PTC) for cabin heating. The second topology "+ Waste Heat Utilization" allows to utilize the cooling system's waste heat for cabin heating, along with an electric heater (Air PTC). The third topology "+ Heat Pump" performs the cabin heating function using waste heat recovery from the cooling circuit in combination with a heat pump and an electric heater (Air PTC). For these three thermal topologies, the energy consumption has been calculated for the vehicle described in Table 1 for the use cases Short Distance Commuter, Privately Owned Vehicle and Robo Taxi - in this case at a constant ambient temperature of 9 °C. For the use case Robo Taxi, the additional energy consumption of the autonomous driving kit is not taken into account for better comparability. The WLTP energy consumption is set as a reference. In the WLTP energy consumption values there is no influence of the HVAC system, due to the definition of the test procedure.

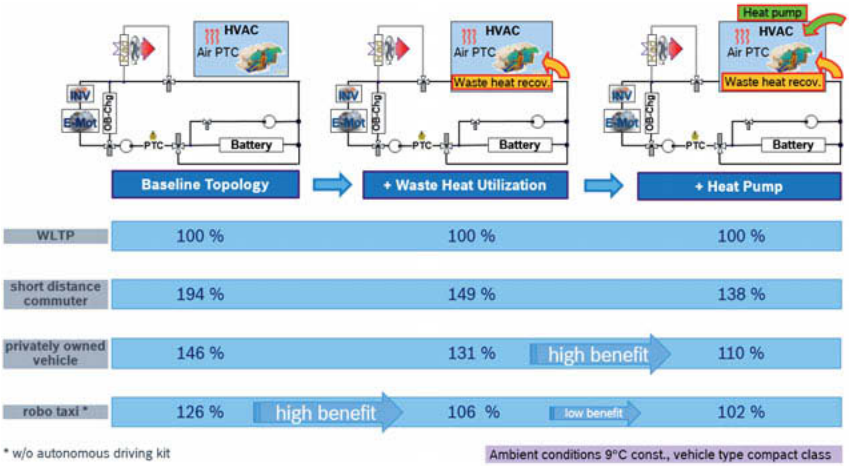


Fig. 8: Influence of thermal topologies on vehicle consumption at constant ambient temperature 9 °C for the use cases “short distance commuter”, “privately owned vehicle” and “robo taxi” [2]

Even though the results shown in Fig. 8 reflect special conditions and depend for instance heavily on ambient temperature, the following conclusions can be drawn:

- The WLTP has been designed such that it reflects worldwide average usage under worldwide average climate conditions, and accordingly are design targets of most vehicles today. However, in case of a fleet operator for instance, who focuses on a specific use case in a specific location and climate zone, a customized vehicle design offers significant efficiency and therefore TCO benefits.
- Especially an EV’s thermal and HVAC system has a significant impact on vehicle efficiency, offering multiple degrees of freedom to optimize a vehicle for a specific use case and climate condition.
- It is worthwhile to design an EV’s thermal system such that waste heat recovery, e.g. of powertrain components, is possible.
- For use cases with a small annual mileage (e.g. use cases like Short Distance Commuter and Privately Owned Vehicle), a heat pump can be highly beneficial.
- For use cases with a high annual mileage, (e.g. use cases like Taxi or Robo Taxi), an appropriate waste heat recovery system is typically much more attractive to support cabin heating than a heat pump.

4. Conclusion

The definition, description and analysis of use cases is key to develop optimized powertrain and vehicle solutions, and to derive optimized component requirements of current and future eMobility solutions. The use cases described in this study consist of commercial applications (Taxi and Robo Taxi) and privately owned vehicles (Short Distance Commuter and Privately Owned Vehicle). Relevant vehicle conditions have to consider not only driving conditions, but also vehicle states like parking and charging. Special focus has to be on the use case location and climate, since heating, ventilation and air conditioning have a significant impact on energy consumption.

Automated driving as referred to by the use case Robo Taxi has a significant impact on vehicle design, for instance only about one third of today's powertrains peak power will be required any more.

The thermal system layout offers a large degree of freedom for an application specific design optimization (e.g. vehicle range, battery life time). For use cases with a low annual mileage, a heat pump can be highly beneficial. For use cases with a high annual mileage, an appropriate waste heat recovery system is typically even more attractive for efficient cabin heating.

The presented methodology for vehicle optimization consists of three layers:

- Description of use case, location, climate,
- Optimization of vehicle architecture and operating strategies,
- Dimensioning of all relevant subsystems and their components, e.g. powertrain, battery, thermal system.

The optimization itself is done in an iterative way using a newly developed simulation and optimization framework.

Future work of the developed methodology and optimization tool chain will take into account additional vehicle classes such as trucks and buses, as well as fuel cell vehicles.

Abbreviations

AD-Kit	Autonomous driving kit
BEV	Battery Electric Vehicle
BMS	Battery Management System
EV	Electric Vehicle
FCEV	Fuel-Cell Electric Vehicle
HVAC	Heating, Ventilation and Air Conditioning
PTC	Positive Temperature Coefficient
SoC	State of Charge
TCO	Total Cost of Ownership
WLTP	Worldwide harmonized Light vehicles Test Procedure

Bibliography

[1] Kampker, A., Gerdes, J., Günther, S.: Think Big, Start Small. Streetscooter die e-mobile Erfolgsstory: Innovationsprozesse radikal effizienter. Wiesbaden: Springer Vieweg 2017

[2] Schindele, L., Schütz, D., Le Hen, G., Müller, N.: Future e-mobility and the change in system requirements, 19. Internationales Stuttgarter Symposium. 2019

[3] Schwarzer, J., Thulfaut, C., et al.: Automated Vehicles – Powertrain Challenges and Concepts, 27th Aachen Colloquium Automobile and Engine Technology. 2018

Modeling and identification of electrochemical energy storage for drive train development

Review and evaluation

Philipp Gesner, **Frank Kirschbaum**, **Florian Landenberger**,
Jonas Scheiffele, Daimler AG, Stuttgart;
Lutz Morawietz, **Bernard Bäker**,
Technische Universität Dresden, Institut für Automobiltechnik (IAD), Dresden

1. Abstract

The technology of energy storage plays a significant role in the electrification of vehicles. Consequently, it is among the higher priorities of many car manufacturers. Modeling the behavior of batteries is thus an essential part of the necessary development effort. For this reason, the present study addresses models of the electrical impedance of automotive batteries. In the first part, model requirements are derived from real driving data. These requirements specify lower and upper bounds regarding the range of temperature, state of charge and current as well as a limitation of frequencies included in the voltage response. In the second part of this paper, available model structures and identification algorithms for the electric battery behavior based on measurement data are presented. The literature indicates a recent emphasis on electrochemically-motivated equivalent circuit models, which describe the inner cell processes using fractional transfer functions. Identifiable parameters of those models are usually estimated by rudimentary least square methods in the frequency and time domain.

2. Introduction

From a technical point of view, the ongoing transformation towards electric vehicles offers advantages regarding the absence of local emissions, the use of renewable energies and particularly the high efficiency of their drive trains. The power losses no longer occur primarily during the energy conversion but instead arise during the charging and discharging of the

battery [1]. This puts the electric storage with its other properties such as energy density, charging time and lifespan in the center of current driving technologies.

A cost-efficient development of drive train systems with lithium-ion batteries (LIB) is often achieved by using computer models. For example, those models deliver first calculations for the decision process on the most suitable cell materials in an early development stage. Other important applications of such digital batteries are model-based software functions. They are mainly used to determine the current *state of charge* (SoC) and *state of health* (SoH) [2]. On a system level of the drive train, the behavioral models are used to simulate the interplay of the components. In later development stages, the models replace real batteries for the design and dimensioning tasks on engine test benches. The modeling for such a system-level and particularly real battery simulator are in the focus of this study and upcoming work.
Obtaining an accurate model requires two different tasks. Firstly, a description of the nonlinear impedance behavior $Z(j\omega)$ has to be found. This function in the frequency-domain is defined by the current $I(j\omega)$ and the resulting voltage $U(j\omega)$ of the battery:

$$Z(j\omega) = \frac{U(j\omega)}{I(j\omega)} \tag{1}$$

Secondly, suitable identification algorithms have to be selected and developed. Fig. 1 shows the main idea [3] behind the estimation of the electrical behavior of individual cells or entire batteries based on input and output measurements.

The error $e(t)$ between the model and the measurement is used to iteratively determine its parameters θ. In addition to the withdrawn battery current $i(t)$, the temperature $\vartheta(t)$ and the state of charge $x_{SoC}(t)$ are inputs of the electrical model. This takes into account the most important nonlinear influences on the impedance and thus the voltage output $u(t)$.

This work focuses exclusively on the description of the electrical behavior and does not regard any models on thermal effects, cell aging or any functions of the *battery management system* (BMS).

3. Requirements on battery models

For an efficient modeling of the impedance (cf. Eq.1), considering the actual battery operation in the vehicle is essential. This reduces the generation and computational effort significantly. Hence, it is promising to formulate model requirements in terms of the current, temperature and state of charge ranges as well as the voltage accuracy. Those models are a starting point for the selection of suitable model structures, necessary measurement data and efficient identification algorithms.

Furthermore, the type of the drive train plays an important role for this analysis, because the electrochemical energy storage is stressed differently in a mild hybrid, a plug-in hybrid or a full electric vehicle (EV). In fact, the type of drive train determines not only the operating range of

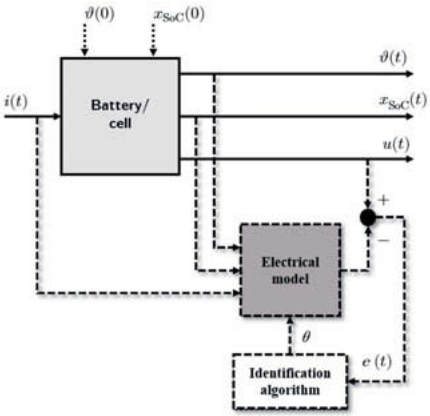


Fig. 1: Identification of the impedance of an electrochemical storage (battery/cell)

the temperature and the state of charge, but it also causes a great difference in the dynamic current load.

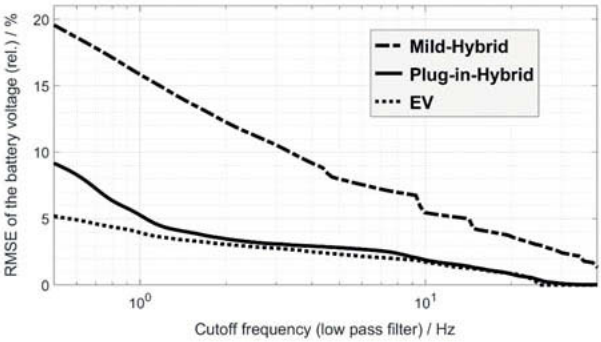


Fig. 2: Effect of reduced frequencies in the voltage response on a relative modell error for different drive train concepts

This investigation of the electric battery behavior within vehicles aims to define a frequency spectrum for the models. A reduction of the frequencies in the voltage response has an immense effect on the measurement and simulation effort.

For this reason, measured battery voltages are filtered with a *low pass filter* (LFP). This filtering of the vehicle data is performed for different drive trains and thus different batteries. The relative error between the measured and filtered voltage is shown in Fig. 2 as a function of the cutoff frequency. The sampling rate of the voltage signals was 100 Hz. It was analyzed up to 40 Hz.

It becomes evident, that ─ even for particularly dynamic loads of batteries in drive trains with low electric power portion (mild-hybrid) ─ voltage frequencies above 30 Hz have hardly any influence on the overall behavior. The error of 2.9 % is calculated by standardizing the *Root Mean Square Error* (RMSE) of 0.013 V with the standard deviation of the measured voltage. As a result, the error for this 48 V battery is negligible.

In addition to the illustrated voltage error, it is necessary to consider the deviation in the mean battery power. The reason for this is the targeted application of the models for the design and calibration of driving strategies, drivability functions and lifespan extension. In the case of the 30 Hz cutoff frequency, the mean power of the measured and filtered battery signals only differ by 0.12 %.

One explanation for the low influence of high-frequency components is the low-pass behavior of electrochemical cells. Impedance spectroscopy data show that batteries transfer high-frequency currents only strongly damped into the voltage signal. For example, typical automotive cells damp, at a temperature of 15 °C and state of charge of 80 %, the amplitude of a sinusoidal excitation with 30 Hz about -60 dB. Accordingly, an alternating current with an amplitude of 200 A results in an voltage amplitude of less than 0.2 V with the same frequency. This low-pass effect of the battery occurs at all operating points regarding current, temperature and state of charge.

Therefore modeling the impedance within a **frequency range from 0 to 50 Hz** seems to be sufficient for most of the development tasks from a drive train perspective. High-frequency applications of electric battery models, for example the simulation of the interplay between the inverter and the battery [4], are not included in this study.

The dynamic transfer function of the battery is non-linear concerning the temperature, the state of charge and the current itself. It is thus necessary to define the operating point ranges to be modeled.

A starting point for this discussion is the secure operation of the electrochemical cells in a vehicle. The so-called operating window is usually defined by the cell manufacturer and includes lower and upper voltage limits, the maximum currents and required cooling concepts. This specification essentially determines the use of the battery in a vehicle [5]. The impedance behavior is therefore to be modeled over the entire **state of charge range** in open-circuit voltage limits of the battery and thus between **0 und 100 %**.

Being monitored by the BMS, the current is limited by the operating window of the cells. This results in limitations for the charging and discharging currents as a function of the temperature, the state of charge and the previous energy throughput. Additionally, the restrictions are specific for each cell or battery and have to be adapted in every modeling process.

For the **temperature range**, it is difficult to establish clear boundaries due to the dependency on the location. The temperature is strongly dependent on the cooling concept and thus difficult to determine. The modeling of the battery impedance from **-40 °C to 60 °C** appears to be expedient.

These requirements provide global boundaries for the model and will need to be further complemented in future investigations with verifiable, local model accuracies.

4. Model structures

The starting point of every successful modeling is the underlying mathematical structure of the model. An initial important categorization of the available models can be made using integrated system knowledge. This results in different modeling depths, ranging from the detailed description of individual ions within the battery cell (*white box*) to fully abstracted model structures (*black box*) that represent the behavior without any physical knowledge.
Based on these distinctions in modeling depths, the literature [6, 7, 8] provides an overview of existing battery models. Furthermore, many of those publications on the identification of batteries introduce the additional requirement of the application in the BMS (*online*). This focus on model complexity and computing effort is not part of the study, due to a preferred use in simulations (*offline*). Instead, the inseparable link between the model structure, the necessary measurement data and the identification algorithm (s. next chapter) moves into focus.

Important research foundations on impedance behavior of rechargeable cells were laid out long before the massive electrification of the drive train in the last decade. For example, Doyle, Fuller and Newman [9] published an extensive and spatially-resolved LIB model in 1993. They particularly used experimental findings on the charge transport and thereby established a standard for many subsequent physically motivated cell models. Afterwards, stochastic impedance models became more popular supported by the success of mobile communication devices. For example, in 2001 Chiasserini [10] began to use a so-called *markov chain* to simulate the discrete state transitions of batteries caused by current loads.

In an automotive context, the impedance is mainly modeled using *electrical equivalent circuit models* (ECM). Their advantage is a graphical separation of the individual effects causing the power loss. Additionally, they cover a wide range of model approaches [11].
An important representative of the ECM is the *Thévenin model* (TM). It describes the complex cell impedance with a resistor and any number of RC-circuits in series. Publications are demonstrating, that models with one [12] or two [13] RC-circuits approximate the linear transfer function at an operating point sufficiently enough for their particular application. With such low-order TMs, the parameter estimation and the prediction effort are quite small, but many electrochemical effects are ignored.
In addition to these phenomenological, simplified ECM, there are electric circuits that integrate an extensive knowledge of the internal processes of LIB cells. These approaches combine the various cell phenomena in one model and often result in high accuracy and globally valid

model. For example, a parallel connection of a *constant phase element* (CPE) and a resistor are used to model the double layer resistance between the electrolyte and electrode [14], and low-frequency diffusion processes are described by so-called *Warburg elements* [15, 16]. The result are extensive ECM [17], that model the dynamic voltage at an operating point. Besides, other physical effects, such as the *Arrhenius equation*, are used to model the impedance between the measured operating points [18].

An additional abstraction in the battery modeling process takes place using dynamic model structures, that are able to capture time series in general, such as ARX [19] and ARMAX models [20]. The ARX (Auto Regression with eXagenous input) approach, for example, provides a mathematical relationship between the input $u(t)$ and output $y(t)$ with the polynomial functions $A(q)$ and $B(q)$ of the discrete shift operator q:

$$y(t) = \frac{B(q)}{A(q)}u(t) + \frac{1}{A(q)}\varepsilon(t) \tag{2}$$

This equation of a *linear time-invariant system* (LTI) integrates an error term $\varepsilon(t)$ [21] and approximates the electrical impedance behavior at an operating point [19]. An extension of these model structures for the battery impedance is accomplished using fractional differential equations [22]. They are part of many elements of physically-motivated ECM. For example, the impedance of a CPE can be calculated with a factor Q and a rational exponent α, which is equivalent to a non-integer derivation in the time domain, as follows:

$$Z_{CPE}(j\omega) = \frac{1}{Q(j\omega)^{\alpha}} \tag{3}$$

These abstract models are often used in online application as a part of BMS functions [23]. The reason for that are well established, efficient estimation algorithms for their parameters. However, these models do not allow any direct conclusions on the physical system and thereby complicate the expansion towards global battery models.

To integrate the entire non-linear and time-varying battery behavior into a model, *machine learning* methods like *support vector machine* [24] and *neural network* [25] are applied. Based on large amounts of data, they promise to capture the current-/voltage behavior of LIB at all operating points but are time-consuming in the generation and simulation.

5. Identification methods

Based on the introduced modeling approaches for the dynamic LIB-behavior, algorithms, which use the measurement data for a parameterization of the model, must be applied (cf. Fig. 1). Here, the possibilities range from an estimation of individual model parameters up to an

alternation of the entire model structure, for example, defined by the number of RC-circuits in the TM or by the layers in a neural network.
Furthermore, the measured battery signals may be used either directly in the time domain or through a transformation in the frequency domain.

Identification in the time domain

Most identification methods in the time domain rely on the voltage relaxation behavior caused by charging and discharging pulses [26, 27, 28]. Excitation signals with a purpose of model identification, such as *pseudo random binary sequence* (PRBS) [29] or *chirp* [30] are less common in the battery field.
A well-known profile that uses pulses to measure the battery impedance is the *Hybrid Pulse Power Characterization Cycle* (HPPC) [31]. It first charges the battery within the operating window and then discharges the system for 10 seconds. Afterwards it waits an additional 3 minutes to record the relaxation phase as well. During a HPPC, the state of charge and the temperature of the battery change. Thus, it leaves the operating point, which is important for the identification of a linear transfer function.
An alternative method that works with discharging pulses is the so-called *small signal excitation* [32]. Similarly, to the HPPC the duration of the pulses is only a few seconds, but the subsequent relaxation period is considerably longer.
The literature provides several methods for parameter identification depending on the model structure and the excitation of the system. For example, the time constants and of the RC-circuits in a TM (2nd order) can be determined graphically [26]. Firstly, this approach calculates the internal resistance via the current-/voltage-ratio immediately after the current step. Secondly, the slower time constant is determined based on a pre-defined time window of the relaxation phase and is then re-used for the calculation of the fast time constant. The method relies on the assumption, that the battery dynamics can be directly linked to two different effects. However, the various electrochemical phenomena overlap and have similar time constants. Hence, this approach might be erroneous.
The linear regression of the impedance behavior is commonly performed using least squares methods [12, 33]. Therefore, the squares of the residuals between the model and the measurement are minimized. As a result of this optimization problem, the parameter vector p of a TM (any order) can be calculated using the following formula:

$$p = \underbrace{(A^T A)^{-1} A^T}_{A^+} y \tag{4}$$

The sampled measurement data of the input (current) and output (voltage) are thereby expressed as the output vector y and the regression matrix A. However, it can be proven, that the resulting model always has a bias for measurement signals with colored noise [34].
Many other identification methods rely model structures with fractional transfer functions. Alavi [27] uses parallel circuits, consisting of an ohmic resistor and a CPE (cf. equation (3). He converts his ECM into a mathematical structure, similar to an ARX model, and can thus estimate the parameters efficiently. The determination of the rational exponents takes place afterwards through approximation in the time domain.
Besides, Alavi applies the *instrumental variable method* (IV). The IV method attempts to minimize the error of the model parameters (*bias*) by reducing the coupling between the disturbance dynamics and the system dynamics in the measured output signal y [3]. This is achieved by a multi-staged identification process, which takes account of the residuals between parameter estimation and measurement in later process steps [33].

Identification in the frequency domain

The second important technique in the identification of the behavior of the electric battery is based on impedance data. Obtaining this data, phase shift and magnitude of the voltage response are measured for different alternating currents. An extensive investigation of this transfer function ─ depending on the excitation frequency ─ is subject to the so-called *electrochemical impedance spectroscopy* (EIS) [35]. Such an analysis of the linear transfer function of the cell impedance is only useful at one operating point regarding temperature, state of charge and current amplitude [32].
The result of the EIS is usually visualized by a locus in the complex plane (cf. measurement in Fig. 3) and is the starting point for modeling the dynamic voltage response in the frequency domain.
Most modeling approaches, especially ECM, are expressed as transfer functions in the frequency domain (cf. last chapter). For example, the RC-circuits of the TM can be interpreted as a simple semi-circle in the complex plane.
An estimate of the model parameter vector p - similarly to the time domain - is carried out by minimizing the error squares between the impedance values of the measurement Z_{Meas} and the model Z_{Mod} for the N_f recorded frequencies f [36]:

$$\min_{p} \sum_{i=1}^{N_f} \sqrt{\left(Re\{Z_{Meas,i}\} - Re\{Z_{Mod,i}(p, f_i)\}\right)^2 + \left(Im\{Z_{Meas,i}\} - Im\{Z_{Mod,i}(p, f_i)\}\right)^2} \quad (5)$$

For a Thévenin-model (2nd order) this identification algorithm provides the approximation represented in Fig. . The result shown also includes an approach based on Birkl [13], who determines the initial parameters directly from the impedance curve.

The Fig. makes it clear that the underlying model structure hardly meets the real physical phenomena. For that reason, many methods found in the literature, that focus on modeling in the frequency domain, integrate fractional models such as ZARC- (cf. equation 3) or Warburg-elements [38, 39].

Each data point in the EIS is produced by its own sinusoidal excitation. Consequently, there are methods of determining the locus in the frequency domain with fewer effort and only single current excitation signals such as PSBR [39] or pulses [32].

For this, the N_t-times sampled current i_n and voltage signal u_n must be converted into the frequency domain via the *discrete Fourier transform* (DFT):

$$Z_k = \frac{U_k}{I_k} = \frac{\sum_{n=0}^{N_t-1} u_n e^{-\frac{j2\pi}{N_t}kn}}{\sum_{n=0}^{N-1} i_n e^{-\frac{j2\pi}{N_t}kn}} \quad (6)$$

This formula transforms the time domain signal with a sampling rate Δt over the condition $\Delta\omega\Delta t N = 2\pi$ in the complex transfer values Z_k with the frequency resolution $\Delta\omega$. To obtain an impedance curve from discharging pulses that is equivalent to EIS results, the

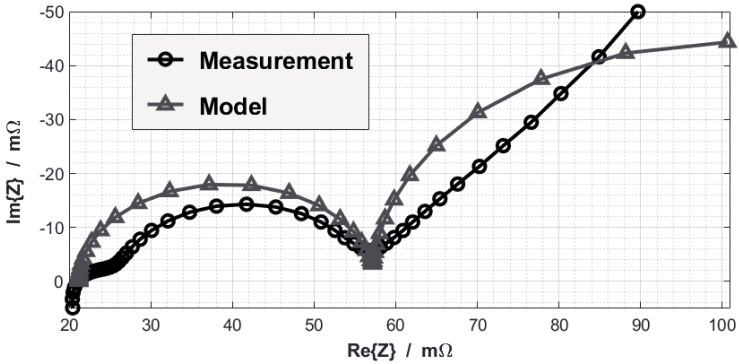


Fig. 3: Identification of the measured impedance behavior (2.9 Ah Panasonic 18650PF [37]) at a temperature of 27 °C and a state of charge of 95 %

leakage effect and system noise have to be taken into account. A promising approach [7] is the application of the current and voltage derivative, instead of a window function. This method is based on the assumption that at the beginning and after the relaxation both derivatives are zero. Thus, there is no leakage effect [32].

6. Conclusion

In this study, demands on battery models for the drive train development were successfully stated. In addition to a limitation of the model ranges regarding current, temperature and state of charge, this paper demonstrates that a simulation of the frequency components in the voltage signal up to 50 Hz is sufficient for most development tasks.

There is a strong focus on physically-motivated ECM when it comes to available methods for the identification of the global battery behavior. They are mostly parameterized with frequency data derived from operating points.

However, in the time domain, rather abstract dynamic models like ARX, ARMAX- or fractional-models are used to reflect the local impedance behavior. An extension of these methods beyond the identified operating points is hardly considered in the literature. In general, many applications focus more on the mathematical model structure than on robust and efficient estimation algorithms. This situation is the starting point for future research and publications.

7. References

[1] J. Wang and I.J.M. Besselink, “Evaluating the energy efficiency of a one pedal driving algorithm,” in *European Battery, Hybrid and Fuel Cell Electric Vehicle Congress*, Brussels, 2015.

[2] P. Shen and M. Ouyang, “State of Charge, State of Health and State of Function Co-Estimation of Lithium-Ion Batteries for Electric Vehicles,” in *IEEE Vehicle Power and Propulsion Conference*, Hangzhou, China, 2016.

[3] L. Ljung, System identification: Theory for the user: Prentice Hall, 1999.

[4] K. Uddin and A. D. Moore, “The effects of high frequency current ripple on electric vehicle battery performance,” *Applied Energy*, vol. 178, 2016.

[5] R. Korthauer, Lithium-Ion Batteries: Basics and Applications. Berlin: Springer, 2018.

[6] M.R. Jongerden and B. R. Haverkort, *Battery Modeling*: Design and Analysis of Communication Systems, 2008.

[7] Jinhao Meng and Guangzhao Luo, “Overview of Lithium-Ion Battery Modeling Methods for State-of-Charge Estimation in Electrical Vehicles,” *Applied Sciences*, vol. 8, 2018.

[8] S. Barcellona and L. Piegari, “Lithium Ion Battery Models and Parameter Identification Techniques,” *Energies*, vol. 10, 2017.

[9] M. Doyle, “Modeling of Galvanostatic Charge and Discharge of the Lithium/Polymer/Insertion Cell,” *Journal of The Electrochemical Society*, vol. 140, 1993.

[10] Carla-Fabiana Chiasserini and Ramesh R. Rao, “Energy efficient battery management,” *IEEE Journal on selected areas in communications*, vol. 19, 2001.

[11] J. Wehbe and N. Karami, “Battery equivalent circuits and brief summary of components value determination of lithium ion: A review,” in *Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering*, Beirut, Lebanon, 2015.

[12] A. I. Pozna and A. Magyar, “Model identification and parameter estimation of lithium ion batteries for diagnostic purposes,” in *19th International Symposium on Power Electronics Ee*, Novi Sad, 2017.

[13] C. R. Birkl and D. A. Howey, “Model identification and parameter estimation for LiFePO4 batteries,” in *IET Hybrid and Electric Vehicles Conference*, London, UK, 2013.

[14] Heiko Witzenhausen, “Elektrische Batteriespeichermodelle: Modellbildung, Parameteridentifikation und Modellreduktion,” Ph.D. thesis, RWTH Aachen, 2017.

[15] M. Schönleber, “Verfahren zur Charakterisierung des Niederfrequenzverhaltens von Lithium-Ionen-Batterien,” Ph.D. thesis, Karlsruher Institut für Technologie, 2017.

[16] M. Oldenburger and B. Bedürftig, “Investigation of the low frequency Warburg impedance of Li-ion cells by frequency domain measurements,” *Journal of Energy Storage*, vol. 21, 2019.

[17] Ulrike Krewer and Fridolin Röder, “Review - Dynamic Models of Li-Ion Batteries for Diagnosis and Operation: A Review and Perspective,” *Journal of The Electrochemical Society*, vol. 165, 2018.

[18] Johannes Schmalstieg, “Physikalisch-elektrochemische Simulation von Lithium-Ionen-Batterien: Implementierung, Parametrierung und Anwendung,” Ph.D. thesis, RWTH Aachen, 2017.

[19] M. Zhang and Z. Miao, “Battery identification based on real-world data,” in *North American Power Symposium*, 2017.

[20] F. Conte, “Characterization Methods for the State of Charge Estimation of Lithium-Ion Batteries,” in *3rd International Hybrid Power Systems Workshop*, Tenerife, Spain, 2018.

[21] Frank Kirschbaum, “Modellbasierte Applikationsverfahren,” Vorlesungsskript, Karlsruher Institut für Technologie, 2014.

[22] Oliver Stark, “Einführung in die fraktionale Analysis,” Skriptum, Karlsruher Institut für Technologie, 2017.

[23] Ngoc-Tham Tram and Kim-Hung Nguyen, “SOC/SOH Estimation Method for AGM VRLA Battery by Combining ARX Model for Online Parameters Estimation and DEKF Considering Hysteresis and Diffusion Effects,” in *9th International Conference on Power Electronics*, Seoul, Korea, 2015.

[24] I. A. Majid and R. F. Rahman, “Electric vehicle battery dynamics modelling using support vector machine,” in Proceedings of the 2013 Joint International Conference on Rural Information & Communication Technology and Electric-Vehicle Technology, Bandung, Indonesia, 2013.

[25] R. Zhao and P. J. Kollmeyer, “A compact unified methodology via a recurrent neural network for accurate modeling of lithium-ion battery voltage and state-of-charge,” in *IEEE Energy Conversion Congress & Expo*, Cincinnati, OH, 2017.

[26] A. Hentunen and T. Lehmuspelto, “Time-Domain Parameter Extraction Method for Thévenin-Equivalent Circuit Battery Models,” *IEEE Transactions on Energy Conversion*, vol. 3, 2014.

[27] S.M.M. Alavi and C. R. Birkl, “Time-domain fitting of battery electrochemical impedance models,” *Journal of Power Sources*, vol. 288, 2015.

[28] R. Jackey and M. Saginaw, “Battery Model Parameter Estimation Using a Layered Technique: An Example Using a Lithium Iron Phosphate Cell,” in *SAE Technical Paper Series*, 2013.

[29] J. Sihvo and T. Messo, “Online identification of internal impedance of Li-ion battery cell using ternary-sequence injection,” in *IEEE Energy Conversion Congress and Exposition*, Portland, OR, 2018.

[30] J. C. Forman and S. J. Moura, “Genetic parameter identification of the Doyle-Fuller-Newman model from experimental cycling of a LiFePO4 battery,” in *Proceedings of the 2011 American Control Conference*, San Francisco, CA, 2011.

[31] Suguna Thanagasundram and Raghavendra Arunachala, “A Cell Level Model for Battery Simulation,” in *European Electric Vehicle Congress*, Brussels, Belgium, 2012.

[32] M. Oldenburger and B. Bedürftig, “A new approach to measure the non-linear Butler–Volmer behavior of electrochemical systems in the time domain,” *Journal of Energy Storage*, vol. 14, 2017.

[33] B. Xia and X. Zhao, “Accurate Lithium-ion battery parameter estimation with continuous-time system identification methods,” *Applied Energy*, vol. 179, 2016.

[34] R. Isermann and M. Münchhof, *Identification of Dynamic Systems*. Berlin: Springer, 2011.

[35] M. E. Orazem and B. Tribollet, *Electrochemical impedance spectroscopy*. Hoboken, NJ: Wiley, 2008.

[36] D.-I. Stroe and M. Swierczynski, “Generalized Characterization Methodology for Performance Modelling of Lithium-Ion Batteries,” *Batteries (MDPI)*, vol. 2, 2016.

[37] P. Kollmeyer, “Panasonic 18650PF Li-ion Battery Data,”

[38] H. Piret and P. Granjon, “Tracking of electrochemical impedance of batteries,” *Journal of Power Sources*, vol. 312, 2016.

[39] R. al Nazer and v. Cattin, “Broadband Identification of Battery Electrical Impedance for HEV,” *IEEE Transactions on Vehicular Technology*, vol. 62, 2013.

Condition monitoring for failure monitoring of power electronic assemblies

Stefan Wagner, **Felix Wüst**, **Stefan Trampert**, **Frederic Sehr**, **Andreas Middendorf**, **Olaf Wittler**,
Fraunhofer Institut für Zuverlässigkeit und Mikrointegration (IZM), Berlin
Martin Schneider-Ramelow, Technische Universität Berlin

Abstract

Ensuring safe and reliable functionality of electronical components, especially power electronic assemblies, over the life of the vehicle, will be an ever greater challenge for highly automated driving (applications) in the future. This makes condition monitoring and failure forecasts for power electronics in electric vehicles all the more important. The circuit approach presented in this study of an independent electronic development, as part of the power electronic module level, and the results of the realized measurements already achieved in previous studies are dedicated to the ambitious goal of ensuring the functional reliability of power electronic systems in the field of electro-mobility by continuously determining the state over the entire service life. This is realized by the symbiosis of two fundamentally different concepts of condition monitoring (condition indicator, parameter determination).

The first concept presented here continuously measures Vce of a power electronic module during operation and thus enables status monitoring. Previous studies [1] have shown that by evaluating the change in Vce of the IGBT, statements can be made about the state of degradation and thus about the remaining service life. This allows for a continuous condition monitoring of the power electronics in the application.

The second concept, which is presented here, is based on the measurement of the collector-emitter forward voltage (Vce_on). [2]

Condition Monitoring

There are three concepts of condition monitoring (Fig. 1). The first concept is parameter monitoring, where a parameter inherent to the system that is sensitive to the aging process is constantly monitored (Fig. 1a). The second concept is known as the canary device (Fig.- 1b). This refers to a component unnecessary to the devices functionality that is loaded just as the regular components but has a weaker structure. Therefore it will fail prematurely. Different structural

layouts can be used to implement different warning stages so that a prediction becomes quite accurate. The third concept, the so-called life cycle unit (Fig. 1c), constantly monitors the load upon a device and predicts the remaining life-time via an underlying life-time model [24].

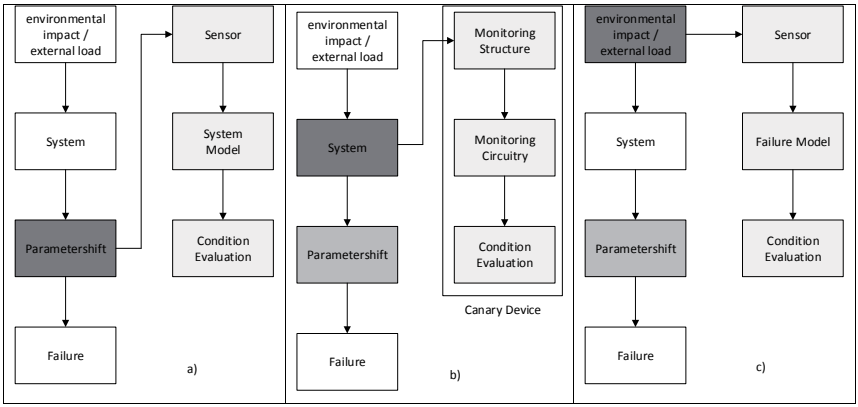


Fig. 33: Concepts of condition monitoring: a) parameter monitoring b) canary device c) life cycle unit [3]

As the IGBT's die temperature is dependent on both load and state of damage, a combination of the first and the third concept is feasible for usage in this research. Temperature swings, being one of the most critical aspects regarding lifetime. These fast temperature changes should be counted as well as a continuously increasing die temperature under similar working conditions might be observed and related to a progressed damage.

Failure mechanism

The thermomechanical stress during operation usually leads to damage of power electronic modules and in the long run to their failure. The failure of the module can be caused by different error mechanisms in different areas of the module.

The following Failure mechanism dominates the degradation of power electronic modules. [2]

- Bond wire fatigue
- surface reconstruction
- solder fatigue

Temrature sensitive electrical parameters (TSEP)

Since direct temperature measurement is not suitable for determining the chip temperature during operation, it must be carried out indirectly. This is usually done by measuring a temperature-sensitive electrical parameter (TSEP) and then converting it into temperature. The TSEPs suitable for this purpose show a dependence on the junction temperature Tj, which corresponds to the temperature inside the chip.

- threshold voltage
- Maximum gate current
- trans conductance
- Duration of the Miller Plateau
- collector-emitter forward voltage

Table 1 sums the boundary conditions of TSEP.

Table 1: Summary of the TSEP boundary conditions

TSEP	sensitivity	time resolution	influencing variables
threshold voltage	high	ns	Tj
Maximum gate current	high	ns	Tj
trans conductance	high	ns	Tj, Vge
Duration of the Miller Plateau	high	ns	Tj, dVce/dt, Ic
collector-emitter forward voltage	very high	µs	Tj, Ic

Concept 1: Measurement of Vce

This concept is based on the interpretation of the temperature dependent behavior of the Miller Plateaus, which is described in the previous chapter.

A measurement circuit is developed and realized, which is shown in Fig. 2.

The gate driver takes care of supplying sufficient energy to the IGBT's gate. Three high speed comparators monitor the gate voltage. These define the start and two stop events, of which only one is evaluated at the moment. The time to digital converter (TDC), an Acam GP 22, measures the time between start and stop with a sampling accuracy of up to 22 ps. The resulting time is communicated via SPI over an isolation barrier to the evaluation unit. In addition, current and collector-emitter voltage are measured by analog digital-converters (ADC) and communicated to the evaluation unit. To ensure that the measurements taken by the digital

acquisition unit (DAQ) are within reasonable limits, the temperatures on the heatsink and the module will also be measured using thermocouples.

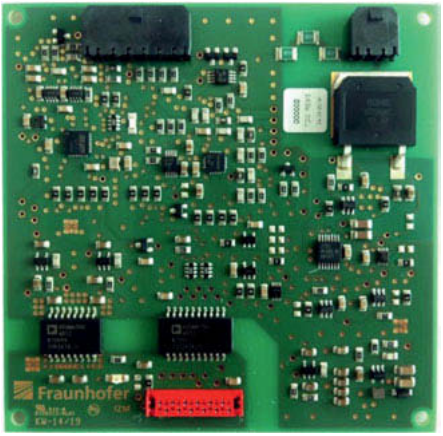


Fig. 2: Image of prototype board of integrated delay measuring on gate driver.

The selected IGBT is an IXGN 200N60A2 in a SOT-227B package with a current rating of 200 A and a maximum blocking voltage of 600 V. Its field of application is motion control, DC choppers and uninterruptible power supplies. The IGBT is turned on for a pulse duration of 30 ms and turned off afterwards. This allows for a test with relatively low self-heating. This setup allows testing the measurement method with close to real life application blocking voltages and high currents. Fig. 3 shows an exemplary development of the gate emitter voltage during turn-off with threshold voltages ($Vstart$ = 12.8 V and $Vstop$ = 0.73 V) and time measurement. The parasitic inductivity of the gate circuit is quite high due to the fact that the board is not directly mounted onto the IGBT, but connected via wires of approximately 6 cm length. This leads to a short oscillation peak between 10 ns and 150 ns. The Miller Plateau ends approximately at 300 ns where the gate voltage decreases again. The second threshold voltage has to be lower than the Miller Plateau level to trigger the end reliably. The starting threshold voltage however may be set in a greater range due to the high dv/dt during the first nanoseconds.

To compare and benchmark the TDC measurement system, the electrical parameters are also monitored with a 12-bit oscilloscope. The evaluation is done equally with the same threshold voltages. The corresponding collector-emitter voltage is shown in Fig. 4. For a current of 50 A the turnoff voltage should result in 200 V with a snubber resistor of 4 Ω. Due to the parasitic inductance of the snubber circuit the switch-off voltage rises above 300 V.

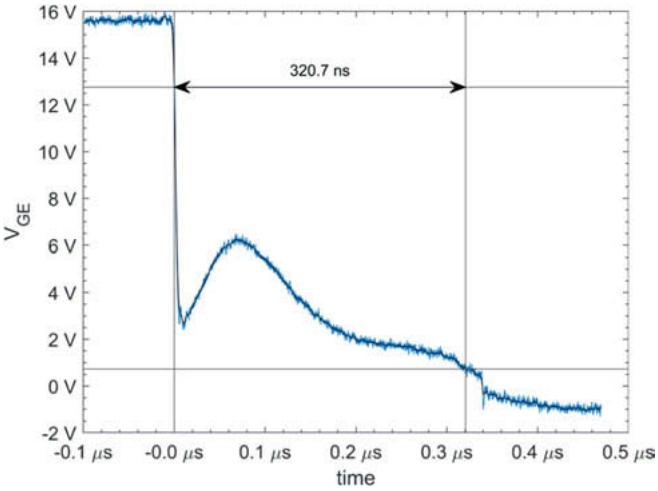


Fig. 3: Exemplary gate emitter voltage during turn-off and corresponding time measurement at threshold voltages.

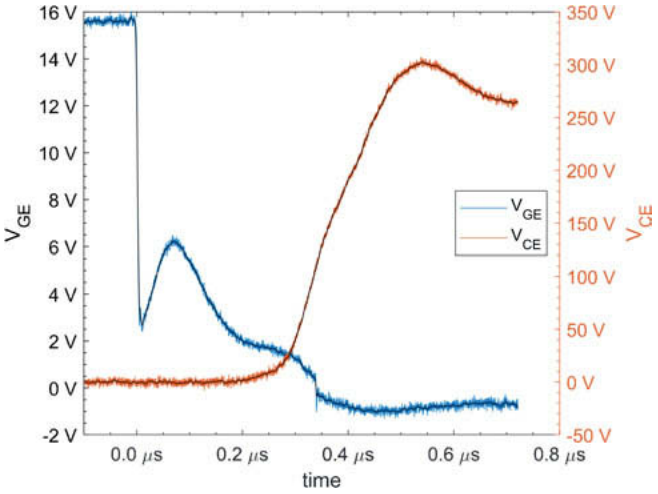


Fig. 4: Exemplary collector-emitter-voltage with corresponding gate-emitter voltage during turn-off.

The switching losses, i.e. the non-linear behaviour of the semiconductor, result in a current dependent attenuation of the snubber. This results in a non-linear curve between the maximum breaking voltage and the breaking current as shown in Fig. 5. However, a linear function is still a good approximation.

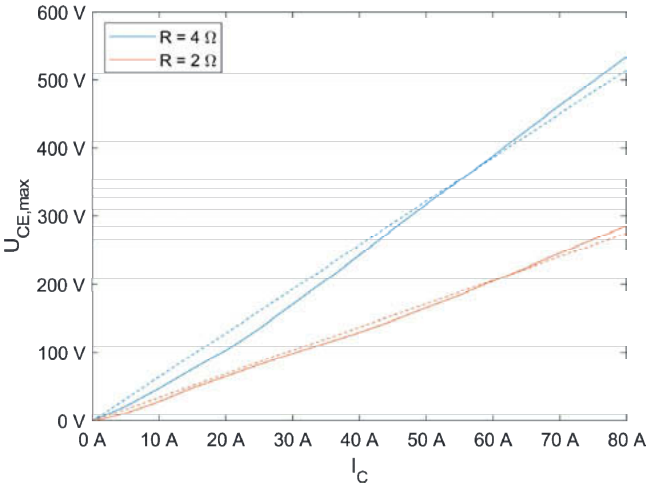


Fig. 5: Maximum collector-emitter-voltage peak during turn-off with linear fit (dashed line) at room temperature.

Fig. 6 shows the comparison of the turn-off duration versus temperature under different DC currents with a snubber resistor of 2 Ω. This shows an obvious difference The linear fit is depicted with a strong line whereas the 95 % error estimates are painted with the corresponding lighter colour.

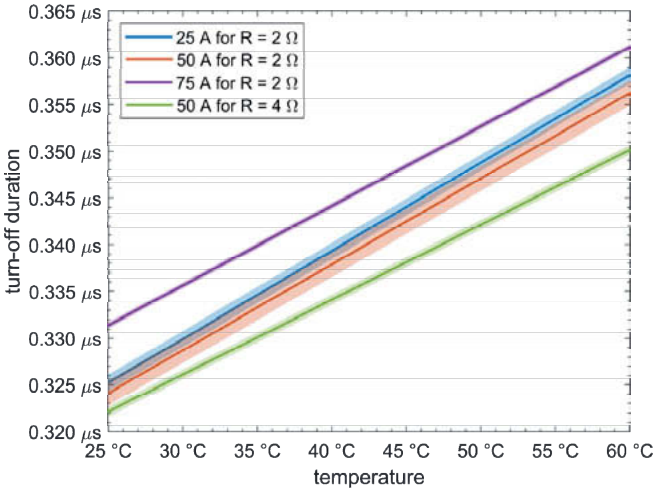


Fig. 6: Comparison of influences of different currents onto the temperature dependency for two snubber resistor values.

Table 2 sums up the gradients of the different curves.

Table 2: Temperature coefficients for different currents

Current [A]	R [Ohm]	Gradient [ns/K]	Offset [ns]
25	2	0.9429	301.62
50	2	0.9193	301.11
75	2	0.8516	310.08
50	4	0.8022	302.02

Concept 2: Measurement of Vce_on

In a second concept study the collector-emitter forward voltage is used as a standard indicator to determine the junction temperature of IGBTs. Due to the required sensitivity of the measuring system of a few mV, the large difference between forward and reverse voltage and the short duty cycle of the IGBT in the µs range, measurement in power converters during operation is complex, but offers the possibility of measuring the temperature more accurately than with commonly used temperature sensors. A major challenge with this concept is that the measurement is on the high-voltage side and the measuring system must be designed to be voltage safe.

Based on the application, the requirements shown in table 3 are realized.

Table 3: Requirements for the measuring system and values achieved

Requirements	required value	achieved value
nominal voltage	1200 V	488 V (measured) > 1200 V (data sheet)
resolution	3 mV	3 mV (Vce_on < 2 V)
minimum measuring duration	2 µs	5 µs
Measurement delay	1 µs – 30 s	4.5 µs – 67 s
Switch-off time of the MOSFETs	160 ns	70 ns – 140 ns

In Fig. 7 the measurement setup is shown.

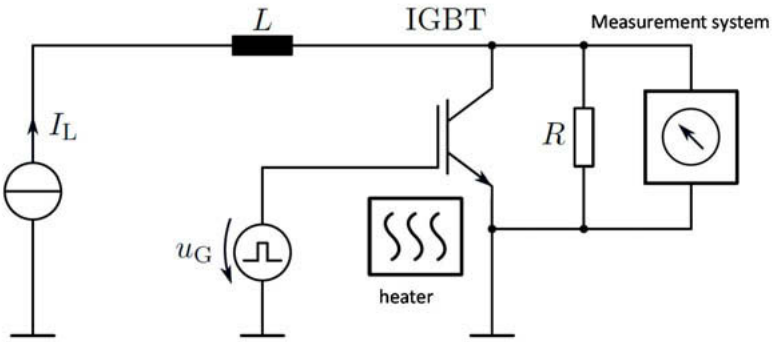


Fig. 7: Measuring stand for recording temperature-voltage characteristics

By an inductance L connected in series to the IGBT and a resistor R connected in parallel to the IGBT, voltage pulses with a current-dependent peak voltage of several 100V are generated at the collector of the IGBT at the moment the IGBT is switched off and the course of Vce during operation of the IGBT with high voltages is simulated.
The required nominal voltage of 1200V can only be checked up to 488V during the application test. Since the MOSFET used for the measuring system has a rated voltage of 1700V and the resistance of the MOSFET to parasitic switching increases with increasing voltage, it can be assumed that the measuring system is also safe with reverse voltages > 1200V.

The resolution of the measuring system is for Vce,on < 2V less than 3mV, for Vce,on < 2.5V less than 8mV. Thus, for IGBTs operated at nominal current with Vce,on < 2V a reliable temperature determination down to below 1K is possible, for IGBTs with Vce,on < 2.5V accurate to 2K to 3K. Since IGBTs of the voltage class < 1700V usually have forward voltages of less than 2V, this does not represent a limitation for the majority of possible application scenarios.

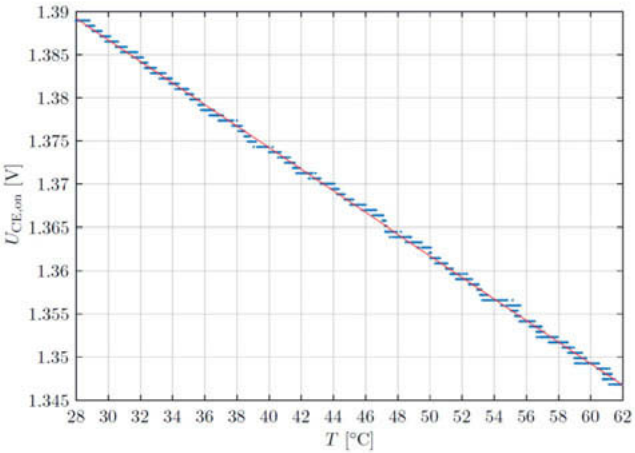


Fig. 8: Exemplary measured temperature-voltage characteristic for Ic = 75A and compensation line

Fig. 8 shows the measured temperature-voltage characteristic for the load case Ic = 75 A. The measuring system fulfills approximately the specified requirements and shows that the determination of the junction temperature of IGBTs can be carried out by measuring the collector-emitter forward voltage while a power converter is in operation and that it is fundamentally superior to measurement by temperature sensors. The temperature resolution achievable with the measuring system is in the range of the limit deviation of common temperature sensors such as thermocouples or NTC thermistors or even below, but the time resolution achievable with the measuring system in the µs range is far below the time resolution of the temperature sensors, which is typically in the ms range or above. At the same time, temperature sensors can only be used to measure the temperature at a spatial distance from the chip, which means that the high temperature strokes occurring on the chip can only be detected to a limited extent, while the Vce,on measurement can be used to determine the temperature inside the chip.

Therefore, rapid temperature changes occurring during peak loads of the IGBT can be detected much better with large temperature strokes of the chip than with temperature sensors.

Conclusion

In this paper two concepts for the measurement of temperature dependent electrical parameters are presented. These electrical parameters allow to interpret a gradient to the initial value as a reduction of the module reliability, which allows to realize a condition monitoring for power electronic modules.

Looking at the results from concept 1, one can see that for most applications, the DC link voltage is quite stable with only small changes in amplitude during operation. Therefore, the proposed method may be used, only being calibrated for currents and temperatures even though the parameter of interest shows a dependence on voltage. In addition, the voltage dependence is small compared to those on current and temperature. Using a double pulse method for calibration, a characteristic curve map can be obtained and embedded into the microcontroller. However, this needs to be done for each semiconductor type that is used with a few parts, so that sufficient statistics can be used.

The measuring system concept 2 allows the measurement of the junction temperature at a forward voltage below 2V with a minimum measuring duration of 4.5 µs at an achievable temperature resolution of ±1K and thus almost meets the requirements. This paper shows that the measurement of the collector-emitter forward voltage is well suited for determining the junction temperature of IGBTs in power converters during operation and is superior to temperature measurement by commonly used sensors. The achieved temperature resolution of the measuring system is comparable to the temperature resolution of conventional temperature sensors, but the achieved time resolution is much higher. The necessary measurement technology can be implemented with small dimensions, low energy consumption and without changes to the investigated IGBT power module and is therefore very well suited for long-term measurements and for condition monitoring of the IGBTs.

With this method, detailed information of the chip temperature and therefore its load can be recorded. The temporal temperature resolution is very important as previous research has shown [1]. With the proposed method integrated into a gate driver board the chip temperature can be monitored continuously during operation. Life-time models can be implemented using this information.

Acknowledgements

As part of the AMWind project, the Federal Ministry of Economics and Technology funded the research leading to the results, which are described as in concept 1.

References

[1] F. Wüst et al. „Comparison of temperature sensitive electrical parameter based methods for junction temperature determination during accelerated ageing of power electronics“, ESREF 2018: 29th European Symposium on Reliability of electron devices, failure physics and analysis.

[2] F. Sehr. “Entwicklung eines Messsystems zur Ermittlung der Kollektor-Emitter-Durchlassspannung als Basis für Lebensdauerbewertungen von IGBTs“. Master thesis, TU-Berlin (2019)

[3] K. Jerchel, M. Krüger, A. Middendorf, N. F. Nissen und K.-D. Lang, „Reliability Improvements in Electronic Systems by Combining Condition Monitoring Approaches,“ in International Conference on Electronics Packaging (ICEP) 2014, Toyama, Japan, 2014

[4] M. Ciappa. „Selected failure mechanisms of modern power modules“. In: Microelectronics Reliability 42.4 (2002), S. 653–667

Holistic Energy Management of 48V Mild Hybrid Vehicles

Philip Griefnow, Prof. Dr.-Ing. **Jakob Andert**,
Michael Engels,
RWTH Aachen University, Aachen;
Dr.-Ing. **Johannes Richenhagen**, **Dejan Jolovic**,
FEV Europe GmbH, Aachen

Abstract

The 48V technology plays an important role in the electrification strategy of many automobile manufacturers. With moderate technical effort, this technology enables short-term CO_2 savings in the vehicle fleet and also significantly reduces real driving emissions (RDE). With the variety of functionalities, such as brake energy recovery, load point optimization and engine stop sailing, as well as electrification options in the areas of turbocharging, vehicle dynamics, air conditioning or exhaust aftertreatment, it is already foreseeable today that the power and energy reserves of competitively designed 48V systems quickly reaches its limits. In addition, a rising number of 48V components increases the electrical load dynamics as well as the degrees of freedom with regard to the operating strategy. While system complexity as well as dynamic interactions take rule-based operating strategies to their limits, predictive approaches currently concentrate only on individual domains of the 48V system, thus wasting valuable potential. A holistic energy management that optimally distributes the available electrical energy and power in the 48V power net is therefore a promising approach, since it enables the best possible operation of cost- and resource-saving dimensioned 48V systems.

Against this background, this investigation presents a non-linear, model-predictive approach that allows to take a holistic view of the different time horizons of electrical energy and powertrain management of 48V hybrid vehicles. The innovative optimization-based operating strategy anticipates driving situations, shifts the electrical load demands of time-uncritical consumers, and can thus optimally utilize the limited performance of the 48V power net.

48 Volt

In addition to pure electric and high-voltage hybrid vehicles, the 48V technology also plays a decisive role in the electrification strategy of many automobile manufacturers. One of the most important advantages of this technology is the CO_2 fleet emission reduction with comparatively little effort and development time. In addition to the CO_2 advantage, the 48V electrification offers considerable potential to significantly reduce real driving emissions (RDE). Already today it is foreseeable that in future the number of implemented 48V functions and components will increase distinctly. Besides typical hybrid functions like brake energy recovery, load point optimization or engine stop sailing, many other high-performance electrical consumers, which support air conditioning, turbocharging, driving dynamics, exhaust gas aftertreatment or chassis systems are conceivable. This suggests however that competitively designed 48V systems will be limited in terms of power and energy reserves. The comparison with high-voltage hybrid systems in Fig. 1 illustrates that the operating range of 48V mild hybrid systems quickly reaches its limits.

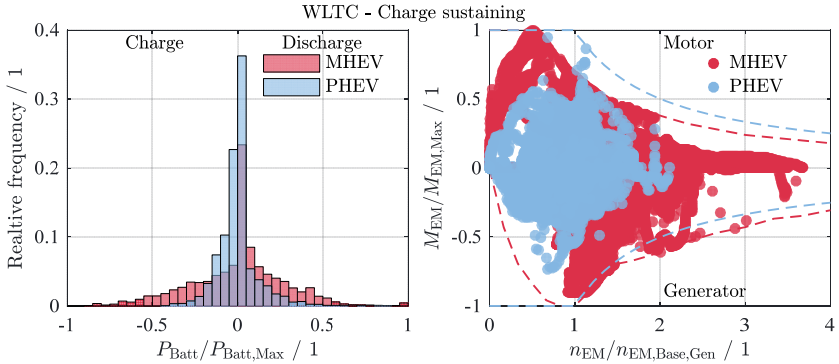


Fig. 1: Comparison of the operating ranges of a high-voltage plug-in hybrid (PHEV) and a 48V mild hybrid (MHEV) in the WLTC [8]

The expected multitude of new 48V components and functions not only increases the dynamic load demand on the supply network, but also the degrees of freedom that result for the operating strategy. Together with a high system complexity, dynamic boundary conditions and interaction possibilities, this is pushing rule-based control approaches to their limits. Hence, a promising approach is a predictive, optimization-based energy management, which optimally distributes the available electrical energy and power in the 48V power net and thus enables the best possible operation of cost- and resource-saving dimensioned 48V systems. [8]

FEV Concept Vehicle

In cooperation with RWTH Aachen University, FEV has developed a 48V mild hybrid concept vehicle. It is based on a Mercedes-Benz AMG A45 with all-wheel drive and a 7-speed dual-clutch transmission. The series vehicle is equipped with a turbocharged 2-litre gasoline engine with a specific power output of 133 kW/l. This remarkable performance is achieved by the use of a large exhaust turbocharger (TC), which despite twin scroll technology significantly limits the maximum torque in the lower engine speed range and leads to a noticeably delayed response. [1, 2] In this context, electrified supercharging and/or electrical torque assist can significantly improve elasticity, particularly in the low speed range where the fuel consumption is low. [3, 4] The 48V mild hybrid powertrain of the concept vehicle is shown in Figure 2.

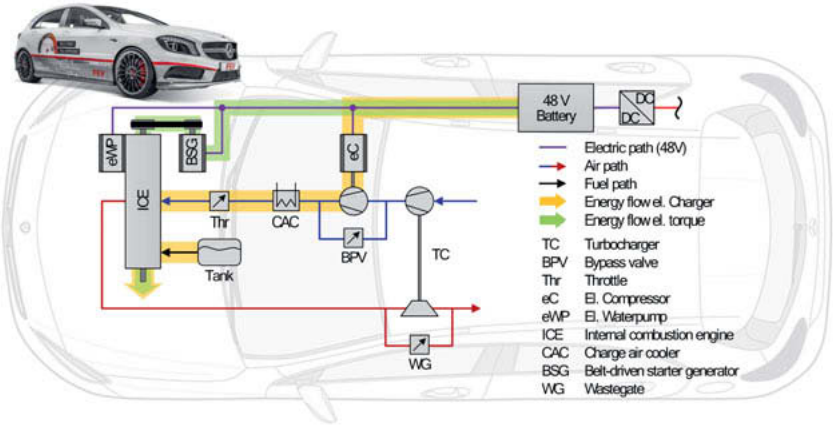


Fig. 2: 48V mild hybrid powertrain of the FEV concept vehicle

Central element is the belt starter generator (BSG) within the accessory belt drive of the internal combustion engine (ICE), which replaces the original 12V alternator. The P0 topology enables hybrid functions such as energy recuperation, load point shift and electrical torque assist. Since the maximum power that can be transmitted by the belt is limited and there is a permanent coupling to the ICE, the system is not predestined for purely electric driving. In order to improve the dynamic response behaviour and increase the torque output at low engine speeds, an electric compressor (eC) is positioned in the charge air path upstream of the charge air cooler. The 12V power net is supplied by a bi-directional DC/DC converter with a peak power of up to 3.5 kW. Due to the modifications on the belt drive, the mechanical coolant pump replaced by an electric 48V variant. A lithium iron phosphate battery module with a capacity of

8 Ah and a maximum discharge power of 15 kW is used for energy storage. The concept vehicle is operated using a Rapid Control Prototyping (RCP) control unit and a model based control software which is based on the FEV PERSIST® architecture. [3]

Rule-based Control Strategy

The electrical supercharging via the eC and the electrical torque assist of the BSG are controlled via a performance oriented rule-based operating strategy with a priority-based power distribution. The approach is illustrated in Fig. 3. It consists of the torque assisting functions in the powertrain management and the higher-level power distribution in the electrical energy management.

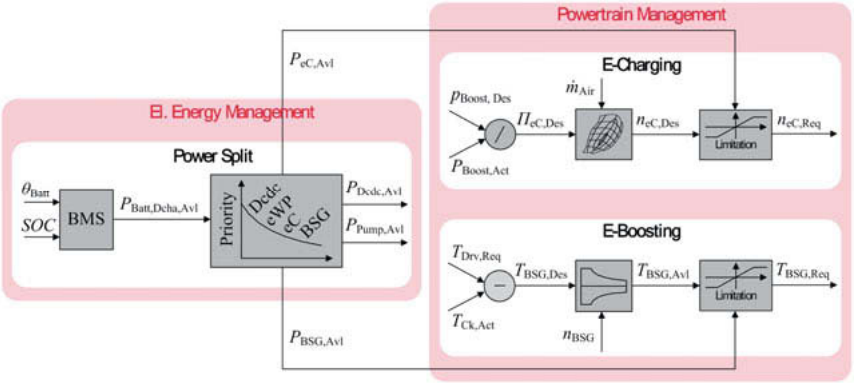


Fig. 3: Performance oriented rule-based operating strategy with priority-based power distribution [8]

The electric supercharging control is based on the pressure ratio of desired and actual boost pressure in the intake manifold. As long as the waste gate (WG) controlled stock TC does not deliver the desired boost pressure, the pressure in the air path is additionally increased by the eC. The required speed is calculated from the compressor map of the eC and then adapted to the available electrical power. In contrast to electric supercharging, in which the propulsion energy results from the additional air and fuel mass, the BSG converts electrical energy directly into mechanical propulsion energy, which supports the ICE (see Fig. 2). The torque required by the BSG results from the difference between the actual torque delivered by ICE and the driver's request. When the accelerator pedal is pressed, this difference is positive and the BSG will transiently compensate for the lack of torque. The BSG torque is then limited according to the available electrical power.

The electrical power limits of the individual 48V components are determined by the electrical energy management (see [5]). During acceleration, in addition to the eC and the BSG, the 48V battery must also supply the coolant pump and the 12V system via the DC/DC converter. The available battery discharge capacity is specified by the battery management system (BMS). Thereto, it is necessary to prioritize the different 48V components according to the driving situation. The available electrical power for each 48V component is then calculated depending on its priority and the actual power consumption of higher-priority consumers. To ensure reliable vehicle operation, engine cooling and the 12V system have a high priority. The remaining power is made available to the eC and the BSG under consideration of a calibrateable power ratio. Even if such rule-based approaches can be improved by further dependencies, there are disadvantages due to the principle. For example, the operating strategy only reacts to the current system status and adapts the actuators independently of the expected load scenario. However, since the temporal behaviour of the torque build-up and the efficiency depend significantly on the load scenario, the selected operating strategy of the electrified powertrain (ICE with TC, eC and BSG) and the electrical system limits, this control approach is usually suboptimal [3, 6].

Predictive Optimization-based Energy Management

Predictive optimization-based energy management strategies (see Fig. 4) use dynamic route information from an electronic horizon for long-term optimization of the route guidance and speed trajectory [7].

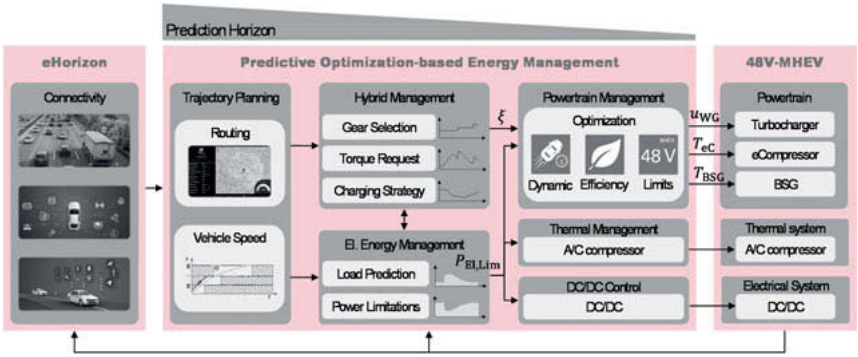


Fig. 4: Predictive optimization based energy management of a 48V-P0-Mild-Hybird powertrain with additional electrical supercharging

Based on this information and suitable vehicle sensors for environmental recognition, the hybrid management calculates optimal trajectories for gear selection, drive torque and charging strategy over a medium-term horizon, taking into account electrical power limits and load prediction. From the predicted states, it is also possible to derive an expected state of charge (SOC) profile of the electrical energy storage device, which adapts an energy weighting factor $\xi(t)$. This factor represents the significance of electrical energy in the energy balance and has a direct influence on energy optimization in powertrain management. At the same time, the response behaviour is optimized by controlling the drive torque, which is composed of the internal combustion engine torque and the BSG torque, while maintaining the dynamic system limits of the 48V system. The predictive optimization-based energy management is investigated in a validated co-simulation of a B-segment 48V mild hybrid powertrain with turbocharged gasoline engine, electrical compression, P0-BSG and electrified air conditioning compressor (eAC) [6].

Electrical Energy Management

The electrical energy management (EEM) is responsible for the electrical load prediction and power limitation. It calculates the power limits out of the available battery power according to the expected loads and submits it to the powertrain management, the thermal management as far as components are electrified and of course to the low level control of components with a pure electrical domain such as the DC/DC converter. In this case, it controls the individual power limits of the DC/DC, the eAC and a combined limit for the BSG and the eC, which are part of the powertrain management and can assist the crankshaft output torque. Since, the 12V system is additionally buffered by the 12V battery and the thermal system has a comparable strong delayed time behaviour, it is potentially possible to partially reduce the load of these components in order to release more potential for torque assist using the BSG and eC. This is actually optimized by a nonlinear model predictive control (NMPC) with a medium time horizon of at least one minute to account for the impact on the thermal and the 12V system.

The behaviour of the NMPC is exemplary shown in Fig. 5. The velocity profile is derived from the high load part of the WLTC. It is assumed that the driver's cabin has heated up to the ambient temperature of 40 °C during parking and must be cooled down immediately after starting. Thus the EEM provides maximum power to the eAC. In acceleration phases, where the torque request significantly increases, the EEM reduces the power supply of the eAC and the DC/DC in order to allow a better torque response. During the optimization the NMPC considers the AC condenser temperature as well as the SOC of the low voltage battery in its cost function

in order to keep it in the range of the desired trajectory. In this way it is possible to intelligently distribute the limited power in the 48V system and ensure maximum system functionality.

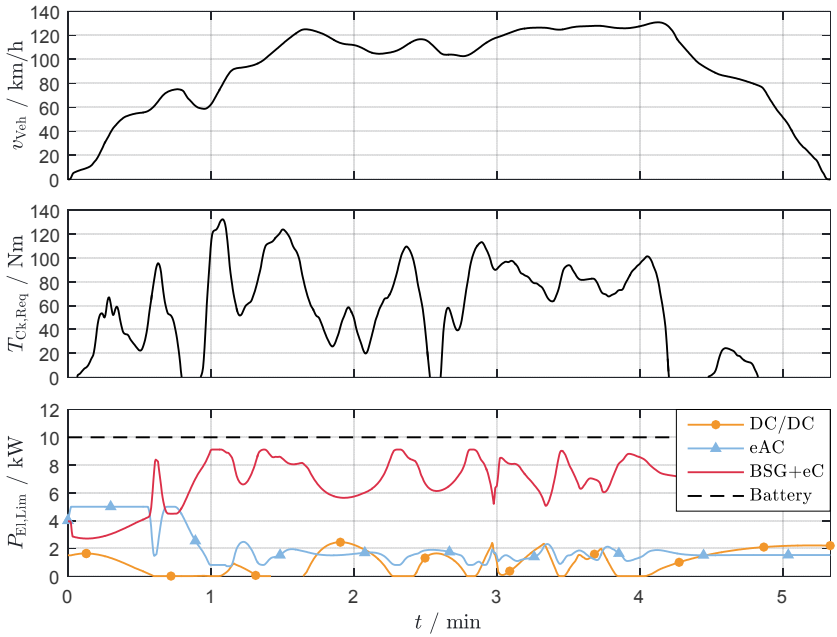


Fig. 5: Functionality of the EEM in the high load part of the WLTC with special temperature boundary conditions

Powertrain Management

Powertrain management is also based on a NMPC, which uses a real-time capable, simplified process model of the 48V mild hybrid powertrain and operates with a time horizon of a few seconds and a time step size in the hundredth to tenth of a second range to map the nonlinear system dynamics. [6] The NMPC calculates the optimum control variables of the WG and the eC, which influence the internal combustion engine torque via the air path, as well as that of the BSG, whose torque is added via the belt drive. In this way, both the differences in the time behaviour of the charge air path and the BSG torque and their influence on the overall efficiency of the electrified powertrain are taken into account in the optimization [8]. Fig. 6 shows a comparison of the simulation results for the NMPC and the rule-based approach at full load

acceleration for different energy weighting factors ξ. An energy weighting factor of four is equivalent to a total charge efficiency of 25 %, while the electrical energy is free in the limit case zero. This may occur, for example, due to a high battery SOC and an imminent downhill drive. Due to the lack of foresight, the rule-based operating strategy reacts identical in both cases, while the NMPC adapts the control variables for WG, eC and BSG depending on the situation to achieve the desired crankshaft torque. In addition, the variation of the optimization parameters shows that the NMPC reduces the crankshaft torque with increasing weighting of the energy ($\tilde{h}_{\mathrm{NMPC}} \uparrow$) in order to minimize energy consumption. If the electrical energy is free of charge ($\xi = 0$), the crankshaft torque is shifted to the BSG, while the eC builds up boost pressure when the WG is open in order to reduce the pumping losses. In contrast, at $\xi = 4$, the NMPC only shortly assists with the BSG in order to exploit the fast dynamics of the electric machine and subsequently save electrical energy [8].

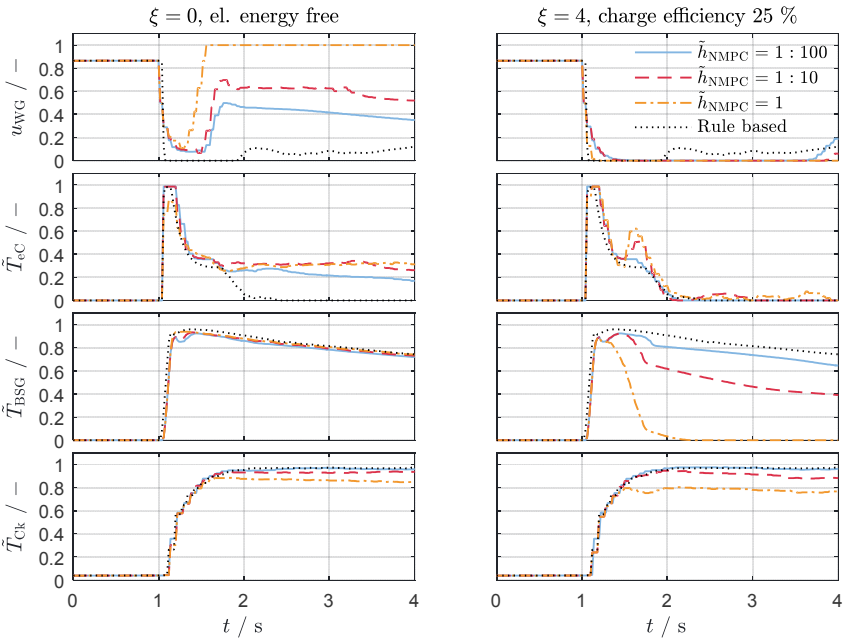


Fig. 6: NMPC optimization for different energy weighting factors ξ at a full load acceleration in 5th gear compared to the rule-based approach (weight ratio energy to response behaviour $\tilde{h}_{\mathrm{NMPC}} = h_{E_{\mathrm{Tot}}}/h_{\Delta T_{\mathrm{Ck}}}$) [8]

Conclusion and Outlook

Due to the multitude of functions, components, degrees of freedom and limited power capacity, competitively designed 48V systems will frequently be pushed to their system limits. In addition, rule based control approaches lead to suboptimal operation and cannot exploit the maximum system potential. In order to keep the electrical system at reasonable dimensioning while achieving a high availability of 48V powered vehicle functions, a predictive optimization based energy management is used for an intelligent power distribution. The nonlinear model predictive control (NMPC) takes a holistic view of the different domains of the 48V system. It anticipates driving situations, shifts electrical load demands of time-uncritical consumers such as the electric A/C compressor and can thus optimally utilize the limited performance of the 48V power net to improve the torque response. In comparison to a rule based strategy, the NMPC can better solve the trade-off between response behaviour, comfort functions and energy saving potential. This shows that NMPC based energy and power management contributes to efficient, sustainable and competitive 48V systems of the future.

Literature

[1] Gindele, J., Ramsteiner, T., Fischer, J. et al.: Der neue 2,0-l-Hochleistungs-Vierzylindermotor von Mercedes-AMG. In: MTZ 74 (2013), Nr. 9, S. 664-671

[2] Uhlmann T., Baumgarten H., Franzke B., Scharf J., Thewes M., Birmes G. (2016): Extreme downsizing for gasoline engines – fun to drive with extremely low emissions. In: Liebl J., Beidl C. (eds) Internationaler Motorenkongress 2016, Proceedings, S. 91-107. Springer Vieweg, Wiesbaden

[3] Griefnow, P., Andert, J, Engels, M., Hülshorst, T. et al. (2018): Advanced Powertrain Functions and Predictive Operating Strategies for 48 V Mild Hybrid Vehicles. In: Aachener Kolloquium Fahrzeug- und Motorentechnik, S. 1669-1694. Aachen

[4] Lüpkes, K., Pillas, J., Pätzold, R., Kirschbaum, F., et al.: Fahrleistungsoptimale Ansteuerung einer elektrischen Maschine und eines elektrischen Verdichters auf 48 V Spannungslage. Electric & Electronic Systems in Hybrid and Electric Vehicles and Electrical Energy Management (EEHE) Conference. Bamberg, 2017

[5] Schäfer, V.: Adaptives Energiemanagement für 48V-Kraftfahrzeugbordnetze. Schriftenreihe des Lehrstuhls Fahrzeugmechatronik der TU Dresden Nr. 15, Verlag Dr. Hut. TU Dresden, Dissertation, München, 2016

[6] Griefnow, P., Andert, J., Xia, F., Klein, S., Stoffel, P., Engels, M.: Real-Time Modeling of a 48V P0 Mild Hybrid Vehicle with Electric Compressor for Model Predictive Control. SAE Technical Paper 2019-01-0350, doi: 10.4271/2019-01-0350

[7] Graf, F., Lauer, S., Baensch, S., Knorr, R.: Konnektivität als Schlüsseltechnologie für den 48-V-Mildhybrid, In: ATZ Extra (2018), Nr. 23 (Suppl 1), S. 12-15

[8] Griefnow, P., Andert, J., Birmes, G., Pischinger, S.: Optimierungsbasiertes Energiemanagement für 48-V-Mildhybrid-Antriebe, ATZ Extra (2019) 24(Suppl 2): 42. https://doi.org/10.1007/s35778-019-0015-5

Abbreviations

BMS	Battery management system
BSG	Belt starter generator
eAC	Electrified air conditioning compressor
eC	Electric compressor
EEM	Electric energy management
ICE	Internal combustion engine
MHEV	Mild-Hybrid electric vehicle
NMPC	Nonlinear model predictive control
PHEV	Plug-In hybrid electric vehicle
RCP	Rapid control prototyping
RDE	Real driving emissions
SOC	State of charge
TC	Turbocharger
WG	Waste gate
WLTC	Worldwide harmonized light vehicle test cycle

Acknowledgments

- This work was partially performed as part of the research project “Advanced Co-Simulation Open System Architecture” (ACOSAR) within the European EUREKA cluster program ITEA3.
- This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 769935.

Easy Integration of 48V Mild Hybridization by Dual Voltage Battery Management

Realizing CO_2 saving potentials by low implementation efforts

Dipl.-Ing., Dipl.-Wirt.-Ing. **Bernd Fähnrich**,
Dr.-Ing. **André Körner**, HELLA GmbH & Co. KGaA, Lippstadt

Abstract

Before the full breakthrough of full electric driving, 48V Mild Hybrid market penetration will increase dramatically within the next decade. This powertrain concept offers high CO_2 saving potentials by moderate system modifications.

The modification efforts based on existing vehicle architectures will be even lower by implementing the HELLA Dual Voltage Battery Management (2VBM) to realize a 48V Mild Hybridization, especially for small and mid-size vehicles.

Introduction

Practically all automotive manufacturers are currently looking for ways to save CO_2 in order to comply with the strict limit value of 95 grams of carbon dioxide emissions per kilometer, that will apply to fleet consumption in Europe from 2021 onwards [1]. The targets for the future are set to be even more ambitious. Further measures are also urgently needed in the USA, China and Japan in order to comply with the relevant national regulations.

Meanwhile, the ongoing public discussions on the subject might well lead to the idea that the only solution is to get rid of the internal combustion engine immediately and switch everything over to full electrification. When looking at the matter more closely, however, it soon becomes clear that – in the short term at least – a wholesale switchover is not feasible.

A whole range of factors – the lack of charging infrastructure, the availability of battery cells and the fact that the electricity mix is currently far from CO_2-neutral and still contains a large proportion of fossil energy sources – mean that a differentiated approach is required when it comes to the subject of electrification. If a car driver is looking to purchase an electric vehicle, they will quickly realise that, although there are some excellent large vehicles available on the market, there is a real lack of attractive mass-mobility solutions to meet the high levels of demand in the compact and mid-size segment. And this is precisely where intelligent solutions are needed in order to reduce CO_2 emissions.

Over 2/3 of the cars sold worldwide currently come under the category of micro, small and mid-size vehicles [2]. This figure is expected to be the same in ten years' time. When it comes to the type of powertrain technology used in these vehicle segments (A-C), it is likely that the proportion of Electric-only vehicles and Full Hybrids will remain relatively low due to the amount of money it costs to implement this technology. It is therefore expected that around 45% of these vehicles will still have a conventional internal combustion engine or will feature Micro Hybrid technology with a Stop/Start system by the end of the coming decade (Fig. 1: Powertrain distribution for vehicle segments A-C, (Source: IHS 2019)). If further CO_2 reductions can be achieved for these drive concepts, it will have a huge impact on fleet emissions and will help to reduce fines for carbon dioxide emissions or even eliminate them altogether.

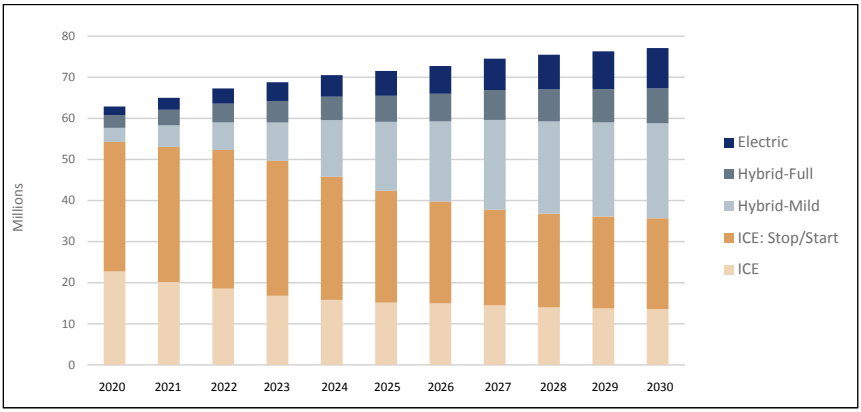


Fig. 1: Powertrain distribution for vehicle segments A-C, (Source: IHS 2019)

This is where the Dual Voltage Battery Management (2VBM) comes in. This bridging technology makes it possible to convert vehicles to a 48V Mild Hybrid with minimal reworking based on an existing 12V vehicle electrical system. The use of recuperation energy alone can save five grams of CO_2 per kilometer at the 48V level. And there is no need for the vehicle to be equipped with 48V consumers. These benefits are particularly important for small and mid-size vehicles:

One of the major challenges when it comes to converting a vehicle to a Mild Hybrid is finding space for everything: as well as the conventional 12V lead acid starter battery, the vehicle also needs to be able to accommodate a 48V lithium ion battery and a 48V/12V DC/DC converter. The battery is often installed in the trunk, which means that extra weight, time and money need to be considered for the wiring of the three components.

The Dual Voltage Battery Management is the ideal Mild Hybrid power storage solution in this case, as it combines the functions of the 12V battery, the 48V lithium ion battery and the DC/DC converter within the package space of the existing 12V lead acid battery (between battery size H5 and H7 depending on power requirements). The system is combined with a 48V belt-driven starter generator to complete the basic configuration for the P0 Mild Hybrid architecture (Fig. 2). The absence of the 12V lead acid battery results in a significant reduction in weight, which in turn leads to extra CO_2 savings. Furthermore, the Dual Voltage Battery Management also offers a solution to the potential ban on lead in starter batteries, as the whole battery is made up of lithium ion cells which store energy.

In concrete terms, the small amount of vehicle integration work required can be summarised as follows:

- Maximum package space required corresponds to that of a H7 battery (with up to 20Ah capacity for 48V)
- P0 architecture with starter generator control system adapted for 48V
- CAN communication for control and diagnostics
- Software adjustment in the body control module or in the engine controller in order to control the Hybrid behaviour

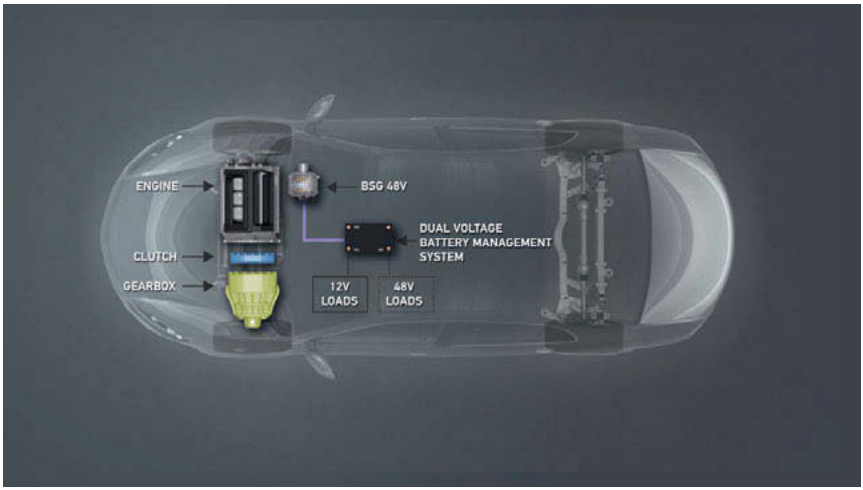


Fig. 2: Integration of the Dual Voltage Battery Management in a Mild Hybrid vehicle

How the Dual Voltage Battery Management works

The major advantage of the concept is the fully variable distribution of the total vehicle's battery capacity to the 12V or 48V on-board power supply systems, depending on the current demands on both vehicle electric system voltages. In the basic configuration, the Dual Voltage Battery Management consists of three strings which work independently of each other and which each consist of three cell blocks. Each individual cell block corresponds to the 12V vehicle electric system voltage (BN12). The 48V vehicle electric system voltage is provided by connecting three blocks in series (BN48). It may seem illogical that connecting three 12V blocks in series would produce a voltage of 48V, but this is down to the fact that the 12V operating voltage range actually ranges from 9V and 16V and the operating voltage range for the 48V vehicle electrical system is defined as 36V to 52V in accordance with VDA 320 [3]. It is therefore possible to ensure compliance with the voltage range for both the 12V and 48V vehicle electrical systems by selecting the right lithium ion cell chemistry. Redundant semiconductor switches are used to switch between parallel and series connection of the lithium battery blocks in any given string in order to comply with safety criteria (Fig. 3).

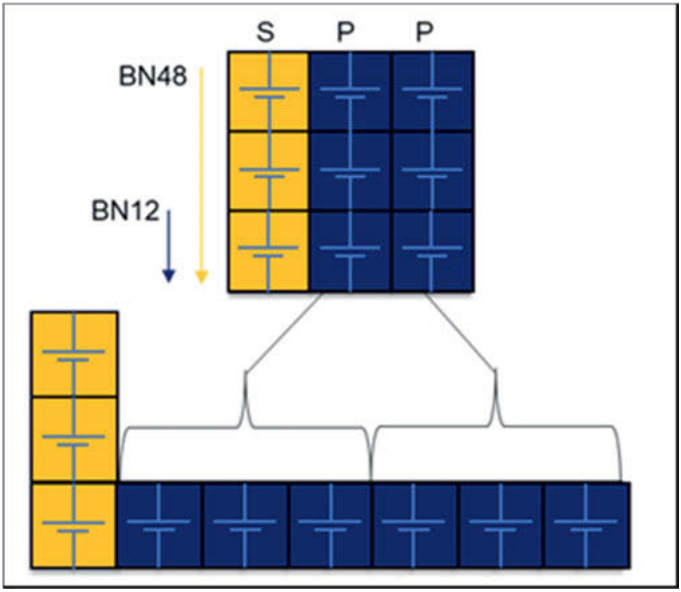


Fig. 3: Variable distribution of the battery capacity to 12V vehicle electric system voltage (BN12) and 48V level (BN48). In this picture the SPP (Series, Parallel, Parallel) configuration of three strings is visualised

There are essentially three different modes that are used during operation:

PPP configuration

In this case, all three strings are connected in a parallel configuration, which means that all nine cell blocks are used to provide the maximum available capacity for the 12V vehicle electrical system. The Dual Voltage Battery Management therefore acts like a 12V starter battery in order to, for example, support cold cranking from the 12V vehicle electrical system or to supply all available capacity to the quiescent current consumers during parking, if no 48V quiescent current consumers need to be supported. During parking, usually 48V Systems are switched off, so this is no functional limitation.

SPP configuration

Once the vehicle has been started, the system really comes into its own. When the driver brakes, electrical energy can now be recuperated at the 48V level and fed back into the Dual Voltage Battery via the belt-driven starter generator. With recuperation at 48V, much more power can be recovered than with a conventional 12V system, and greater efficiency can be achieved.

The three blocks of one of the strings are connected in series in order to store the recuperation energy. The available capacity for the 12V vehicle electrical system is therefore reduced to two strings, but this is perfectly sufficient to keep supporting the 12V system.

At the same time, potential consumers in the 48V vehicle electrical system can be supplied from the battery in this mode.

The three strings alternate between series and parallel connection (SPP - PPS - PSP) in rapid succession, typically within seconds or faster depending on the required energy transfer. This ensures uniform charging and discharging of the strings with only very slight differences in the state of charge. The micro-cycles that this operation mode results in, only have a relatively low impact on the durability of the lithium ion battery cells. This degree of loading is acceptable for this application. The control algorithms ensure that at least one string is always connected in series to the 48V vehicle electrical system and ensure smooth switching between the strings without interruptions or noticeable voltage jumps. This guarantees the required level of stability for both the 12V and 48V vehicle electrical systems.

SSP configuration

If the starter generator supplies more recuperation energy or the consumers in the 48V vehicle electrical system require more power, the three blocks of a second string in the Dual Voltage Battery Management can be operated in a serial configuration. This doubles the available capacity on the 48V side, while the capacity on the 12V side is reduced to 1/3 of the original total capacity compared to the PPP configuration. Assuming an available capacity of 60Ah on the 12V side, this leaves 20Ah capacity for 12V consumers in this configuration. The string configurations alternate in rapid succession as in the case of the SPP configuration, with the same objectives and effects.

SSS configuration

The SSS configuration is a special case. In this mode, all three strings are operated in the serial configuration, meaning that all the available battery capacity can be used to support the 48V vehicle electrical system. This mode can be used to store extremely high quantities of recuperation energy at 48V or for short-term driving support (creeping, boosting). As the initial aim of using the dual voltage battery management system for the mild hybridisation of small and mid-size vehicles is to reduce CO_2, driving support – which corresponds to one of the highest power requirements in the system – is not the primary focus of the application and can only be supported to a limited extent. This must be assessed in detail in the respective vehicle configuration.

In the SSS configuration, the bottom block of each of the three strings is also used to supply the 12V vehicle electrical system by means of a coupling arrangement, which results in non-uniform discharging of the blocks. The utilisation period for this configuration is therefore to be limited. This will not impose any restrictions in practice.

Cooling concept

To ensure simple, straightforward integration into an existing 12V vehicle architecture, work is continuing to utilize a non-fluid-based cooling system. This is the only way to keep system costs low, as small and micro vehicles generally do not have a separate water-cooling circuit for power consumers that could be used to cool a high power Mild Hybrid battery.

An air-based cooling concept therefore is the only way to keep temperatures below the critical level for the cells. In principle, it would also be possible to do away with the cooling system completely, but this would lead to a deterioration in the performance.

Cranking capability

With integration of lithium ion as a starter battery chemistry the cranking capability becomes one of the big concerns. Fig. 4 shows the comparison of the cranking behaviour of a 1.6l, 77kW Diesel engine at 8°C ambient temperature with different battery types. The blue graphs show the cranking performance with a standard AGM lead acid battery with 68Ah capacity at an SOC of 95%. In comparison a 12V series lithium ion (LFP) 69Ah starter battery shows, driven by the lower internal resistance a smaller voltage drop during the cranking event (yellow curves).

The performance with a Dual Voltage Battery Management prototype with 56Ah (NMC) capacity on 12V and approximately 52% SOC shows even much higher performance reflected in a more than 100ms shorter cranking cycle and a lower voltage drop (brown curves).

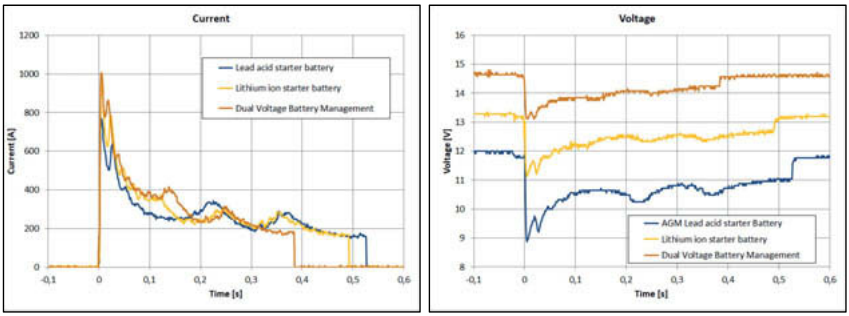


Fig. 4: Current and voltage curves while cold cranking of a 1.6l Diesel engine using different battery types

Cranking events at lower temperatures show higher voltage drops due to higher internal resistances of the cells, but based on the installed capacity in a Dual Voltage Battery with all strings are switched to parallel configuration (PPP-mode), even cold cranking events can be supported after longer parking modes at low environmental temperatures of -25°C. Considering an LN4/H7 housing for a Dual Voltage Battery, this packaging format provides space for an installed capacity of approximately 55Ah for the 12V vehicle electrical system. The performance of a lithium ion battery is in general better than that of a lead acid battery with the same installed capacity [Ah]. The installed 55Ah capacity in a Dual Voltage Battery corresponds to approximately the power and usable energy of a 75Ah lead acid battery for typical automotive application, which supports even more the cranking capability of the 2VBM approach.

Voltage dynamics from alternations of strings

Another important question while switching between the serial and parallel configuration is the impact on the voltage dynamics. Fig. 5 shows a dynamic driving profile with high BN48 and BN12 electric load consumptions based on a two string 2VBM prototype. The voltage changes in the BN12 system (red curve) are caused due to string alternations and power demand dynamics. The voltage increases in the range of 50mV are caused by string changes. The highest voltage drop of approximately 0,5V at the BN12 occurs while power system coupling during a boosting phase. Based on a three-string approach, these couplings are not needed any longer to support shorter boosting phases.

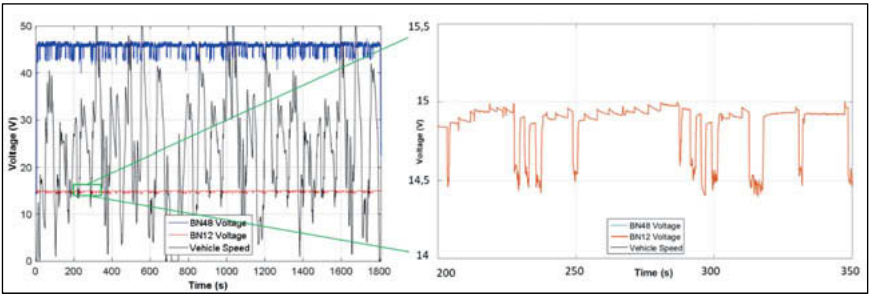


Fig. 5: Voltage dynamics while string alternation within a prototype of a two string 2VBM with high electrical load consumptions on BN12 and BN48

Implementation and potential CO_2 savings

During the development phase, the potential CO_2 savings were simulated specifically for vehicles from the B segment. A vehicle with a mass of 1200kg, a petrol engine with 1.2l displacement, 82kW power at 176Nm torque and manual transmission was used by way of example. The simulations were carried out by two independent institutions and produced the same results.

With only recuperation at 48V, a CO_2 saving of about 5 g/km can be achieved in the WLTC. This assumes a capacity of 15Ah for the Dual Voltage Battery Management on the 48V side. Increasing the battery capacity further does not provide any significant added value with regards to CO_2 savings – the CO_2 savings reach a saturation point as the capacity increases, because the 12V power system consumption as the only consumer of the recuperated energy is limited in WLTP to the base consumption. In real driving with more electrical energy consumption like for comfort loads, the potential will be higher, what allows to reduce the gap

between test and real emission. The 15Ah capacity is therefore the optimum level in terms of potential savings and size.

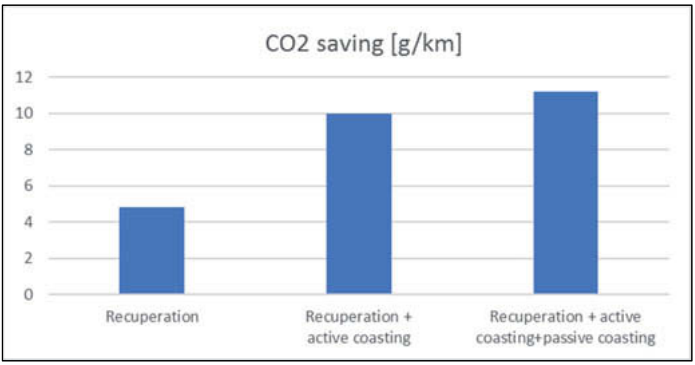


Fig. 6: CO2 saving potentials by implementing the Dual Voltage Battery Management while different driving conditions

If recuperation is considered in the WLTC in conjunction with active coasting phases, where the internal combustion engine in a P0 Hybrid varchitecture is dragged during coasting phases, the potential CO_2 savings increase to 10g/km, because more of the recuperated energy can be consumed in a fuel- and emissions saving manner. Only consuming recuperated energy in the 12V power system could not activate this high recuperation potential of a 48V System.

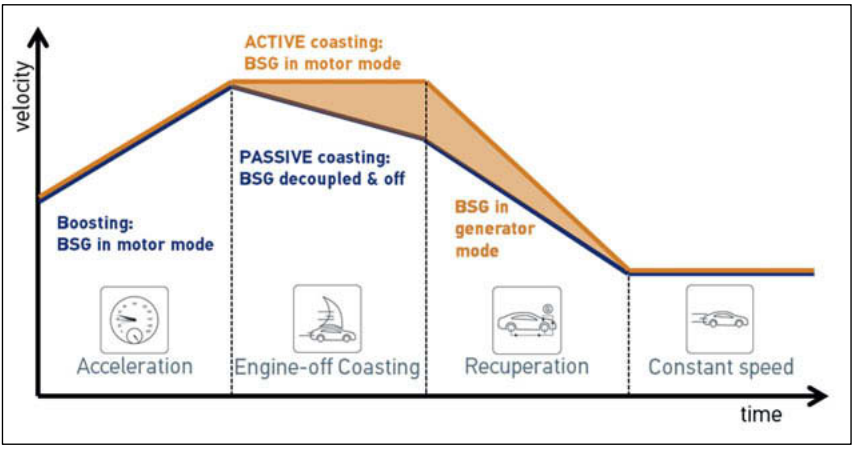


Fig. 7: Visualization of different driving modes

Further CO_2 savings can be achieved by making another adjustment to the vehicle architecture: by utilizing an electronic clutch to decouple the internal combustion engine during coasting (passive coasting), which eliminates mechanical engine torque losses. In this case, the potential CO_2 savings increase to more than 11g/km, when recuperation is combined with active and passive coasting (
Fig. 6). The different driving phases are visualized in
Fig. 7.

The distribution of the phases within the WLTC are shown in
Fig. 8. It shows, that about 50% of the driving cycle time one of the Hybrid functions Stop/Start, active coasting, passive coasting and recuperation is active.

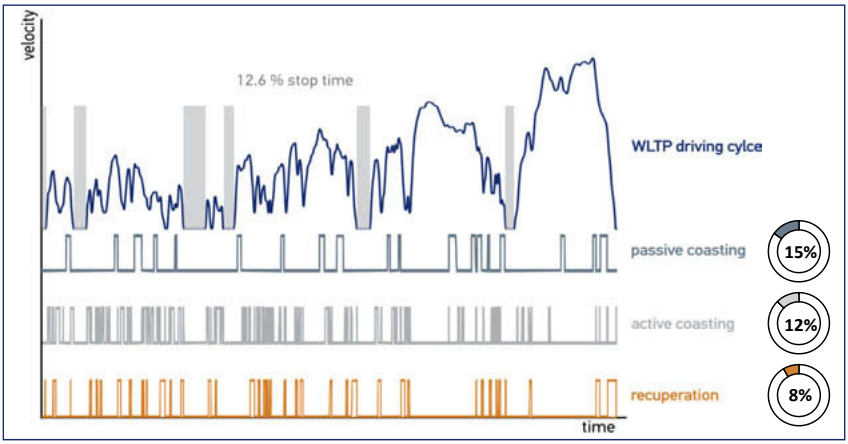


Fig. 8: Distribution of Hybrid operation phases within the WLTC

Recuperation into different battery types and the associated CO_2 emission effect

Independent simulations have been conducted to visualize the different CO2 saving potentials with different recuperation approaches into different storage techniques. Fig. 9 shows the different results for the CO_2 emissions according to WLTP for a B-segment vehicle.

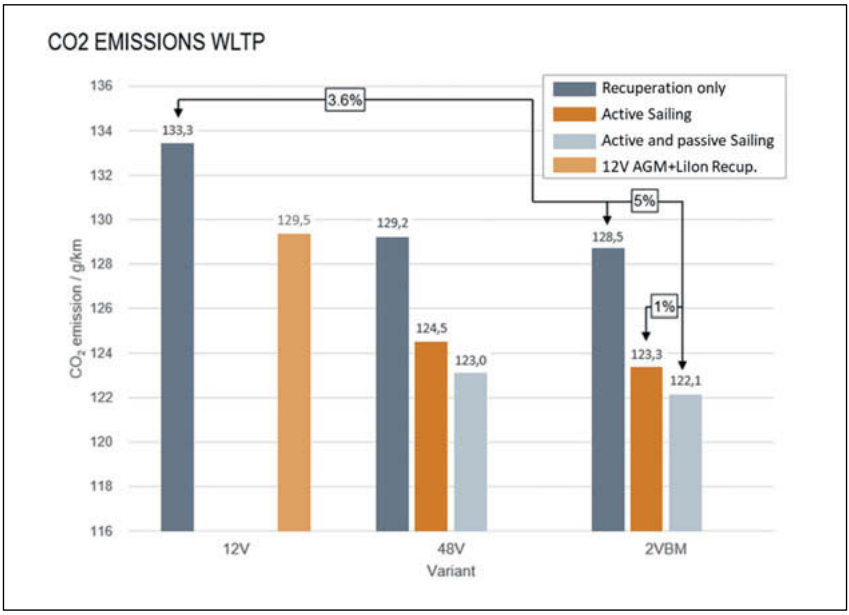


Fig. 9: Recuperation into different battery types and the associated CO2 emission effect based on simulation data

The recuperation into a 12V AGM lead acid battery can be set as a baseline with 133.3g/km CO_2 emission. Recuperation on 12V into a lithium ion battery which is coupled to the AGM lead acid battery and a 3kW generator reduces the CO_2 emission already by 3.8g/km to 129.5g/km. Further CO_2 reductions in the WLTC can be gained by combining recuperation with coasting phases but only by implementing a 48V Mild Hybrid. With a classical approach considering a 48V lithium ion battery with 15Ah capacity in combination with a 12V/48V DC/DC, the CO_2 reduction can be reduced in the simulation down to 123g/km while the approach with the Dual Voltage Battery Management offers an additional 1g/km saving down to 122.1g/km, which demonstrates an additional functional advantage of the 2VBM compared to classical 48V Mild Hybrid storage concept in addition to the architectural advantages.

Best "cost per gram CO_2-saving" ratio for the Dual Voltage Battery Management

Beside the overall CO_2 saving potential for a new device as the Dual Voltage Battery Management the system costs versus benefits have to be taken into account so that the implementation of such a device is also economically reasonable.

Considering again a base system setup with a 12V AGM lead acid battery, 12V starter and 12V generator with recuperation into a 12V lead acid battery, the CO_2 emission is simulated to 133,3g/km according to Fig. 9. With setting the costs for this configuration as baseline, the add on costs per gram CO_2 saving can be estimated as shown in Fig. 10.

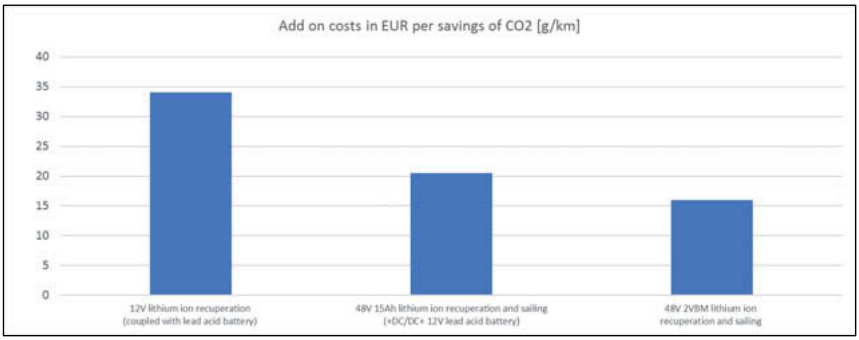


Fig. 10: Add on costs per CO2 saving for different recuperation concepts

Based on recuperation on 12V using a lithium ion battery which is coupled to a 12V AGM lead acid starter battery and a generator with a power of 3kW, the add on costs per gram CO_2 saving can be estimated to approximately 34EUR/g with the limitation, that an additional overall CO_2 saving is simulated to only 3.8g (Fig. 11).

An even better ratio of approximately 20EUR/g add-on costs per saved gram CO_2/km can be realised by implementation of a classical 48V Mild-Hybrid application considering a 12V/48V DC/DC converter and a 15Ah 48V lithium ion battery besides a 12V AGM lead battery. To realize recuperation on 48V and support, the add-on costs for a 48V belt driven starter/generator are taken into account.

The lowest add on costs per gram CO_2 saving can be realized by implementing the Dual Voltage Battery Management under the same conditions. Based on this approach the ratio can be reduced to 16EUR/g savings. This figure supports the strategy to implement the 2VBM besides the package and weight advantages especially for small and mid-size vehicles. Due to the lower implementation effort and the removal of the lead acid battery including cables and the

battery sensor, it also offers lower system costs compared to a classical 48V Mild Hybrid approach with 48V lithium ion battery and 12V/48V DC/DC converter. Due to the implementation of three independent strings within a Dual Voltage Battery Management, redundancy aspects can be fulfilled at least equal or even better than compared to a standard 48V Mild Hybrid power system approach.

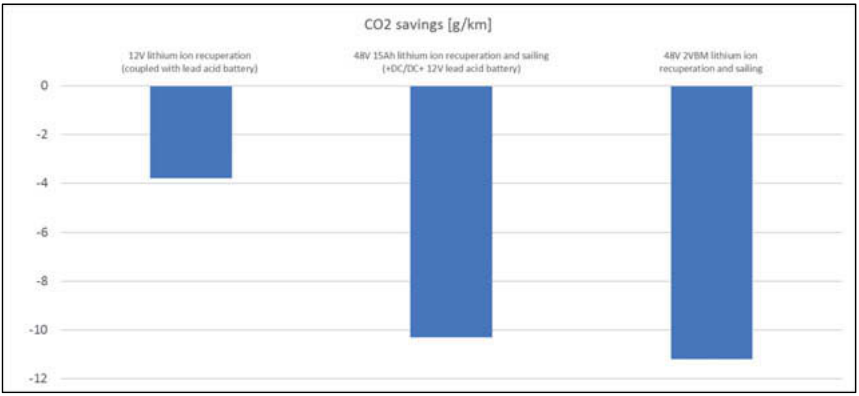


Fig. 11: Comparison of CO2 saving potentials for different recuperation concepts

Outlook

So far, three different prototype versions have been developed which confirm the basic functional concept on the HIL test stand and in the vehicle. Subsequent development work will focus on increasing the degree of integration to move closer to the target package space of existing 12V lead acid batteries and on the design of the cooling concept. Intensive discussions will be held with various OEMs to ensure that the product is perfectly aligned with the needs of the market. Development projects with selected OEMs will work towards integration and demonstration in potential target vehicles. In conjunction with this, demonstrating the CO_2 savings in vehicle tests will also be on the agenda for the coming months. The aim is to bring the Dual Voltage Battery Management to series production by 2023.

In conclusion, the Dual Voltage Battery Management will serve as a valuable bridging technology in combination with the existing internal combustion engine – particularly for small and mid-size vehicles – and will help to meet the ambitious CO_2 targets set for the mass segments over the coming years.

[1] https://www.vda.de/de/themen/umwelt-und-klima/co2-regulierung-bei-pkw-und-leichten-nfz/co2-regulierung-bei-pkw-und-leichten-nutzfahrzeugen.html

[2] IHS 2019

[3] VDA Empfehlung 320, Elektrische und elektronische Komponenten im Kraftfahrzeug, 48V-Bordnetz, Anforderungen und Prüfungen, August 2014

48 Volt High Power: Electric Drive for Excellent CO_2 Emissions & Electric Driving Features

Friedrich Graf, **Simon Baensch**, **Thomas Knorr**,
Dietmar Ellmer, **Christophe Marechal**,
Division Powertrain, Continental Automotive GmbH, Regensburg,
Continental France SAS, Toulouse, France

Abstract

Future motor vehicles regulatory requirements for CO_2 emission reduction require a significant increase in hybrid vehicles in the fleet mix of manufacturers.

The use of the internal combustion engine is increasingly replaced by the electric drive in hybrid vehicles and the challenges for exhaust aftertreatment increase, as the thermal supply of the internal combustion engine for thermal management decreases. The electric machine must have sufficient performance to enable corresponding driving patterns not only in the test cycle but also in real-life operation.

Results from Continental's 48V ECO-Drive System have shown that for P2 compared to a P0 hybrid, the challenge for emissions is significantly exacerbated by an increase in pure electric driving capability, especially for plug-in hybrid vehicles. At the same time, technological advances allow 48V drives to penetrate areas of application that were previously reserved for high-voltage solutions.

This paper presents a compact design 48V 30kW electric machine solution, suitable for a P2/P4 configuration intended for Full-Hybrid applications or even a plug-in hybrid, and characterized by compact design with a very high power density. At the same time, the system approach of exhaust aftertreatment is illustrated. A 48V 30kW "electric" means not only less CO_2, but also increased electric driving and more total driving performance with significantly less effort and complexity due to 48V as compared to a high-voltage solution.

With the help of simulations and vehicle measurements, CO_2 potentials as well as various emission-relevant driving conditions of a 48V 30 kW high-performance hybrid are examined. At the same time challenging environmental conditions and the potential for improvement is determined by the advanced Emicat® exhaust aftertreatment system.

CO_2 performance is also the result of intelligent operating strategies. In the future, they will have to make their contribution robustly to the lowest CO_2 in every driving cycle . For this purpose, connectivity and optimization algorithms in real time are enablers [6].

Keywords: 48V P2 topologies, High Power and Torque density, Permanent Magnet Synchronous Machine, Embedded PCB inverter, Cost Efficient Plug-In Hybrid

1. Introduction

Since the European Parliament has agreed beginning of 2019 to reduce by 2025 CO_2 emissions by 15% and by 2030 CO_2 emissions by 37.5% from new passenger cars (31% for new vans) in comparison to the already existing target of 95 g/km of CO_2 from 2021, the European automotive industry needs to achieve a significant step towards decarbonising and modernising the European mobility sector.

In addition, the main pollutant emissions limits are drastically reduced from 2025 onward in comparison to 2020 especially for Nitrogen Oxides (NOx), Particle Mass (PM) and Particle Number (PN).

Finally, a strengthened market monitoring and surveillance system will be set up with the adoption of the new test cycles (e.g. WLTP [Worldwide harmonized Light- Vehicle- Test Procedure] and RDE [Real Driving Emissions]) to enhance the representativeness of the official test procedure for determining the emissions with respect to real-world driving (Fig. 1).

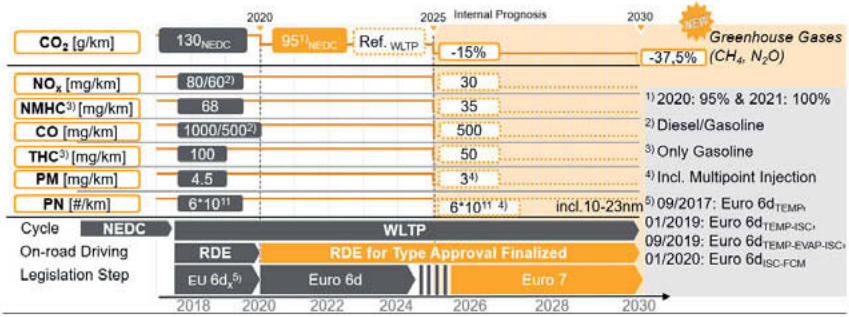


Fig. 1: Europe further consumption, emission limits & test procedures

As the combustion engine has already reached a very high level of technical sophistication and further efficiency improvement potential is more and more limited, it is widely accepted that (48V) electrification will be necessary to comply with the future CO_2 requirements in and in respect to an OEM CO_2 reduction strategy,

As consequence, there are an important number of 48V P0 mild HEV applications already in development or even in serial production, since the technology provide many advantages with following characteristics and contexts [2]:

- Main focus is CO_2 reduction,
- Mainstream architecture is P0 where the 48V Belt Starter Generator (BSG) can be integrated easily as a replacement for existing alternators with minimal impact on the existing vehicle designs,
- 48V Belt Starter Generator are optimized to deliver requested performance for current architectures, with dedicated DC/DC, 48V battery and belt starter generator as main ingredients
- Diesel scandal accelerated serial applications: from niche to standard
- Technical standards have been introduced and are effective

In a next step, much higher CO_2 reduction can be achieved by so-called Px hybrids, which are obtained by integrating the 48V drive system into the gearbox or a hybrid module (P2, P2.5, P3) or by adding a second electrically driven axle (P4).
These Px-systems have in common that the drag losses of the combustion engine are avoided during braking, coasting and electric driving, which all result in remarkably low CO_2 emissions and the experience of pure electric driving features.
Therefore, the 48V electric drive system has to be consequently designed towards high power density and maximum efficiency to optimize not only the energy recuperation but also to maximize the traction performance.

2. Requirements for future 48V electric drive system

The current generation of 48V electric drive addresses mainly the P0 architecture is designed to provide up to 60 Nm and 15 kW peak power. With such system, the CO_2 reduction on WLTC cycle is limited to a maximum of 10% (depending of the component sizing, vehicle size, transmission type and others powertrain characteristics) and the following hybrid functions can be enabled: Start/Stop, Recuperation, Torque assistance, Coasting, Electric Creeping and reduced emission for Gasoline or Diesel based internal combustion engines thanks to the combination of 48V system and electrically heated catalyst – EMICAT [3] [4].

Now with the usage of 48V electric drive system in Px- hybrids, additional CO_2 reduction is expected up to 20% and also electric driving features on a C segment vehicle with P2 – P4 architecture (Fig. 2):

- An acceleration of 1.5 m/s² up to 30 kph in full electric mode with
- A continuous electric sailing up to 55 km/h

These requirements translate to an increase in power requirements for the 48V electric drive in a power increase as well as an enhanced efficiency (>90%) to enable an extended recuperation potential:

- For engine restart, 30 kW (mechanical) for 5 seconds at 2200 rpm crank shaft speed
- For electric driving, 20 kW (mechanical) for 20 seconds
- For electric sailing, 15 kW (mechanical) continuous at 2200 rpm crank shaft speed
- For electric boost and Diesel cold crank, 70Nm for 5 seconds

	Key Requirements
SOP	2023+ incl. Commercial Vehicles
Hybrid architecture	Px
Performance	>20 kW peak
Hybrid function	Same as 2nd generation: + Enhanced recuperation + Electric driving features
CO_2 reduction (WLTC*)	Up to 20 %

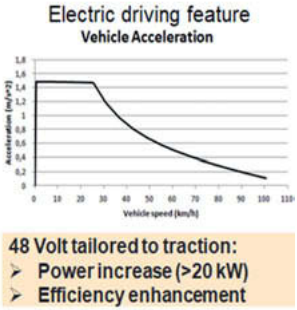


Fig. 2: Key requirements for future 48V electric drive system

3. '48V High Power' electric drive component

Based on these requirements, Continental has developed a '48V High Power' Electric Drive system with a peak performance of 30 kW and an excellent efficiency. Parking, active electric sailing and following the traffic in city areas can all be achieved pure electrically with this '48V High Power' Electric Drive. Its compact design with a very high-power density flexibly allows the integration into a P2-hybrid module, into a gearbox (P2.5, P3), as well as the realization of an electric axle drive (P4) and to enable 4WD capability.

Technology: It consists of two main sub components:

- An electric permanent magnet synchronous motor (PSM) with an I Pin stator winding technology (Continental patent) able to reach a maximum shaft speed of 20.000 rpm. The permanent magnet rotor provides many advantages in comparison to an AC induction machine (asynchronous machine). The volumetric torque and power density

are higher, the motor's efficiency is higher, less current is used (as no magnetizing current are necessary), lower temperature is reached, and ramp-up time is shorter.

- An inverter with 6 phases based on Embedded PCB technology which includes up to 3 MOSFET's to increase the provided current (Continental Patent).
  The inverter PCB technology ensures a high-power density of 25 kW for 0.79 liter/1.51 kg.

In terms of cooling, a shared cooling concept has been designed between inverter and electric machine: the inverter is liquid cooled as the electric machine's design benefits from a water jacket.

Performance: The development of the first sample of electric machine and associated inverter has been finished since mid 2018 resulting in a compact design (diameter 175mm) with an integrated inverter. Sample performances have been measured on an 48V electric drive test bench with water cooling temperature below 85°C and water-cooling flow of 3 l/min (Fig. 3) marked by solid points. Efficiency is high ~ 90 % in motoring mode 14.5Nm, 10kW, 36 V at 6600rpm.

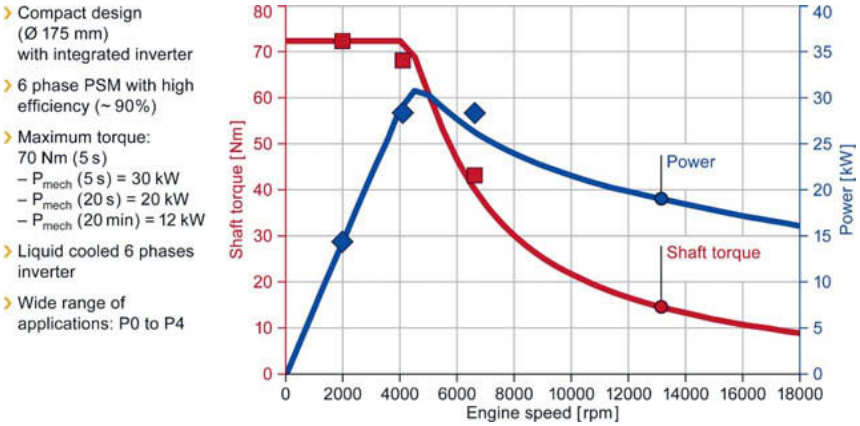


Fig. 3: '48V High Power' electric drive sample, power and torque in function of shaft speed

For a hybrid vehicle, it means than we can benefit from a theorical maximum additional torque of 70 Nm (multiplied usually by 2 – 3 gear ratio) provided by the '48V High Power' Electric Drive to boost the already available torque from the Internal Combustion Engine (ICE). Therefore, the vehicle offers more dynamic acceleration which can be useful for instance when merging into traffic. Generally, the low-end engine behavior is significantly improved, which

combines very well with highly charged, power-dense gasoline engines. The transmission input torque capacity has to cover this torque growth.

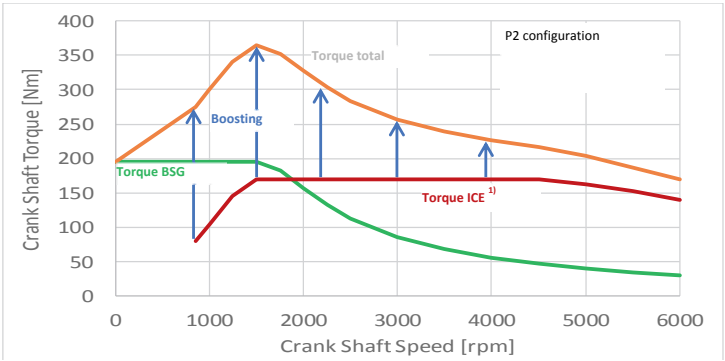


Fig. 4: Combined ICE and BSG torque in P2 configuration

4. Utilized powertrain system architecture

The next step was to integrate and to assess the performance of the '48V High Power' Electric Drive in a real use case on a demo car. The choice has been made to re-use as a basis the P2 hybrid architecture already developed in Continental [5] [6], and then to replace the previous generation of 48V electric drive motor by the new '48V High Power' Electric Drive. The 48V Li-Ion battery has been also replaced by a storage with more capacity at 6.5 kWh and providing more max. current for (dis)charge.

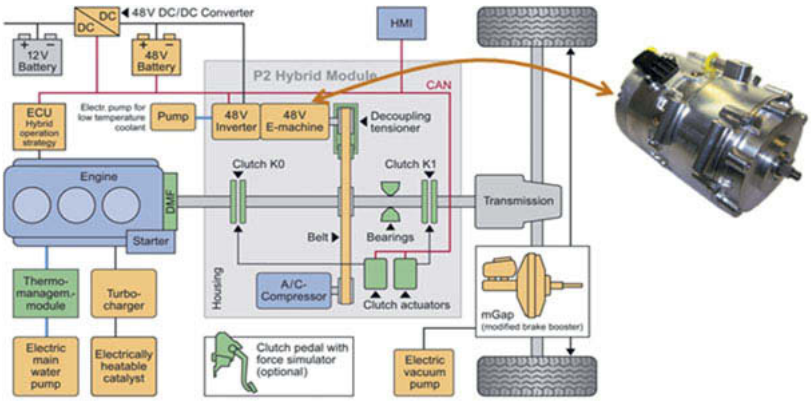


Fig. 5: System architecture of the 48V P2 full hybrid vehicle

This so-called P2 hybrid architecture shows the electric traction machine and the additional K0 clutch being installed between the Internal Combustion Engine (ICE) and the transmission (Fig 5). Hence drag losses in hybrid states such as recuperation, coasting (rolling with open powertrain), sailing (driving electrically at constant speed) and pure electric driving are eliminated. The combustion engine is the Ford 1.0L EcoBoost three-cylinder gasoline engine. It is enhanced by a matching RAAX turbocharger [1], a 200bar fuel pressure injection system, an electrically heated catalyst (EHC) and an engine control unit coordinating the entire powertrain. A further key engine component is the Continental RAAX turbocharger with a low inertia radial-axial turbine, which generates high boost pressure and fast pressure build-up even at low engine speeds.

As required for a P2 hybrid vehicle a small electrical water pump enables electric machine operation when the ICE is shut down.

Another key subsystem is the high level ‘hybrid operation strategy’, which is executed by the engine control unit. It is implemented in a model-based manner and controls the entire hybrid powertrain.

As shown in Figure 5 the vehicle is principally able to operate in all hybrid modes which are known from high-voltage hybrid vehicles. In particular, it is possible to drive purely electrically which can be used to perform electric creeping and launch (in combination with ECM) as well as to maintain constant vehicle speed up to 80 km/h (‘sailing’).

5. Simulation of vehicle performance with ‘48V High Power’ electric drive

The vehicle hast been revealed to Journalists at the Continental Technology Show in July 2019. Before performing real tests on vehicle, Continental has used its Powertrain System simulation tool chain called Vehicle Simulation Suite (VSS) to evaluate the performance of the vehicle equipped with the ‘48V High Power’ Electric Drive.

Fig. 6: System evaluation vehicle at Conti Tech Show 2019

Vehicle Simulation Suite is based on a model in the loop approach with a physical plant model and control functions working in co-simulation. This complete tool chain includes various powertrain architectures, libraries of components, functionalities for automated simulations and analyses. The simulation results have been correlated to real experimental measurement to ensure the required level of accuracy.

CO_2 reduction on WLTC cycle: the 48V P2 mild hybrid demonstrator already revealed a measured CO_2 reduction of 15% on the WLTP cycle. By replacing the previous generation of 48V Electric drive motor with the new '48V High Power' Electric Drive, the simulation showed an enhanced potential of CO_2 reduction up to 19% under same conditions (Fig. 7).

These results are similar to the CO_2 level reached by state-of-the-art high-voltage hybrids vehicles, demonstrating the high potential of 48V technology in regard to high voltage solutions.

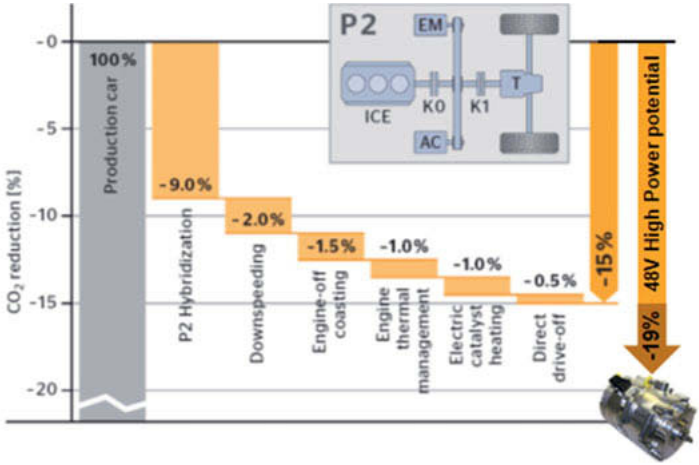


Fig. 7: CO_2 reduction on a WLTP cycle for a P2 full hybrid vehicle

CO_2 reduction focus in urban area: Thanks to the high efficiency of the '48V High Power' Electric Drive, the CO_2 reduction can be even be more relevant when the vehicle is driven in urban areas. To illustrate this, a real-world city cycle, compliant to RDE and with a total distance of 33km, has been defined around Regensburg in Bavaria (Fig. 8).

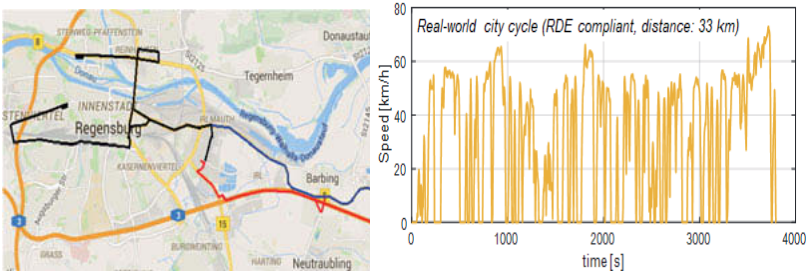


Fig. 8: Real-world city cycle (RDE compliant)

The simulation of the fuel consumption as Full Hybrid on this real-world city cycle shows an average fuel consumption of 3.9 litres and a CO_2 emission of 91.7 g/km which is 14% better than the homologation value on WLTP. It is obvious than 48V technology especially with this power specification and efficiency fits very well to city driving.

Electric driving: The driving mode simulation of the 48V P2 mild hybrid vehicle on a WLTP cycle shows than the first start of combustion occurs only after 12.75 minutes which means an All-Electric Range (AER, EU type approval 2017/1151) value of 4.7 km (assuming no battery capacity limitation). Indeed, a very significant section (94% of time) of WLTC can be driven without the use of combustion.

Following the definition of the 'Equivalent all-electric range' (EAER) the driving range is 34 km with a battery capacity of 5.3 kWh net (useable).

These results underline that the usage of '48V High Power' Electric Drive in a P2 configuration provides extended electric driving capabilities, which are very especially beneficial in urban conditions.

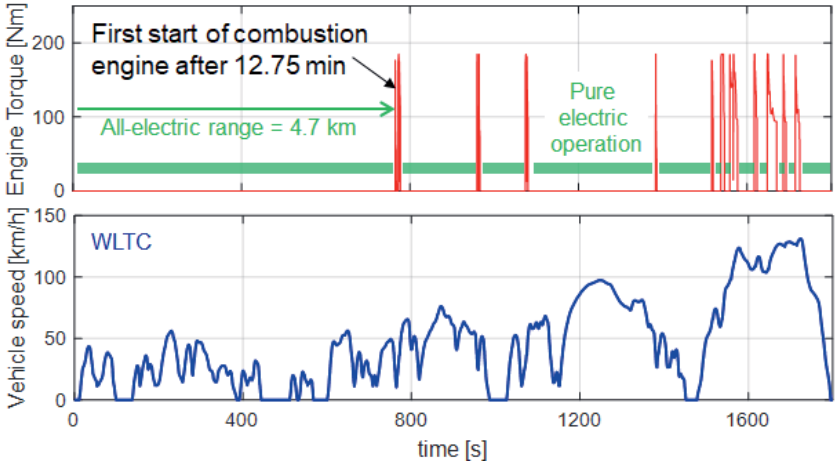


Fig. 9: Electric driving capabilities on WLTP for a P2 full hybrid vehicle

6. Emission management

Optimized temperature management for the combustion engine and the exhaust chain is essential for hybrid vehicles, especially in view of RDE and future emission regulations. The use of more powerful electric machines like the 48V HP machine and larger batteries makes it possible to unlock further potential for CO_2 reduction but requires strategies within the framework of comprehensive temperature management. For example, the 48V higher power drive enables purely electrical, real-world short-distance operation up to 5 or more kilometers without the activation of the ICE. In that case the exhaust aftertreatment system may remain

cold, any heating measures represent an unnecessary expenditure of energy. On the other hand additional power demand beyond the electric drive, spontaneously requested by the driver, requires the activation of the ICE without delay and without pre-heating steps of the catalyst.

Here, the catalytic converter from EMICAT® with two heating discs represents a logical step towards heating the largest catalytic volume possible. Various heating strategies that have been optimized for efficiency are presented from the stationary engine to high-load and high-speed operation [3].

Taking the example of the WLTC, in depletion mode operation and CO_2 optimized following the COMMISSION REGULATION (EU) 2017/1151, we illustrate the emission behavior for a vehicle with 1565 kg test weight at 23°C ambient temperature without any heating measures (fig. 10, left). The ICE is only activated at a driving power request above 20kW.

The tailpipe emission behavior is clearly beyond the EURO 6d limits.

In the following step the electrical heating (EMICAT®, 2 x 4 kW) is activated, synchronously to the activation of the ICE and without optimization of the total sum of heating durations. The principal potential of this simple electrical heating measure is obvious and shown in fig. 10 (right):

- NO_x emission below 50% of the EURO 6 d limit
- NMHC: slightly above the EURO 6d limit (15 %)

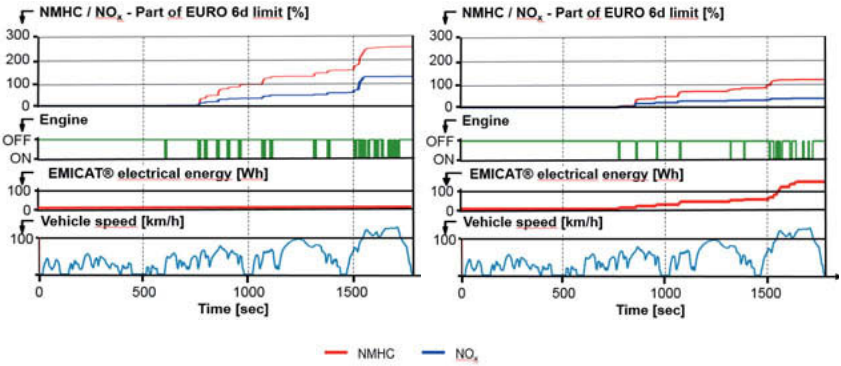


Fig.10: Emission level in WLTC without and with EMICAT® application

Essentially, these very first results show that even in such an intermittent engine operation, emissions appear to be manageable with additional measures without a relevant CO_2 penalty.

7. Cost comparison

Besides its very good performance, the '48V High Power' Electric Drive in a P2 configuration is also very well positioned cost-wise. Continental has compared its solutions with a Power-split high voltage system using a Multi Point Injection (MPI) as ICE and an electric Continuous Variable Transmission (eCVT). The study based on non-binding estimation shows a delta cost of (Fig. 11):

- -15% for a '48V High Power' Electric Drive using a Gasoline Direct Injection (GDI) as ICE and a Dual Clutch Transmission (DCT),
- -28% for a '48V High Power' Electric Drive using a Multi Point injection (MPI) as ICE and a Dual Clutch Transmission.
- Main contributors to cost saving:
  Usage of 48V technology for the electric drive & power net (e.g. DC/DC, electrical AC compressor) instead of more expensive technology with high voltage components
- Cost efficient electric machine technology
- Low impact on engine & transmission technology

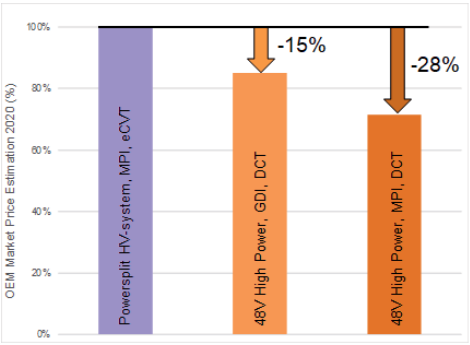


Fig. 11: Price estimation of powertrain & components related to electrification

This comparison demonstrates than 48V High Electric Drive technology is suitable for mass market relevant car segments and could be a strong basis for high market penetration in the future.

8. Conclusion

With the innovative ‘48V High Power’ Electric Drive, Continental has greatly improved 48V electrification in different aspects. This electric motor delivers peak output of up to 30 kW, thus offering more torque to support the combustion engine electrically to a high extent, leading to a performance at low end rpm known so far only from Diesel engines. Furthermore, ‘48V High Power’ enables quiet, purely electric inner-city journeys at significantly reduced cost compared to state-ofthe-art HV hybrid solutions, which could bring such 48V Full Hybrids to more popular vehicle segments.

The entire unit comprising electric motor and integrated inverter is extremely compact and delivers an efficiency of over 90%. This facilitates mechanical integration for different Px architecture options and ensures high CO_2 reduction benefits.

By utilizing these ’48V High Power’ Electric Drive in the described P2 hybrid architecture, Full Hybrid performs at a high level – with up to 20% CO_2 reduction - are achieved while providing enhanced pure electric driving experiences. In a next step, the Plug In Hybrid Vehicle (PHEV) option will be explored. Simulation calculation results show less than 50 g/km (New EU Type approval 2017/1151), which would mean a classification as a Low-Emission vehicle.

9. References

[1] MAIWALD, Oliver; SCHAMEL, Andreas; WAGNER, Uwe: *"48 V P2 hybrid vehicle with an optimized combustion engine – Fuel economy and costs at their best combined with enhanced driving behaviour"* in 37th International Vienna Motoren Symposium 2016

[2] SCHÖPPE, Detlev; KNORR Thomas; GRAF, Friedrich; KLINGSEIS, Bernhard; BEER, Johannes; GUTZMER, Pete; HAGER, Sven; SCHATZ, Axel: *"Downsized gasoline engine and 48 V Eco Drive – An integrated approach to improve the overall propulsion system "efficiency"* in 35th International Vienna Motoren Symposium 2014

[3] KNORR, Thomas; ELLMER, Dietmar; BAENSCH, Simon; SCHATZ, Axel: *"Optimization of the 48 V Hybrid Technology to Minimize Local Emissions in the RDE"* in 27th Aachen Colloquium Automobile and Engine Technology, Aachen 2018

[4] AVOLIO, Giovanni; BRÜCK, Rolf; GRIMM, Jürgen; MAIWALD, Oliver; RÖESEL, Gerd; ZHANG, Hong: *"Super Clean Electrified Diesel: Towards Real NOx Emissions below 35 mg/km"* in 27th Aachen Colloquium Automobile and Engine Technology, Aachen 2018

[5] LAUER, Stefan; GRAF, Friedrich; SPRINGER, Moritz; WECHLER, Stefan: *"48 Volt Hybrid with e-drive features - Excellent fuel efficiency and drivability"* in Electric & Electronic Systems in Hybrid and Electrical Vehicles and Electrical Energy Management, Bamberg, 2017

[6] GRAF, Friedrich; LAUER, Stefan; BAENSCH, Simon; KNORR, Rainer and Dr SANS, Mariano: "*48 Volt Technology in the Light of the Connected Vehicle and Electrical Board Net Advancements*" in 26th Aachen Colloquium Automobile and Engine Technology. Aachen, 2017

10. Glossary

AER: All Electric Range

BSG: Belt Starter Generator

CH_4: Methane

CO: Carbon monoxide

CO_2: Carbon Dioxide

DCT: Dual Clutch Transmission

EHC: Electrically Heated Catalyst

EVAP: Evaporative Emissions

FCM:Fuel Consumption Monitoring

eCVT : electric Continuous Variable Transmission

GDI: Gasoline Direct Injection

ICE: Internal Combustion Engine

ISC: In-Service Conformity

LIVC: Late Intake Valve Closing

mHEV: mild Hybrid Electric Vehicle

PHEV : Plug In Hybrid Vehicle

PM: Particular matter

PN: Particle number

MPI: Multi Point Injection

NEDC: New European Driving Cycle;

NMHC: Non-methane hydrocarbons

NO_x: Nitrogen oxide

NOX: Nitrogen Oxides

N_2O: Dinitrogen monoxide

RDE: Real Driving Emission Test

THC: Total hydrocarbons

VSS: Vehicle Simulation Suite

WLTP: Worldwide Harmonized Light duty Test Procedure

Efficiency Advantages of SiC in Electric Drive Train Applications

Thomas Grasshoff, **Oliver Tamm**,
SEMIKRON International GmbH, Nürnberg

1. Abstract

Power electronics is a key technology in the area of car electrification. Vehicle applications and the related environmental conditions create challenges to meet the requirements for the power electronic conversion. Technically most demanding are different ambient temperatures ranges, a high number of thermal cycles with wide temperature steps, strong demand to minimize weight and space consumption, high resistance to shock and vibration as well as rugged design to protect against environmental impact by dust or liquids. To optimize system costs and customer acceptance there is an increasing demand for extending the electrical driving range and efficiency increase of the inverter. Silicon Carbide MOSFET´s are seen as the most promising semiconductor material to meet these requirements.

Today the majority of electric drive train applications operate in the 400V battery voltage range requiring 650 or 750V blocking voltage of the semiconductor. In this voltage area IGBT based solutions are well established and will go into volume in the next 2 years. The alternative use of 750V SiC offers in the higher power range an very attractive option to increase the travel range in the same design or boosting the drive train power in the same space. In the 800V battery voltage range the

usage of 1200V SiC MOSFET´s provides a big efficiency improvement compared to 1200V IGBT´s. IGBT´s in this voltage class have three time the dynamic losses compared to the lower blocking voltage.

To achieve the optimal performance of SiC MOSFET´s, the power module's mounting technology and connection to the DC link must be optimized for fast switching and high temperature operation. The paper focuses on demonstrating the performance of SiC-MOSFET compared to IGBT´s in conjunction with an appropriate module design based on the WLTP driving cycle and real engine data.

2. Introduction

Vehicle applications require efficient and reliable energy conversion. The typical power range is between 40 and 250kW. The cost split of a reference converter shows that the power module represents approx. 30% of the costs. In order to reduce the size and weight of the inverter, the system and in particular the cooling system has to be optimized, as it counts for 15% of the costs. Most electric drive solutions contain a multichip power module. It can include IGBT´s or MOSFET´s. The typical topologies are half bridges or six packs. In every power module, the efficient utilization of power semiconductor capabilities is the key to cost efficiency.

3. Advantages of SiC

The efficiency of the entire system is an important criterion for the selection of the power semiconductor. SiC can be an alternative to silicon if the semiconductors reach their limits in terms of power dissipation and necessary switching frequencies. SiC is an extreme hard material with a high electrical breakdown field. It is a fast switching unipolar device which enables low switching losses. Figure 1 shows the material characteristics and the effect to the usage in power electronic applications.

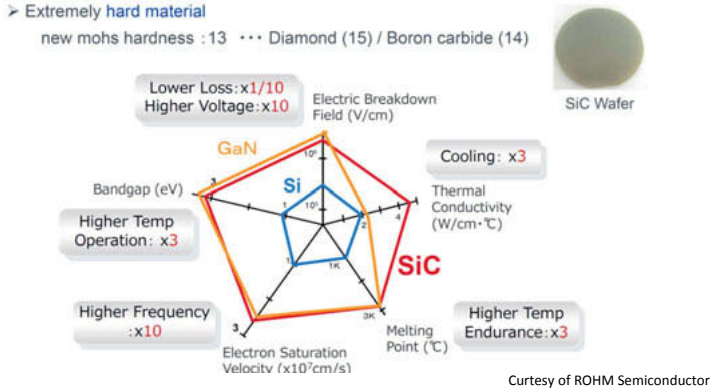


Fig. 1: Characteristics of Silicon Carbide

SiC MOSFETs can be the best choice for losses, but their use also poses major challenges for the power module and system design. Fast switching means steep current edges and high di/dt values. The module's and system's parasitic inductances cause voltage spikes due to the

high di/dt, resulting in voltage overshoot across the chips. If the current slope is too high, this overvoltage might exceed the maximum blocking voltage, e.g. 750V/1200V, of the SiC device.

4. Packaging requirements

Decreasing the switching speed or the DC link voltage will reduce the overvoltage but compromise the SiC power module's performance. A module and system design focused on low commutation inductance is therefore essential. The module's commutation inductance is mainly provided by the DC bus terminals with 8 to 10nH, depending on the power module design. The DBC design, i.e. DBC tracks and wire bonds, contribute another 1 to 6nH [2]. Adding the stray inductance of DC Bus bar and DC Link capacitor on top overall stray inductance easily counts up to 14-20nH.

Fig. 2: Traditional Low stray inductive designs: left to right: Bondwire DBC, Direct lead bond

To achieve the required power switching speeds of 25-50kV/µs at 450V/800V DC link voltage, including sufficient margin between the reverse voltage of the SiC MOSFET and the overvoltage measured across the MOSFET chips, it is necessary to reduce the total leakage inductance below ~ 10nH. Possible options for minimizing the parasitic effects include reducing the geometric length and area of the current loop, laminated power terminal and bus bar arrangement or planar chip interconnection using metal lead or other layers.

Figure 3 shows layouts with minimum stray inductance. In the half-bridge substrates, the commutation path and area through DC+ and DC− are reduced to small levels. Bonding wire-free concepts, such as direct lead bond, double-side soldering/sintering or top-layer substrate are available solutions to lower the stray inductance. By continuing the laminated DC path throughout the DC links capacitor, low stray inductance significant below 10nH for designs up to 600A-1000A can be achieved. With that snubber capacitors become superfluous.

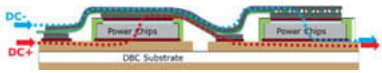


Fig. 3: Ultra Low Stay inductance DBC layout using a flex foil as additional commutation layer

In addition to the electrical power, the thermal power of the entire system plays a decisive role. The power density of SiC chips is higher than that of silicon-based chips. The entire thermal system is optimized in such a way that the heat generated by SiC components is optimally dissipated and the thermal resistances from the chip to the cooling medium are matched. While traditional water-cooled silicon based power modules use alumina oxide ceramics and solder joints, this design is unsuitable for SiC. AL_2O_3 ceramics offer limited thermal conductivity compared to the 3-4 more higher power density of SiC devices, while standard solder joints, on the other hand, do not have sufficient reliability for SiC chips operating at Tj=175°C and beyond. Si_3N_4 ceramics, for example, in combination with sintered connections or direct contact systems such as pressure contacts, offer a suitable solution that brings out the best in expensive SiC MOSFETs.

The solder layer between substrate and cooler or base plate also contributes significantly to the overall R_{th}. Novel technologies such as Direct Pressed Die (DPD) completely eliminate this resistance by adding only a small value due to the very thin TIM layer.

By thermal stack simulation (Fig. 4) the performance of different Packaging Technologies can easily being quantified. Optimized architectures save more than 60% of the junction-case thermal resistance compared to traditional solder packaging technologies.

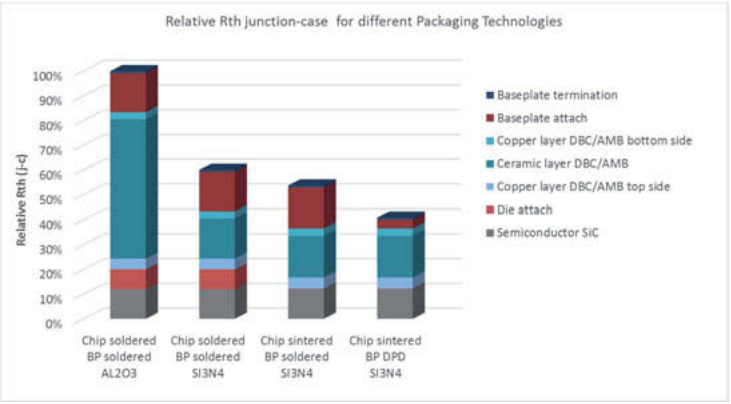


Fig. 4: Thermal performance for different Packaging Technologies

SiC has other material properties than silicon. The modulus of elasticity of SiC is up to factor 4 higher and SiC chips are about 100% thicker than standard 1200V silicon IGBT´s. This requires a suitable material selection that corresponds to the thermal expansion in order to

achieve the required reliability. SiC and Si_3N_4 are therefore good partners (Fig. 5). Without changing the assembly technology, the expected SiC performance cycle capability is only 1/3 of silicon in standard packages. [1]

Material Substrate/Chip	Al2O3	Si	AlN	SiC	Si3N4
Coefficient of Thermal Expansion (ppm/K)	8.3	4.1	5.7	3	2.5

Fig. 5: Thermal expansion coefficients of chips and ceramics

5. Performance benefits of SiC

SiC-MOSFETs generally exhibit significantly lower switching losses and in particular lower voltage drops in the partial load range than silicon IGBTs in comparable applications. The key to this is the linear output characteristic of MOSFETS. Under high load conditions, the static losses of SiC can be higher than the static losses of IGBT. The reason for this is the temperature coefficient of SiC devices (Fig. 6). In automotive applications, this only plays a role for peak power operating points, which, on the other hand, do not contribute significantly to the energy efficiency of the vehicle.

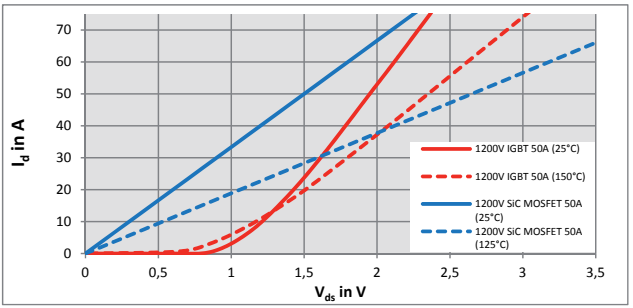


Fig. 6: Static losses of Si IGBT and SiC MOSFET at 25°C and 125°C

With all optimization options from pure Si-IGBT to hybrid SiC to full SiC power modules, switching losses can be reduced by 80%. A comparison of the switching losses was carried out with different options of the chip technology (Fig. 7). The reference was a 62mm standard module package and the optimized package was a module based on DPD technology. [2]

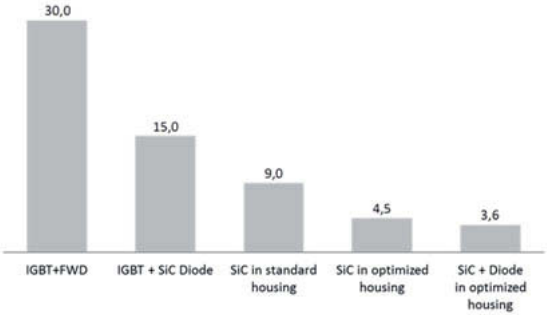


Fig. 7: Nominal switching losses of 100A rated chipset at Vdc=800V

An essential design goal of Electric Drive Systems (EDS) is the optimization of the electrical efficiency with simultaneous optimization of the overall system costs. In addition, energy efficiency is increasingly determined by legal regulations that depend on vehicle classes. Higher EDS efficiency extends the range of vehicles or reduces the battery size and thus the vehicle weight, which gives vehicle designers more freedom. In addition to the lower dynamic switching losses of SiC components compared to IGBT components, MOSFET forwarding properties play a decisive role. Power modules must be dimensioned to support the motor's peak power of up to 60 seconds. From power module perspective, vehicle peak power is a static case as power modules reaches thermal equilibrium usually after 1-3 seconds. The rapid temperature gradient is mainly caused by the limited cooler mass in the order of 1 kg per six-pack copper cooler-based power module.

To quantify the energy savings of SiC-based inverter solutions compared to IGBT(+FWD)-based solutions, an Electric Drive System was developed for a premium BeV with 300kW peak power. Alternatively, SiC-MOSFET and IGBT (+FWD) chipsets based on the latest 750V generations were equipped. Both options have identical package contours and identical water cooling characteristics, while the SiC option has been upgraded to Si_3N_4 substrate vs. Al_2O_3 to compensate for the higher power density of the SiC-MOSFET. The globally harmonized cycle of the Light Vehicle Test Procedure (WLTP) has been applied. Furthermore, the switching

frequency was varied and the switching speed of the SiC application was limited to ~20kV/us. The EDS efficiency was derived for these four options.

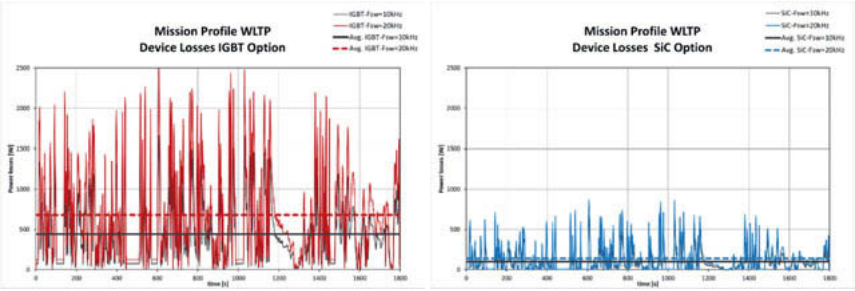


Fig. 8: Comparison Inverter losses IGBT vs. SiC

For 10kHz switching frequency, SiC-based inverters show only ~23% of the loss profile of an IGBT-based solution. At 20kHz, the loss is only ~20% of the IGBT solution. For example, by using SiC-MOSFET as the switching element, more than 75% of the losses in the power module based on WLTP can be saved.

For peak power profiles, the loss savings of SiC vs. IGBT are mainly limited to lower dynamic losses, since the forward resistance of the SiC-MOSFET itself reaches a level of similar or even higher voltage drop, relative to the value of the collector-emitter saturation voltage of the comparative IGBTs. Thus, SiC and IGBT solutions show a different efficiency curve than output powers with a significant advantage in the low power range that dominates the WLTP profile.

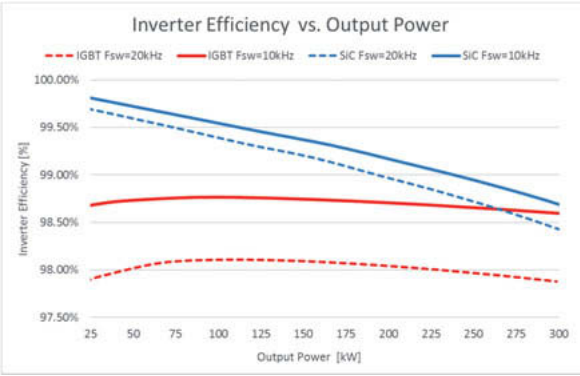


Fig. 9: Inverter Efficiency vs. Output Power

SiC based converters have a 1% higher efficiency compared to Si based converters even at the lower switching frequencies.

6. Summary

For traction inverters in the automotive industry, SiC-based power modules show a significant improvement in EDS efficiency. The average losses based on WLTP are reduced to approx. 25% compared to IGBT-based inverter architectures. However, the requirements for packaging technologies are much more demanding than for IGBT due to higher chip temperatures, higher power density and lower leakage inductance. As a result, new package technologies for power modules that enable faster switching and higher power density are becoming increasingly important.

7. References

[1] Power cycling capability of Modules with SiC-Diodes, Christian Herold et.al., CIPS 2014

[2] 400A, 1200V SiC power module with 1nH commutation inductance, Peter Beckedahl et.al., CIPS 2014

The Transition of EV Applications from Silicon to Silicon Carbide

Helping the power electronics design community overcome reliability challenges for EV applications that use silicon carbide

A. Kashyap, **A. Gendron-Hansen**, **D. Sdrulla**, **B. Odekirk**,
Microchip Technology Inc., Bend, Oregon, USA

Abstract

Although electric vehicles (EVs) are not new to the automotive industry, the target rate of adoption is much higher than historical rates. Initiatives such as the EV30@30 campaign set the benchmark for EV sales goals, which target 30% of all auto sales to be electric by the year 2030. This requires all automotive original equipment manufacturers (OEMs) that serve global markets to rethink and reshape their roadmaps and strategies. Typical EV applications such as on-board charging (OBC), DC-DC converters, and EV traction inverters must consider established and new technologies that enable achievement of long-term goals and strategies. This paper will explore EV high-power application requirements, today's application solutions, and why many of these applications are targeting wide bandgap technologies such as silicon carbide (SiC) and gallium nitride (GaN) for future designs. The authors will also discuss the reliability and ruggedness requirements critical for automotive applications, and will outline the means of achieving them for SiC devices — beyond standard qualification metrics.

1. Introduction

There is no doubt in the industry that the SiC vertical MOSFETs are the devices of choice for high-power applications [1]. A high blocking voltage, a high current handling due to the drain contact on the backside, an excellent heat dissipation from the backside and the frontside (with proper design) and a voltage controlling circuitry compatible with the already existing driving circuits all make the SiC MOSFETs by far the leading power device for the automotive industry (Fig. 1).

A wealth of information has been already published on almost every aspect of this emerging technology, very similar to the dawn of the vertical power MOSFETs in 1980. And as it happened then, this emerging industry has reached the critical mass that will transform it into a widely accepted and lucrative high-tech manufacturing.

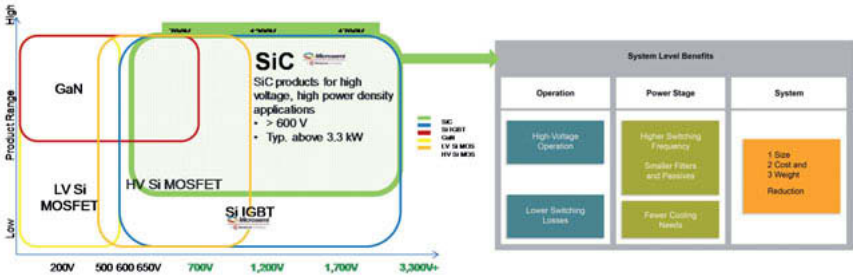


Fig. 1: Wide bandgap SiC technology positioning

2. Requirements of EV high-power applications

The high-power electronics requirements for a plug-in hybrid EV include the external DC charger, the on-board charger, a DC-DC converter and the traction motor control (Fig. 2). Today's specifications for the power, voltage, and current required to charge an EV battery are summarized in Table 1. In the next three to five years, the voltage level is expected to increase with the move from 400 to 800 V battery systems.

In the case of house-based chargers (single phase, three to eight hours charging time), a discrete solution with SiC MOSFETs and Schottky barrier diodes (SBDs) clearly satisfies the power and current specifications. A full bridge with either SiC discrete parts or SiC modules provides a viable AC-DC converter architecture. At and above 120 kW power levels, the silicon insulated-gate bipolar transistors (IGBTs) are today's devices of choice, and the task on hand for SiC power device manufacturers is to commercialize SiC MOSFETs with superior electrical performances and higher reliability. The switching frequency for these applications is in the range of 20 to 100 kHz, and IGBTs' reverse recovery has a high contribution to the total power loss. The adoption of SiC MOSFETs with extremely fast reverse recovery is an irreversible trend (Fig. 3), providing multiple benefits such as lower switching losses, a higher efficiency and a smaller footprint of the equipment.

For on-board chargers with power requirements above 3.3 kW, SiC-based modules provide a better trade-off between efficiency and cost than do silicon IGBT-based modules. Similarly, SiC devices enable high-efficiency DC-DC converters to down-convert the battery voltage to 48 and 12 V. For the traction motor control, SiC modules are very attractive thanks to a high-power density and excellent thermal performances.

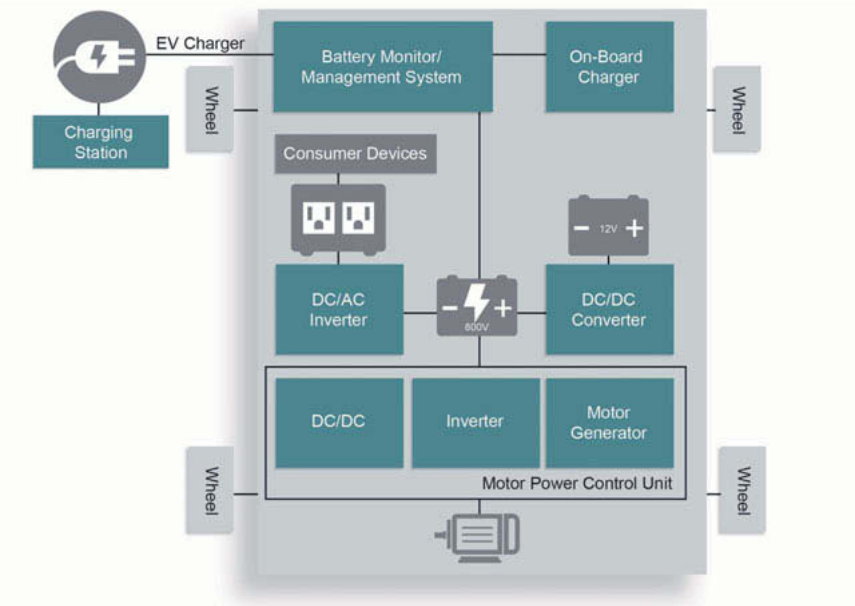


Fig. 2: High-power electronics requirements for a plug-in electric vehicle

Table 1: Power, voltage and current requirements to charge an EV battery

Charging time for 100 km of the battery	Power supply	Power	Voltage	Max. current
6–8 hours	Single phase	3-4 kW	230 V AC	16 A
3–4 hours	Single phase	7-10 kW	230 V AC	32 A
2–3 hours	Triple Phase	10-20 kW	400 V AC	16 A
1–2 hours	Triple Phase	20-40 kW	400 V AC	32 A
20–30 minutes	Triple Phase	40-60 kW	400 V AC	63 A
20–30 minutes	Direct Charging Current	>60 kW	400–500 V DC	100–125 A
10 minutes	Direct Charging Current	>120 kW	500–800 V DC	300–350 A

3. Long-term reliability burn-in stresses

In this section, we will review the long-term reliability of SiC power devices and present a comprehensive analysis for vertical MOSFETs and SBDs. We will also discuss specific stresses and evaluations to improve the screening of parts with lower mean time between failures (MTBF).

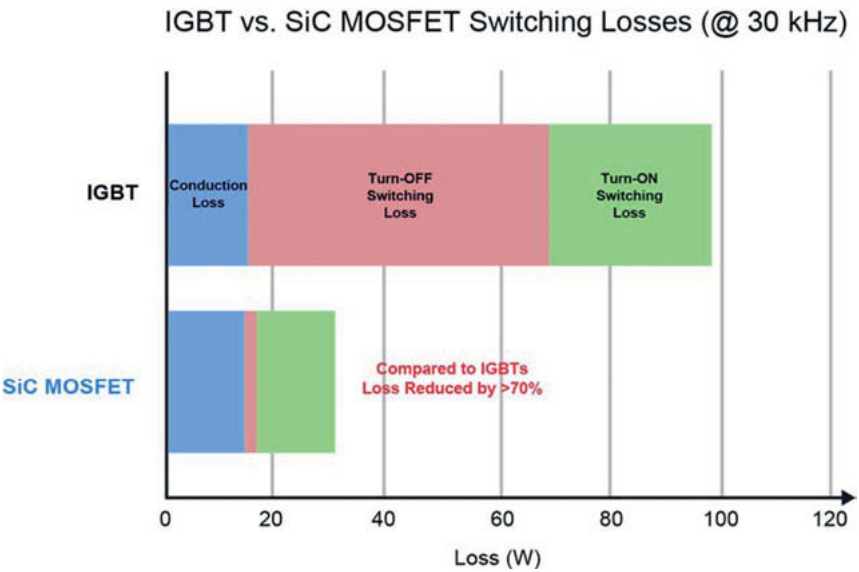


Fig. 3: Effect of the reverse recovery time on the power losses at high frequency

SiC electrical and thermal properties lead to a significant improvement of the continuous and dynamic electrical performances, even though the channel mobility is much lower than that of silicon devices. Continual efforts from SiC device suppliers brought the manufacturing cost down to levels acceptable to the automotive market. The co-foundry model, with silicon and SiC lines sharing the same tools, was key to fully leveraging the benefits from high-volume manufacturing. The performance and cost are essential, but not sufficient for the successful commercialization of SiC devices. The long-term reliability has to be guaranteed as well, and new challenges come with the qualification of SiC devices based on historical burn-in stresses for silicon devices (Fig. 4).

Die-oriented tests

• Positive high-temperature gate-bias (PTGB) (MOSFET): This test checks the stability of the gate oxide, including positive mobile ion contamination, and is significantly more stressful for SiC than silicon MOSFETs. SiC MOSFETs have a much thinner gate oxide with a gate voltage rating close to the onset of the Fowler-Nordheim tunnelling. It is highly recommended to perform a time-dependent dielectric breakdown (TDDB) test to determine if the gate oxide was degraded during the stress.

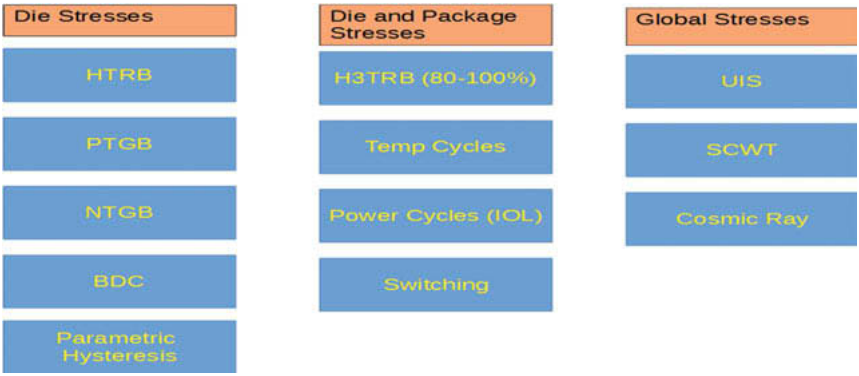


Fig. 4: Overview of the reliability tests for SiC devices

• Negative high-temperature gate-bias (NTGB) (MOSFET): This gate stability test is specific to SiC MOSFETs, for which a negative voltage of the driving circuit is highly recommended. Therefore, the gate oxide reliability under negative bias and high temperature must be proven. As for positive HTGB, a TDDB test post-stress is recommended.

• High temperature reverse-bias (HTRB) (MOSFET and SBD): This is a stability test for the high-voltage blocking capabilities. Historically, it was intended to check for positive mobile ion penetrating through the final passivation. In modern semiconductor technologies, the density of mobile ions in the dielectric layers was reduced to the point where they do not affect the long-term reliability. However, the moisture penetration due to the delamination of the molding compound is a serious problem for SiC devices, which have a very narrow termination with a high risk of arcing.

• Body diode stability under high current (BDC) (MOSFET): This test is mandatory for all SiC MOSFETs, even if an external diode is used as a free-wheeling diode. Typical current densities vary from 100 up to 1000 $A.cm^{-2}$. If the gate is shorted to the source, a significant part of the body diode current flows through the channel and charges the traps near the gate oxide interface. In addition to the body diode forward voltage ($V_{SD,ON}$), the on-state resistance ($R_{DS,ON}$) and the threshold voltage (V_{TH}) should be monitored.

• Threshold voltage hysteresis (MOSFET): This test characterizes the quality of the interface between the gate oxide and the SiC epitaxy. The density of interface traps varies widely between SiC device suppliers who rely on proprietary process flows to form the gate oxide.

Die and package-oriented tests

• High temperature, high humidity, high reverse bias (H3TRB) (MOSFET and SBD): The intent of this test is to prove that the top part of the die, in contact with molding compound or with the sealing gel, has the required reliability. The recommended reverse voltage bias is at least 80 percent of the rated voltage and could be up to 100 percent in the most stringent qualification standards.

• Temperature cycles (MOSFET and SBD): This test verifies the quality of the interface between the die and the package.

• Intermittent operating life (IOL), also known as power cycles (MOSFET): This test verifies the reliability of the die attach to the package's header, and the reliability of the wire bonds.

Global tests

• Single-pulse and repetitive unclamped inductive switching (UIS) (MOSFET and SBD): This is a stringent ruggedness test that drives a high current density while the device is in avalanche breakdown. Many different types of design and process weaknesses could induce UIS failures, from the quality of the starting material to the quality of the processing or design.

• Short-circuit withstand time (SCWT) (MOSFET): This is another stringent ruggedness test in which a high voltage (typically 80 percent of the voltage rating) is applied on the drain while the MOSFET is in on-state. The time-to-failure results provide some insights into how well the part can dissipate heat under very stressful electrical conditions.

4. Design trade-offs and impact on reliability

Schottky barrier diodes

A critical requirement for SiC SBDs is to have a low forward voltage (V_F) to minimize the conduction losses. SBDs from all major SiC device suppliers are based on a junction barrier Schottky (JBS) design that consists of P-doped regions interspaced with Schottky contacts. This design reduces the field-induced barrier lowering (Fig. 5) and hence guarantees a low reverse leakage current up to the maximum temperature rating of the part (typically 175°C). However, the P-doped regions are lost for the Schottky conduction, and the low V_F target means a large Schottky area, which comes with a high total capacitive charge (Q_C).

Rugged SiC SBDs should survive forward current surges lasting from a few microseconds to a few milliseconds but well above the maximum continuous current rating. The JBS design could be leveraged to improve the surge capability of a part, as the resistance of the drift region (and the power dissipation) decreases thanks to the conductivity modulation. Obviously, there is a trade-off, and one figure of merit (FOM) for SiC SBD is the ratio of the maximum continuous

forward current (I_F) or the maximum surge current (I_{FSM}) over Q_C. This FOM allows the end user to easily rate various suppliers and pick up the best device.

UIS rating should be imposed for all SiC SBDs to ensure a high ruggedness against unwanted voltage spikes (single or repetitive events, Fig. 6). This rating is of paramount importance for a device to be successful in the automotive market. Almost everything in the process and the design of a power device has a certain level of impact on the UIS capability: quality of the starting material, defectivity level of manufacturing, proficiency of the design of the high voltage termination, etc. The blocking capability of the termination should be much higher than the blocking voltage of the active area, and this gap should remain relatively the same under a wide range of temperatures or levels of contamination from the surrounding elements [2].

The very high electric field in the high-voltage termination could lead to reverse leakage current instabilities. The temperature has a strong effect on these instabilities, and the leakage current could increase by several orders of magnitude (from a few microamps to a few milliamps) during an HTRB test. To achieve a low and stable leakage current, the dielectric layers on top of the termination should be of high quality, and designers should pay special attention to the quality of the molding compound or the insulating gel on the top of the die.

Requirements on the heat dissipation lead to stringent specifications on the thermal resistance. Even though SiC has a thermal conductivity three times higher than that of silicon, the smaller size of SiC dies drives the thermal resistance up. To counterbalance this increase, thin dies and thin die bonding layers are mandatory requirements for SiC power devices. From a reliability point of view, the thermal resistance should remain stable through temperature cycle and power cycle stresses.

Power MOSFETs

To compete with silicon IGBTs, a low specific $R_{DS,ON}$ ($R_{DS,ON}$ multiplied by the die size) is a critical requirement for SiC power MOSFETs. Designers can choose to increase the packing density and reduce gate oxide thickness. However, both approaches come with trade-offs. A higher packing density increases the capacitances and slows down the MOSFET. A thinner gate oxide makes the part less rugged for gate voltage spikes. It also limits the value of the gate resistor in the driving circuit, which slows down the part even more. Another option is to reduce the drift resistance by optimizing the thickness and the doping concentration of the epitaxy. Lower resistivity epitaxy designs have lower breakdown voltages, which shrinks the margin to the voltage rating of the part. In this case, the yield could be affected, and the SiC MOSFET is more susceptible to degradation from cosmic rays. Finally, a trench technology

eliminates the JFET component of the $R_{DS,ON}$ but brings new challenges for the reliability of the gate oxide in the trench.

The body diode of a SiC MOSFET has reverse recovery characteristics close to those of an SBD with a fast recovery time (trr), a low reverse recovery charge (Qrr) and a limited increase of Qrr with temperature. The end users are inclined to take advantage of this feature and use the body diode as a free-wheeling diode in their applications. For this design to be safe, the long-term reliability of the body diode has to be thoroughly validated. Surge tests must also be performed with the channel fully turned on and fully turned off.

During a UIS test, the temperature of the die can be as high as 700°C and the voltage across the device significantly higher than the voltage rating of the part. Under these very demanding conditions, the gate oxide is submitted to an extremely harsh stress [3]. For rugged SiC MOSFETs, the design of the JFET region is optimized to keep the electric field across the gate oxide below 3.0 MV.cm^{-1} (Fig. 5). The integrity of the gate oxide after repetitive UIS tests should be validated, and TDDB tests are highly recommended (Fig. 6) [4].

As for SiC SBD, a low thermal resistance is an important requirement for SiC MOSFET. The low specific $R_{DS,ON}$ allows the designers to shrink the die size. However, it adversely impacts the thermal resistance of the part. And as the thermal resistance goes up, the maximum current capability of the MOSFET is also degraded.

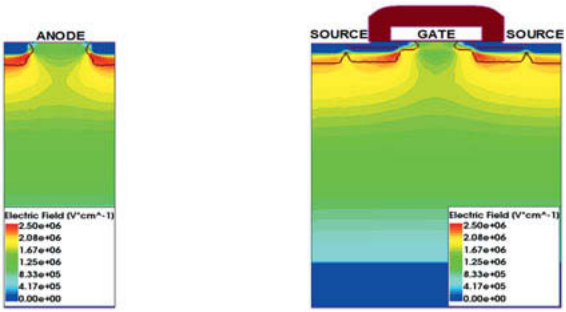


Fig. 5: Electric field distribution at maximum voltage rating for a 1200 V SBD (left) and a 1200 V MOSFET (right)

5. SiC failure precursors

Depending on the application, it is sometimes possible to monitor the main electrical parameters of SiC MOSFETs and SBDs. Variations in these parameters could serve as failure precursors, which allow us to react before the part becomes totally non-functional.

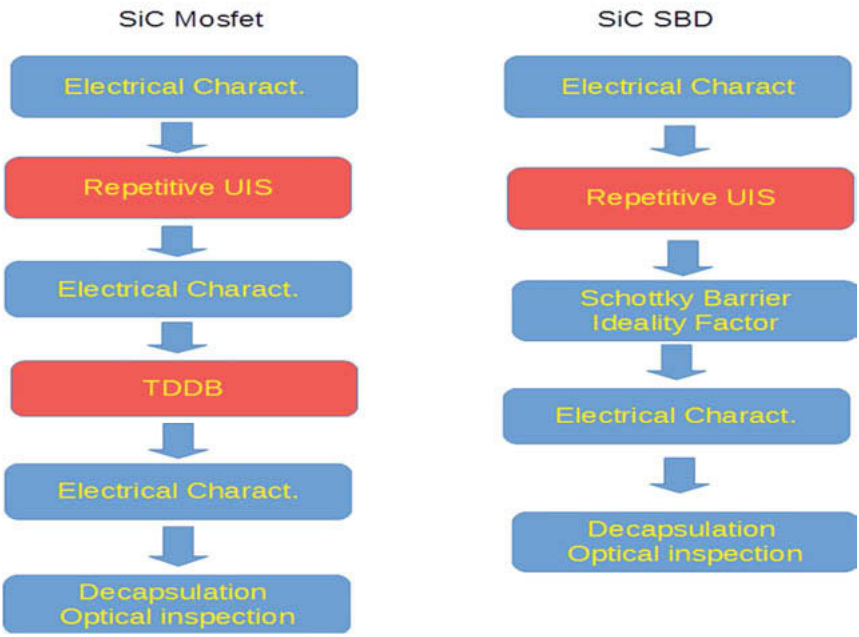


Fig. 6: In-depth characterization of the impact of stresses on SiC devices

Schottky barrier diodes

A drift in the leakage current is an indication of a degradation in the blocking capability of the part. Several degradation mechanisms could be investigated. Moisture penetration and its effect on the passivation layers could affect the high-voltage termination. If the leakage current increase takes place in the active area, a likely root cause is the failure of the Schottky barrier due to the alloy of the front-metal into the barrier metal layer.

An increase in forward voltage through time is typically the signature of a failure in the wire bonding scheme. It has been documented that aluminum frontside metallization and aluminum wires could be prone to failure. If this is the case, a corrective action would be to switch to copper or gold metallizations, which have significantly better MTBF through power cycles. Solderable frontside metal with clips is another option to increase the reliability of the part.

Power MOSFETs

The leakage current is a failure precursor for power MOSFETs in the same way as for SBDs. A specific failure mechanism for MOSFETs would be if the increase in leakage current is driven by a downward drift of the threshold voltage. This effect is expected to be stronger at increased temperatures where the threshold voltage is lower.

Mobile ions in the gate oxide push the threshold voltage down and could create the conditions for a thermal run-away. As the threshold voltage drops, the leakage current increases and so does the self-heating, which further reduces the threshold voltage, until the part is permanently damaged. The threshold voltage could also go up as traps get charged near the interface between the SiC and the gate oxide. The increase in threshold voltage degrades $R_{DS,ON}$, which results in higher conduction losses and a higher junction temperature. In this case too, the rise in temperature could lead to the permanent failure of the part.

The forward voltage of the body diode is an important parameter to monitor, even if it is not used as a free-wheeling diode. An increase in forward voltage is a strong indication of the progression of crystal defects (basal plane dislocations) into the drift layer. Over time, these defects affect the reverse breakdown of the MOSFET to the point where it loses its ability to block the voltage. If a forward voltage increase happens in conjunction with an $R_{DS,ON}$ increase, the likely root cause is a degradation of either the die attach interface or the frontside wire bonding scheme.

6. Conclusion

These are exciting times in the power semiconductor industry. The fast adoption of SiC devices in the automotive markets opens many opportunities for design wins and huge returns in investment. SiC device suppliers are very well-positioned to benefit from expansion of the EV market. The performances, reliability and cost of both SiC SBDs and MOSFETs make them the devices of choice for EV applications, dislodging the incumbent silicon IGBTs. Plenty of work lies ahead for us all, as we work to further improve SiC device performance and develop the next generations of SiC technologies and designs.

Bibliography

[1] Kashyap, A.S., Gendron-Hansen, A., Ji, I. H., Sdrulla, D., Odekirk, B., Meyer, D. Hong, C. Brower, W.: Beyond the Datasheet: Commercialization of 700 V - 1.7 kV SiC Devices with Exceptional Ruggedness for Automotive & Industrial Applications. PCIM Europe (2018), pp. 434-440

[2] Gendron-Hansen, A., Sdrulla, D., Odekirk, B., Kashyap, A.S., Starr, L.: 4H-SiC 1200 V Junction Barrier Schottky Diodes with High Avalanche Ruggedness. Material Science Forum 924 (2018), pp. 585-588

[3] Ji, I. H., Gendron-Hansen, A., Lee, M., Maxwell, E., Odekirk, B., Sdrulla, D., Hong, C., Kashyap, A.S., Faheem, F.: Highly Rugged 1200 V 80 mΩ 4-H SiC Power MOSFET. International Symposium on Power Semiconductor Devices and ICs (2017), pp. 371-374

[4] Gendron-Hansen, A., Sdrulla, D., Kashyap, A.S., Odekirk, B., Brower, W., Thornhill, L.: 4H-SiC Junction Barrier Schottky Diodes and Power MOSFETs with High Repetitive UIS Ruggedness. Energy Conversion Congress and Exposition (2018), pp. 850-856

Modular DC-DC converter for high-performance fuel-cell systems in trucks and buses

Thorsten Bürger, **Kunal Goray**,
AVL SFR, Regensburg;
Falko Berg, **William Resende**,
AVL List GmbH, Graz, Austria

Summary

Currently the industry is facing a massive push towards PEM (Polymer Electrolyte Membrane) fuel-cell based powertrains for heavy-duty trucks and buses [**1**]. These applications are putting additional and new requirements on the DC-DC converter and other electronics of the PEM fuel-cell system.

The biggest challenge for a DC/DC converter in a fuel-cell based powertrain for heavy-duty trucks and buses is the requirement of 800V output voltage in addition to the power level which is 2-3 times higher as compared to passenger car powertrains. Coupled to that, the number of fuel cell systems (FCS) in a truck could reach up to 3 units. Ideally one DC-DC converter could be used for these 3 FCS units, but this requires a change in DC-DC converter connection and currently this is not present in the market. In addition, DC-DC converters have to meet further technical requirements, mainly efficiency, durability or degradation, vehicle integration and life-time.

These requirements in combination with the requirements on the operation of the fuel cell accessories on 400V result in a new architecture for the DC-DC converter, which is currently not available on the market.

AVL will present in this paper the latest results of its internal R&D project for a modular and scalable DC-DC converter that is targeted for commercial vehicle applications and provides additional special features. The DC-DC converter technology is intended to be flexible so that it can be customized for different power levels and also operating voltages as per the vehicle application.

Introduction

Fuel cell systems for automotive applications have been known for several decades but it is only in the recent past that there is renewed interest in this technology for commercial vehicles

such as trucks and buses **[1]**. The focus of this paper is to describe in some detail the development efforts inside AVL for a modular DC-DC converter targeting the fuel cell commercial vehicle segment but with flexibility to also meet passenger vehicle applications.

The functionality of the DCDC converter in a fuel cell vehicle is to unidirectionally transfer power from the output of a fuel cell stack to a high voltage battery in order to charge it **[2]**. The fuel cell itself converts hydrogen into electricity onboard the vehicle; the DC-DC converter thereby enables power transfer from the stored Hydrogen into usable battery power onboard the vehicle. The power transfer needs to be unidirectional but depending on the overlap between battery voltage and the fuel cell voltage the DC-DC converter mode needs to be either Boost (battery voltage is higher than fuel cell voltage) or Buck- Boost (Battery voltage overlaps

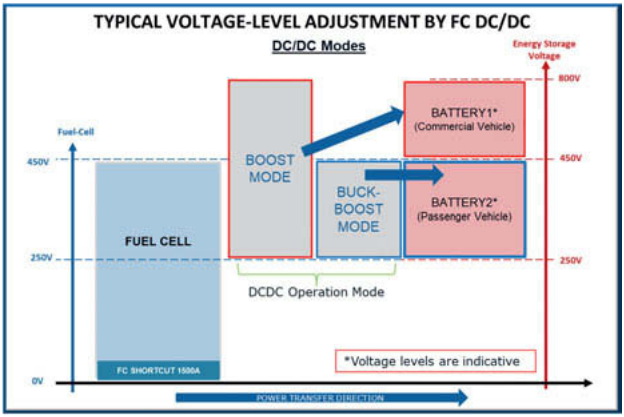


Fig. 1: Fuel Cell DC DC Converter operation modes

with the fuel cell voltage). This is illustrated in the **Fig1** for a commercial vehicle and passenger car with typical assumed battery voltage levels. It is assumed that for a commercial vehicle the battery voltage would be higher and in the 800V range whereas for passenger vehicles the battery voltage is in the 400V range. A DC-DC converter in a commercial vehicle has hence a different operation mode as compared to a passenger vehicle.

Key Electrical Components in a Fuel Cell System

A typical Fuel Cell System topology is shown in **Fig 2** . It consists of the Fuel Cell Stack which is the energy conversion device itself; a Hydrogen system that provides the fuel at the right temperature, massflow, humidity and pressure levels and an Air system that consists of a compressor and a humidifier to provide the air at the right temperature, massflow, humidity and

pressure. These Fuel Cell Systems (FCS) are available today in different passenger car applications **[3]**:

- Toyota Mirai
- Honda Clarity
- Hyundai Nexo
- Mercedes-Benz GLC F-Cell

Just to name a few.

The interesting components for the High Voltage (HV) integration are the fuel cell stack itself and the air compressor which are both HV components and play a role in the design of the DC-DC converter. Also, a couple of other parasitic loads are driven off the HV-Bus. These are the coolant pump, the fans in the intercooler of the vehicle cooling system and sometimes cabin heaters, but, their implementation is less critical. So, this paper will use the air compressor as major focus for all parasitic loads.

As is clear from **Fig2** the DC-DC converter is just one of many electrical components in the fuel cell system (FCS). As mentioned the FCS also has an electrically driven air compressor inverter and a cooling pump. Due to special packaging and dynamic flow requirements the air compressor and hence the electrical motor driving it rotate at 100000 RPM or even higher. This motor is controlled using an electrical inverter that is connected to a stable DC voltage source. This DC source can be either the battery or the DC-DC converter. At the moment of writing this paper the commonly available air compressor motor/inverter in the market are rated for 400V and this is not compatible with the 800V battery voltage for the commercial vehicle application. The air compressor is necessary for the operation of the fuel cell system and the voltage mismatch between the motor voltage and the battery voltage (namely 400V vs. 800V) leads to non-optimal workaround solutions for system integration that are not mature for a series product.

It is important to note that the output voltage of the fuel cell stack is lower than the voltage of the battery and hence there needs to be protection implemented in the system to prevent the high voltage of the battery from getting imposed on the lower voltage components of the fuel cell system such as the stack, BOP components etc. This situation can occur for instance in case of a short circuit within the DC-DC converter

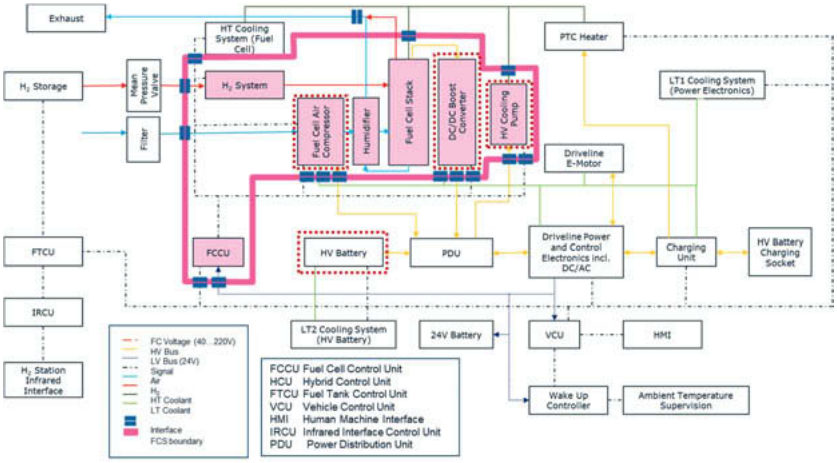


Fig. 2: Illustration of a Fuel Cell System (FCS)

Difference Between Fuel Cell Systems in Commercial vehicles and passenger vehicles

Typical fuel cell systems for automotive applications today are designed for system power levels of 70 to 120 kW depending on system and vehicle. These power levels are usually achieved with one fuel cell system. Commercial vehicles on the other hand require higher power levels (up to 350 kW for HD Trucks). To achieve these power levels today, several fuel cell systems are used and are installed “in parallel” which requires different electrical architecture. As it is ideal to operate each system separately to maximize efficiency and durability, the DC-DC converter needs to allow this operational flexibility. Voltage and current levels also pose a significant change. Typical truck applications are being designed for a HV DC bus(or battery) voltage of 800 V (driven by traction motor requirements), whereas passenger vehicles tend to start at around 400V.

Due to the focus on passenger cars for the last decade or more, fuel cell balance of plant (BOP) components that are available today are typically designed for 400V.Therefore the DC-DC converter for a commercial vehicle(CV) needs to be able to supply 2 HV DC bus voltages in order to allow free choice of BOP components, one for the powertrain and a second one for the balance of plant components.

In addition, the current levels required are also high. Fuel Cell stacks optimized for cost operate at currents above 500 A, most likely 600A. This allows for the cost reduction of the fuel cell

stack by reducing the required number of cells, but it requires DC-DC converters that can easily support these current level and these are hard to find in the market.
Durability will also play a role in the design of the cooling of the DC-DC because trucks are more constantly operating at max power than other vehicles. We expect the DC-DC converter to run at higher temperatures for longer times and this might lead to accelerated degradation in comparison to passenger vehicles.

Requirements of the DC-DC converter for commercial vehicles

AVL has been in continuous discussion on the topic of fuel cell systems with multiple automotive stakeholders globally and based on these discussions the **Table 1** lists the key requirements of the DC-DC converter for a commercial vehicle application

Table 1: DC-DC Converter Requirements

DC-DC Converter Requirement	Value
Input Voltage Range	3 FC stack Systems: 200 to 400 V 1 System: 600 to 1200 V
Output Voltage Range	800 V
Max System input current (continuous)	600 to 700 A
Peak System Current (peak, 30 s)	900 A
BoP Voltages	400 V
Input/output voltage during freeze start	80/250 V
Efficiency at max power	> 97 %
Durability	< 1% efficiency degradation at 8000 hrs
Coolant Flow rate	10-15 litre/minute nominal or lower
Max Coolant temperature	85^0C without derating
Galvanic Isolation for LV	Implemented between control system and HV
Galvanic Isolation for HV	Depends on vehicle requirements and application (may not be necessary)
Lifetime Requirement	Depends on customer but 2-3 times higher than passenger vehicles
Integration with Stack	Not necessary
Safety Protection	Implemented to protect Fuel cell stack

Challenges of meeting the DC-DC converter commercial vehicle requirements:
AVL has assessed the requirements listed above and a short explanation of the key challenges is documented below. The potential solution to these challenges as implemented in the AVL DC-DC converter design is explained in later sections of this document

1. Durability of DC-DC – Because the efficiency of the FC system directly influences the range of the vehicle and operating costs this needs to be maintained as high as possible over the lifetime of the product. Durability here refers to minimizing degradation in performance over lifetime of the DC-DC. Key driver for the durability is thermal and power cycling of the DC-DC components over lifetime. In addition, it is important to ensure that key components inside the DC-DC age at same or similar rate to ensure optimal life consumption
2. Lifetime of the DC-DC converter- Automotive qualified power electronics are typically designed to meet 8000hours lifetime, and this Fig. is significantly lower than the desired lifetime for Commercial Vehicle (CV) application. Because CVs drive longer distances and have higher usage hours than passenger vehicles one way to achieve the lifetime is to allow repair and replacement of the key components such as the power switches, naturally this needs to be supplemented with a limp home capability and/or prognostics. Another option that is also being explored by AVL can be to use custom packaged electronics or non-automotive components that are inherently designed for higher thermal cycling reliability (e.g for power generation applications)
3. Efficiency – The full load and part load efficiency of the DC DC converter needs to be higher than 97% and this needs to be achieved through a combination of efficient hardware and control design. A strategy to manage the switching losses, conduction losses and losses in the passive components needs to be implemented
4. Galvanic Isolation on HV – The Galvanic isolation requirement for the DC-DC converter is a safety feature to avoid unintended high voltage and currents from appearing on the fuel cell stack or the battery due to a failure in the electronics. The size of a galvanically isolated DC-DC converter can be roughly twice that of a non-isolated DC-DC converter with a similar impact on cost. Hence a decision to go with a galvanically isolated design can have a major impact on the cost of the final product
5. Mission Profile- As mentioned earlier the drive cycle or mission profile for a truck is significantly different from a bus and both are quite different from passenger vehicles. The mission profile needs to be converted into currents so that lifetime calculations can be carried out for the individual components. Especially in CVs with multiple DC-DC

converters and fuel cell stacks connected to a battery this conversion of mission profile to power device currents can be tricky

AVL Modular DC-DC Converter

The AVL DC-DC converter is a result of a global effort across AVL teams in EU and North America and builds upon the extensive fuel cell system knowhow and power electronics knowhow that was already existing inside AVL. In addition to the electrical and electromechanical design the team also referred to existing fuel cell vehicle solutions **[3]** to ensure a unique and state of the art solution that is practical for the industry.

Electrical Design: The DC-DC converter is based upon an interleaved topology that spreads the overall power and current flowing in the DC-DC converter across multiple phase legs. The interleaved design [**4**] is scalable in power and in addition provides flexibility to the mechanical designer with each phase leg being a building block that can be repeated multiple times to achieve an overall DC-DC converter power. The interleaved DC-DC converter topology is a classic topology known for its simplicity and reliability and is also in series production **[5,6,7]**. Each phase leg is also provided with an internal solid state switch that can very quickly cut off a phase leg from the rail if needed.

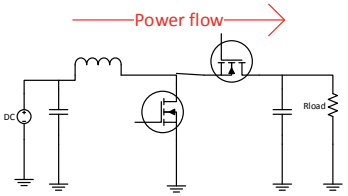
The DC-DC under development is not galvanically isolated between the input and output voltages but the solid- state switch can provide a level of protection to the system without the additional weight and complexity of a galvanically isolated system

In order to ensure high efficiency and performance of the electrical system Silicon Carbide power switches are used in the design.

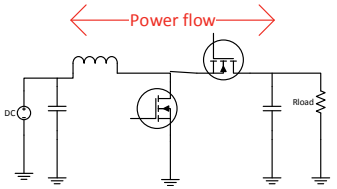
To ensure practical implementation in automotive application the design utilizes commercially available automotive qualified electrical components.

As explained in an earlier section the DC-DC converter has different modes of operation. The phase legs for the DC-DC converter can be conFig.d to allow for the below operation modes:

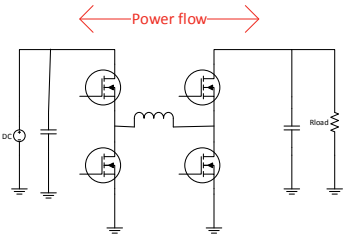
1. **Boost Operation** - A unidirectional DC-DC converter, which boosts the input voltage from fuel cell to the battery



2. **Buck or Boost** = The phase leg can be modified to operate as an unidirectional converter which is able to act as a boost from Fuel cell to battery or act as buck in opposite direction e.g from battery to Fuel cell BOP



3. **Buck-Boost** = A Buck-Boost bidirectional converter can be be realized by joining two phase legs and this is enabled through special measures implemented in the electrical and mechanical design. This will enable the DC-DC to decrease and increase voltage in both direction e.g as needed in a passenger vehicle application



Thus the phase legs enable flexible design options for the electrical operation of the DC-DC converter

Mechanical Design: The realization of the interleaved converter mechanically is done using an open frame structure with two phase legs placed on two sides of a liquid cooled heatsink

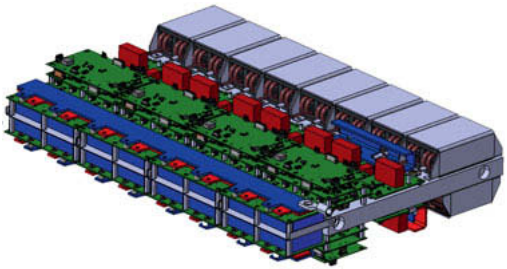


Fig. 3: One of the Mechanical concepts under evaluation for a 160kW DC-DC converter

Fig3 shows one of the multiple mechanical concepts under evaluation at the time of writing this document. The number of boards shown in the **Fig4** will get reduced in the later stages of the development as many test structures and analog components included on the PCBs get eliminated. Due to the interleaved topology the number of phase legs can be varied to achieve a desired rated power. The coolant flow rate of nominal 10-15 litre per minute is assumed as is standard practice in commercial vehicles.

Control System: The control of the DC-DC converter is implemented in an automotive grade Digital platform. As is standard practice the control system and the HV power sections are galvanically isolated for safety and reliability considerations.

The control scheme is also scalable like the phase legs. In addition the control software has the provision to include or exclude the phase legs based on the current demand and safety considerations to ensure that the efficiency of the DC-DC converter stays as high as possible even at partial loads and that the durability and lifetime goals can be achieved. In addition to the control of the DC-DC the controller on the DC-DC will also be able to implement the THDA (Total Harmonic Distortion Analysis) algorithm on the phase leg to enable a health monitoring of the fuel cell stack. THDA is an AVL patented technology **[8]** that enables health monitoring of the fuel cell stack based on imposition of specific voltage and current signal patterns and associated measurements. The advantage of this technology is that it can provide prognostics and thus help in avoiding failures through preventive actions **[9]**.

Innovative Aspects of AVL DC-DC converter

The AVL DC-DC converter solution is scalable and modular in its design and is unique in the fact that it is designed for the commercial vehicle application right from the concept stage. The various innovative elements in the design are:

1. Scalability – The DC DC converter can be scaled in steps of 40kW until 160kW in a single unit and multiple units can be connected in a master slave configuration to provide a 320kW FC DC DC converter solution
2. Integrated Blower inverter- The air compressor inverter for a 400V blower unit can be included within the DC-DC converter housing as an option
3. 800V and 400V components in one package– The DC-DC converter can step up to a 800V DC bus voltage but can also be conFig.d to allow power transfer from 800V to 400V. This allows usage of 400V BOP components in a 800V battery system
4. Safety- Each phase leg is provided with a solid state switch that can be actively controlled to quickly isolate a single phase leg in case of a failure on the phase leg
5. Efficiency and Durability- Depending on the current demand the central controller can define the number of phase legs in operation thereby ensuring >97% efficiency at part and full loads. In addition the controller can ensure all the phase legs get utilized equally over lifetime
6. Redundancy- In case of a failure in a phase leg it can be completely eliminated from the electrical circuit thereby ensuring continued operation of the Fuel cell system
7. THDA Ready- AVL has unique IP for Fuel cell stack health monitoring and the DC-DC converter is designed to be THDA(Total Harmonic Distortion Analysis) ready. This is achieved by using the DC-DC converter to generate the THDA excitation signals that are later processed in the fuel cell control unit

Summary and Next Steps

A DC-DC converter designed for the 800V commercial vehicle application is described in the document, The DC-DC converter can be directly connected to a high voltage battery thereby simplifying the electrical architecture for the CV application. This also means that the solution can be much more easily translated to a series production. The hardware of the phase legs is designed to operate in various modes to transfer energy from the input to the output and allows the connection of 400V BOP components on a 800V system bus to avoid the need of new 800V BOP components and help reduce time to market. In addition the mechanical design is envisaged to be flexible allowing for greater freedom in vehicle integration of the component.

The goal of the internal R&D effort is to have a partially tested DC-DC converter by Q4'2019 that will then be used to demonstrate the technology in a fuel cell vehicle application. In the near term this will be offered commercially.

References

[1] Hunter, Chad, and Penev, Michael. Market Segmentation Analysis of Medium and Heavy Duty Trucks with a Fuel Cell Emphasis United States: N. p., 2019. Web.

[2] Resende, W., et. Al, Fuel Cells: A Profitable Zero-Emission Solution for Heavy Duty Trucks, 2019 AVL ICPC Conference

[3] AVL Internal Reports on benchmarking and Teardown of Fuel cell systems in passenger vehicles

[4] Khalilzadeh, et. Al. A Novel Interleaved DC-DC Converter with Reduced Loss for Fuel Cell Vehicle Application: IEEE 2017

[5] Cheng, el al, Development of a PEM Fuel Cell City Bus with a Hierarchical Control System: Energies 2016

[6] Miwa , Dissertation - Interleaved Conversion Techniques for High Density Power Supplies, MIT 1992

[7] Okura, Development and Future Issues of High Voltage Systems for FCV, Proceedings of the IEEE | Vol. 95,No. 4, April 2007

[8] Patent A51008/2016, Verfahren zur Diagnose eines technischen Systems, 2016

[9] Ramschak, E., et. al, Detection of fuel cell critical status by stack voltage analysis, *Journal of power sources, 157*, 837-840, 2006

Future electric/electronic architecture

Sustainable design of a digital in-vehicle backend infrastructure

Dr. **Matthias Traub**, **Hans-Ulrich Michel**, BMW Group, München

Abstract

Through the cooperate strategy “Number One> Next”, relevant action fields for the BMW Group's digital transformation are set. This implies not only to “re-invent” the car itself but further the automotive electric/electronic architecture (e/e) in particular. Essentially, the aim is to modify existing and future vehicles as well as services, to build customer-perceived and personalized digital premium products. The design of all products and services is based on their "life cycle", with the goal to optimize them economically over their lifetime period. A further goal is a fast implementation of applications and their provisioning to customers, similar to the CE industry e.g. for customer features and security updates.

With these objectives in mind, the BMW Group is working on a future e/e architecture vision, which includes a sustainable base for the implementation of an integrated digital e/e infrastructure (vehicle - BMW backend - 3rd parties) for all applications and services. In addition legal requirements, competition and technology trends are also relevant input parameters. For the future e/e architecture, BMW Research develops the necessary e/e building blocks in close cooperation with the product development departments and external partners. The objective is to develop a solution in which BMW can concentrate on its core competencies. Main priority is a standardization and cost optimization. Moreover, one aim is to implement a "continuous deployment" by a standardized and homogenous e/e architecture and thus to offer BMW customers a product which is always up-to-date over its whole lifetime.

Based on the mentioned input factors, precise requirements for a future e/e architecture are formulated and goals are defined. These requirements and goals are used for the evaluation of each specific e/e building block. The goal achievement for the solutions will then be compared to the current serial implementations with a 360° target matrix. As an example, two solutions and their evaluation are addressed in this paper:

- The zonal architecture enables a modular design of the on-board supply system. The main advantage is the abolition of a customer specific wiring harness. A gradual transfer to an automatic cable harness production will be supported with the modular framework. An automated test is afterwards possible.
- The scalable modular computing platform is generating a homogeneous basis for integration of applications across different domains. One of the key element is a free partitioning of functions (vehicle and/or backend).

Starting point and strategic goals

Long-term social trends such as "aging society", "green mobility", "urbanization", "accident-free driving", "social media" and new technologies are leading to changing usage scenarios and to new, distributed customer functions. Safety and security requirements are being increased by completely new functions such as autonomous driving and connected mobility at the same time. These mega trends, legal requirements and BMW's corporate strategy "Number One Next" set the relevant fields of operation for the digital transformation of the automotive industry. The main story points are the realization of future cars and services as inspiring and personalized digital premium products. The design of these products and services is based on their life cycle. A further point is the fast development of customer functions and the need for forward looking security patches, which must be deployed to the vehicle without delay during its lifecycle.

Driven by these story points, the research department is working on the electric/electronic architecture vision (e/e), which realizes a seamless digital infrastructure (vehicle – OEM cloud – cloud of 3rd parties) that supports a sustainable base for all applications and services. Figure 1 shows the derivation of the resulting game changer based on the mentioned objectives.

- The e/e architecture covers the vehicle, the OEM cloud and all relevant system parts of 3rd parties (applications and their interfaces (APIs)). The main goal is a partitioning decision-independent application development. The subsequent use of edge computing or dynamic functional allocation is thus also possible.
- The e/e architecture integrates AI-based decision systems and supports an adaptive, personalized and customer specific behaviour.
- The e/e architecture enables a situational connectivity of systems and a dynamic repartitioning of functions plus an optimal usage of resources at any time.

- A flexible access to all in-vehicle data and its mirroring in the backend (so-called Digital Twin) leads to an improvement of the product range and is the basis for the realization of intelligent functions.
- The e/e architecture supports energy-efficient behaviour at system and component level.

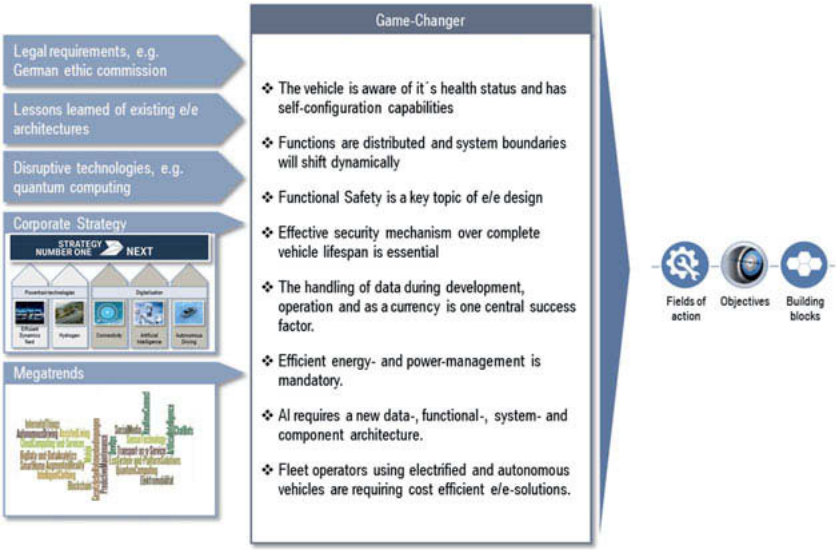


Fig. 1: The next generation and vision of the e/e architecture are driven by several inputs. The concrete fields of action, objectives and building blocks can be derived over the so-called game changers.

The main fields of action for a sustainable e/e architecture are depicted below. These topics are derived from the game changers, particularly from the areas of connectivity, digital services and autonomous driving.

- The vehicle e/e architecture no longer ends at the ECU respectively at the vehicle boundary, it also comprises the connectivity path (including edge), the OEM backend and cloud parts of OEM-independent third party vendors.
- An independent partitioning of applications between cloud- and vehicle-infrastructure is one of the key patterns. This generates benefits for an optimal resource sharing, for a cost-effective fail-safe/fail-operational solutions and for a seamless connection of off-board services or applications.

- The perception of the vehicle as a digital product (analogous to CE devices) toward the customer is absolutely necessary. This means that individual applications can be updated continuously and ideally over the whole lifecycle of the product.
- The consequent usage of seamless connectivity based on 5G for autonomous driving and connected services.
- Establishing a sustainable and optimal economic solutions, which are taken the complete lifecycle of the vehicle in to account.

Based on these fields of action, an evaluation against the 360° objectives is performed. The definition of each objective is based on the motivation and concrete customer experiences. In the following, a concrete objective by the example of “energy-efficiency” is described.

Objective: Energy-efficiency

The average electrical power requirement (for all voltage levels) should be minimized. A continuous reduction of power consumption for each e/e architecture generation with the same functional scope should be striven for. E.g. the goal on low-voltage side for the next generation is at least a 15% reduction. Customer experiences and motivations are:

- Fuel consumption is significantly improved. The electrical energy is used optimally, this leads to an increased range.
- The legal requirements are fulfilled.
- Each saved Watt reduces the CO_2 footprint of the fleet.

Fig. 2: The next generation and vision of the e/e architecture are aligned over the main objectives (360° view).

Figure 2 depicts all objectives, which are relevant for the evaluation of the vision e/e architecture. For deriving the vision and relevant building blocks for the next generation of e/e architecture, a well-defined system structure is necessary. This system structure contains the overall digital e/e infrastructure for automotive systems and services.

Structuring of the overall system

The digital e/e infrastructure for automotive products consists of the in-vehicle e/e architecture, the OEM cloud and shares of the third-party cloud. These parts need to be linked up together. Figure 3 shows two core areas. The application area and the infrastructure area (onboard platform, mobile communication infrastructure (so-called edge), offboard platform and the IT-infrastructure of third parties). Up to now, most of the automotive applications are being developed for a specific platform. Especially for the onboard platform, there is a high variety of different realizations (classic Autosar, adaptive Autosar, GENIVI …) with heterogeneous developments including processes, methods and tools. In the future, most of the applications should have a platform-independent implementation. This gives the freedom to run applications onboard or offboard without the need of platform specific adaptions.

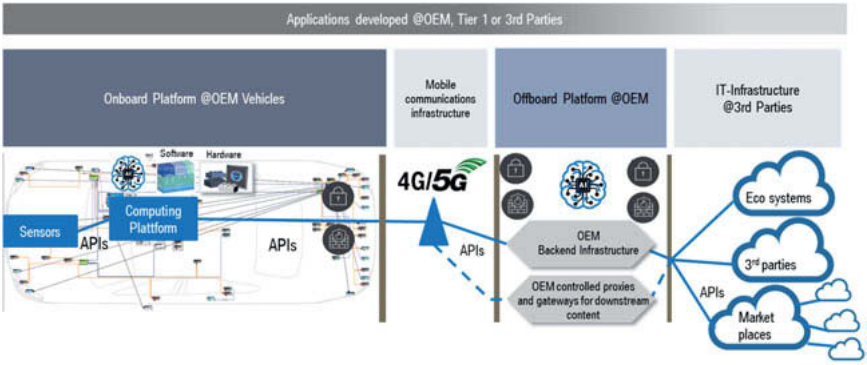


Fig. 3: Seamless digital e/e infrastructure as nucleus for future innovations [1].

Based on the two areas in figure 3, the specific subsection of a digital e/e infrastructure is shown in figure 4. The upper part of figure 4 covers the logical architecture. This part is technology-independent and consists of three subsections:

1. Customer functions and legal requirements: Abstract description of applications as well as legal, functional and non-functional requirements.

2. Functional architecture: This level includes the decomposition and structuring of customer functions including their requirements in individual function blocks as well as the necessary interfaces between the function blocks.
3. Service-oriented architecture: The mapping of the function blocks to concrete services including a description of the service interfaces is described on this level.

The modeling and description of the individual subsections in the logical architecture does not require any information about a subsequent technical realization. Only requirements for a real implementation are formulated. The lower part of figure 4 contains the subsections of the technical architecture. These are the basis for the derivation of individual concrete technical solutions. Safety, security and privacy by design become more and more important as a cross-cutting task over both areas (logical and technical architecture).

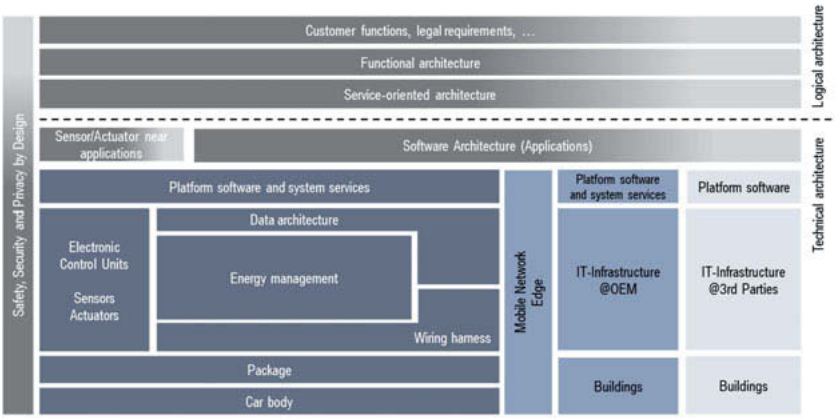


Fig. 4: The logical and the technical architecture of a seamless digital e/e infrastructure consists of several subsections [2].

Building blocks for a sustainable digital e/e infrastructure

Based on the game changer and goals more than 30 architectural building blocks addressing subsections of the logical and technical architecture shown in figure 4 are the framework for the development of a future-proofed e/e architecture. Figure 5 shows an extract of the overall building blocks. Every building block is structured in focus topics and subtopics with a timeline, an indicator for the technical maturity that can be achieved and accompanying projects that will support key issues.

The building blocks “Network Architecture” with zone orientation and “Computing Platforms”, which are sketched out below, are two major factors that will influence the concrete architectural design.

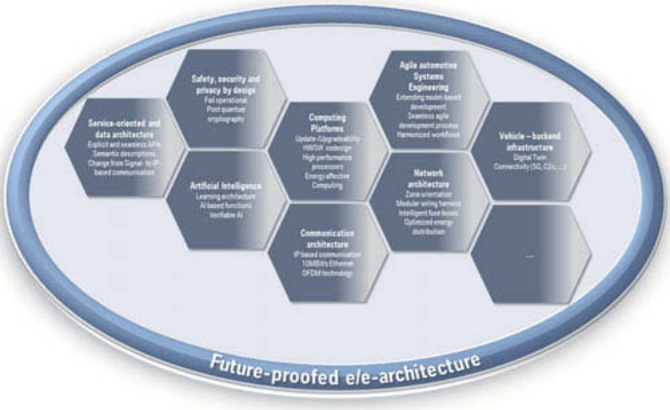


Fig. 5: Derivation of necessary building blocks for a sustainable digital e/e infrastructure for automotive systems and services.

Building block: Zonal Architecture

In the current e/e architectural design, the networking of electronic control units (ecu) and the resulting wiring harness is an important element to achieve the goals of a sustainable, upgradable and cost-efficient product. In the past, the wiring harness was mainly the subsequent result of a functional partitioning taking backward compatibility, ecu-development and options-scaling into account. Because of the high volume of options BMW introduced the concept of a customer specific harness. This customization results in a high variation of harnesses which must be manufactured, delivered just in time and integrated into the car, all within the production cycle. Because of the high portion of manual assembling, the production of the wiring harness is mainly done in low-wage countries, which increases the difficulty more and more, since the security of supply is indispensable for the car-manufacturer. In addition to producability, weight and available space are important factors which must also be taken into account.

In order to address these challenges, a more holistic approach in e/e development is necessary; reducing cabling effort, harness-variants and supporting a step to automatable wiring

harness production. Automation will also increase the test- and safeguard-quality which is necessary to meet the safety-requirements of highly automated driving functionality. To address this challenge, a zone-based architecture was developed, figure 6 shows a general view.

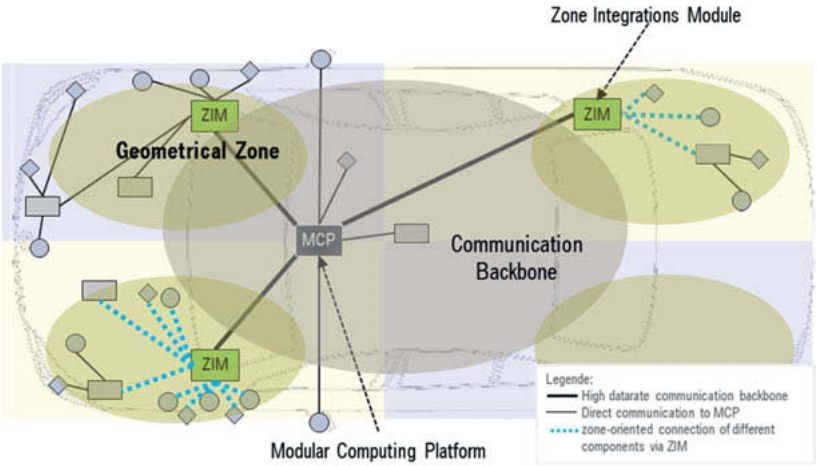


Fig. 6: General view on a zone based e/e architecture [2].

The vehicle will be subdivided geometrically in zones, this number of zones may vary depending on the vehicle type and configuration. Zone-Integration-Modules (ZIMs) act as a gateway for data networking, connecting e/e components within a specific area with the modular computing platform (MCP). At the same time ZIMs assume the task of an intelligent central power supply with semiconductor switches for all zone components.

Amongst others the benefits of a zone oriented architecture are:

- Significant reduction of the overall harness length (>100m) concerning analog and bus based communication wiring with communication aggregation.
- Supports the separation of independent harness modules and with that the number of automatable wiring harness elements.
- Semiconductor switches allows fast detection of shortcuts and enables the reduction of conductor cross sections.
- Support of safety measures concerning a safe energy supply for automated driving functions by enabling intelligent degradation concepts.
- Reduces the burden on MCP by hosting power driver components.

Building block: Modular Computing Platform

Actual in-vehicle integration platforms are developed in a heterogeneous way. Each domain has usually his own technical design patterns and specific processes, methods and tools (PMT). The design of these components is strongly oriented towards the optimum of manufacturing costs and the technologies available at the starting time of development, usually 5 years before start of production. As a result, at start of production the components are typically operating on the limit of what is possible regarding performance and thus no or only small updates over runtime are possible.

The building block "Modular Computing Platform (MCP)" provides an important and sustainable solution in the sense of an "infrastructure as a service". A homogeneous modular system for the cross-domain integration of applications is offered with the MCP, including technology, PMT as well as cooperation and business models. Another feature of the MCP is its update and upgrade capability, to bring leasing returns technically up to date, e.g. for new innovations, security or safety topics. Figure 7 shows the two core action fields "technical design" and "systems and software engineering" which are necessary for a successful implementation.

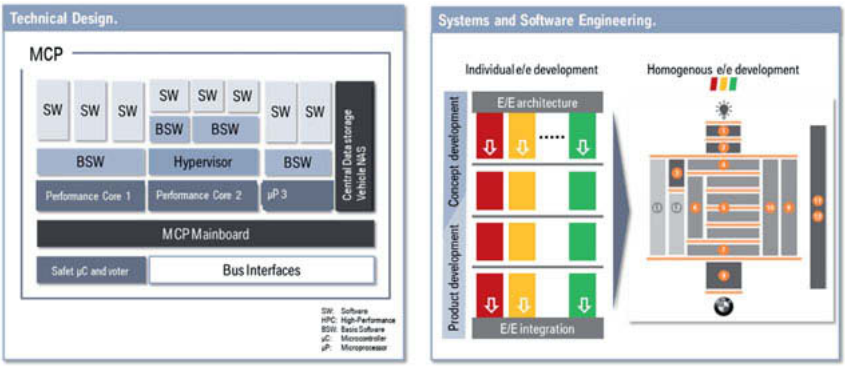


Fig. 7: The technical design and a homogenous systems and software engineering are essential for an overall integration approach.

The MCP also achieves a reduction of ECUs over high integrations. Furthermore, the manufacturing costs and the efforts during the development process are reduced. Figure 8 depicts the overall solutions of the MCP approach. In the vision e/e architecture (big picture) all in-vehicle integration platforms are derived from the MCP. The high number of integrations-ECUs (ECU class 1 and ECU class 2) in the actual e/e architecture can be significantly reduced. In

the big picture only 1-3 integration platforms are necessary. The remaining part is realized with standard components.

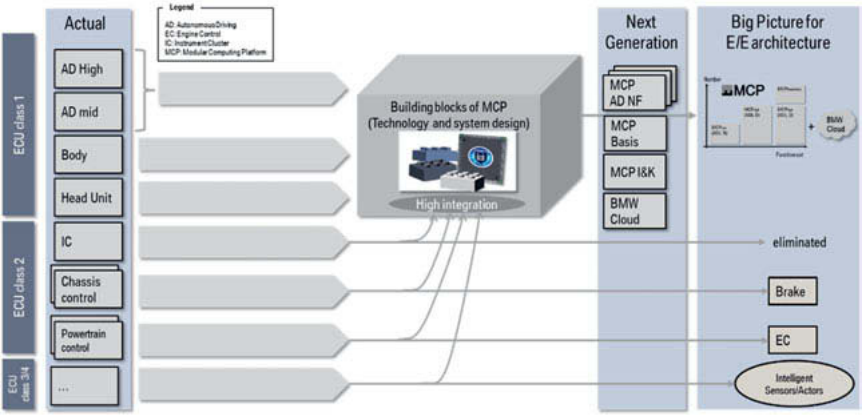


Fig. 8: The modular Computing Platform (MCP) creates synergies and offers the possibility for a homogenous system structure.

Overall View – future e/e architecture vision

Based on the objectives described above, figure 9 illustrates a first approach of a zonal based architecture with a centralized computing platform. Redundancy because of safety related functionality is not shown. Important attributes of this approach are:

- Central compute cluster MCP: Use of a scalable, high-performance computing platform which ensures a clean separation of competitive, BMW specific high level functionalities and applications. Standardized interfaces to low level functionality allows fast update/upgrade cycles while more vehicle specific sensor/actor related functionality remains stable over the same period of time.
- Zone-Integrations-Modules ZIMs: Each of them acts as an aggregator of sensor, actor or embedded control unit data and at the same time as an intelligent power supply based on physical location in the vehicle. An Ethernet-Backbone provides a single interface to host computing.
- A central access point implementing a firewall for authentication and encryption based on TLS, provides a safe and secure access to the BMW Backend and 3rd party service providers (Internet).
- Backend infrastructure: The BMW Backend supports fast, safe and secure remote upgrade/update cycles for vehicle functionality and security or safety patches. At the same

time, controlled and limited internet access is enabled. Based on 5G networking and guaranteed quality-of-service functionality, functions will be partitioned across the BMW Backend and in-vehicle in a flexible way.

- Power Distribution Center: Power Distribution point for all ZIMs.

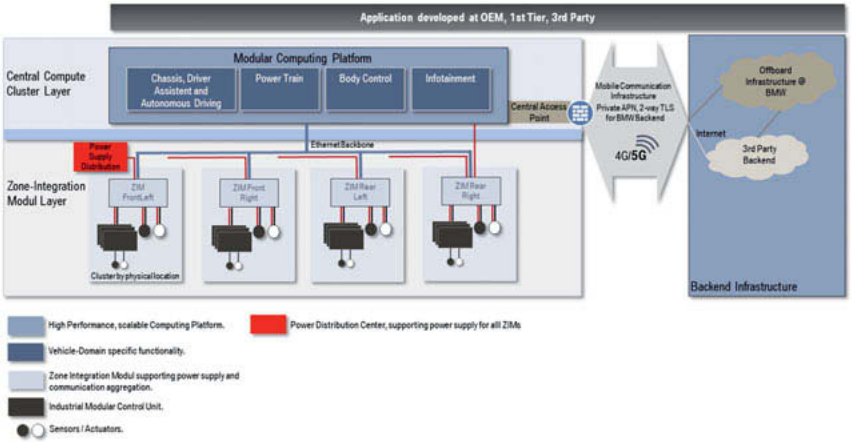


Fig. 9: Overall view on a future e/e architectural vision

Conclusion

In this paper we described game changers, action fields and architectural building blocks to address a future-proofed e/e architecture. We outlined, in an exemplary way, the scalable modular computing platform (MCP) and the zonal architecture approach as possible solution elements. The MCP reduces the number of control units and offers a homogeneous basis for integration of applications across different domains. It supports the separation of functionality with short updates/upgrade cycles and the free partitioning of functions between vehicle and the BMW backend infrastructure. The zonal architecture approach with zonal integration modules (ZIMs) leads to a significant reduction of the overall harness and supports the automated wiring harness production. The ZIMs host power driver components and with that, relieves MCP resources. They act as an intelligent power distributor and provide support to fulfill safety requirements concerning the power supply using degradation concepts. Future work and challenges covers inter alia:

- Detailed safety and security concepts on e/e-system and component level. One goal for safety is to reduce the effort of physical redundancy by enabling intrinsic fail-opera-

tional behavior of the MCP. This results in additional challenges concerning power supply and cooling concepts on component level. Also freedom from interference must be guaranteed between safety and non-safety related functionalities.

- Energy efficiency is a central topic concerning legal requirements, fully electric driving and will reduce the effort concerning safety measures for power supply and cooling concepts. For this reason, one main focus for the MCP within this topic will be on hardware/software co-design.
- The zonal architecture creates new challenges by changing the view on current existing functional bus systems, which must support a zonal approach in the future. The current focus is on a zonal LIN-Bus approach.
- Another topic is to substitute the use of different bus technologies by using OFDM based Ethernet transmission over copper twisted pair. To support a migration path for existing bus technologies (CAN, CAN-FD, Flexray) the gateway concept on ZIM-level is important to guarantee latency requirements and reduce the configuration effort.
- Data-compression solutions to transport raw sensor data are necessary if this sensor components should communicate via ZIM to the modular computing platform.

Mentioning all this topics for future work we believe a discontinuous transmission within the e/e architecture is necessary to meet the requirements of a future proofed design.

References

[1] “Building blocks of a sustainable digital automotive infrastructure”, Matthias Traub, Philipp Obergfell, Vector Congress 2018, Stuttgart.

[2] “Electric/Electronic Architecture as an Enabler for Connected Mobility and Automated Driving”, Matthias Traub, Vehicle Electronics am Connected Services Conference, Gothenburg, 2019.

Function- and Service-Orientation – a Game Changer for the E/E-Architecture of Tomorrow

Rüdiger Roppel, Dr. **Matthias Görber**,
Dr. Ing. h.c. F. Porsche AG, Weissach

Abstract

A new continuous and tool-based function-oriented development approach for E/E-architectures is needed. This is due to the automotive megatrends as well as the challenges of the digital transformation.

In today's automotive industry, it is required to constantly generate new features in time-decreasing cycles of development. This can only be achieved with the consequent abstraction of functionalities from the concrete technical realization.

A new development paradigma is needed to control and handle the increasing complexity in communication. It concentrates highly dynamic functionalities on only a few extremely high-performant ECUs. It also allows OEMs to integrate functionalities from external partners.

In addition to the classical signal-based communication, a new service-oriented approach will be established.

An essential factor of success will be well-defined Application Programming Interfaces (API) that can be reused for all application functionalities.

In the upcoming paper, the function-oriented development approach for the E/E-architecture of the next generation at the Dr. Ing. h.c. F. Porsche AG will be introduced.

Next, the development approach for the service-oriented communication will be shown. Followed by its processes, new roles and tools.

Finally, insights will be given about the methods and measures how the function- and service-oriented development approach at Dr. Ing. h.c. F. Porsche AGs EE-department is realised. And, no less important, how it will be spread amongst other development domains.

1. Challenges in modern E/E architectures

The requirements of a modern E/E architecture are becoming increasingly diverse. Firstly, it has to model the current and future customer requirements of a vehicle and its ecosystem. Secondly, the implementation of the functional requirements is expected in ever shorter time windows and with greater business potential.

What's more, development has long ceased to focus solely on the vehicle itself – instead, the vehicle is an integral component of an ecosystem, in which the back-end connection for providing cloud content is playing an increasingly important role.

A further requirement involves the scalability of an E/E architecture. The aim is to model the largest possible range of vehicle projects – in the extreme case, from cost-optimized subcompacts to the innovation-driven luxury segment.

This spectrum of requirements produces central premises for a modern E/E architecture. The start involves abstracting function development from the hardware development. As such, a modern E/E architecture also has to be separated from the dominating domain character. In addition, a successful separation of the hardware and software requires clearly defined interfaces at the system boundaries.

Lastly, the implementation of reusable services represents a further paradigm shift for a successful E/E architecture. It makes it possible to increase the robustness of the system, avoid redundant developments, and reduce error rates.

This manuscript describes the importance of function development for future E/E architectures, as well as addition of service orientation. In particular, it will describe how Dr. Ing. h.c. F. Porsche AG (hereinafter Porsche) is approaching the paradigm shift toward a service orientation with regard to the technological, organizational, and process related challenges.

2. Functional orientation

2.1 What is functional orientation?

Functional orientation separates the individual functions from the respective hardware. To do so, the internal and external requirements of the function and its integration are described first. Based on these requirements, the subfunctions are partitioned on the technically logical target hardware environments and ultimately implemented there – integrated, tested, and released. This is explained further below using the "Wet mode" function as an example.

The function-oriented development approach makes it possible to plan and implement even complex functions in a structured way. The impact of errors and changes are easier to assess. With the introduction of the calculation level in the end-to-end electronics architecture at the latest, the complexity of the high-performance computers (HCP) can only be managed through consistent functional orientation.

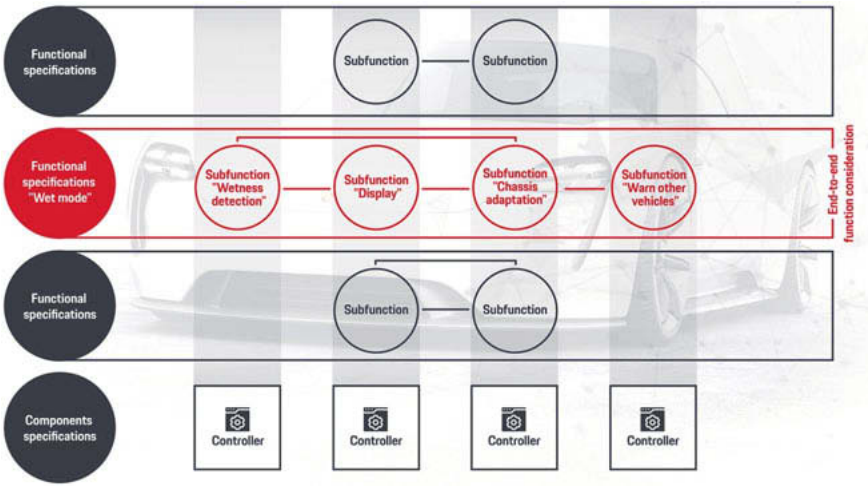


Fig. 1: The function-oriented development approach in detail

2.2 What is functional orientation?

As mentioned in the introduction, the complexity, number, and degree of distribution of functions continues to increase rapidly. Many controllers have now become hosts for a large number of subfunctions, which can only perform an end-to-end customer function through perfect interaction. As a result, precisely planned function hubs are required in development, above and beyond the controller clusters. Under a strict component orientation, this overall complexity is no longer manageable, resulting in the following development risks:

- High risk of errors and late error discovery
- Difficulty of assessing the impact of errors
- Redundant development of functions

2.3 What are the challenges of functional orientation?

Many automotive OEMs are still dominated by a domain-oriented development structure and organization. This means classic domains such as infotainment, powertrain/chassis, and so on bear responsibility for developing their own functions.

In light of the increasingly complex, networked functions, this invariably leads to questions regarding expertise and responsibility at the domain borders. In addition, there is a risk that the respective domains could develop the same or similar functions multiple times – perhaps without even knowing it.

This makes it necessary to develop a consistent, domain-independent alignment of the company organization in favor of a function-oriented development approach.

The high complexity and level of distribution in function-oriented development, in turn, leads to a high complexity of the communication and networking structure.

These motivations, among others, make the supplementary introduction of the service orientation and service-oriented communications essential.

3. Service orientation

The aforementioned functional orientation can only achieve its full potential when combined with service orientation. Together, both development paradigms support the methodological development approach toward an open, future-enabled E/E architecture.

3.1 What is service orientation?

Service orientation is a development paradigm for structuring and using distributed functions that are the responsibility of different owners.

In the example of the "Wet mode" function, this means the "Wetness detection" subfunction is provided to a wide range of users across all domains. (“wet mode” is a vehicle assistant function that detects wet conditions and adapts the vehicle into a safer mode e.g. PSM, PTM, adaptive aerodynamics etc.)
The key of this new development approach is the encapsulation of existing subfunctions to form services. In addition, the activities of services have to be bundled to create "higher services".

The services generated as a result are made available to all functions, both vehicle-internal and external, through a marketplace. The services are used based on a subscription concept in which no further discrete connections are established between individual components.

A decisive factor for this concept is a standardized interface that is stable and reusable. It is realized using an API (application programming interface). The definition of the API is the central communication-relevant and function-relevant factor.

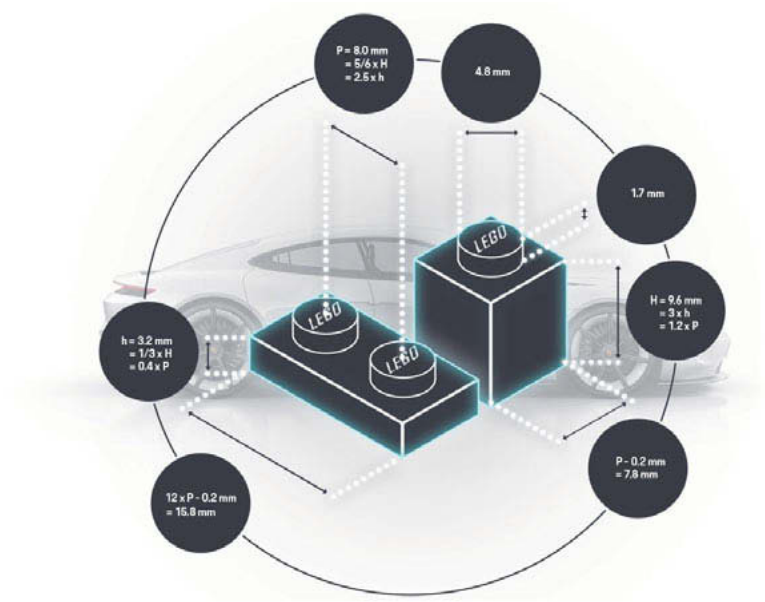


Fig. 2: The API as standardized interface

As a result, the user of the services knows only that this service is offered, which inputs it requires, and which type its result has. The details of how the result was determined are not known, however.

This approach makes it possible to simplify collaboration across area boundaries, with a clear distribution of development responsibility and avoidance of redundant development activities.

The aims of the service orientation:

- Reduce complexity under function-oriented development
- Increase robustness and minimize errors
- Structuring of the functional orientation through reusability of (sub) functions and services, with the resulting reduction of development costs and greater flexibility

3.2 What is service-oriented communication?

Service-oriented communication is based on a service-oriented architecture (SOA) of IT systems. The SOA structures the services so they can be used by different sources independently of one another.

The orchestration of services opens up new possibilities for designing and maintaining functions. In this approach, the service orientation forms the new paradigm that enables a client to trigger and control the exchange of information with services.

The independence of the services provides for a high degree of flexibility for their use – and it does not matter where the communication participants are actually located.

3.3 Examples

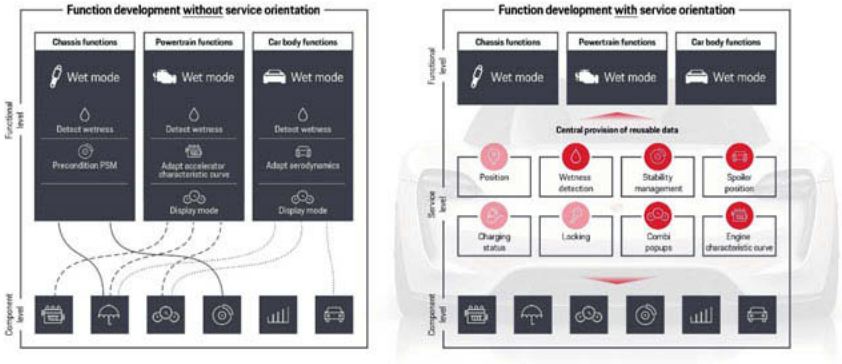


Fig. 3: Functional orientation with and without service orientation

For the "Wet mode" example, this illustration emphasizes the difference between pure function-oriented development on one side and the addition of the service orientation on the other side. The path on the left still faces the risk of multiple development of subfunctions, such as wetness detection. This risk is minimized by introducing a "Wetness detection" service.

The following Fig. emphasizes the approach of a domain-independent vehicle API, that is, the total of all individual APIs. The functions can access and reuse the various services through the known interfaces. This also applies to functions from external partners.

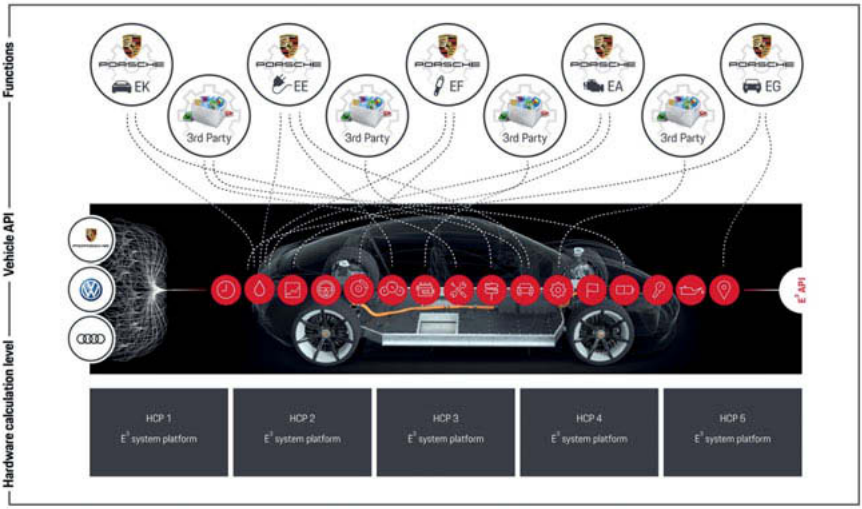


Fig. 4: The connection between vehicle API and function development

4. The path to service orientation at Porsche

The fundamental characteristics and contents of service orientation were described in the previous chapters.

Porsche is defining the contents of the service orientation for the coming generations of the E/E architecture together with its development partners, VW and Audi. However, Porsche is following a separate, individually designed implementation concept in the brand and the specialist departments.

The following chapters describe what the process-based, technological, and organizational implementation looks like at Porsche.

4.1 Process-based implementation of the service orientation at Porsche

The process-based implementation of the service orientation at Porsche is divided into an extensive E/E architecture emergence process and intermeshes closely with this process.

At first, the function catalog of the E/E architecture and function lists of the projects are used to produce a function architecture model. This model describes the local communication relations.

Based on this model, the service and signal requirements are derived and incorporated in the existing Group-wide communication coordination and approval process.

Once all the function participants have agreed to the communication route through signal or service, the relations are transferred to a communication matrix, which is then provided to the specialist development departments for technical integration.

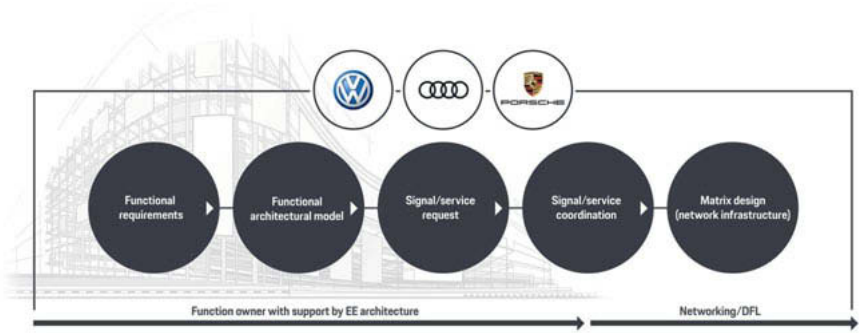


Fig. 5: The process for function-oriented development

4.2 How is the service orientation at Porsche implemented technologically?

On the technology side, the described process for implementing the service orientation requires the following:

- A capable tools landscape for an end-to-end, model-based architecture and function development approach
- Clear, comprehensible design rules for services and APIs
- Consideration of the service orientation on the hardware platform by creating resource pools

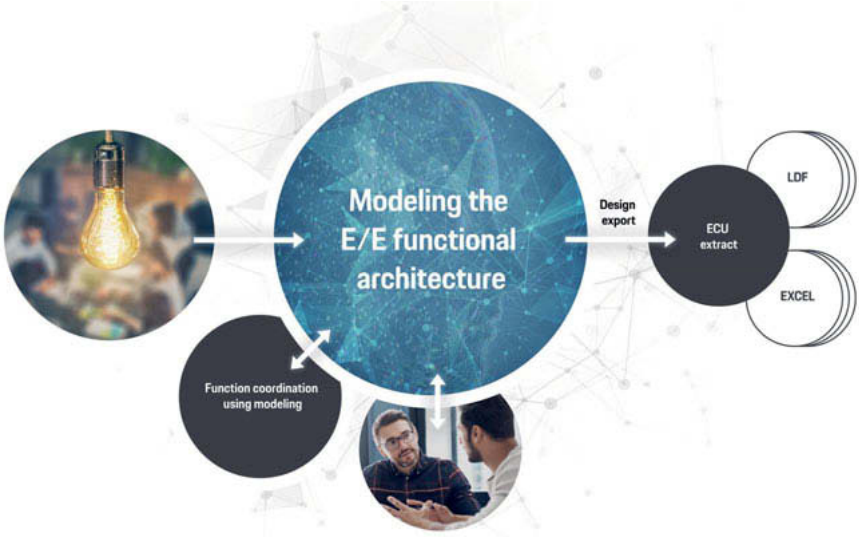


Fig. 6: Tool chain at Porsche for the technological implementation of the service orientation

4.3 Organizational implementation of the service orientation at Porsche

Since the service orientation represents a paradigm shift in the development process, as mentioned above, Porsche has defined different organizational success factors for the understanding and rollout of the service orientation.

They include:

- Close intermeshing of the network data definition, E/E architecture development, and the specialist departments that develop the functions

- Central requirements management for the service-oriented communication in the data definition as a consequence of the model-based architecture emergence process
- Extensive concepts for training, marketing, and communications
- Support and help desk availability to provide technical support to the specialist departments

4.4 Target vision at Porsche

As already mentioned in the previous chapters, close integration of the specialist departments that develop the functions is a key success factor in the service oriented development process.

In the end, the specialist departments should be capable of describing the services themselves, using the provided tool chain, and offering them at the service marketplace.

The contact persons from the respective specialist departments will accompany the service development largely autonomously and help to spread the new development paradigm through their function as multipliers.

An incentive system can be helpful here to establish the service mindset in the specialist departments.

As an interdisciplinary function, the data definition will be responsible for describing the service interfaces together with the specialist departments.

Providing suitable training courses, communication documents, and support offerings is another interdisciplinary responsibility.

The expected result is that the initial expense for developing the services in the specialist departments and in interdisciplinary work will be higher than under a strict functional orientation. These higher initial costs will be more than compensated for during the further development and change process, however, due to the expected reusability of the services.

4.5

Rollout in the development departments

Since the implementation and dissemination of the service orientation at Porsche involve interdisciplinary areas and functional areas, the rollout is being accompanied by a communication and marketing concept. An external marketing company was engaged for support in these areas.

Central goals of the concept:

- Internalize the function-oriented and service-oriented development approach as the success-critical mindset in all specialist departments, starting already with the brainstorming process for new functions
- Emphasize the advantages and benefits of the service orientation
- Create transparency regarding all information and training content, as well as for the applicable processes, rule sets, and tools
- Provide information in particular to the multipliers in the individual specialist departments
- Abstract the complex technical content to the key new features and emphasize the central paradigm shift
- Produce regularly available information events, such as training courses, webinars, posters, flyers, intranet, videos, and so on

4.6 Partnering

Due to the extreme novelty of service orientation in the automotive sector, collaboration with strong development partners is a crucial success factor.

Partners who have strong competencies in the area of software and platform development and can provide references for the successful implementation and support of the service orientation are especially relevant.

The focus of this partnering lies on sharing knowledge and understanding of the counterpart's demands and working methods. On one hand, automobile manufacturers like Porsche can learn from the software and communications industry how to apply established technologies and mechanisms in cars and realize megatrends there. On the other hand, the IT companies benefit from access to a car manufacturer's market and customers.

A key aspect of partnering results from function-oriented development and from the separation of hardware and software development.

Where previous models involved industrializing and commercializing functional developments in the form of controllers, the development partners now have to adapt to completely new business models.
In the future, the software product itself will be the product that is marketed with the OEM or the customer – independently of which hardware components are involved. Particularly for this business case, it is essential for OEMs to provide suitable APIs that partners or other third-party vendors can use to program their applications. The OEM share will lie primarily in application integration in this case.

The better the OEMs and partners reach understanding on API descriptions, thus creating an open, configurable interface, the greater success this model will have.

A win-win situation will arise for both sides if the OEM can benefit from other partners' dynamic development and competencies and they, in turn, get the opportunity to make their products accessible to a large group of customers.

5 Conclusion and summary

In conclusion, we can ascertain the following: When pursued consistently, a function-oriented development approach offers enormous opportunities for car manufacturers to get a handle on the functional requirements and complexity of ever-shorter development cycles.

The additional step toward service orientation structures the function-oriented development, by defining reusable subfunctions as services and making them available to other functions.

This development approach, which is new for automotive manufacturers, requires the use of powerful, adapted processes and tools. In the opinion of the authors, however, this aspect is already well-managed.

The greatest challenge in implementing service orientation is the service-oriented mindset of the individuals involved – and this is the true game changer.

The idea of creating reusable services that others can use and enhance with little effort requires a service mindset in the organization in the literal sense.

Opening up to external partners will also make it possible to capture entirely new business models. The development and sale of hardware components will have a diminishing effect on differentiating sales. Instead, software functions that are created based on defined APIs will open up new revenue models for both sides – OEM and partner. Collaboration and capturing synergies will continue to increase in importance in the future.

This article describes the basic idea of functional orientation, as well as its combination with service orientation.

It does so in particular by describing the technological, organizational, and process-based implementation at Porsche. It also describes the special features and importance of partnerships for using functional and service orientation as an opportunity for creating new collaboration models for both sides.

Vehicular RF Architectures

Managing integration of next generation automotive wireless systems

Thomas Zipper, Continental Automotive GmbH, Regensburg;
Robert Gee, Continental Automotive Japan KK, Yokohama, Japan

Abstract

As an increasing number of wireless connectivity technologies are integrated into vehicles including cellular, Wi-Fi, Bluetooth, broadcast, GNSS, and V2X, the larger number of antennae and related electronics are demanding, and the integration becomes more challenging. This is primarily because in-vehicle use cases are becoming richer, with quality of service requirements that may necessitate different connectivity solutions within the vehicle.

This paper will explore the different wireless services, their properties, and potential automotive applications. Particular focus will be placed on 5G cellular and possible MIMO configurations with the introduction of an Intelligent Antenna Module (IAM) for overall connectivity integration.

Introduction

Various wireless technologies can be found in today's vehicles starting with AM/FM for broadcast radio which has been available for several decades, to the upcoming cellular technology 5G. Furthermore, the connected car offers various applications, such as emergency call (E-Call), software flashing Over The Air (OTA), Stolen Vehicle Tracking (SVT), Remote Vehicle Control (RVC), and applications based on Vehicle to Everything (V2X). The market trend shows that ever more cars will be connected in the future, resulting in a CAGR (Compound Annual Growth Rate) of cars with network connectivity of 22%, and a total of 70% of vehicles sold worldwide after 2024 should be "Connected-Cars" [1]. This growth requires an optimization of the connectivity implementation. In addition, new technologies, higher frequencies, new and advanced use-cases, and changes to overall vehicle properties will all require changes to the wireless architecture. This is not to mention ADAS related wireless technologies such as Radar, Lidar, and others, which will not be addressed within the scope of this paper.

Driving factors for Connectivity

Ongoing and future market trends are turning connectivity into a de facto standard feature for light passenger vehicles. One of those trends is the legislated, mandatory emergency call in Europe (eCall), bringing basic emergency calling functionality to all new type-approved vehicles starting in April 2018 – it is a dedicated feature to save lives, even to the extent that the call for emergency assistance can be triggered by crash detection and airbag deployment, despite the driver being unable to act themselves. Another important trend is software flashing over the air, paired with remote diagnostics, not only to upgrade vehicle features or fix issues, but also to provide continual cybersecurity protection. From the commercial perspective, vehicle connectivity allows vehicle manufacturers and, in the future, perhaps other third parties, to offer consumers new services and applications, creating a new consumer market in the evolving passenger vehicle market. As Automated Driving (AD) vehicles become prevalent, the interior of the vehicle, like the interiors of many airplanes, buses, and trains today, will become a information and entertainment hub as vehicle occupants will eventually not need to be concerned with controlling the movement of the vehicle.

But the upcoming generation of AD and highly automated driving (HAD) vehicles will also create their own new requirements on the connectivity capabilities of the vehicle, demanding additional and enhanced wireless services, including improved and potentially guaranteed Quality of Service (QoS) for safety-related driving functions. The high speeds and low latency of 5G and future wireless technologies, together with V2X either based on cellular (C-V2X) or DSRC will be pivotal to support the safety, efficiency, and comfort of highly automated driving. These same V2X services are also not standalone, as good driving is a precision art, and precision by the machine will require precise and up-to-date knowledge of its environment and its exact position within that environment. Therefore high precision positioning and dynamic, high-definition (HD) maps such as eHorizon (Electronic Horizon) will be necessary to consider and further influence the wireless data connectivity considerations of the vehicle.

Different Wireless Services and Status

The following discussion will briefly touch on the primary wireless technologies found in today's or next generation vehicles and will consider the wireless properties.

For nearly 90 years, terrestrial broadcast has been one of the oldest services available in vehicles, and survives nearly unchanged in many regions today, such as AM and FM radio in the United States. But this is slowly changing, sometimes by replacement of technologies by digital

communications, such as DAB in Europe, and sometimes by addition of digital services to the old analog broadcast services (for example, HD Radio in the USA, which does not replace AM and FM radio). New services such as DAB+ or CDR (China Digital Radio) require changes and a broader coverage of reception technologies. The most promising solution would be to move to a Software-Defined-Radio (SDR) approach to enable support for the different radio standards [1]. In addition, satellite based SDARS (Satellite Digital Audio Radio Service) such as SiriusXM is available for 48 of the 50 states in the United States and part of southern Canada. Compared to consumer electronics, automotive broadcast reception technology is rather more advanced, and not only for the differing geographic standards. Another factor might be the comparatively stationary consumer electronics usage, compared with automotive in which receive diversity and higher sensitivity are needed.

Furthermore, access services for remote keyless entry and keyless entry/ keyless go are becoming more popular today and are common in mainstream vehicles.
TPMS (Tire Pressure Monitoring System) often uses capabilities of the access wireless system to communicate with the wheel units. For these functions, proprietary implementations operating in the ISM (Industrial-Scientific-Medical) frequency band are usually found. The frequency range for ISM and access implementation starts at 315 MHz and goes up 915 MHz.
The security and accuracy acceptance level has been very high for current implementations, but the new PAK (Phone As Key) concept is changing the setup. From the classical ISM band implementations, next generation access may use Bluetooth Low Energy (BLE) and Ultra-Wide-Band (UWB) as the base technologies. For tire pressure monitoring BLE as wireless technology type is also being discussed.
With PAK, consumer electronics and automotive access systems will be interconnected. In this instance, the technologies, consumer devices and automotive access systems, are complementary in that each technology exists separately without a deep level of common integration, but rather a common interface is defined.

Positioning by GNSS has also been available for over 30 years, but also here there will be upcoming changes. Currently, positioning is achieved using a combination of satellite families – GPS, GLONASS, Galileo, BeiDou, and potentially others in the future -- along with an optional correction service. This can be improved in various ways. First, the existing single band (L1) reception can be changed to multiband (L1 plus L2 or L5). This would provide additional robustness against atmospheric disturbances and distortions. In addition, position accuracy can be improved to the decimetre or centimetre range using technologies such as Real-Time-

Kinematic (RTK), Precise-Point-Positioning (PPP), or a combination. Each would require an additional data link to a correction service. But compared to consumer electronics, positioning is different for automotive. One reason is that for automotive, the position is typically enhanced by dead-reckoning algorithms utilizing additional car sensor information, such as gyro data, wheel position pulses, and speed information. It should also be noted that for future automated driving solutions, the position generation function may be required to support higher automotive safety levels, such as the Automotive Safety Integrity Levels (ASIL) defined under ISO 26262.

Finally, it should be noted that V2X will not only require high positioning accuracy for only the single vehicle but will also include requirements for high relative positioning accuracy between the vehicle and surrounding dynamic objects such as other vehicles and VRUs (vulnerable road users, including pedestrians, bicyclists, motorcyclists, and persons with disabilities). While part of the relative positioning requirements will be addressed by on-board vehicle sensors such as cameras, Radar, and Lidar, wireless technologies become important with the capability to “see” around obstructions and to provide data at further distances. This leads to the possibility of additional wireless services being added in the future, meaning additional antennae and transceivers, to address the different weight, size, and power consumption requirements for devices carried by VRUs.

Local device connectivity is done via Wi-Fi. For Wi-Fi, improvements to the standard are regular releases to provide higher throughput, such as 802.11ac or 802.11ax, in combination with an enhanced antenna system with Multi-Input-Multi-Output (MIMO) configuration. Furthermore, the use of this wireless technology is broadening from interior connectivity (such as for a mobile hotspot) to exterior connectivity for various applications, such as connectivity between the vehicle and the home, connections for inventory management and software updates at a car dealership, and connections between the vehicle and factory equipment for tracking, software installations, and testing while the vehicle is on the production line. Wi-Fi for automotive is typically similar to the consumer electronics version, but with some differences in the combination with BLE, quiescent current requirements, operating temperature, and reliability requirements.

Of primary importance in vehicles is the cellular connectivity, currently launching in new vehicles with 4G/LTE and 5G. Cellular connectivity is the de facto standard for car-to-network con-

nectivity today and will soon be approaching implementation in more than half of all new vehicles sold worldwide. LTE is the most common cellular technology and will likely remain so for the next several years as 5G is gradually deployed over each of the geographic regions.
5G includes a broad range of promises and service types, such as the next generation New Radio (NR), that includes sub 6 gigahertz frequencies (FR1, Frequency Range 1) and the high frequency mmWave (FR2), above 24 GHz. There are also technology enhancements related to Cellular V2X (C-V2X), particularly focused on vehicle-to-vehicle communications, and low-power IoT (Internet of Things). Apart from benefits and changes to the cellular infrastructure, automotive will initially benefit primarily from the latency improvements and C-V2X extensions included in next generation 5G versions as Release 16. The low latency part is described by the Ultra-Reliable Low Latency Communications URLLC feature for the first 5G version (Release 15), with improvements in the further releases. There are also discussions related to safety and quality of service for cellular networks as it may pertain to future automated driving vehicles that are anticipated to use connectivity as a key sensor.

The throughput enhancements in 5G would be realized by the new mmWave frequency bands for which data rates in the range of 10 Gbit/s are planned. However, research is still ongoing regarding the usability and business benefits of such high speeds for vehicle connectivity, as it is not the vehicle-to-base station connection that is under question, but rather the need and capability to process such high amounts of data. Nevertheless, the speed of even the initial release of 5G (Release 15) shows great promise to enable vehicles to share on-board sensor data, such as from forward-looking cameras, to eliminate blind spots and provide earlier warning capabilities such as for vehicles in the back of a queue and without a line-of-sight to an obstacle ahead that may require sudden braking. It is the promise of such safety benefits that are driving the research and need into the inclusion of advanced cellular capabilities on vehicles.

And as is typical for automotive, the cellular implementations for next generation cellular in vehicles must take into account the increased requirements for temperature, vibration, reliability, and functionality compared to consumer device implementations.

Applications and Requirements for Automotive Wireless

Implementation of wireless technologies in vehicles must therefore include existing wireless services, the evolution of those existing services, and the addition of the new technologies, all in a cost-effective, durable, and compatible way. It must further consider the applications for

which those services will be used, and in particular the effects on angle, range of coverage, reliability, and safety.

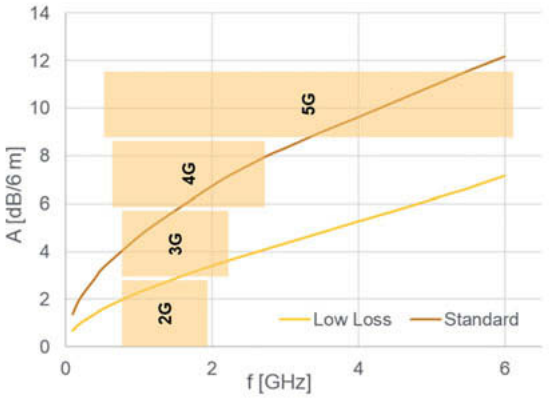
For vehicle access functions, different technologies can be used for communication and localization. For communication in the case of BLE, a fast and early connection is important to provide the best reaction times for keyless entry. Therefore, an antenna position on the rooftop is preferred to enable 360° coverage and to improve range of this low-power technology. Furthermore, cable losses need to be minimized. In addition, both exterior and interior communication is required for this function, suggesting the implementation of the BLE transceiver together with exterior and interior antennae near the peak of the roof.

A similar situation occurs for Wi-Fi uses-cases which also require interior and exterior communication. For interior use, the application could be as a standard hotspot, providing Internet access to mobile devices inside the vehicle by utilizing the vehicle's cellular system and more optimally-placed external rooftop antennae. For exterior Wi-Fi usage, there are additional possibilities which may affect the design. The vehicle may act as Wi-Fi device in workshop mode which would allow access to the diagnostic data or allow a software update, which might not have high requirements on the external antenna configuration as such a vehicle workshop could be anticipated to be an enclosed and controlled environment. In another use case, network offloading using an external hotspot connection would be a possibility, which would trigger higher antenna performance requirements, since the external environment is uncontrolled from the perspective that the external hotspot could be at any angle and distance from the vehicle. And as for BLE, the best external coverage is important, so cable losses should be avoided; therefore, the Wi-Fi antenna and transceiver would ideally be implemented at the roof top.

For better energy efficiency or optimized electric vehicle range, due to reduced air conditioning usage, IR (Infra-Red) reflective auto glass is discussed and already implemented in certain instances. Assuming all windows are equipped with such IR protection based on reflective metallization, wireless signal transmissions between the interior of the vehicle and external transceivers will be extremely attenuated. Considering BLE and Wi-Fi services, for which both external and internal communications may be desired, such windows will determine whether two separate transceivers with corresponding antennae. Alternatively, the roof-top implementation with a single transceiver using dual antennae can also support such in-vehicle and external communication.

Another aspect of the application is the high potential data speed of next generation connectivity, which will eventually enable hundreds of megabits or even gigabits per second. For this, the primary question will be to determine the necessary data sources or sinks for such high-speed data, which in turn would be determined by the in-vehicle use case. On the other hand, the connectivity paths within the vehicle are more obvious: one main data channel in the car infrastructure will be Ethernet, available in a variety of speeds and supporting different vehicle architectures, including redundant pathways, backplanes, and switches if needed for safety applications. Also, in-vehicle Wi-Fi can be a main path for the connectivity speeds offered by 5G or Advanced LTE, although the source and sink endpoints are still to be considered. This leads to the potential architecture that the cellular modem and Wi-Fi transceiver might be tightly coupled and used in combination to enable flexibility for in-vehicle and extra-vehicle communications.

Maybe one of the most important changes in wireless is that cellular, V2X and Wi-Fi will operate at higher frequencies compared to today's configurations. New cellular frequency bands can go up to 6 GHz, even for the lower FR1 range. Also, V2X will operate almost world-wide at 5.9 GHz independent of whether the implementation selected is Cellular V2X or DSRC. In addition, Wi-Fi may also use 5 GHz for automotive applications, and there are ongoing discussions by the Federal Communications Commission in the USA for additional bands between 5 and 6 GHz. However, current vehicle RF architectures use coax cables with signal attenuation that makes implementation difficult or impossible at 5 GHz and above.



Finally, for technologies such as LTE, 5G, and Wi-Fi, MIMO schemes are used. Furthermore, higher order MIMO techniques are increasing from 2x2 (two antennae) to 4x4 (four antennae) and will be implemented for automotive applications.

5G Network (4x4 MIMO / 1x4 Diversity) [3]

Multiple-Input-Multiple-Output or MIMO is used with Spatial Multiplexing (SM) with the goal to increase capacity. The data is divided into separate streams and each stream is transmitted independently via multiple antennas - using the same resources in both frequency and time. Also, MIMO can be used for diversity with the goal to increase reliability. The same data would be transmitted (or received) redundantly over multiple antennas (e.g. transmit diversity and receive diversity). Spatial Multiplexing may also be combined with diversity coding.

With the introduction of 5G, MIMO communication will become mandatory. According to 3GGP release 15, there will be an obligation for standard equipment to use a 2x2 receive MIMO scheme for frequency bands below 2490 MHz and a 4x4 receive MIMO scheme for higher frequency bands. However, for automotive implementations an exemption to use 2x2 MIMO has been defined.

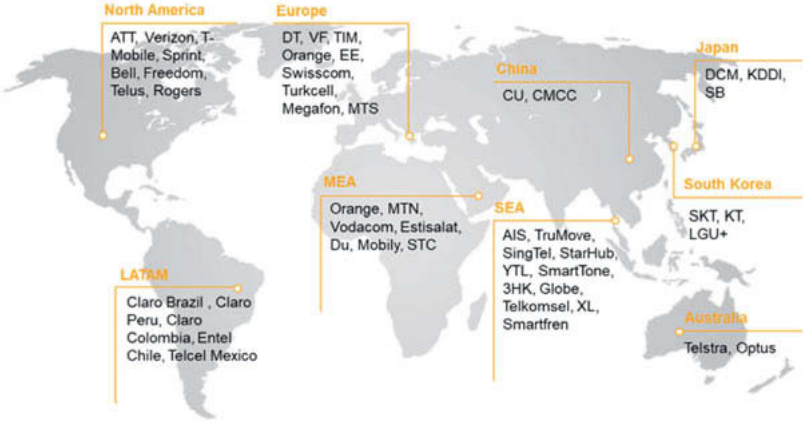


Fig. 1: Mobile Network Operators (MNOs) with 4x4 MIMO

Table 1: Deployment status of 4x4 MIMO / 5G NR

	4G LTE 4x4 MIMO	5G NR
EU	9 MNO / 4x4 / 2017	8MNO / 3.5GHz+mmW / 2019
US	8 MNO / 4x4 / 2017	4MNO / 2.6GHz+mmW / 2019
CN	2 MNO / 4x4 / 2017-2018	3MNO / 2.6-3.5-4.9GHz/ 2019
JP	3 MNO / 4x4 / 2017-2018	4MNO / 3.7-4.5GHz / 2019

※ number of MNOs / Specification / Start of service

One usage scenario considered by 5G networks is URLLC. Multi-antenna architectures deployed in vehicles are key elements in this scenario, as they enable robust connections at cell boundaries via diversity schemes or save signal power under good reception conditions.

The performance of a multiple-antenna system depends on the location of the antenna elements and the way they are connected to the modem [5]. Compact arrangements as in a shark fin housing on top of the vehicle roof may suffer from higher correlation between the antennas. Distributed arrangements, however, may suffer from cable losses. An alternative could be solutions consisting of compact arrangements of some of the antenna elements and distributed arrangements of the remaining elements. Yet this leads to imbalances between the antennas, what can be considered as a loss in efficiency, causing performance degradation in terms of throughput and reliability.

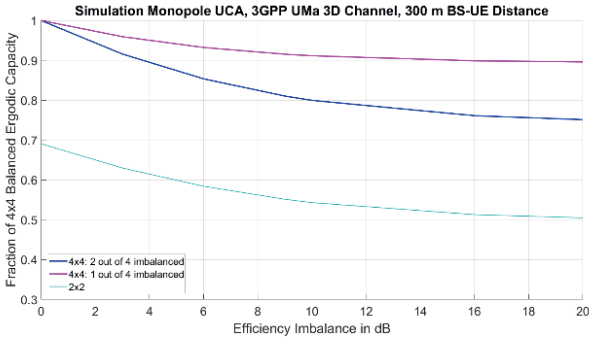


Fig. 2: Channel capacity plotted against antenna imbalance in an urban macro-cell scenario

Example Next Generation Wireless Architecture as Intelligent Antenna Module [3]

An Intelligent Antenna Module (IAM) is the integration of antennae, receivers, and transceivers such as a cellular modem and a car network interface such as Ethernet or CAN. As a result, two components, for instance the shark fin antenna array and the telematics control unit (TCU), are merged into one unit. The mechanical packaging may also be a flat module (hidden by being flush or nearly flush with the external roof) or other designs. For optimal reception of satellite services such as GNSS and satellite radio, the antenna should be located on the vehicle roof.

In traditional architectures, antennae and the TCU can be very far apart. The RF signal connection is via an RF coaxial cable. Under these circumstances, the RF coaxial cable length between shark fin (commonly mounted on the rear of the roof) and TCU (which may be mounted in the boot, near the centre console, or within the cockpit module) may be 6 meters. Depending on the cable length and the signal frequency there would be deterioration of the RF signal that is sent and received. There are possible countermeasures to compensate for this effect. However, such countermeasures increase material costs and engineering effort. In addition, it is difficult to restore the SNR (Signal to Noise Ratio) as it degrades. Since the IAM does not use RF coaxial cable, there is minimal degradation of the RF signal, and the SNR can be optimized.

Fig. 4: Comparison between classical and IAM architecture

The IAM systems offer several key benefits, most of which are due to the removal of RF coaxial cables. In classical architectures, a coaxial cable is needed for each antenna element. As explained in future connectivity systems based on 5G and V2X, there will be an increasing number of antenna elements such as 4x4 MIMO (4 antennae), V2X (2 antennae, if front and rear are necessary), GNSS (up to 2 antennae, L1 plus L2 or L5). and broadcast (1 to 2 antennae, SDARS). Therefore, many RF coaxial cables are needed for a future connectivity system in a vehicle. This means high cost, increased weight and wiring design pose major problems for

the classic architecture. The IAM can solve this problem because all antenna elements and the communication unit are directly connected via an RF connector and not coaxial cables. Cost and weight reduction through simple cabling design are achieved because RF coaxial cables are expensive and the cost of each cable for 5G frequency bands can be higher. In addition, RF coaxial cabling is much heavier than the UTP (Unshielded Twist Pair) cable used for the Ethernet connection to an in-vehicle network. Therefore, the large number of RF coaxial cables would be minimized by replacing them with an UTP connected IAM. Performance improvements can also be achieved by lowering RF signal losses due to reduced or eliminated RF coaxial cabling lengths.

Conclusion

The next generation automotive wireless system will be accompanied by various changes and challenges. Many new wireless functions are being implemented inside vehicles and RF performance requirements are becoming more demanding. We must also consider the most suitable antenna locations with the best combination of good RF performance and optimized system cost. As described, the IAM is a possible solution and can be an essential component for the automotive industry in the “Connected Car” era.

[1] Strategy Analytics, Q2 2018

[2] Advanced Digital Radio: HD Radio, DRM, DAB & CDR, September 2015, GatesAir, Tim Anderson

[3] Contribution of Intelligent Antenna Module to 5G 4x4 MIMO System Performance, Hiroyuki Matsumoto, JSAE Annual Congress on May, 2019

[4] 3GPP Release 15: https://www.3gpp.org/release-15

[5] Vehicular MIMO Antenna Investigations in LTE Networks, Dipl.-Ing. Thomas Lankes, Kathrein – Werke KG, Rosenheim, Dipl.-Ing. (FH) Frank Mierke, Kathrein – Werke KG, Rosenheim, Dipl.-Ing. Peter Turban, Continental, Regensburg

Going from an Electronic Unit Centric Development to Application Software Centric Requires a Different Architecture Mindset in Automotive

Anders Magnuson,
Volvo Group Truck Technology, Gothenburg, Sweden

Abstract

There are many incentives for a higher degree of automation for commercial vehicles to gain productivity, while at the same time facing very different demands on final transport applications. In addition, the environmental impact drives the need to reduce fossil fuel usage by introducing electrified torque generation, which could be distributed over several vehicle units in a vehicle combination. Electronics and especially software play a fundamental role for commercial vehicles in order to achieve energy/power balancing, assist a driver to manually operate vehicles effectively in combination with various degrees of automation and doing that dependable in different transport applications. Although the overall design thinking in the commercial vehicle industry is still very much oriented towards a geometric perspective and thus physical modules, which for software means binaries related to physical electronic boxes (ECUs) – classical ECU-oriented mindset. In this paper a supplementary perspective is added to the traditional geometry-oriented perspective – a functionality perspective, which facilitates reasoning about functionality and thus application software. The paper proposes a reference architecture that is based on four horizontal and two vertical layering of functionality.

Introduction

One of the cornerstones in the automotive industry has been and still is to achieve large scale reuse of manufactured entities in order to provide the market with mass-produced cost efficient vehicles. Thus the overall design thinking, at least within AB Volvo, is characterized by modularization of the products viewing them from a geometric perspective and thus geometric modules and how geometric modules are wired together are in the forefront. All these “modules“ are formed in a “platform“ which can be looked upon as a gigantic shopping bag full of pieces. At Volvo Group Trucks Technology (VGTT) this is known as the Vehicle Module Structure (VMS) and Common Architecture & Shared Technologies (CAST) highlighting generic (geometric) vehicle modules as in Fig. 1 and geometric interfaces. Transferring this perspective

into the software landscape; this is similar to a physical view [1], which nowadays is commonly called deployment view.

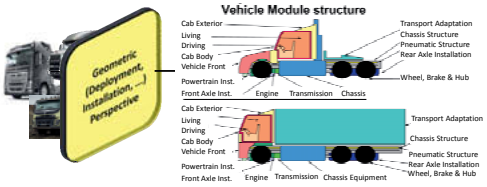


Fig. 1: Main perspective and mindset is geometry-oriented.

This physical thinking has been transferred into the area of electronics where Electric and Electronic Engineering has been focus and still is on where to place mechanic boxes enclosing some electronics and then wire these mechanic enclosures together. These mechanic enclosures are commonly known as Electronic Control Units (ECUs). In early days this wiring was about having dedicated wires going back and forth between ECUs. When these dedicated wires were replaced by communication technologies the focus has continued on wiring the ECUs together, now via Controller Area Network (CAN) and Local Interconnect Network (LIN) links. If you ask for the EE Architecture an automotive company most engineers would provide you with a PowerPoint showing how the ECUs are connected via physical data links as in Fig. 2, but few would present anything similar for the software.

However, already long time before electronics were introduced, when monitoring and control logic was achieved through relays, there has been a structural element known as Electrical Distribution System (EDS). EDS had a focus on how these relays were wired together and the wires carries signals. What has been experienced during the years is that an EDS nor the network topologies fit properly anywhere in a geometric-centric modularization. They so to say give another kind of perspective! Furthermore, the geometric and ECU centric way of looking at the solution has led to that when we are talking about software the focus has been on the ECUs and the binaries that are flashed into their memory.

Also, traditional focus on wires has led to that automotive electric system engineering as a discipline focus on wiring these binaries together via a Signal Data

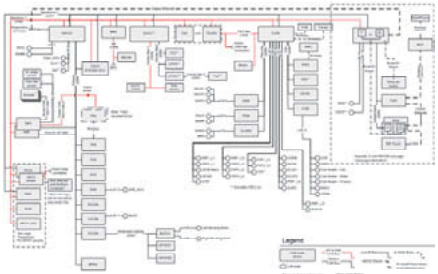


Fig. 2: Network topology is what most look upon as THE architecture in the automotive.

Base (SDB). The SDB and EDS are rather similar, where EDS focus on packaging hard wires (signals) into harnesses and SDB focus on packaging soft wires (signals) into frames. Also, the SDB is hard to find in a geometric-centric structure. The software elements that really does something, in Classic AUTOSAR - the Software Components [8], are not really visible and treated, but it is primarily the binaries that are installed in the assembly line that are counted and registered in our product data management tools. All other kind of “systems” are too a large invisible in our product life cycle management tool, e.g. a “system” as in Fig. 2 is not officially released, the description is, but the system with all ECUs is not released as a “system”.

System complexity in automotive is increasing

Complexity of a system is rather subjective and is impacted by many things. Already 1986 Fredrik Brooks reasoned about essential and accidental complexity [2] and such as people skill, organization, tools and processes and the system itself are all contributing to the overall complexity. Essential system complexity is a result of “Complex behavior that arises from the inter-relationship, interaction, and interconnectivity of elements within a system and between a system and its environment” [3]. Looking at Fig. 2 the complexity looks quite moderate, but what is missing in this picture is the application software. So when the network topology is supplemented with application software structure, as in Fig. 3, the complexity increases. Thus the traditional view on EE Architecture Fig. 2 gives a rather false picture of the truth! System wide complexity has moved from electronics and wires into the software! What we have experienced is that we can no longer proceed with just the ECU-thinking as system behavior is not clear from the properties of its individual parts.

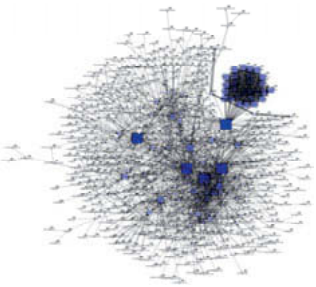


Fig. 3: The complexity increase that comes with interacting software elements.

Continuous software increase to facilitate automation, electrification and connectivity

Nowadays, it is when the application software comes into focus the complexity shows up. That brings us into the topic of where the changes will comes. For example, new legislations on NO_x and CO_2, zero emission zones and noise zones will be enabled through electrification of “powertrain devices“ but we must also monitor and control the operation these together with

various other high power auxiliaries on trucks such as air fans and air compressors, which thus also needs to be electrified. Thus for example “brake blending” suddenly becomes an energy balancing topic rather than a plain braking topic reducing brake pad wear. From an overall energy management perspective, usage of different energy sources and buffers requires monitoring as well as prediction to minimize the energy consumption for a certain transport mission. This is related to the overall vehicle mission as a whole rather than to a particular “mechanic device” device and thus such functionality should not be part of the device. This overall coordination only be handled through application software!

Software has become a major enabler for improving old features as well as providing new features in the automotive industry. Many of these features are not directly linked to the geometric modularization as in Fig. 1. Cars and trucks have continuously evolved with software enabled features in a relatively moderate pace over the last three decades. It is not only high end vehicles but also low end vehicles that have a quite impressive amount of software in order to manage things like a vehicle’s perimeter, seat adjustments, acceleration, braking, etc. It has been estimated that more than 80 percent of new vehicle innovations are enabled through software [4]. It is important to highlight that not even old software entities such as a Cruise Speed Controller or Cruise Distance Controller are linked to geometric modules as in Fig. 1 like a combustion engine and its associated Engine Control Module (ECM) which today host the Cruise Speed Controller according to SAE J1939/71 [5], [6] (there is an implicit deployment built into this standard). The trend in the Volvo Group is that the pace of innovations through software is accelerating. Also, it will be hard for standards such as SAE J1939 to keep up with this acceleration as it restricts the solution space.

As in enterprises, software is a major contributor to automation – replacement of human performed activities. This of course also goes for various levels of vehicle automation, where vehicle automation is synonymously with a huge amount of application software. As commercial vehicles are used in B2B operations, the interest in automation is perhaps of higher incentive than for cars as it contributes to operational margins. Roughly 1/3 of operation cost is related to having a human behind a steering wheel. The automation will also go hand in hand with an increased level of connectivity in order to operate logistics of unmanned vehicles, maintenance and in some case remotely drive a malfunctioning vehicle to get it into the roadside. It will be the application software that drives the need for powerful electronics, i.e. flexible and reconfigurable computers! [7]

The way of working with commercial vehicles is far from adapted to looking upon them as software intensive products or service providers, which vehicle automation, connectivity and

also electro mobility is about. The traditional geometry and deployment perspective is not feasible any longer!

Supplement the Deployment Centric Perspective with a Functionality Perspective

To achieve strategic large scale software reuse it is a necessity to apply a product line engineering approach [9], which sometimes is referred to as “platform development“. As commercial vehicles operates in a diverse set of transportation applications there is a need to build in many and different kinds of variation points (variability) to enable vehicle feature tailoring, while at the same time aim for reusability and changeability in such a product line. One can think of it as a “one software code branch only“ [10] with built in variability. To manage the transition that software is mainly about mechanical component control, e.g. engine, transmission, braking, to actual vehicle feature control, it is a must to not just talk about the physical modules but rather more abstract entities – some kind of entities of functionality. Therefore we are adding an additional perspective on the vehicles – here defined as the functionality viewpoint, Fig. 4 [1], also identified by [11] as a link between overall customer demand and physical structure. This will make a shift to focus on reasoning about and reuse of modules of functionality rather than “physical” modules – in this sense a product line is a set of modules of functionality shared across multiple end-user products [12]. The intention is that we look upon structural entities showing up in this perspective as “products” in a similar way as structural entities in a geometric viewpoint are treated as “products”.

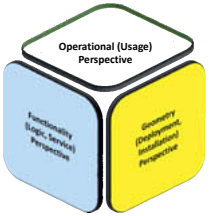


Fig. 4: Multiple viewpoints to deal with separated and unrelated concerns.

With higher and higher demands on dependable vehicle operation, driven by more advanced features but also by the functional safety standard ISO 26262 [13], there is a necessity to more clearly structure and through that separate different concerns such as different levels of criticality of functionality from each other in order to guarantee that the lower criticality elements cannot interfere with the functioning of the higher criticality elements.

Organizing the Functionality Perspective

So, *how to think when looking upon commercial vehicles from a functionality perspective?* In principle one has to take a full vehicle perspective on this and also include functionality handled through mechanics, pneumatics, electronics and not just software. Based on the thinking of

separation of concern the architecture approach made here is based on that functionality dealing with monitoring and control things is separated from functionality handled through electricity, diesel, mechanics, pneumatics, hydraulics, etc. as in Fig. 5. For example the combustion engine from a functionality perspective would then reside in the Torque Supply & Distribution System but when it is installed (deployed, packaged) it can be under the cab at the truck, in front of the cab in a conventional truck, in rear or at the center in a bus. So this thinking does not just support software. We thus at the highest level apply a layered approach. As can be seen in Fig. 5 there are dependencies from the Vehicle Monitoring & Control System (VMCS) to various "systems", which will mainly be handled by the Device Abstraction Layer introduced in Fig. 7. Furthermore, as the nature of the functionality dealt with in the VMCS is very different, ranging from converting an analogue value to a digital value, forwarding this digital value from a converter circuit to an application software entity where it might become a temperature, leads to that this is structured into three major "system" entities as in Fig. 6, which is a kind of layered architecture style [21]. As these are separated from each other the reasoning and focus in these also varies a lot. Another thing worth to highlight is that VMCS focus on single vehicles. However, to gather statistics of wear or energy consumption trends of for example all Volvo, Renault and Mack trucks, there is a need to add another layer on top as in Fig. 5, here called Vehicle Fleet Monitoring & Control System (not in the scope of this paper).

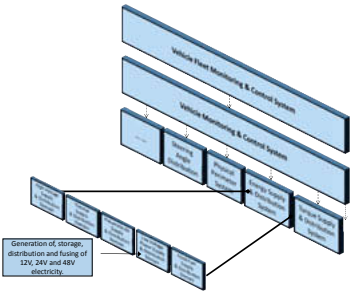


Fig. 5: Separate the things that are to be monitored and controlled from the one monitoring and controlling them

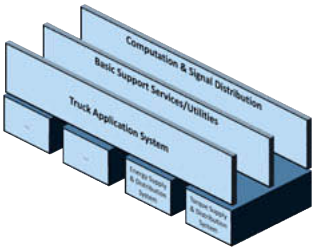


Fig. 6: The Vehicle Monitoring and Control System is divided into three major blocks of functionality.

Computation & Signal Distribution System: the focus is on electronic functionality interfacing various actuators and sensors that are part of the device functionality such as interfacing

a Fan Motor, an Inlet Air Pressure Sensor, and Brake Pad Wear Sensor etc. It also focuses on processing and memory capacity as well as network structures its performance. We can say that this system is the classical way of looking at an “EE system“ – the ECUs and their network connections as in Fig. 2. That is the diagram in Fig. 2 is a “document” that only describe the internal design of this system and that released when the system and all its content is released.

Vehicle Application System: the focus is on managing entities that owns and deals with information about the vehicle and thus used to monitor and control the operation of the vehicle. These entities are relevant to talk about in the application domain such as the *inlet air* and its temperature, humidity and mass flow (speed) or a *climate comfort* service. It is this system that holds all application software functionality all the way down to source code modules.

Basic Support Services/Utilities System: The world in-between these is some kind of middleware functionality that bridges these two worlds, where AUTOSAR Classic Platform [8] is just one such middleware framework, Linux another one. Graphic Engines, Voice Interprets also belongs to this system.

Except for the nature of the functionality another really important reasoning for this separation is that cycle-time for developing functionality in the Computation & Signal Distribution is very different from the one in the Vehicle Application System. Furthermore, as the hype around agile/lean based development process frameworks like SAFe [14] also has reached the automotive domain with hope for more software-based features both faster and more continuously developed, integrated and distributed to customers, there is really a necessity to even more strongly make this separation happen at the top level.

Vehicle Application System - layer application functionality

By work and experiments conducted during 2009-2018 at Volvo GTT it has been concluded that there is a need for a slightly different layering than in [15], among other things the architecture presented herein has added clear separation of HMI and the ego vehicle’s environment and operational functionality has been divided into two layers. The

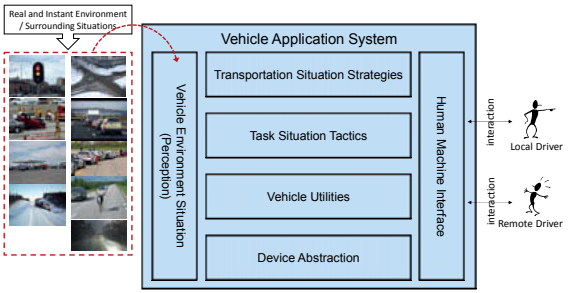


Fig. 7: Basic layering of the application software inside Vehicle Application System.

reference architecture presented in this paper proposes four horizontal layers as in Fig. 7 where two lower layers address “operational” functionality; one layer deals with tactic functionality; and one layer deals with strategic functionality, but also two vertical layers are added. Each horizontal layer raises the abstraction level from the “physical” functionality that is to be monitored and controlled. All with the perspective of a single vehicle, but the vehicle may at the vehicle utility layer be looked upon as two track model, as single track model in the task situation tactics layer and as a particle in the transportation strategies layer!

The architecture approach defined here combines a strict hierarchical style [16] with a heterarchical style [17] where all modules communicate with each other – it becomes a layered style. The idea of a layered style is to deal with the disadvantage of a strict hierarchy as it introduces inflexibility and long response chains but also the problems associated with a strict heterarchical style as it introduces many problematic couplings all over and lowers the possibilities to reuse and by that achieve a variable product line. A similar layered architecture for platooning feature of commercial heavy vehicles has been presented in [18], which also contains strategic, tactical and operational layers and it can be seen that the focus has been solely on platooning. In the presented approach here, the platooning planning is managed in Transportation Situation Strategies and the actual joining and leaving is taken care of in Task Situation Tactics, e.g. the “adaptive cruise controller“ is becoming a “platoon cruise controller“. In [18] it is not revealed how full automation is going to be approached nor how transition of transport automation which can include manual and automated driving.

Device Abstraction Layer: The overall thinking in this layer is that application software entities shall be representations, device abstractions (DA), of mechatronic devices such as a *Fuel Tank*, *Wheel*, *Clutch*, *Wheel Brake*, *Windshield Wiper*, etc. to *localize the knowledge about the characteristics of a particular device* as shown in Fig. 8.

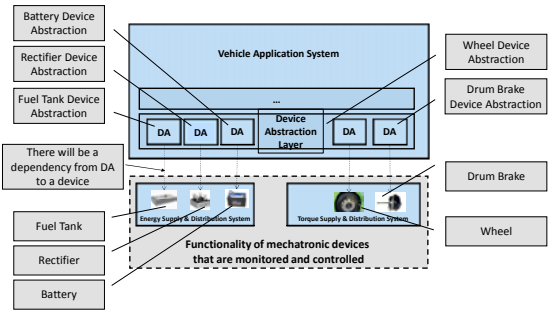


Fig. 8: Device abstractions are supposed to be representations of modules of the functionality that is monitored and/or actuated.

DAs are carriers of information elements that represent various properties of a mechatronic element that is to be monitored or controlled. For example *current tire pressure*, *current wheel speed*, and *current tire temperature*, and *nominal tire*

dimension are organized into a software entity representing the *Wheel*. And data such as *current wiper position*, *current wiper speed*, *wiper operation*, etc. is data that together forms a software entity representing a *windshield wiper*. The important here is that Das are looked upon as service providers and does not necessarily run on its own ECU. We are not there yet, but having such entities looked upon as individual deployable entities would ease a continuous evolution of our product line as well as ease upgrades in the fields. Furthermore, this thinking is also related to that there might be different variants of these devices such as a Disc Brake and a Drum Brake and as these share many properties it is good if they are defined in a single point, Friction Brake, as in Fig. 9. Unfortunately many design tools popular together with AUTOSAR such as Matlab/Simulink and DaVinci Developer do not support this kind of design. Instead it would be beneficial to apply the object-oriented inheritance mechanism for this.

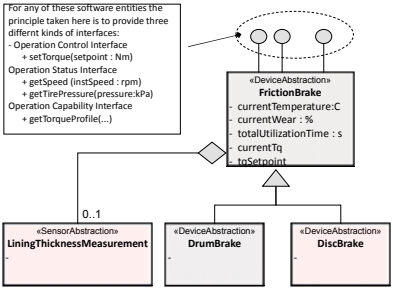


Fig. 9: A DA with two variants and three different types of interfaces.

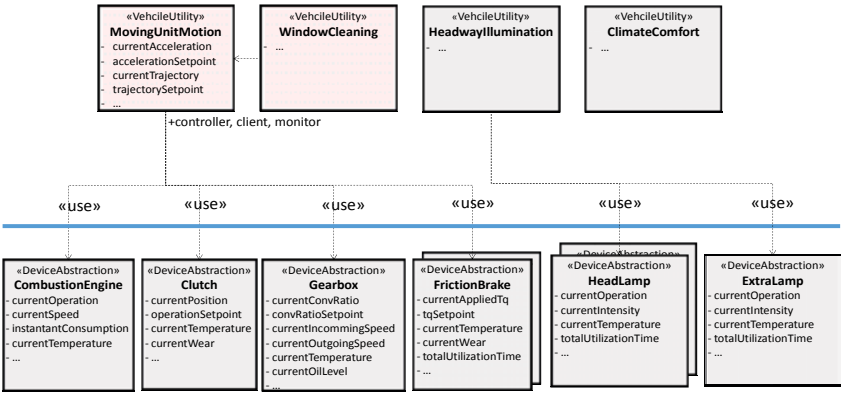


Fig. 10: Example of Vehicle Utilities that are acting as controller, clients and encapsulate coordination between different device abstractions.

Vehicle Utilities Layer: The purpose for this layer is first of all to raise the abstraction level from individual device abstractions into so something we call vehicle utilities (services). These shield away individual devices. A vehicle utility is defined as some kind of useful vehicle wide service that enables a client to perform one or many of its activities via an HMI in a manually operated truck or when automated by elements in the Task Situation Tactics Layer. Hence *a Vehicle Utility represents a goal experienced* by the consumer e.g. a Moving Unit Motion (for trucks there can be many units that form a vehicle combination), Power Situation, Headway Illumination, Window Cleaning, Climate Comfort etc. as in Fig. 10. When a Vehicle Utility is operational it is utilizing lower level DAs to achieve its desired goal and thus it will take on many different roles towards DAs such as controller, client, monitor, and coordinator and towards elements in the HMI or Task Situation Tactics layer Vehicle Utilities will works as servers.

What we have done is that this layer will in practice be replaced through a number of domain packages organizing a set of vehicle utilizes as seen in Fig. 11. We intend to look upon these

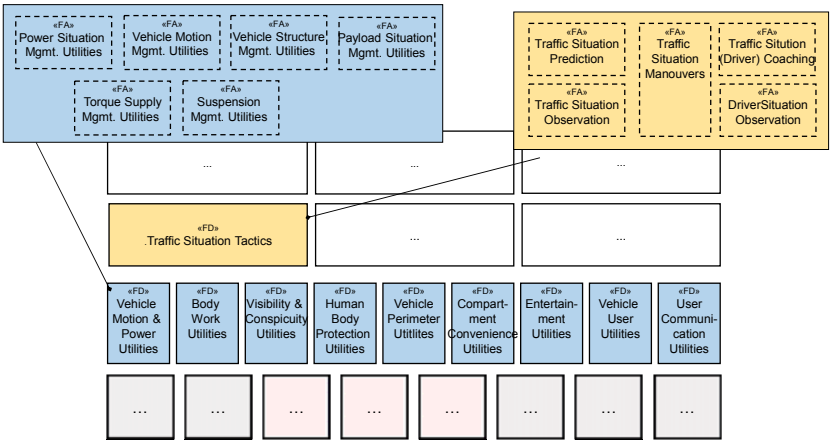


Fig. 11: In practice, a layer will be a set of application domains looked upon as “products”.

as “products“ that are released, maintained, evolved, etc. and thus they enable us to point out “product manager/owners“, although its internal elements might be distributed and executed in a number of ECUs. Furthermore, it is these domains that facilitate our organization to talk about the application software as the network topology and its ECUs has supported automotive organizations in the past. As there can be quite many software elements in these, these are

further decomposed into some intermediate products which we denote areas as seen to upper left in Fig. 11.

Task Situation Tactics Layer: When a human user performs a use case such as “transport payload from A to B” or “load/unload payload” is actually about the performance of a coordinated set of tasks in a given situation. In a situation where a physical user performs a task (a non-automated situation), the human user is acting as both a monitor and controller. The purpose of the task situation tactics layer is to enable introduction of more and more automation of these tasks and coordination of tasks traditionally performed by drivers and other local operators. In order to achieve this, it “consumes” services offered by various vehicle level utilities. So *basic entities that show up inside this layer are entities that represent tasks a user performs (today but not tomorrow)* and apply some tactics to perform these. For example, in the case of a driver in the control loop, the driver is actually acting as a *driving controller, speed controller*, a *trajectory controller* (steering), *headway sight controller*, *forward sight controller*, etc. and therefore such tasks shall also be represented as application software entities. Further breakdown of this layer into domains can be made such as *Traffic Situation Tactics* in Fig. 11. As seen in Fig. 11 and in [19] and [20] the *Traffic Situation Tactics* is broken down into smaller functionality areas (FA) such as *Traffic Situation Observation* (a local world model [16]), *Traffic Situation Prediction*., *Traffic Situation Maneuver (*plan, perform, and perform maneuver’s both L2 to L5 automation), Traffic situation (*Driver) Coaching* and *Driver Situation Observation*.

Transport Situation Strategies Layer: This layer raises the abstraction one level further, and instead of reasoning about it as a “physical” truck or user activities, it instead focuses on what kind of different transportation (mission) that can be offered (still in the single vehicle perspective) such as: How many containers and sizes are expected to be transported and assigned by a vehicle or gravel transport assignments, soil transport assignments, sand transport assignment, or frozen food transport assignment. This is about dealing with strategic decisions such as “optimize” for operation cost or delivery speed among assignments.

Ego vehicle's environment situation: Another kind of functionality that is different compared to the concerns addressed by the previously defined horizontal layers is the functionality that deals with creating a picture of an ego vehicle’s environment situation. In this architecture this is an orthogonal layer to the horizontal layers as seen in Fig. 7. For example a *Weather Situation*, *Traffic Jam Situation, Environment Traffic Situation* and a *Road Characteristics Situation* are software entities located herein and provides services to the horizontal layer. Thus they abstract away how that is gained from making of other more internal software entities, e.g. a Weather Situation entity can make use of detections in an image from a camera, clutter from a radar and/or detection from dedicated temperature sensors as well as communication with

another vehicle to get a proper picture of the complete weather situation in the ego vehicle's environment.

Human Machine Interface "layer" - think Model-View-Controller: Although automation is a hot topic for commercial vehicles the transition will take time and a human will need to interact with various kinds of functionality and that human, acting as a driver, which can be either locally or remotely as in Fig. 7. This is especially important to consider in product line that intend to enable trucks within the range from manual to full automation. The approach taken here is to clearly separate Human Machine Interface (HMI) as in Fig. 7, where View and Controller according to MVC pattern [21] resides herein. As the Volvo Group is dealing with multiple brands in its product portfolio this separation enables HMIs to look very different between the brands while maintaining stable core knowledge. This also opens up for a possibility to localize what kind of control a user can do and when that can be done as well as addresses the issues of managing multiple users to control from multiple places simultaneously and not having that issue rippled down into the core knowledge, e.g. having direct access buttons mounted e.g. in a door panel and in a handheld device as well as via soft buttons visible in an thin GUI app running on a smart phone. In order to deal with this a generic architecture component known as a User Input Controller (UIC) has been defined which is responsible for WHAT control is offered to a human user and WHEN that control is possible. The architecture also has a similar User Output Controller that provides feedbacks valid for a human user.

Think Object-/Component-/Service-Oriented instead of Function-Oriented

It is important to understand that any kind of software system such as a gearbox control system, navigation system, telecom base station or an order management system all deal with a massive amount of information and most of the times in real-time. Having this said, ownership of information that represents various kinds of states such as a *speed limit*, *max vehicle speed*, *current acceleration*, *current curvature*, *instantaneous fuel consumption*, *average power consumption*, etc. is essential. Also, *current vehicle acceleration* is completely different than *instantaneous fuel consumption per time unit* which is an issue for a combustion engine device abstraction and *instantaneous fuel consumption per travelling unit* (km or m) is different and the scope of a vehicle utility. Therefore the layered architecture is supplemented with the mindset of an object/component-oriented architecture style, as highlighted in Fig. 9 and Fig. 10, as the lowest level in software modularization within the layers. These objects are acting as service providers and service consumers. The intention is also that these will be individually deployable entities in a SOA-environment. This will in turn facilitate a scalable product line but

also ease agile self-going development teams having smaller deployable entities than deploying a big binary monolith as of today. Achieving a directed dependency, e.g. as depicted in Fig. 10, *requires a careful design of the interfaces at the end of the dependency arrow*. That is, the dependency is realized by Operation Interfaces of one or more public Objects residing in these domains, areas, units, etc. (a successive hierarchic structure) Thus we propose that software elements provide three different kinds of interfaces, as visualized in Fig. 10: Operation Control, Operation Capability (instantaneous capabilities), and Operation Status. This is very well supported by a component-based approach although current version of the component-based AUTOSAR framework does not have ownership of information at its heart. This is a quite big difference from current design paradigms in automotive which often favor the function-oriented paradigm focusing on algorithmic decomposition where data is flowing around among the functions as parameter passing or as global data.

Conclusion

It has been recognized that an architecture has major impact on the easiness to cooperate among teams. In 1997 it became clear that Microsoft divide development work in a way that mirrors the structure of its products, which helps teams, create products with logical and efficient design and with efficient groupings of people [22] and we conclude the same but all recognize that is really hard to establish such thinking a large global organization. But, it has also been recognized that the design of any system is significantly affected by the communications structure of the organization that develops it, i.e. any organization that designs a system will produce a design whose structure is a copy of the organization's communication structure [23] such internal line organization structure, structure of agile release trains, and various tiers of suppliers (Conway Law). Thus, going in this direction putting the application software in the forefront rather than the ECUs it will have large impact on many suppliers' business models and their current intellectual property investments as intellectual property very often lies in the application software rather than in the electronics as well as our internal organization and such transitions are painful. Thus, there is no need to debate that development of software functionality for commercial vehicles would be vastly different.

Furthermore, taking a functionality-oriented perspective is more or less a must to move towards more service-oriented solutions [24] and focusing in "logical structures" and "logical relationships". Service-oriented solutions will also be an enabler for more self-going agile teams. How to take full advantage of that we also needs to move towards more IP-based solutions instead of CAN/LIN-oriented protocols.

References

[1] P B Kruchten. The 4+1 View Model of Architecture. IEEE Software November 1995

[2] Fredrick P. Brooks Jr, No Silver Buller – Essence and Accidents in Software Engineering, TBD

[3] Mitleton KE, Land F (2012) Complexity & Information Systems. Blackwell Encyclopaedia of Management.

[4] Manfred Broy, Ingolf H. Krüger, Alexander Pretchner, and Christina Salzman. Engineering Automotive Software. Proceedings of the IEEE. Vol. 95. Issue. 2. Feb. 2007.

[5] SAE J1939/71 Vehicle Application Layer (The data parameters (SPNs), messages (PGNs) and reference figures and information previously published within this document are now published in SAE J1939DA, Published October 25

[6] SAE J1939 Digital Annex (has been broken out from the SAE J1939/71), Published April 22, 2019

[7] EEA for the CONNECTED AUTONOMOUS FUTURE. January 2017

[8] Virtual Functional Bus, AUTOSAR Classic platform, Release 4.3.1

[9] Frank van der Linden. Software Product Families in Europe: The Esaps & Café Projects. IEEE Software July/August 2002

[10] Hans Aerts and Han Schaminée, How Software Is Changing the Automotive Landscape, IEEE Software November/December 2017

[11] Kevin Baughey. Functional and Logical Structure: A System engineering Approach, SAE International 2011-01-0517

[12] Agus Sudjianto and Kevin Otto. Modularization to support multiple brand platforms. Proceedings of the DETC: ASME Design Engineering Technical Conference September 2001. DECT2001/DTM-21695

[13] Road Vehicles – Functional Safety. ISO 26262 http://www.iso.org/iso/catalogue_detail?csnumber=43464

[14] Scaled Agile Framework (SAFe). http://www.scaledagileframework.com/

[15] Sagar Behre and Martin Törngren. A Functional Reference Architecture for Autonomous Driving, Information and Software Technology, Vol 73. May, 2016 p136-p150

[16] 4D/RCS: A Reference Model Architecture For Unmanned Vehicle Systems. Version 2.0. The Army Research Laboratory Demo III Program. Aug. 2002, National Institute of Standards and Technology, Gaithersburg, Maryland 20899

[17] Kimon P. Valavanis, Denis Gracanin, Maja Matijasevic, Ramesh Kolluru, and Georgios A. Demetriou. Control Architecture for Autonomous Undererwater Vehicles. IEEE Control System. Vol. 17, Issue: 6, 1997, p48-p64

[18] Magnus Adolfson, ON-BOARD SYSTEM FOR TRUCK PLATOONS - Cooperative mobility solutions for supervised platooning (Companion), Final project conference, Applus IDIADA Technical Centre, Sep. 2016.

[19] Peter Nilsson, Leo Laine, Bengt Jacobson, and Niels van Duijkeren. Driver Model Based Automated Driving of Long Vehicle Combinations in Emulated Highway Traffic, Proceedings of the IEEE 18th International Conference on Intelligent Transportation Systems, Spain, 2015.

[20] Sachin Janardhanan, Mansour Keshavarz Bahaghighat, and Leo Laine. Introduction of Traffic Situation Management for a rigid truck, tests conducted on object avoidance by steering within ego lane. Proceedings of IEEE 18th International Conference on Intelligent Transportation Systems. Spain. 2015.

[21] Frank Buschmann, Regine Meuiner, Hans Rohnert, Peter Sommerlad, and Michael Stal, Pattern-Oriented Software Architecture (POSA) - A Systems of Patterns, ISBN 0-471-95869-7

[22] Michael A. Cusmano. How Microsoft Makes Large Team Work Like Small Teams. Sloan Management Review. 1997 Fall. p9-p20

[23] Melvin Conway. How Do Committees Invent. Datamation Magazine. 1968 April

[24] Phil Bianco, Rick Kotermanski and Paulo Merson, Evaluating a Service-Oriented Architecture, TECHNICAL REPORT, CMU/SEI-2007-TR-015

Using Cloud-Based Electronic Horizons to Enable Distributed Driving Functions

Peter Engel, Dr. **Alexander Geraldy**, Dr. **Jan Wolter**,
Robert Bosch GmbH, Hildesheim

Abstract

More and more vehicle functions are shifted into the cloud to make use of up-to-date data or to use the computational power of cloud-computing. Many functions rely on map-based information that could be provided by an electronic horizon. In the following we propose a system partitioning for a “cloud-based electronic horizon” which allows introducing distributed horizon-based functions even as low-cost solutions for smaller cars. On the other hand, it enables more sophisticated driver assistance and future automated driving functions.

We will introduce our cloud-based electronic horizon and our requirements, we will reason the usage of the cloud-based electronic horizon for distributed functions. After that we will present our system partitioning and hybrid localization solution and give an overview of the evaluation results of our system.

1 Introduction

1.1 Motivation

Nowadays, customers often pay one-time fees for vehicle functions and services when buying a new car, especially for built-in systems. To reduce the purchase costs, they often deselect functions with reduced or unknown benefit. A cloud-based horizon allows to obtain services as needed, for a short period of time or with a special extent. Such try and pay-per-use functionalities could convince customers to use more driver information and driver assistance systems (DIS/DAS) when needed. Aside from that, additional (pay-per-use) data services could be offered much easier and directly by integrating those into a horizon sent to the vehicle.

At the same time, the cloud-based services can benefit from enabling synergies, by, e.g., computing similar horizons for multiple cars in the same regions. The in-vehicle hardware can be scaled down, since the stored data volume (data storage) as well as the computation power (CPU) can be reduced. This allows to sell cost-effective vehicles with basic features and to

book additional services and benefits as required. When only the required horizon is transferred to the vehicle, this can even lower the data transfer volume compared to in-vehicle map systems with over-the-air updates.

Map-based data for DIS/DAS applications is currently provided by the navigation-system itself, based on a local map. This implies an onboard navigation system which supports the DIS/DAS applications. With our proposed solution the data needed for such applications will be provided by a cloud service. This enables DIS/DAS applications even for vehicle segments that are typically not equipped with a navigation system.

To offer up-to-date information when required, a cloud-based horizon can be provided by a cloud-based service. The provided horizon consists of relevant data along the most probable path (MPP) or along possible branches attached to intersections in the MPP. In the following we call the horizon along the MPP “1D-Horizon” and the branching parts of the horizon “stubs”. In the following, we call a horizon containing a 1D-Horizon as well as stubs a “1.5D-Horizon”.

Additionally, the cloud-based horizon should be adapted to the current situation and requirements, e.g., by adapting the horizon length, the number and size of contained stubs, and the contained attributes to the vehicle speed, mobile communication status and the subscribed vehicle functions. Hereby, the horizon transmission can be reduced to the required minimum to save transmission costs and data rate.

1.2 Exemplary applications

Vehicle applications can be clustered according to the level of support of driver: Driver Information Systems (DIS) support the driver by delivering information, Driver Assistance Systems (DAS) provide assistance for dedicated driving function and at Highly Automated Driving (HAD) the vehicle will drive fully autonomously. As representative, we list one application per cluster, which can be supported by the cloud-based horizon.

DIS: Hazard Spot Warning

The hazard spot application warns about upcoming hazardous points (slippery road, black ice, speed cameras, deer crossing, sharp bends, and hazardous crossings) along the future driveway. The cloud server knows different kinds of hazard spots and includes them into the horizon data structure that is delivered to the vehicle. Inside the vehicle the electronic horizon is evaluated and an application warns the driver about the hazardous event in an appropriately way.

In case a vehicle is equipped with sensors that detect hazard spots, the cloud-based horizon system can also be used to transmit raw data of hazardous events to the cloud. In the cloud,

the crowd-sourced data is then aggregated and provided to other vehicles within their cloud-based horizons.

DAS: Coasting assistance

The coasting assistance supports the driver by adapting the longitudinal control of a vehicle to speed limits, curves or slopes. This can be useful to increase the safety of the vehicle and to safe fuel. The necessary information, e.g., speed limits and curvatures, will be provided within the horizon from the cloud and delivered to an Adaptive Cruise Control (ACC).

HAD: HAD Attributes

A precise positioning solution is essential to have good working HAD functions. This will not be achieved sufficiently by traditional positioning based on global navigation satellite systems (GNSS). For a precise positioning required for highly automated driving, the horizon can provide the necessary HAD information like video or radar-based landmarks (e.g., road markings, and objects) to be used as position references in the vehicle.

1.3 Requirements

When relying on or providing a cloud-based horizon, we have to keep in mind several requirements influencing the usability and generated value of the overall system.

Our cloud-based horizon must be scalable to a huge number of clients. As stated before, this could be achieved by a clever clustering of regional horizons to same instances of the cloud.

To support DIS/DAS applications, an exact and reliable position of the vehicle in the digital map is essential. This requires on the one hand that the map-based vehicle position is well known in the cloud and on the other hand that the current position on the horizon is known to the car. Otherwise, the in-vehicle functions are provided with incomplete or imprecise information leading to a reduced usability and availability of the function.

Therefore it is essential, that branching from the horizon by a client is detected or predicted early and an updated horizon is delivered quickly**.** To achieve an adequate quality of service (QoS), the client has to deal with connection loss and has to bridge over the time until reconnect or safely degrade the functionality – up to a hand-over to the driver. A suitable system partitioning and a tailored positioning mechanism solving this problem will be presented in this paper.

1.4 Outlook to this Paper

The remainder of the paper is structured as follows. The next section gives a brief overview of the state of the art regarding electronic horizon systems. In section 3 we present our approach

of a cloud-based horizon in detail. We focus on aspects like hybrid localization, prediction of the most probable path (MPP), and present a sketch of our architecture with particular emphasis on the partitioning between light-weight clients and the backend. Section 4 reports about different evaluation aspects and section 5 concludes the paper and gives a short outlook.

2 State of the Art

2.1 Electronic Horizon

The advanced driver assistance system interface specification (ADASIS) standard [1] aims at connecting map-based data providers (e.g., navigation systems) with devices requesting map-based data via standard on-board busses like CAN. ADASIS defines the functional architecture as well as a standardized data format of the horizon. Typical applications using ADASIS horizons are “predictive driver assistance” [2] systems like Curve Speed Warning, Enhanced Adaptive Cruise Control, Adaptive Light Control, or adaptive truck systems for optimized fuel consumption [3] [4]. ADASIS allows to connect devices of different manufacturers based on the common standard.

When the on-board navigation system is used as horizon provider, its outdated map data can be a problem for applications relying on that information. Missing or wrong attributes can lead to misbehavior or low performance of the ADAS systems.

2.2 Cloud-based horizon

At CES 2015, Continental has presented the “Dynamic eHorizon” [3] to provide vehicles with HERE map based data and data from IBM connected car cloud. Continental sees their Dynamic eHorizon as an enabler for a wide range of new services and products. While their announcement emphasizes the aspects of the dynamic cloud-based data, we will in the following focus on technical aspects of the collaboration between vehicle and cloud. Other suppliers also presented connected horizons, e.g., Infoware [5], and Bosch [6].

In [7], Burgstahler et al. present a bandwidth-adaptive “Dynamic Cloud-based eHorizon”. The required MPP approach uses road class and turning angles to decide which road segments are included in the horizon. When the network bandwidth is not sufficient, the horizon is adapted to the available bandwidth, leading to a trade-off between horizon size, and functionality.

3 Our Cloud-Based Horizon Approach

In this chapter, we present a distributed system consisting of a client inside the vehicle, which communicates with a powerful backend system. The client provides sensor data from the vehicle to the backend, the backend processes the data and returns an electronic horizon to the client.

3.1 Hybrid localization

The localization of the vehicle follows a hybrid approach, containing a localization on a digital map in the cloud and a position tracking on the horizon at the vehicle (cf. Fig. 1). This distributed approach was chosen to enable the vehicle to overcome connection problems. After receiving a suitable horizon and a map matched position from the cloud service, the component *Position Tracking* allows to follow this horizon and to update the current position in the vehicle even if the connection is lost.

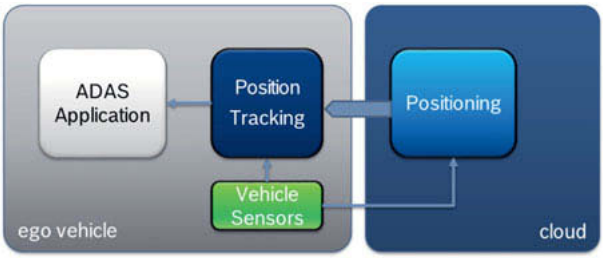


Fig. 1: Component distribution with focus on hybrid localization

Cloud Part

The localization in the cloud (component *Positioning*) matches the periodically received vehicle positions to a digital (topological-geometric) map. The map matched position is given as a road section (link) and a position on the road section (offset on link).

Each map matched position will be used for a vehicle tracking in the cloud to detect, if the vehicle has left the current MPP. When necessary, a recalculation and transmission of a new horizon is triggered. The same happens, when a vehicle with 1.5D-Position-Tracking requests a recalculation (cf. next chapter).

Vehicle Part

The *Position Tracking* uses high frequent measurement data (odometer and other vehicle sensors) to estimate the progress of travel and to calculate the position within the electronic hori-

zon. The quality of used measurement data will be improved by using sensor fusion. The starting point of the position tracking is the derived map matched position within the horizon from the cloud. The position derived from the cloud contains errors due to the time, needed for transmission and processing. Thus, the *Position Tracking* has to correct the received position by compensating the time delays.

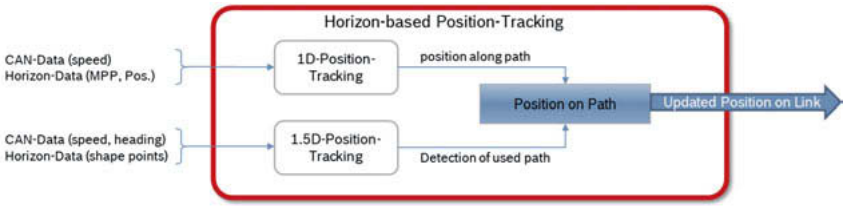


Fig. 2: Position tracking on horizon

In the so-called *1D-Position-Tracking* (cf. Fig. 2), the position within the horizon will only be tracked along the current 1D-horizon (i.e., along the MPP) using odometric information. The main advantage is, that the algorithm is very simple to implement and has low demands on processing power. The main gap of the algorithm is that the *1D-Position-Tracking* is not able to detect if a vehicle is leaving the MPP (e.g. unexpected turn-off at an intersection).

To detect a leaving of the MPP, an additional *1.5D-Position-Tracking* is needed, which considers also the stubs of the horizon. To identify the used path-segment, it correlates the direction profiles from derived electronic horizon representing the road network (so-called Reference-Profiles) with the direction profiles based on current sensor data of the ego vehicle (so-called Sensor-Profile). Therefore, the *1.5D-Position-Tracking* needs more computing power and more geometric information within the horizon (like shape points), which can increase the amount of data to transmit. On the other hand, a vehicle using *1.5D-Position-Tracking* can actively request a new horizon, if the MPP is left. So the periodic position update message from the vehicle can be omitted, reducing the network load again.

In *1D-* or *1.5D-Position-Tracking*, each derived map matched position is used to synchronize the position tracking, i.e., to compensate the incremental error of the tracking using vehicle sensors processing. The synchronization also takes into account the timing of the processing chain (i.e., sensor data transmission, processing of map matched position and transmission of the map matched position).

3.2 Prediction of the Most Probable Path (MPP)

Goal of our cloud-based horizon is, to provide vehicles with data for their most relevant path. While the most relevant path can be predicted in most cases easily and reliably with an active navigation guidance, this is much more challenging, when no target and driven route is known. According to an evaluation of Krumm [8], drivers only use navigation assistance in 1% of their trips. In all other situations they drive without any assistance, because they already know the route to their destination, e.g., when a person drives from his home to the workplace multiple times a week. Therefore, our approach does not rely on a navigation system and just reuses the driven road segments (links) from the server-based map matching (cf. section 3.1) to learn a prediction model that is able to predict a user-specific most probable path (MPP). This MPP is then extended by stubs and augmented with attributes to build the horizon. The stubs shall serve as fallback, when the vehicle branches spontaneously from the MPP.

When a drive is completed, the backend performs a map matching of the complete drive based on the sensor data that was transferred to the server. The result is a sequence of road links representing the map matched positions of the drive. This sequence is used to learn a driver individual prediction model. During the next drive of the user, this model predicts for the recently driven links $<\ldots, s_{i-2}, s_{i-1}, s_i>$ the most probably upcoming link sequence $<s_{i+1}, s_{i+2}, \ldots, s_{i+l}>$ of length l. To perform this task, a set of sequence prediction algorithms are known from literature [9] and used with some adoptions to fit them to our application. Most of the algorithms make use of a Markov chain of order K, which means that the prediction of a link is only based on the last K links and all previous links are not taken into account.

The prediction model can of course only predict a MPP in case for the recently driven links some possible successors are provided by the model. When a driver crosses an area for the first time, a prediction is not possible. To deal with such situations, we use a second prediction approach based on a heuristic which is able to predict a MPP in even completely new environments. The heuristic operates on properties of the underlying map, e.g., turn angle of streets, link type, or road class.

3.3 Partitioning / Architecture

Our distributed End-2-End system (cf. Fig. 3) is partitioned into a vehicle subsystem, which communicates with the cloud subsystem. As transport protocol the User Datagram Protocol (UDP) was chosen. Thus, acknowledgements, error checking and retransmissions can be performed in the application in an application-specific and optimized way. An automatic retransmission within TCP (Transmission Control Protocol) could, e.g., be senseless, or even harmful, if the retransmitted information is already outdated.

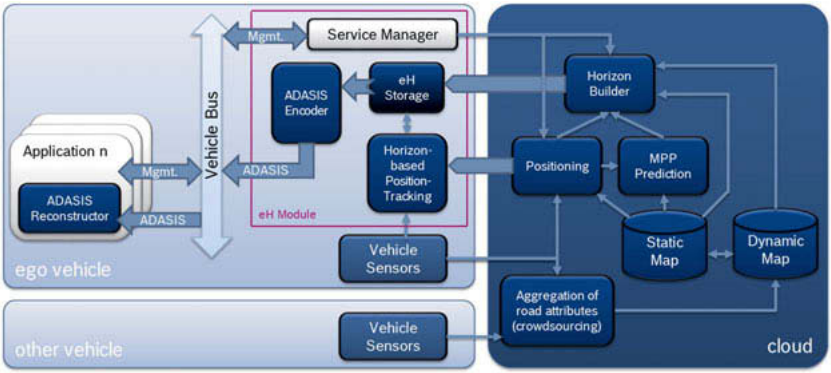


Fig. 3: Functional architecture of distributed cloud-based horizon system

A vehicle application needing a cloud-based electronic horizon registers at the cloud with application-specific requirements, e.g., required minimum forecast length, and attribute types (geometry and optional attributes like curvature, slopes, speed limits, road class …). The demands will be collected by the *Service Manager*. Additionally, the *Service Manager* gets information about the vehicle like the vehicle type and identification of the driver. The vehicle type (car, bike, truck …) is relevant to improve the estimation of the vehicle's behavior within the component *Positioning*, the identification of the driver is needed to enable the calculation of the user-specific MPP, and the needed length is relevant to reduce the transmission data.

The component *Vehicle Sensors* collects sensor data from different sources like Global Navigation Satellite System (GNSS) or other sensors connected, e.g., through the Controller Area Network (CAN). To gain a unified base of time, the different timestamps of different sources have to be synchronized. At CAN, some message types are delivered with higher frequency than used by the electronic horizon system. Therefore the high frequent data will be aggregated or thinned out to reduce the data volume to transfer to the cloud. To have a valid position information even without GNSS reception, the position is corrected and updated using dead reckoning.

This sensor data will be distributed to the component *Position Tracking* and sent to the cloud subsystem of the connected horizon system. In the cloud the sensor data will be processed by the component *Positioning*, which matches each received vehicle position onto the digital map. The *Positioning* can, e.g., be realized as a particle filter. This map matched position will be sent directly back to the vehicle and used there for synchronization purposes in the *Position Tracking* component.

Using the past routes, the *MPP Prediction* will calculate the Most Probable Path including branching stubs as described in section 3.2. Afterwards, the component *Horizon Builder* attaches to the MPP all demanded information from the digital or dynamic map like geometry, speed limits, or traffic jam information. The complete horizon will be compressed and transmitted to the vehicle, where it will be stored.

In the cloud, the current map matched position will be used to detect the validity of the previously calculated electronic horizon, especially of the MPP. When a vehicle branches from the MPP, a new MPP and horizon is computed and transmitted to the vehicle.

To distribute the horizon to requesting applications, the horizon will be converted into a standardized format like ADASIS and transmitted via CAN to the applications. Additionally the tracked position will be also transferred to the applications. The applications can use the horizon and the position within horizon to do their intended job (cf. section 1.2).

Using collected data from more than one vehicle, additional data services can be generated with the crowd source approach. The collected data will be analyzed, aggregated and stored in a *Dynamic Map*.

4 Evaluation

The introduced approach was realized to evaluate the practicability of the described solution and to evaluate the precision of positioning, dynamic configuration and timing constraints. At vehicle, small and cheap consumer hardware (Raspberry Pi, CAN shield, USB-GPS-Receiver, and USB-Modem) was used to evaluate, that such a system can be realized with small processing power. We tested with cars and motorbikes the influence of different vehicle types.

Fig. 4 shows an exemplary horizon. The position of the ego-vehicle is marked by the yellow car-icon. The MPP is drawn in green and the stubs are drawn in light blue. The horizon contains speed limits, which are requested from the coasting assistance application.

Fig. 4: Visualization of an exemplary horizon containing speed limits
Map: Neuhof-Hildesheim, 52.1271N 9.9138E,
© Google, Map Data © GeoBasis-DE/BKG 2009

4.1 Accuracy of the positioning

The accuracy of the positioning was validated in the cloud (*Positioning*) and at vehicle client (*Position Tracking*) using several drives.

In the cloud, the accuracy of the positioning is affected by the problem of inaccuracy of the digital map (outdated information, card offset, and simplification of geometry). Overall (including inaccuracy of digital map), the median of deviation between GNSS raw position and map matched position is less than 3 meters. This accuracy is sufficient for the desired applications. In case that a more precise validation is required, it should start with high precision ground-truth data leading to much higher efforts.

Due to the transmission delays (vehicle to cloud (sensor data), and cloud to vehicle (position data)), and processing time at the cloud, the position received at the vehicle will be outdated. Thus, the position error at the vehicle is higher than the error of the *Positioning* in the cloud.

At vehicle, in situations with good mobile communication coverage, the 1D-Position-Tracking significantly improves the positioning accuracy compared to the backend-based provision of

map matched position data. The deviation of map matched position from the cloud to the current GNSS position depends mostly on the vehicle speed and transmission times. In our evaluation even the 95% percentile of the deviation was not worse than 25 meters without Position-Tracking. This value can be improved to 7 meters using the 1D-Position-Tracking.

Our evaluation shows that the 1D-Position-Tracking is well suited to extrapolate the vehicle position with a desired update rate along a known route to overcome mobile communication outages.

The 1.5D-Position-Tracking in the vehicle has detected the leaving of the MPP during our evaluation with a mean accuracy of 98% at a distance of 10m after the related junction.

4.2 Handling of communication interruptions

The cloud-based electronic horizon system needs a reliable communication, which unfortunately cannot be guaranteed for every location. Thus, the applications has to deal with communication losses, e.g., missing or lacking horizon information.

During our evaluation using Long Term Evolution (LTE) communication we Fig.d out, as expected, that the network coverage strongly depends on the network provider. Thus, the latency and transmission failures (up to the total loss of connection) differ among individual network providers. We observed (as expected) that the latency of UDP is smaller and more predictable than TCP communication.

To handle lacking communication, we have established some measures: Communication interruptions must be detected to adapt the transmission strategy like data caching while interrupted communication and (re-)transmission in case of (re-)established communication. An interrupted or lossy communication can be detected by missing response messages (timeout) from the receiver. Then a retransmission can be triggered.

The sensor data from the client are cached and transmitted as soon as the connection is reconstituted. The *Position Tracking* takes over the localization, whenever updated position information from the cloud is missing. The horizon is calculated in the cloud with a suitable length to overcome potential connection insufficiency and/or to degrade the ADAS applications in a safe time. This horizon length is configurable by the vehicles application demands and can be adapted to the communication status.

The time used for recalculation and transmission of horizon after communication interruptions, explicit request by the client or detected leave of the MPP will take less than 500ms.

4.3 Evaluation of MPP Prediction

We performed a statistical evaluation to determine the performance of the different prediction algorithms mentioned in section 3.2. We evaluated the algorithms with respect to four dimensions: the runtime of the learning phase, the amount of memory the prediction model allocates after learning, the accuracy of the prediction, and the runtime of the prediction.

As input data we used 120 real drives, performing a *k-fold evaluation*. The number of drives is randomly partitioned into *k* equal sized subsamples. Of the *k* subsamples, a single subsample is retained as the validation data for testing the prediction model, and the remaining *k-1* subsamples are used as training data. The process is then repeated *k* times, with each of the *k* subsamples used exactly once as the test data.

Regarding the learning runtime, most of the algorithms need less than 40 milliseconds to learn 120 drives containing up to 1385 links each (computed on a performance notebook with Intel i7 2.7GHz CPU and 32GB RAM). Only the All-K-Order-Markov model needs much more time. The allocated memory usage of the prediction model after learning phase of 120 drives lies in a range of 6-15MB for most of the algorithms.

For the evaluation of the prediction accuracy, we performed the evaluation with a varying parameter of the MPP length (5, 10, 20, and 100km). For 5km MPP length the median value indicates 100% correctness, which means that at least half of the evaluated drives were predicted completely correct. For 100km length the median is around 50% correctness for most of the algorithms.

For the evaluation of the prediction runtime 4 out of 6 prediction algorithms show a very good performance: They need, even for a long MPP of 100km length, not more than 3 milliseconds for the complete prediction. The other two algorithms had with up to 1000 milliseconds a much worse performance.

5 Conclusions & Outlook

In this paper we have presented a system partitioning for a cloud-based horizon which allows to provide up-to-date driving-related information from the cloud to light-weight clients. We have motivated the challenge of disconnections and introduced the Position Tracking as a solution to bridge temporary disconnections reasonably while tracking the position of the vehicle on the horizon. Therefore, the resulting system is suitable to provide in-vehicle applications with a cloud-based horizon.

In the future, automated driving use-cases with higher timing and safety constraints could offer new challenges to the cloud-based horizon. To fulfill these requirements, we will analyze redundant communication channels and predictive QoS measures to improve the reliability and to gain time to safely degrade the functionality.

6 References

[1] "ADASIS Website" [Online]. Available: https://adasis.org.

[2] Bosch SoftTec GmbH, "Ahead of the curve – Predictive driver assistance" [Online]. Available: http://www.bosch-softtec.com/connected_horizon.html.

[3] Continental, "CES 2015: Dynamic eHorizon from Continental Points to the Future" [Online]. Available: https://www.continental-corporation.com/en/press/press-releases/2014-12-10-ces-104906.

[4] Bosch Mobility Solutions, "Eco.Logic motion for economic driving" [Online]. Available: https://www.bosch-presse.de/pressportal/de/en/eco-logic-motion-for-economic-driving-41843.html.

[5] Infoware, "Connected Electronic Horizon" [Online]. Available: https://www.infoware.de/en/gps-navigation/automotive/electronic-horizon/.

[6] Bosch SoftTec, "Connected Horizon" [Online]. Available: http://www.bosch-softtec.com/connected_horizon.html.

[7] D. Burgstahler, A. Xhoga, C. Peusens, M. Möbus, D. Böhnstedt and R. Steinmetz, "RemoteHorizon.KOM: Dynamic Cloud-based eHorizon" *Proceedings of the 7th GMM-Fachtagung Automotive meets Electronics (AmE 2016),* March 2016.

[8] J. Krumm, "Where Will They Turn: Predicting Turn Proportions at Intersections" *Personal Ubiquitous Comput. 14.7 (Oct. 2010),* p. 591–599, 2010.

[9] P. Fournier-Viger, J. Chun-Wei Lin, A. Gomariz, T. Gueniche, A. Soltani, Z. Deng and H. Thanh Lam, "The SPMF Open-Source Data Mining Library Version 2" *Machine Learning and Knowledge Discovery in Databases. Springer International Publishing,* pp. 36-40, 2016.

Use of open source software in automotive safety projects

A decision tree for the usage of open source software components in safety projects

Rudolf Grave, Elektrobit Automotive GmbH, Erlangen

Abstract

This paper describes the usage of open source software in today's automotive projects, focusing on the implications on the safety argumentation as well as the topic of qualification of open source software. The following scenarios are considered: separation, qualification, and redevelopment. A decision tree for the usage of open source software components is provided.
Keywords— Open source software, safety, software qualification, separation, ISO 26262

I Introduction

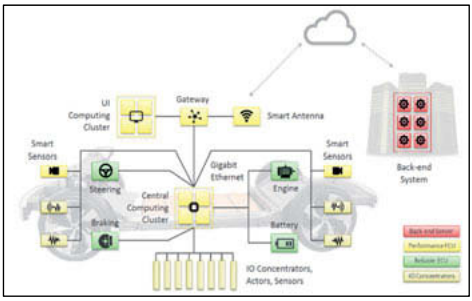
For years Open Source Software (OSS) is being used in automotive infotainment systems without considering safety argumentation, due to the fact that the safety requirements were realized by self-developed proprietary software running on embedded controllers. The subject of OSS licenses is another ongoing topic but will not be discussed in this paper.

The current vehicle E/E (electric/electronic) architecture integrates only one or a few vehicle functions per electrical control unit (ECU). This increases both the number of control units and distributed software functions and the complexity of connectivity respectively. In this context, the E/E architecture must perform an increasing number of driver assistance functions. Estimates of software complexity assume that the over 120 control units in a current premium vehicle contain more than 100 million lines of code.

The availability of higher-performance systems on a chip (SoC, e.g., Renesas' R-Car H3, NXP S32G, or NVIDIA drive platform) suitable for automotive applications and the necessity to save weight, for example by reducing control units or cabling, result in the aim to integrate multiple functions on several domain controller and, in a second step, on a central controller.

This paradigm shift changes the vehicles' E/E architecture considerably. It involves the introduction of service-oriented communication and dynamic operating systems, which, in turn, must meet the requirements for real time, functional safety, and security. Moreover, the use of dynamic control units allows the addition of functions that are not yet available when the vehicle

is launched. Figure 1 shows the probable future E/E architecture. At the heart are one or few central computers that communicate via a vehicle-internal Ethernet backbone.



Since the central computers are covering body and chassis domains, they are required to run software realizing safety requirements that achieve body and chassis functions up to ASIL B in the first step, and in the future should be able to achieve up to ASIL D autonomous driving functions.

OSS is widely used in this centralized performance computer because many standard functionalities are freely available for projects, e.g. firmware provided by chip vendors or libraries such as OpenSSH and OpenSSL. A current development counted around 120 OSS components. In addition, the standardization communities like AUTOSAR are moving towards the open source direction: Adaptive AUTOSAR is requesting a POSIX operating system underneath which might be Linux and is developing an Adaptive AUTOSAR concept in a shared environment between the consortium members.

OSS is usually not developed according to the automotive safety standard ISO 26262 [1] or Automotive SPICE [3]. This does not mean that the software is developed in an uncontrolled manner; many OSS projects implement restrictive change and review processes, automated test suites, and documentation of decisions in mail threads. These OSS projects are basically, applying the same development intent reducing the risk of software problems but using different work products and methods as defined by the safety standards.

The ISO 26262 does not define any special requirements for OSS components – the OSS components are to be handled in the same way as other software elements. The mechanisms defined in the ISO 26262 need to be applied to OSS components. An relevant exit point from ISO 26262 is provided in part 8 chapter 12.4.1 where a software element provider can show

evidence that the software is developed in accordance to an appropriate national or international standard.

II Usage Scenarios

This chapter describes four concepts of how OSS can be used in a safety-related project. At the end a decision tree is provided.

A. Separation

The first aspect is to analyze if the OSS function is used to implement safety requirements. This is usually done by analyzing the safety-affected signal and data flow. If the OSS is not used in the safety critical path it can be separated from the safety critical components. ISO26262:2018-6 Annex D provides informative requirements for “freedom from interference of software elements” which means that the OSS component doesn’t interfere with the safety critical component. Usual mechanism in software are the usage of MPUs, Container and Hypervisors in the spatial domain and execution monitoring in the temporal domain.

This approach was for example taken for a security library that provides access to a certain ECU mode, since no safety requirements were mapped to access itself the library could be used in his own isolated space.

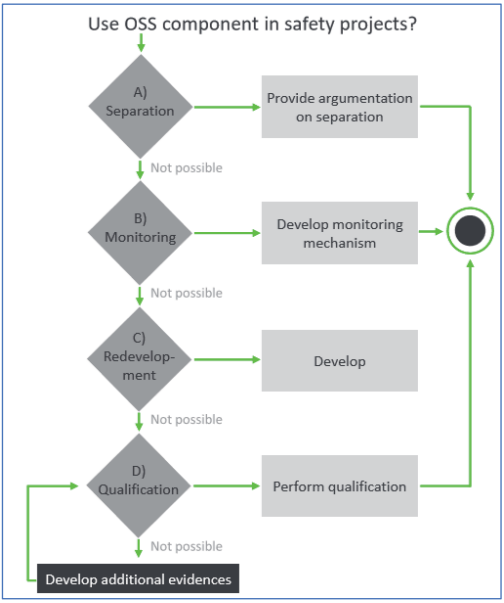
A special case in the separation scenario is the approach to argument that the OSS component itself is disturbance-free. The use case is that no safety requirements are assigned to the OSS component, but it should be integrated along with safety software elements without separation. A reason for this approach might be performance issues. In this case, the project then must show that memory or timing violations occur by the component, so that it can be integrated in the same partition as the safety-related software. For this approach, the verification activities of the OSS component development community can be reused. This special case might be used for lower ASIL levels.

B. Additional meassures / monitoring / protecting

If the OSS function is on the safety-critical path, an argumentation can be to add protective measures that detect or correct failures on the signal or data path. This might be realized by having a self-developed monitor that checks the calculated output boundaries and that lets the fine-grained OSS function calculate the normal control. In this case, separation as described in A is also required since one calculation path should not influence the other path.

C. Redevelopment

To avoid the effort for qualification of a software component it can be considered to reimplement certain parts of an OSS component on its own according to the safety standards. As an example, we might take the Eigen [4] component that is a C++ template library for linear algebra: matrices, vectors, numerical solvers, and related algorithms that were used for data fusion. After analyzing how the library was used, the project discovered that only some functions were used and that it is more goal-oriented to do a self-implementation according to [1].



D. Qualification of software components

ISO 26262:2018-8 chapter 12 explains the qualification of a software component that enables the use of COTS and provides guideline for the usage of "*already existing SW components not developed according to ISO 26262*". In this case the project must show evidence that:

- A specification of the software component is available that basically enables judgement on the next bullet points,
- The software component complies with its requirements,

- The software component is suitable for its intended use, The software development process for the component is based on an appropriate national or international standard (e.g. ISO/IEC/IEEE 12207),
- A plan for the qualification of the software component exists.
- In general, the required argumentation is not available since the OSS community is not targeting this type of development. Therefore, the project needs to deliver the required evidences and work products. This solution might initiate larger development activities to argument that OSS components can be used according the ISO 26262:2018 chapter 12, even if it is considered that the OSS component requires to be continuously developed and new features need to be requalified again. Figure 2 shows a decision graph on how OSS components can be used in safety projects.
- All four approaches have been realized in the last year and the authors recommend that the usage of OSS components is regulary monitored, since software projects might easily add new components for debugging and testing. Late detection in projects will lead to additional efforts and possible project delays.
- One approch was the SIL2LinuxMP [5] project which targeted the qualification of the Linux kernel, based on application-dependent assumptions, i.e. targeted for a certain application. A more generic approach is the successor project ELISA “Enabling Linux in Safety Applications” that started in early 2019 with strong partners from the automotive domain.

III Organisational topics

Apart from the license topics which is usually handled by a central department, there are several organizational tasks for risk mitigation, in most scenarios the supplier takes responsibility for the software that is taken in use, therefore the provider needs to take care that the software is fit for purpose to the intended use. In addition to the development process, a continuous monitoring of the OSS project is required to be informed about issue reports, changes, and in general the activity of the development community. Since this needs to be done for many OSS elements in a project, it might be outsourced to an experienced company/department for OSS qualification.

Additional information other than the issue tracker and verification reports of the OSS element itself, is available on sites like the NIST National Vulnerability Database (NVD) [6]. An additional topic is that the project needs to fix issues in a short amount of time for a certain criticality. This also needs to be done after start of production (SOP). In today’s project the teams are ramped down after SOP. To solve this issue, companies need to establish a

software integration and deployment environment in which software can be efficiently maintained and vehicle software updates can be realized. This opens the business up for companies that are specialized on open source software industrialization and commercialization.

Currently, automotive companies are not that active in OSS development due to the possible mindset of “If I invest in OSS, our competitor will get if for free” – this is something we all need to overcome and rather motivate us to contribute.

IV Summary

The need to use OSS components is there and projects will use them. Finding an argumentation to satisfy the safety case is possible but requires additional work. This work is not only a one-time development, it requires continuous monitoring or better yet active contribution to the open source communities. One motivation should be to get information on issues and activities from the development teams. Another aspect is to play a strengthening role for safety argumentation to become easier for the projects.

Commercial software companies will enter the open source qualification market and will industrialize topics like continuous qualification and maintenance of software products based on open source projects. Elektrobit for example is working closely together with Kernkonzept to industrialize the L4Re Hypervisor for automotive usage. Let us start the journey to combine safety development with the benefits of open source.

References

[1] ISO 26262:2018 Road vehicles — Functional safety

[2] AUTOSAR - https://www.autosar.org/

[3] Automotive SPICE http://www.automotivespice.com/

[4] EIGEN http://eigen.tuxfamily.org/

[5] SIL2LinuxMP
http://www.osadl.org/SIL2LinuxMP.sil2-linux-project.0.html

[6] NIST National Vulnerability Databse https://nvd.nist.gov/

[7] Open Source for business, Heather Meeker, Second Edition Flemming Editorial

[8] Forge your future with Open Source, VM(Vicky) Brasseur, The Pragmatic Programmers

[9] ISO/IEC/IEEE 12207 Systems and software engineering — Software life cycle processes

Trucks as the drivers of connectivity-based innovation

What the passenger car sector can learn from the experience already gained in trucks today

Gilles Mabire, Continental, Frankfurt am Main

Abstract

Connected vehicles are no longer a hype topic, they are already part of our daily life. Although connected passenger vehicles seem to get most of the share of voice in this field, a closer look reveals that commercial vehicles have higher fitment rates of connectivity related functions. This is partly due to legislation which mandates toll collection units in many European countries as well as the storage of drivers' hours of service and vehicle speed data. In addition, connectivity-based services offer efficiency benefits, the key profitability parameter for fleets. High fitment rates are the logical consequence if, as is usually the case, the services offer short to medium term ROIs. Standardized interfaces like rFMS reduce the entry barriers for providers of these services. This paper provides numerous examples connectivity-based services for commercial vehicles as well as the reasoning behind these and thus reveals some factors which could improve the take rate of connectivity-based services for passenger cars.

Connectivity – Trucks with high fitment rates and standardized interface rFMS

There is much talk about connected vehicles. Connected cars seem to get most of the share of voice in this field with topics like connected navigation systems, in-vehicle apps, smartphone connectivity etc. Individual car manufacturers offer advanced solutions such as over the air updating of vehicle functionality. Many of these functions are driven by the desire to offer increased vehicle comfort. Even if all these functionalities are no doubt attractive, they typically only have low to medium fitment rates and are usually part of a list of optional extras. But even this leads to volumes of units per year in the hundreds of thousands.

By contrast connectivity is an essential enabler with the capability of improving commercial vehicle efficiency, THE driver for truck fleets. Connectivity based functions which improve efficiency are attractive to all fleets and thus have high fitment rates. This is the reason why a typical truck in Europe was already connected many years ago. In 2012 the first truck OEM introduced a telematics unit as standard fitment in all heavy-duty trucks. Today almost all trucks produced in Europe contain a telematics/connectivity unit leading to significantly higher fitment

rates than in passenger cars. The telematics units deliver numerous vehicle parameters to the OEM backend where these are the basis for efficiency and reliability-improving service. The services today already include fleet management, remote diagnosis and vehicle & driver performance tracking.

But the trucking industry has already taken connectivity a step further with the rFMS standard [1]. This standardized cloud interface is the vehicle data access point for third parties to offer fleets interesting third-party services. The data available in the rFMS interface is a subset of the data which an OEM would typically extract from their vehicles. Thus, OEMs still have an inherent advantage for services for their own vehicles.

Looking at the truck fleet market we typically see that it contains vehicles from multiple manufacturers. Normally this would be a considerable hurdle for the fleet or service provider thereof to harmonize and aggregate the different data from the different manufacturers. As a cross manufacturer standardization, rFMS also enables OEMs and service providers to offer fleets cross-OEM solutions since all interested parties can access the same standardized data, even if this is a subset of what could be available. This creates an all-important level playing field for the service provider industry while enabling the essential efficiency improving services. In addition, rFMS offers the OEMs the opportunity to monetize their vehicle data while eliminating hardware costs for service providers.

Further Connectivity in trucks

But that is only the beginning of the truck connectivity story. Trucks are already the most connected vehicles on European roads. Further connectivity examples already in use today include the connected digital tachograph, tolling and smartphone apps.

European regulations mandate a digital tachograph which collects the vehicle driving speed over time and thereby also the hours of service of the driver. The driving and rest time of driver are highly regulated to protect the drivers from overwork. The driving and rest times of the drivers must be downloaded from the tachograph to a storage device where they must be stored for at least a year. The download of driver data can be performed while the vehicle is driving. Today the digital tachograph has multiple standardized interfaces for the access to driver data. The following solutions are available:

a. Connection to the telematics unit and download of the data via this
b. Connection to a smartphone via a dongle and download via this
c. Download via a dedicated download device which can be plugged into the tachograph

What all solutions have in common is that there must be a secure upload of the data to a secure storage site since personal data of the drivers is being stored. Continental offers the VDO TIS Web® service package for this task.
Almost all countries in Europe have a truck tolling system as can be seen from figure 1 below.

Figure 1: Tolling map for Europe

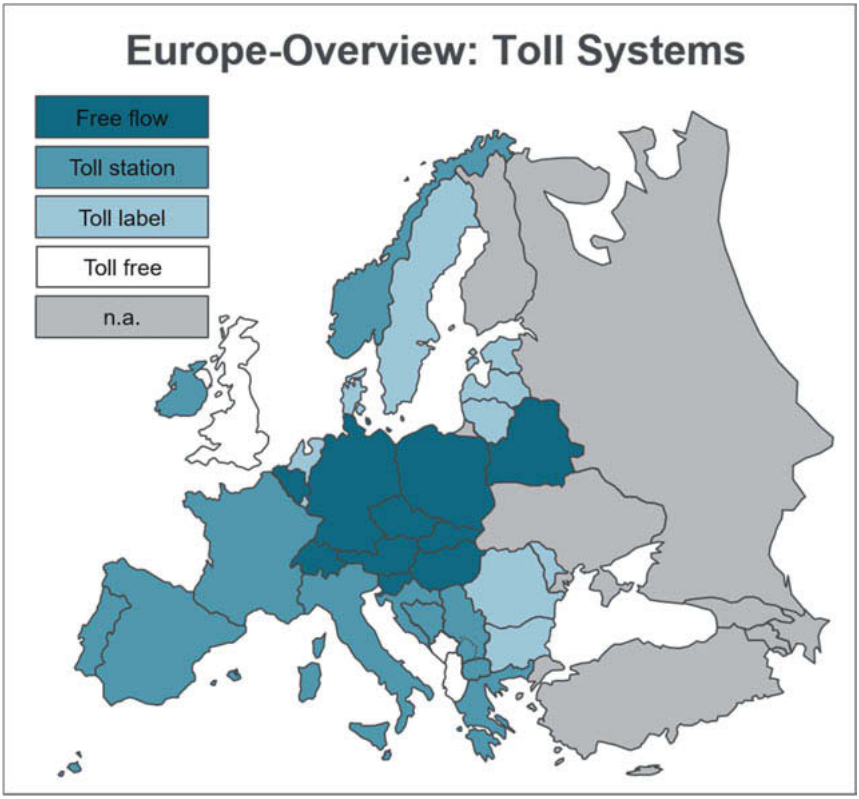


Fig. 1: Almost all countries in Europe have a truck tolling system
(Copyright: IMPARGO GmbH)

The free flow solutions involve either direct high frequency communication to a gantry when the vehicle is driving underneath as is the case in Austria for instance or communication to a

backend as is the case in Germany. Although tolling has been standardized in Europe with the EETS standard, there is considerably legacy in the current solutions. Thus, it is not uncommon to see long distance trucks with 5 or more tolling onboard units (OBU) behind the windscreen. Continental delivers the OBU for numerous tolling solutions in Europe. Here the customers are not the OEM but the tolling operators or solution providers. It could be argued that the tolling connectivity could also be used for other services, but the contractual set-up strict limits the possibilities of tolling operators.

The above service solutions rely on specific hardware in the vehicle. Additional connectivity is available via the driver's smartphone. This opens the door for numerous possibilities ranging from support finding a parking space (Truck Parking Europe), truck navigation (PTV-Navigator), remote control of the tachograph (VDO Driver) and capacity matching (Uber Freight) and fleet management (Fleetboard) in addition to the usual text and voice messaging to the fleet office. Today, most of the truck drivers have smartphone connectivity.

The business case for commercial vehicle services

The motivation for the adoption of any service by a fleet is quite simple – it must pay off. This is essential especially with the very low financial margins with which the fleets must work with. There is little room for long term investments. As a rule of thumb any investment should pay off, i.e. have a return on investment (ROI) of less than 2 years. This obviously means that any service – apart from legal requirements – must offer a financial benefit e.g. reduce operation costs, increase capacity utilization, or improve vehicle availability which can be weighed against the initial investment and/or operational costs. Hence, the majority of decisions will be based on a ROI calculation with a resulting ROI timeframe.

Although the services market is known for more adventurous business models, these are not really seen on the fleet market. While some fleets are technologically advanced, others still operate with basic IT tools. The business models of the service providers, regardless whether they are OEMs or third parties, are quite similar and conservative. Fleets are usually charged per vehicle per month plus any initial installation costs.

The decision process at the OEM is slightly more complex. The business case behind fitment of a telematics unit for instance is not just based on the ROI calculation for the service. The close to real time data from a vehicle are valuable for a vast range of organizations e.g. R&D (early failure detection and countermeasure implementation) and Quality (quality statistics). Increasingly the solutions are also linked to backend open service platforms like MAN Rio®, something which passenger car and truck OEMs have in common.

Interesting services on the horizon

While the solutions on the market today have had a considerable positive impact in fleet efficiency, the transportation industry still has some real pain points. Significant ones are:

1. High number of empty trips (EU approx. 20%) [2]
2. Low utilization of freight transport capacity (62% in Europe) [3]
3. Insufficient number of drivers (45.000 missing in Germany alone) [4]
4. Increasing number of fatal accidents involving trucks year on year in NA [5]
5. Fuel bill/CO2 emission needs to be reduced by 30% by 2030 [6]
6. Ideal future service solutions should address these pain points.

Pains 1 & 2 must be addressed by automated freight matching services in which intelligent algorithms to link available transport capacity to transport demand. This will require automated assessment of the available volume and load of trailers along with their planned route and timing. Similarly, there needs to be a constant stream of information on transport demand, citing the volume, weight destination and required ETA of a to-be delivered load. For highest impact this needs to be done across fleets. The largest fleets constantly perform this exercise. Small fleets need to be aggregated in matching platforms for highest impact. Currently this is the task done manually by freight exchanges, but it is interesting to see first start-ups like LoadFox and Cargonexx already performing intelligent freight matching for many thousands of shippers in 14 European countries. It is reasonable to assume that this will be a field of future activity and even disruption.

Pain 3 can be addressed with vehicle automation. Looking at the current announcements of the start-ups as well as the CV OEMs, there is clear drive for level 4 automation for hub2hub transport. This effectively leapfrogs most of the intermediate levels since these do not provide a sufficiently attractive business case. The long-term target is clear – driverless trucks travelling on defined routes, all of which will be connected. Connectivity will be mandatory to enable download of real time local incident information, download of high definition (HD) map data, over the air (OTA) firmware update of functionality, and real time monitoring of vehicle performance/state. Accident statistics reveal that approx. 90% of fatal accidents are caused by human failure. Replacing humans by machines as is the case in level 4 automation would thus also address pain 4. Continental also sees highly automated driving (HAD) as one of the top priorities in the coming year. Recently Continental announced a development partnership with Knorr-Bremse [7] with the target of developing a system solution for multiple HAD use cases. The partners plan to demonstrate a highway pilot by the end of 2019.

There will need to be a wide range of measures to address pain 5 and to achieve the desired 30% savings result. It will take an holistic approach to achieve this target. All levers must be addressed like powertrain concepts, aerodynamics, rolling resistance, automation. Services will no doubt also play a role. Recently an eHorizon solution including real time traffic information was developed and tested in the European Commission funded research project IMPERIUM [8]. Live traffic data and road topology information for the next 2-4km was provided to predictive powertrain algorithms. The algorithms were tested in simulated traffic conditions and provided up to 5% savings in conjunction with hybrid powertrains on a Vecto cycle.

Success factors for commercial vehicle services

Two factors come to mind when analyzing which factors have led to the high penetration of connectivity-based services; quantifiable justification and standardization.

In passenger cars enormous work goes into identifying customer needs and the associated willingness to pay. Especially for comfort features this is a tricky exercise which can lead to vastly different desires in different countries & regions. This makes go/no-go decisions very cumbersome. To a high degree of certainty, it can be said that the truck world is much more straightforward. If there is a business case with a return on investment (ROI) of less than 2 years, there is a very high likelihood that this feature will be very attractive in the market. Even with the high fragmentation of the fleet market, there is a relatively high homogeneity with respect to the requirements. This is probably also the reason why there are so many providers for fleet related services.

An already fragmented market is not attractive since economies of scale become very difficult. This makes the market less attractive which means there is a lower likelihood of innovative solutions. Standardization like rFMS (and FMS in the past) reduces the entry/investment barriers for third parties thus making the market more attractive especially for SMEs to offer attractive and innovative value-added services. Similar is valid for vehicle communication for which the J1939 standard was developed.

Summary and conclusions

Trucks are the most connected vehicles on the roads today. Telematic units are factory fitted in almost all heavy-duty trucks produced in Europe today. That should not be a surprise since connectivity-based services typically offer return on investments of approx. 2 years, exactly what fleets are looking for. Their development can be supported by standardization like the rFMS interface. The services also address the key pain points of the transportation industry,

namely CO_2 emission reduction, reduction in number of road fatalities, utilization of the available transport capacity and lack of qualified drivers. Many of these pain points are also valid for passenger vehicles (PV); good reason for the PV industry to see where it can learn from the experience already gained in trucks today to provide further value to their customers and society.

[1] http://www.fms-standard.com/Truck/down_load/Technical_Specification_rFMS_V2.0.0_21.09.2016.pdf
[2] https://ec.europa.eu/eurostat/statistics-explained/index.php/Road_freight_transport_by_journey_characteristics
[3] https://de.statista.com/statistik/daten/studie/593243/umfrage/auslastung-der-laderaumkapazitaet-im-europaeischen-transportmarkt/
[4] https://www.logistik-watchblog.de/neuheiten/1543-lkw-fahrermangel-2020-mindestens-150000-fahrer-fehlen.html
[5] https://www.fmcsa.dot.gov/safety/data-and-statistics/large-truck-and-bus-crash-facts-2017
[6] https://ec.europa.eu/clima/policies/transport/vehicles/heavy_en
[7] https://www.knorr-bremse.com/en/media/press-releases/knorr-bremse-and-continental-announce-a-partnership-for-high-ly-automated-driving-in-commercial-vehicles.json
[8] http://www.imperium-project.eu/

Functions on demand – Enabler for digital business with car functions

Challenges of implementation of a high complex security mechanism

Jan-Kristof Landgraf, **Alexander Fabri**,
AUDI AG, Ingolstadt

Abstract

The following paper “Functions on demand – enabler for digital business with car functions” is examining the dares from the perspective of an automobile manufacturer to bring this technical enabler to market. To give an idea about the range of areas, on which functions on demand have an impact on, different scopes were selected: starting with the business model and customer journey, the paper shows, where the idea of demand functions is coming from and how the clients’ experience shall look like. Moreover, the impacts on the financial models are displayed. The chapters ‘Technology’ and ‘Security and Safety’ are focusing on the architectural chain of car and IT systems, as well as on selection criteria of new on-demand functions. The paper concludes by describing the challenges of a wide-spread project team within the organization of AUDI AG.

1. Business Model and Customer Journey

In contrast to earlier times, consumers in the 21st century are increasingly using so called subscription models for the use of Software-as-a-Service (Saas) services, music- or video-streaming services. Software as a Service is defined as a reference model, which offers a standard software solution as a service over the internet [1]. The companies are also responsible for the operations and maintenance of the software. The user pays a user fee on a monthly or yearly base instead of a full license. The big advantage for the customer is a flexible and need-oriented service. Through this business model the companies can build a strong long-term customer relationship and generate recurring revenues.

Best practical examples can be found in different industries. For instance, Microsoft 365 combines web-based office application, different online services and office software subscription for an annually or monthly fee. Netflix offers three different subscription models for a monthly

fee, which differs in the movie quality of the movie (SD / HD / Ultra HD) and the number of devices, which can be connected at the same time.

The development in the context of digitization does not spare the automotive industry. With the new **functions on demand** (abbreviated as FoD) business model, the AUDI AG aims to respond in a flexible way to the customer needs. Up to now, customers have chosen the optional features of their vehicle in the car configurator or make sure that the vehicle had the desired optional features when purchasing a used car. For the first time the Audi customer can book selected functions on a monthly or yearly subscription. With this strategy, Audi wants to flexibly orientate itself to the needs of the customer in terms of software and hardware functions during the vehicle lifecycle in addition to the classic after-sales offers. Especially for leasing or used car customer the new business model gives the customer the advantage that the customer doesn't have to pay the lifetime prices or can book the favourite later on, because the function is already built-in.

In order to make this new business model possible, a highly networked enabler technology is used in the vehicle. In combination with the car backend Modularer Backend Baukasten MBB), state-of-the-art encryption technologies and a new sales channel with an e-commerce shop Audi Commerce Platform (ACP), this service can be offered to the customer simply and intuitively through the myAudi smartphone app (see figure 1).

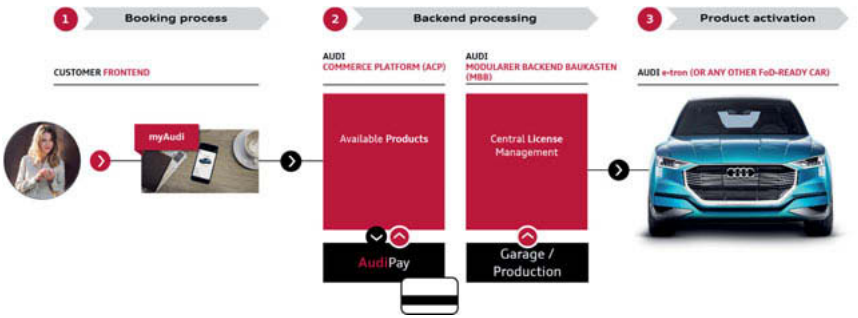


Fig. 1: Functions on demand - customer journey

The overall customer journey is defined in only three simple steps: **booking process, backend process and product activation**. The customer can choose selected functions through the myAudi smartphone app and is guided to the check out like a normal e-commerce process. Therefor the AUDI AG implemented a new IT sales system, called ACP, with a new

payment service to offer a new sales channel for the customers. After the customer has successfully booked a functions on demand function the license will be generated in the car backend MBB and will send over-the-air to the functions on demand enabled vehicle. In the vehicle the customer will be informed within the HMI (Human Machine Interface) about the installation process and the function activation.

But the very important question is what kind of functions are useful and attractive to be activated temporary and have a strong advantage for the customer (see chapter Security and Safety for more information). The final result for the functions on demand portfolio for the first vehicle project looks as follows:

Fig. 2: Functions on demand - portfolio Audi e-tron

The portfolio is divided into three main categories lights, driving assistance and infotainment.

Lights

The lights category offers the matrix LED headlights with dynamic indicators in the LED matrix package. Furthermore, the other light package includes adaptive light, highway-/ cornering-/ intersection light and animation front and rear light.

Driving Assistance

In the first step the single function is the park assist function, which assists the customer with controlling the steering wheel.

Infotainment

The infotainment category has the Audi smartphone interface, DAB and navigation incl. connect available.

The portfolio is individual for each country based on market specific country settings, legal requirements and function availability.

2. Financing Model

As the launch of time-limited functions is totally new business for the automotive industry and is therefore affecting all kinds of processes, also the financing model is concerned [3].

The main topics could be identified in pricing (concepts), cost and income structure, as well as revenue calculation.

Car producers have not been known for a high variability of prices so far and the situation therefore fundamentally different to e.g. gas stations, where prices might vary throughout a day. With digital services and direct marketing coming up, the risk to annoy customers with volatile prices is increasing. As a current survey of A.T. Kearney has shown, 76% of more than 1000 interviewed persons in the U.S. are sensitive to numerous price changes of online providers. [1, p. 10] If price changes are necessary or desired, it is better to start with higher basic prices and lower them by giving discounts than the other way round.

According to A.T. Kearney, other important factors in pricing concepts are (among others): flexible accounting (per timeframe and product) and flexible offers (i.e. a modular product structure with different product bundles).

Figure 7 shows the cost and income factors influencing the business case of functions on demand at AUDI AG.

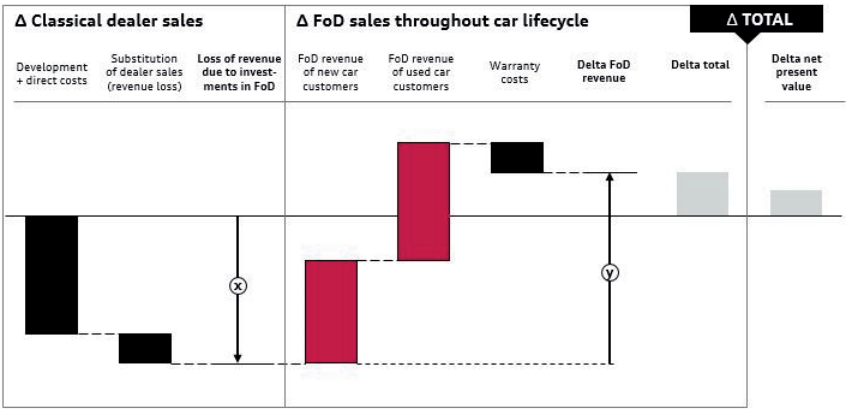


Fig. 3: Business case calculation for Functions on demand

As it can be derived, there are four different cost factors: development, direct and warranty costs, as well as a loss of revenue due to substitution of dealer sales.
Development costs represent the costs to enable a car model with function on demand. Direct costs are additional material costs, which have to be built in a car to enable a certain FoD function (e.g. DAB antenna). Substitutional costs is displaying the expected loss of sales, as purchase of functions in after sales might replace purchases of functions at the beginning (i.e. classical dealer sales). As a fourth cost type, there are (additional) warranty costs, as customers are granted additional warranty time on later-on purchased FoD functions. The warranty time is assigned for the duration of the function, maximum one year (for lifetime purchases).
On the other hand there are two income factors: the earnings of new car customers, as well as the ones of customers purchasing a used car. Especially the target group of customers purchasing a used car is new to the automotive industry, as, up to now, no sales are generated with these customers.
shows moreover the assignment of costs and revenues to the date of their occurrence: development, direct and substitutional costs accrue at the time, when a car is produced. Otherwise, revenues from new and used car purchases as well as warranty costs occur during the lifecycle of the car.
Furthermore, the revenue calculation for on demand functions has evolved from a static evaluation to a dynamic view on costs and revenues depending on the date of their occurrence. Due to the fact that sales income is not only generated once (at the time of the car delivery), but during the vehicle's lifecycle, revenue calculation has to be adopted accordingly: calculations performed with the net present value display time lags of the payments, which are carried out later (compared to the date, when the business case calculations were done). This means, a certain interest rate is taken into account, which ensures that any income generated after some years is evaluated in another way than the income generated today.

3. Technology

The history of function enabling shows mainly three different stages: first, the hedging of the ECUs (Electric Control Unit) parametrization. This is the simplest way of function enabling, as it embodies the static activation of a function, when the car is produced (i.e. in production).
At Volkswagen Group, the next step was to activate functions via software: for so-called SWaP functions (Software as a product) a service operation for the vehicle is necessary and additional hardware has to be integrated for some functions. A good example for these kind of functions is the speed control function: beside the activation itself, the hand gear has to be

retrofitted to the car. Furthermore, the identification of functions is needed and therefore function-enabling IDs have been introduced (German abbreviation: FS-IDs), which has been a change of paradigm, as function enabling is no longer performed by parametrization, but by contribution of FS-IDs to the car. Last, but not least, a basic backend infrastructure has to be set up to store information about the enabled functions for each car.

Today, functions on demand stands for the technical solution to activate functions online (i.e. via over-the-air update). Premises of FoD are that no additional hardware is used and the car has to be “ready” for the function activation, after it has left production. That means, that all calibration and parametrization of ECUs has to be done in the production phase.

Figure 3 shows a rough overview of the networking architecture of functions on demand at Volkswagen Group. The technical chain can be divided into three main parts: first, in order to ensure secure function activations and to prevent misuse a vehicle-specific function enabling code (FEC) containing the FS-IDs to be enabled has to be created by crypto systems. Second, the FEC is transferred from the car backend system to the Master ECU in the car. The secure distribution of the FS-IDs from the Master to the Slave ECUs then conclude the enabling process.

In contrast to this centralized Master-Slave solution, each ECU could also be supplied directly with an FEC. This was not chosen since, among other issues, it would require significantly larger storage and processing resources in the vehicle ECUs. Besides the function enabling process itself, there are further processes, which are taking place: a successful purchase in the shop (called ACP – see chapter ‘business model and customer journey’) is a pre-condition that the FEC is created by the crypto system. Furthermore, production and garages are involved, as functions are also activated or restored during the respective processes. Warranty systems are affected, as the function duration is an important input factor for the calculation of warranty claims (see chapter ‘financing model’).

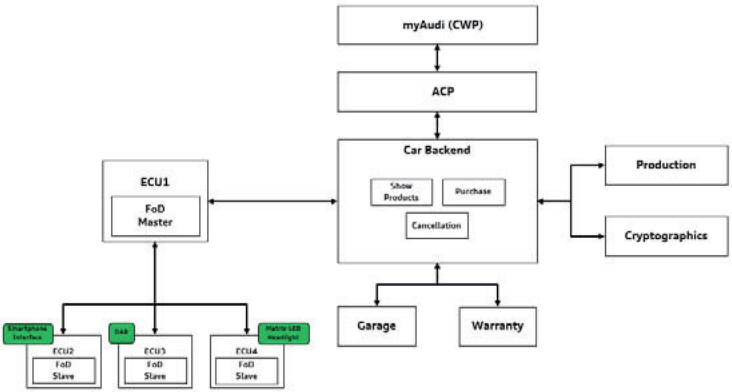


Fig. 4: Technical End2End architecture of Functions on demand

4. Security and Safety

During the concept phase two main topics concerning security have been identified, which needed further investigation: The purchase process and the enabling process.

Therefore the security concept of functions on demand is based, first, on the proof-of-purchase (PoP), which is created in the shop systems. And second, on the component-individual signature of the FEC, which ensures that only the respective car, in which a function shall be activated, can read the enabling code.

The PoP is created after the successful purchase of a function in the shop systems. The most important data in the PoP is the function ID, so that the function can be identified unambiguously.

As the car is not able to process a PoP, a conversion has to take place in the end-to-end chain. As it is explained in the technology chapter, only FS-IDs are readable by the car. This concludes in the creation of the FEC, which can only take place in a safe environment – in this case, this is the crypto system. It is necessary for ensuring an E2E integrity of the data, the PoP has to be kept in the signed FEC data package.

A useful and attractive portfolio is important to be a strong advantage for the customer. For answering the question, if a function is suitable for FoD, a criterion grid has been developed. It considers the user experience, safety, direct costs and potential revenue for each possible function (see figure 4).

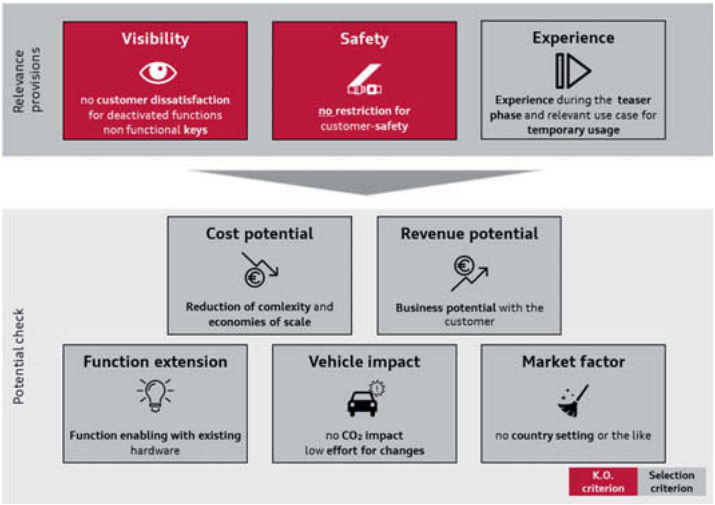


Fig. 5: Functions on demand - two-level criterion grid

The grid is divided into two categories **relevance provisions** and **potential check**. The first category is about the visibility and safety. That means that functions, which have a restriction for the customer's safety or lead to a customer dissatisfaction in a deactivated status, would be classified as a knockout criterion for functions on demand. The criterion "safety" is also about the valuation if the function needs a separate warning screen before expiration. For driving safety functions, e.g. lane assist, the customer will be informed additionally about the expiration of the function by the instrument cluster. The last criterion "experience of a temporary usage of a function" is highly material for the relevance provisions. The big question here is, how the customer can experience the function in the teaser phase.

The second category covers financial topics and hardware specific topics (for the financial topic see chapter financing model above).

The criterion "function extension" covers the check for enabling functions without additional hardware. There should be no need to put more direct costs into the vehicle to raise the revenue potential. The next criterion "vehicle impact" is about CO_2 emissions and the effort for changing software or hardware within the ECUs. The criterion's gain is it's revenue potential and that it does not affect the type approval process. During the development of a function for functions on demand there are discussions about security changes or memory extensions for necessary requirements in the ECU. The last criterion "market factor" deals with the competitor's activities or EURO NCAP requirements for the function.

5. Organizational Challenges

The implementation of a digital function does not only mean to deal with the topic in the R&D centre, to develop sales IT systems or to change the legal conditions, but also to deal with the existing organizational structures in the enterprise [1]. At AUDI AG, development projects are traditionally carried out within the matrix project organization, since projects were previously implemented within a division or department. Due to the increasing complexity of projects that primarily concern digitization, more and more experts from the most diverse departments are required to work on projects either full-time or temporarily. The disadvantages of this form of organization are well known [2]: Potentially conflicting project and department goals, risk of double burden for employees due to their dual role (in project and core organization) and extensive need for coordination between the project manager and department manager. Due to the current end-to-end responsibility of the so-called function owners, complex projects cannot be implemented by a single person, since the function owner is supposed to be the technical contact from the R&D centre. The job profile of a project manager is fundamentally different and is also highly required for this complex digital project function on demand.

As the number and complexity of OEM projects continues to increase, project managers have to work in specialized departments, like a project management department or a project management office (PMO). A consistent further development are the project-oriented or project-based companies. This means that the majority of activities are implemented in projects. So-called project-based firms (PBF) can build up common competences and a strong knowledge management [4]. In addition, there are different approaches to project implementation. While the waterfall model still predominates in the R&D centre at Audi, the internal IT development department is primarily working in an agile way. So the first approach was, to bring the agile method to the vehicle development for the functions on demand project. Retrospectively, to hire a software developer internally for the core project team and to assign the software development company directly has been a more than efficient approach and brought a short development time to the software component for the vehicle side. Although the IT department develops its projects in an agile way, the IT development trades were assigned for all IT car projects, so the scope of the project was in a competition with other projects.

The following figure 5 shows the final project organization for functions on demand. The stakeholder decided to split the project into two main scopes to handle all the relevant project requirements. On the one hand, the car project team focuses on the vehicle architecture, HMI (abbrey. Human machine interface) requirements and the function portfolio. Furthermore, the IT for car and security is also a component in the focus of the car, because it is a direct link to the vehicle. On the other hand, the customer team handles the communication with the dealers

and prepares marketing activities. Also, the IT customer and e-commerce team takes responsibility for the Audi Commerce Platform (ACP) and the relevant customer frontends.

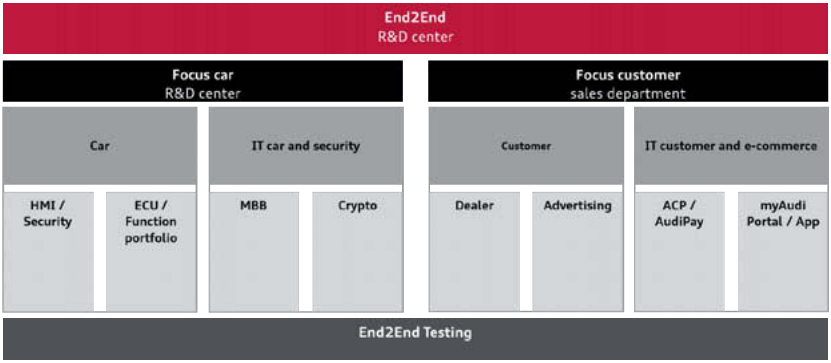


Fig. 6: Functions on demand – project organization

6. Literaturverzeichnis

[1] „Enzyklopädie der Wirtschaftsinformatik," [Online]. Available: http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/is-management/Integration-und-Migration-von-IT-Systemen/Software-as-a-Service/index.html. [Zugriff am 28 07 2019].

[2] GPM, „Gehalt und Karriere im Projektmanagement 2017," 2017.

[3] F. Walter, „Erfolgreich Projekte Leiten," [Online]. Available: https://erfolgreich-projekte-leiten.de/projektorganisation/. [Zugriff am 28 07 2019].

[4] R. Whitley, „Project-based firms: New organizational form or variations on a theme?," *Industrial and Corporate Change,* p. 77–99, February 2006.

[5] A.T. Kearney GmbH, „Schöne neue Kaufwelten?," *Automotive Vol. 7,* pp. 6-11, März 2017.

System of systems structured data for mobility services

Yann Chazal, Renault, Paris, France;
Dr. **Andreas M. Hein**, Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay, Paris, France;
Dr. **Samuel Boutin**, Knowledge Inside, Versailles, France

Abstract

Pressing global sustainability issues such as urban sprawling and congestion, social inclusion, energy and pollution concerns, will likely lead soon to radical transformations. The question of how the automotive industry and other industrial players will move their business in these transformations is at stake. It is very likely that the automotive business of the future will be less product-centric but more service-oriented, collaborative, and tailored to a geographical area. Over the past 3 years, we have been investigating a knowledge-centric and model-driven methodology and tool able to analyse these transformations and to develop transformation scenarios. First, the idea is to model the complexity for a given geographical area, striving to describe and understand how territorial issues are interwoven with economic relationships and systems interacting with humans. Then, from this model, transformation scenarios based on product and service innovations are designed and assessed, respecting territorial requirements and human preferences, changing business models and technical behaviours. Outcomes for economical entities would be the business models of new service/product offerings and related technical requirements. Since these transformations will rely on unusual co-creations between private companies, public bodies and local communities, our platform intends to share, gather and manage knowledge for various users and disciplines, and thus enhance collaboration. First returns on experience we have at Renault with projects both in mobility and electricity domains prompt us to move forward.

1 Introduction

The automotive industry is facing a profound transformation today. The reasons for this transformation are linked to the emergence of new technologies (autonomous driving, electric and connected vehicles) and a changing societal context. For example, in urban areas, local governments implement policies in order to reduce the number of private cars and to foster the deployment of mobility services as a complement to public transport. Fig. 1 illustrates this tendency for Paris, where it can be clearly seen that the percentage of cars as a mode of transport

has decreased from 46% in 2002 to 17% in 2008 [1]. The percentage of personal transport (PT), bike, and walking as modes has significantly increased.

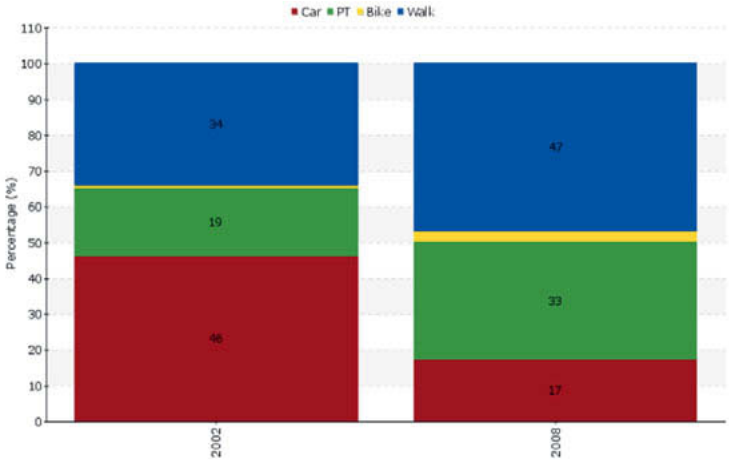


Fig. 1: Evolution of transportation modes in Paris

Car manufacturers should not expect any reversal of this situation. Adding to this trend, the number of privately-owned, individual cars will decrease even more in urban areas, as the population in urban areas does not depend on personal cars. As a consequence, they will be urged to switch to alternatives such as car sharing, ride sharing, and multi-modal mobility offerings. Taking these trends together, it is likely that privately-owned, individual cars are more and more substituted by such services.

As a consequence, car manufacturers are obviously interested in how to enter the business of mobility services. However, these services bring additional challenges.

1.1 Trend from product to complex service development

Car manufacturers are used to the business of designing cars and selling them to their customers, as depicted on the left side in Fig. 2. In this business, the Original Equipment Manufacturer (OEM) does not depend on collaborations outside the automotive industry. This is no longer the case for services involving electric, connected and autonomous cars, as depicted on the right side of Fig. 2. Indeed, their deployment on a territory depends on the coordinated deployment of infrastructures, both tangible (special road equipment, charging infrastructure, etc.) and intangible (communication network and data services, high definition maps, etc.).

Thus, an OEM has now to design and contract services with enabling parties, outside the automotive industry, under the auspice of local public authority.

Moreover, competitors on this market of mobility services are not only other car manufacturers but actors unrelated to the automotive industry. Hence, these competitors might not have the competencies to succeed in the traditional automotive business, but they might be more skilled in developing collaborative services. In particular, mobility services do not only provide value to those immediately concerned such as passengers and operators. They also have the potential to generate indirect value. For example, mobility services might create value for shop owners by increasing the number of clients, or generate savings for public transport, benefitting public stakeholders. Thus, within the market of mobility services, OEMs will compete with new actors outside the automotive industry with different competencies and value propositions, including delivering value to stakeholders that are outside the traditional automotive business.

At the end of the day, new mobility service proposals have to be evaluated. Alignment with societal priorities at the territory level, or at least no negative impact on sustainability are more and more necessary conditions for the success of a new mobility service. Thus, respecting norms is still mandatory but not enough. Mobility service providers need to credibly demonstrate the potential social benefits from the service they plan to introduce.

1.2 Business development process

Some acknowledged practices are applicable to the development of these mobility services. A schematic development process for mobility services is shown in Fig. 3. Companies usually work first on the design of a conceptual business model. Then they focus on territories that are of special interest.

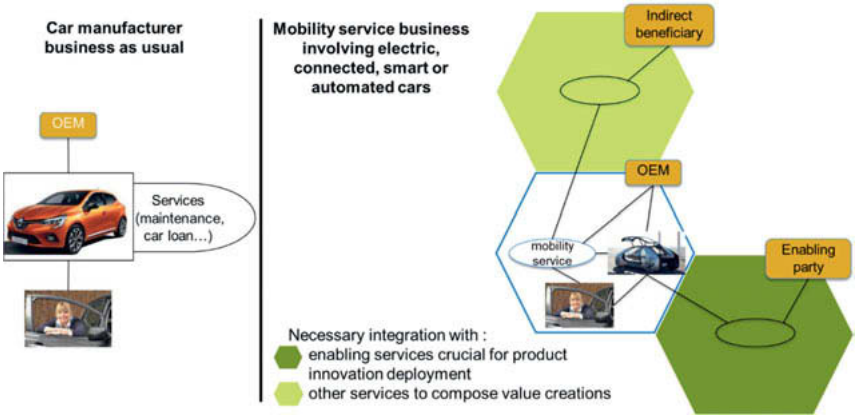


Fig. 2: Traditional OEM business versus future mobility service business

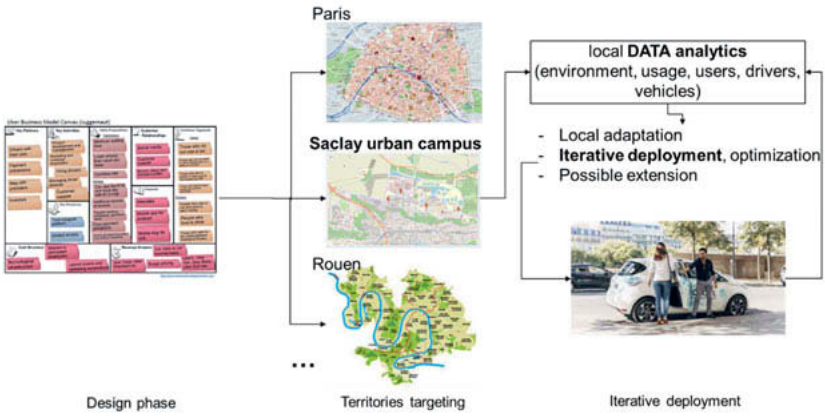


Fig. 3: Schematic development process for mobility services

Collecting and exploiting data about the environment, usage, users, drivers is then the way to understand how to adapt the business model to the local context. An iterative process is set up in order to implement the mobility service and get feedback at each step. Once the service is operated, more data gets available and the service can be optimized.

However, this process faces challenges at each stage. In the business model design phase, there is often a lack of deep understanding of technical aspects on which the business model is based. Moreover, this kind of business model relies on features that are unusually hard to address in the design phase, since it has to be driven by both market opportunities and societal stakes, while relying both on a value chain, enabling services and indirect beneficiaries. This situation is so unusually complex that it is difficult to find any exemplary successful case.

When focusing on territories, mobility services cannot be conceived by looking at the territorial arrangement alone. Territorial mobility issues are specific, entangled with local mobility, and subject to many trends, endogenous and exogenous. Thus, it is not easy to understand if an intended mobility service would be actually suitable for a territory. For example, the introduction of free-floating bike sharing in crowded areas has led to pavement congestion, which soon became unbearable.

Furthermore, local iterative deployment is rather a development in the field. The mobility service provider has to compose the service, taking technical specifics and access conditions of field equipment into account, such as parking lots or charging stations. More generally, the mobility service provider needs to make itself rapidly familiar with all human, technical, organizational, spatial specifics, to deal with in-field operations. The railway industry has developed extensive know-how to deal with that challenge. However, the recent bike sharing operator with docking stations in Paris has been struggling for many months beyond contractual deadline to make its business work.

2 A system of systems structured data model for mobility service development

The system of systems structured data model aims at addressing these challenges and supporting the business development process. The previous shortcomings are caused by ignoring important knowledge and interactions, and failing to use them in a proper methodology. Most of the required knowledge and data already exist and are not out of reach. We've focused on the needed framework, methodology and tool to use them appropriately.

Common perspectives on the system of systems, called "views" are operational, functional, and physical as shown in Fig. 4. The operational view shows the interactions of the system of systems with its context, the functional view shows how data, material, and energy flows are transformed by system elements, and the physical view shows the actual hard- and software to which the functions are allocated. These views are commonly used in systems engineering methodologies and frameworks.

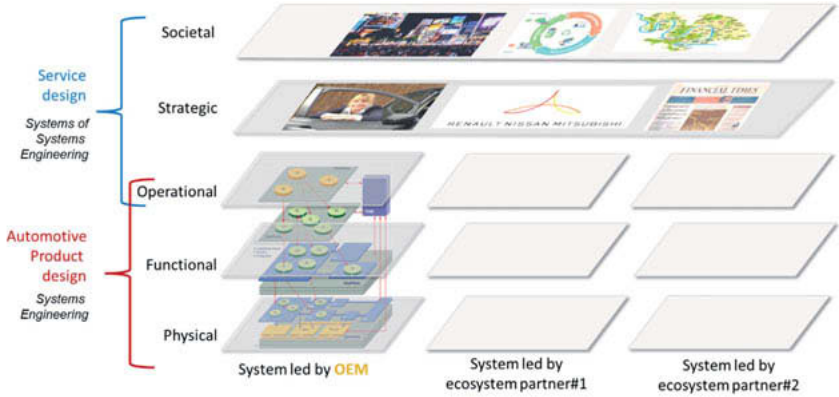


Fig. 4: System of systems views

For the purpose of mobility services, we have to consider how new kinds of cars should behave and interface with systems led by enabling parties, and indirect beneficiaries. This is true from a technical point of view of reusing the Systems Engineering data structure. It also has to make sense from a business point of view of service contracts. Thus, we found it

necessary to model in a strategic and economic level how companies, authorities and users coordinate accordingly with technical behaviour. At the end of the day, business value creation has to be aligned with sustainability concerns and main territorial expectations, which are at the societal level.

Thus, we have defined a systems of systems data structure which is able to address mobility service challenges. It enables to gather and turn knowledges into data feeding 5 levels. How these data elements interact has also been clarified so that consistency within and between levels can be ensured in the design process of mobility services. In the following, we will focus on the first phase of the mobility service process, the design phase, depicted in Fig. 3.

2.1 System of systems structured data enables to understand technical background

Existing approaches for understanding mobility services are mostly focused on the business model, which corresponds to the mobility service design phase, depicted in Fig. 3. Common representations include the business model canvas [2]. For example, for Uber, the business model canvas depicts the main value propositions for customers and drivers (minimum waiting

time, lower prices, cashless ride, etc.), customer relationships (how the relationship with the customer is established), channels (how to reach the customer), and which customer segments the value proposition is addressing. The corresponding business model canvas is shown in Fig. 5, based on the canvas presented by Juggernaut [3].

While the business model canvas is a powerful tool which is in widespread use, it provides limited support for representing the complexity behind mobility services. For example, the service Uber provides is based on a complex network of relationships between actors (drivers, customers, Uber), the resources and systems they operate or manage (cars, smartphones, service platform), and constituent services (driving service, booking service, etc.). All these relationships are missing in the business model canvas.

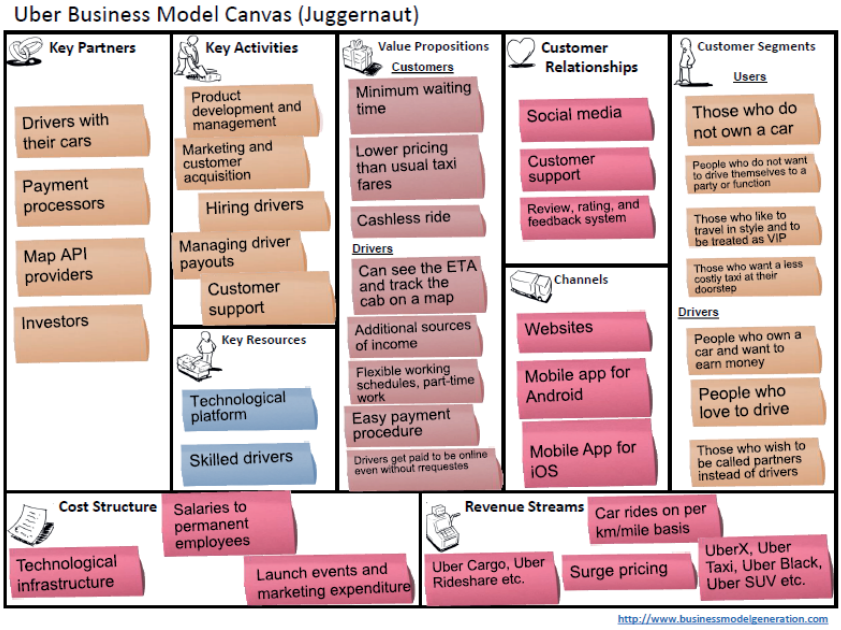


Fig. 5: Uber business model canvas based on Juggernaut image [3]

[AH1]

The system of systems structured data model allows for a more detailed representation of the diverse relationships between actors, systems, and services, as well as a hierarchical

representation, in order to zoom in and out. An example for an application of the system of systems structured data model for Ride Hailing is shown in Fig. 6.

This synthetic view (arKItect) [4] is based on an import of data available in a reference architecture for intelligent transportation solutions, called ARC-IT (Architecture Reference for

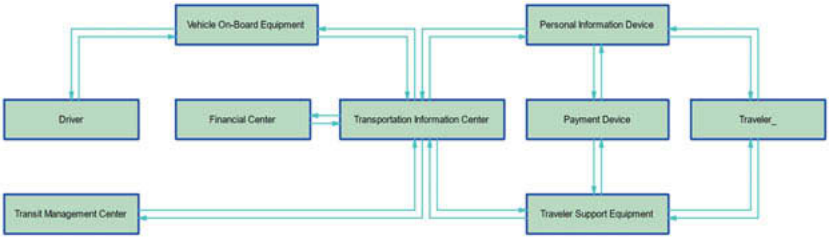


Fig. 6: Ride hailing data model

Cooperative and Intelligent Transportation). This reference architecture provides a standardized representation, using a set of predefined concepts for systems, services, actors, etc. It also gives a first idea of the complexity of the transportation business since 5500 concepts with their relations have been elaborated to describe this reference architecture. Dynamic Ride Hailing is one of the 139 services described in this architecture. It can be seen that the Traveler uses a Personal Information Device (e.g. a smartphone). The Personal Information Device exchanges data with the Payment Device (the Uber app connected to an external Payment service platform). Travelers should also be able to book their ride through other devices (traveler support equipement) and be able to aggregate different service offerings. On the other side, the Driver uses Vehicle On-Board Equipment (Driver smartphone with Uber app) to communicate with the Transportation Information Center (Uber service platform), which will match traveler, drivers and trips. The requested coordination with mass transit (Transit Management Center) is also described. According to Juggernaut [3], Uber has analyzed and selected the specific perimeter they would like to be responsible for.

The example demonstrates that a system of systems structured data model allows for a more in-depth understanding of the relationships between system elements. It also enables to analyze what systems Uber chose to be responsible for in the ride hailing business.

2.2 System of systems structured data enables to design scenarios and assess risks

Mobility services are to be designed by identifying key enablers (technological, regulatory, etc.) and analyzing direct and indirect value creation. For example, a mobility service, as shown in

Fig. 7, might help stimulate local commerce in a city. This would have immediate value for the passengers and shop owners but also for the city via increasing its attractivity or generating higher revenues from corporate tax. On the other hand, many cities do not have the financial resources to invest in new infrastructures. Thus, if the mobility service requires a charging infrastructure, it may be interesting to collaborate with the distribution system operator (DSO) and design a charging service. Such a collaboration would reduce the required investments for installing and operating the infrastructure.

Furthermore, such an analysis would also identify potential conflicts at the political, economic, social, and environmental level. Finding mitigation strategies for these conflicts is also key to designing sustainable mobility services for a specific context.

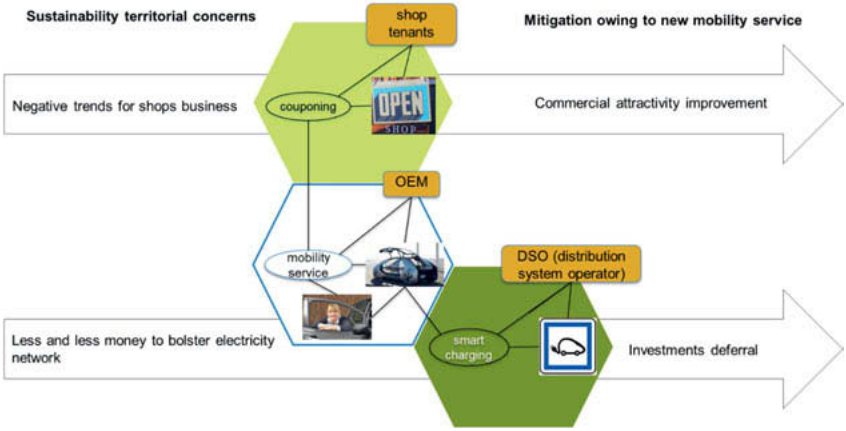


Fig. 7: Integrated mobility service with couponing service and smart charging service

2.3 System of systems structured data enables to anticipate risks with societal evolutions

Rappeler la vue process et zoomer sur la phase 2 et problème associé

After the mobility service design is developed in phase 1, a suitable territory is selected for its deployment, as depicted in Fig. 3. The main challenge is that each territory has unique characteristics such as the existing transportation infrastructure, roads, and the population living and working on the territory. In particular, the evolution of the wider societal context is important to consider, as the mobility service risks to fail if it is not properly adapted to the particular context.

The system of systems structured data modeling approach we are proposing also links the system of systems to a wider societal context. We use the example of the Paris Saclay Urban Campus in the following. The Paris Saclay Urban Campus is a large research cluster in the Paris suburbs, bringing together key actors from French industry and academia together in one geographic location.15% of the French research output has already been generated by the Paris Saclay Urban Campus by 2014.

The evolution of the Paris Saclay Urban Campus context introduces risks to future mobility services. A metro line was originally a key to mass transit on the campus. However, its construction was postponed and will not be on time to cope with the massive growth of the campus. On the other hand, to compensate for the metro line, bus operations are going to be substantially ramped up, including a dedicated bus rapid transit lane. For mobility within the campus, the use of personal mobility devices such as electric scooters, bikes, etc. is also expected to increase. On the other hand, the use of individual cars is likely to decrease due to initiatives for shared mobility and parking space in the surroundings of the campus. Hence, shared mobility solutions such as carpooling are also likely to increase. For a mobility service operator intending to propose ride hailing, the analysis brings another more ambiguous future: They are likely to face barriers to access the campus. However, there might be opportunities for using ride hailing for small trips inside the campus.

The Paris Saclay Urban Campus example shows that local contexts for mobility services are very specific and need to be analyzed individually. A mobility service needs to take this context carefully into account, if it aims at successfully integrating into it.

2.4 System of systems structured data enables to refine use cases and configuration

Within the mobility service development process, shown in Fig. 3, phase 3 involves the iterative deployment of the mobility service on a territory. A key challenge is to rapidly adapt and iterate the mobility service with respect to the local context. However, traditionally, OEMs are not used to adapt and iterate its products rapidly to local contexts. The automotive industry is used to sell cars able to be used almost anywhere. Thus, there's usually no need to take care of special technical modifications in order to adapt to local features of roads, filling stations, etc.

With this respect, the situation for mobility services is similar to that of the railway industry. Trains are very dependent on railway signaling. However, many instances of signals can be found, sometimes with special features. Thus, trains need to interact with equipment in order to achieve the global mission. In the case of mobility services, it is also important to capture features of existing equipment in the field, such as access conditions to parking lots or

connections to EV charging spots. It is also necessary to rationalize the way to use these equipment, taking the road network and typical local use cases into consideration.

Capturing detailed data from the local context is thus essential. Due to the availability of a diversity of open data sources as well as the collection of large amounts of user data, analyzing the local context becomes possible. However, key challenges are the integration of the data from different sources and how the integrated data can help guide design decisions regarding mobility services. Fig. 8 shows an example of the Paris Saclay Urban Campus where data from a data base of registered organizations (for profit, non-profit, charities) is depicted. Furthermore, parking spaces are represented as well. This data can guide, for example, the build-up of an electric mobility infrastructure, where a key element is the potential of installing charging stations on parking spaces.

Fig. 8: Data of the Paris Saclay Urban Campus mapped on a OpenStreetMap map

3 Conclusions

In this paper, we have described pitfalls related to the state of the art of mobility service development and investigated a model-driven approach aiming at limiting the associated risks.

We target several benefits from this modelling approach:

- Get a sufficiently complete understanding of the situation and the model shall help us identifying missing or incomplete information.
- Capitalize analysis information from one case to another. Even if the focus on societal trends will change from one context to another, completing a reference list with each new experiment is again a guarantee of not repeating a mistake.

- Ensure consistency between all layers of the model (Societal, Strategic, Operational, Functional, Physical): the idea is that the requirements identified for a service shall be traceably refined to functional and product requirements.
- Enable efficient impact analysis in the course of the service design or during its life cycle.
- Follow the experience of the railway industry for field application. They indeed came to the conclusion that they could no longer rely only on a requirement or document-based process when a diverse set of equipment needs to collaborate.

As an input to our modelling approach, we could integrate or interface existing structured data: ARC-IT, OpenStreetMap, but also official registers for companies including localization and number of employee, land registry, a lot of available “open data”. Of course proprietary sources can also be very interesting, many applications capitalize traffic information for instance. So one important aspect of the modelling has been the capability to integrate all reliable sources of data. During the modelling, additional information is captured as a result of inquiry and analysis.

Getting an aggregated model that we can trust is the basis for our present and future work. Up to now, we have focused on the design and rigorous risk analysis of new services taking into account all relevant dimensions. Given the challenges of developing mobility services, the main questions we have addressed:

- for what societal benefits
- with whom in a new value network
- on what technical foundations

So far, we are more at the level of qualitative analysis. Our next step is about investigating with enough detail the performance and value analysis and decomposition for each service.

4 References

[1] INSEE, Modal Split: Enquête Nationale Transports et Déplacements, in: Déplacements à Paris En 2010, 2008: p. 3.

[2] A. Osterwalder, Y. Pigneur, Business model generation: a handbook for visionaries, game changers, and challengers, (2010).

[3] Juggernaut, How Uber Works: Insights into Business & Revenue Model, Medium. (2015). https://medium.com/uber-for-x/how-uber-works-insights-into-business-revenue-model-5a52da69e94 (accessed August 3, 2019).

[4] Knowledge Inside, Breakthrough Technology for Data Integration, (2019). http://www.k-inside.com/web/ (accessed August 3, 2019).

Building a Standardized Data Pipeline from the Cloud to All In-Vehicle ECUs and Sensors

A New Opportunity for the Connected Car

S. Acharya, Excelfore, Fremont California, USA;
M. Gardner, Molex, Lisle Illinois, USA;
S. Herz, Hella GmbH, Lippstadt;
C. Hosner, Alpine Electronics, Auburn Hills Michigan;
F. Lesbroussart, ZF Friedrichshafen AG, Friedrichshafen

Abstract

Authors from five companies in the automotive industry describe the challenges of over-the-air (OTA) technology, and explore the benefits that could come from a multi-company standard for a data pipeline reaching all devices in the connected car. We consider the implications from the perspectives of the cloud and four in-car technology domains.

1. Introduction

The term “connected car” suggests a 2-way cloud connection. Within the vehicle, “end-to-end networking” defines data pathways between ECUs and sensors. One of today‘s challenges is linking these, to establish data pathways from the cloud to all ECUs and sensors in the car.

The value of building such a data pathway is clear. Recall costs, life-cycle vehicle health management, diagnostics, predictive analytics, efficiency or autonomous driving all improve with connectivity. Customer retention and new revenue generation through post-sale upgrades and features-on-demand can also benefit.

There are more than 50 major Tier-1 suppliers, serving more than 30 major OEMs. Many Tier-1s and OEMs are developing proprietary OTA mechanisms. More will in time. And the great majority do not address data gathering – that technology comes from an entirely separate and growing set of vendors through an entirely different technical infrastructure. There are already more than 30 different OTA and data gathering solutions. Some OEMs have implemented different solutions by car brand or model, or by geography or even trim level, leading to even more complexity.

But the solution is clear: the automotive industry needs a single standard, a bi-directional data pipeline for reaching all devices in the car. The eSync Alliance is an open trade association that has been formed to promote and develop such a standard data pipeline.[1]

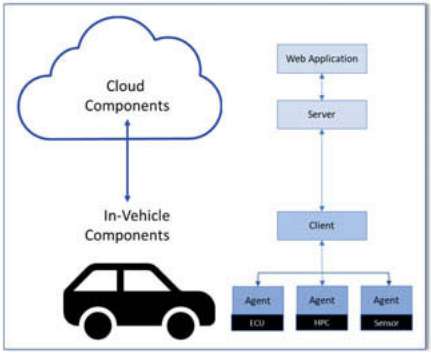


Fig. 1: The eSync Alliance proposes a standard OTA data pipeline

Using a Server-Client-Agent architecture with a minimum feature requirement for each level of the pipeline, as well as standardized messaging and communications protocols for the Server in the cloud, the Client in the car, and the Agent for the device, enable any member company to participate at any level with eSync Compliant services and products.

2. The Server in the Cloud

Many OTA concepts have been developed for non-automotive environments – for mobile devices and the IoT (Internet of Things). Why should it be different for a car? The answer derives from defining what is the “Thing” in IoT. Is it the car? If we focus only on connecting the car and server, this is a highly limiting approach.

If we want to reach many electronic devices in the car, we should define those devices as the “Things” we wish to reach. And if we do, there are substantial implications for the server.

Know before you Update

Automakers today offer many car models, with a variety of options and equipment packages, which may be replaced or upgraded in the field. To send correct updates the server should know the current details of all devices in each car. In an industry with this level of complexity the best approach is a bi-directional data pipeline which standardizes protocols and OTA-related behaviors for each device. If each device can report its presence, what software it is running, its status, and its OTA-related capabilities, then the server can automate tracking and distinguishing between vehicle configurations.

The eSync specification for a data pipeline provides this capability.[2] Within the eSync architecture, an Agent is written to each device. Agents take into account the characteristics

of different devices, while implementing a standard set of behaviors and protocols for the data pipeline. Any two devices may have different resources: for example an ADAS high-performance computing platform will have more processing, more memory and a more sophisticated OS than a seatbelt tensioner. Using standard protocols the Agents for these two devices can report their different capabilities, so that the Server can use different techniques to prepare updates for these two devices.

Similarly, the Client in each vehicle corresponds with the Server in the cloud using a standard set of behaviors and protocols. One of the standard Client behaviors is to report its complete "tree" of all Agents to the Server, and to update the information in the Server when there is a change. This process fully automates the maintenance of an up-to-date vehicle database. At every moment the Server has current information on each device in every car in its fleet.

Standardization of the interfaces in a bi-directional data pipeline also facilitates establishing an OTA solution for worldwide markets. Various regions have different security, privacy and regulatory requirements. The eSync specifications provide for policy-directed OTA campaigns. Policies can be adjusted to change the OTA processes for conformance to localized requirements.

Any eSync Compliant Server will conform to the standard functional behaviors and messaging protocols. Providing standardization at this level allows portability across public and private clouds and facilitates consistent capabilities in multiple geographical markets. eSync Servers have been implemented on Amazon AWS, Baidu, Microsoft Azure and Tencent public clouds, as well as OEM-proprietary servers. A car that is marketed in many geographies can be updated by servers in the local area, according to local policies, even if the server software is hosted on a different cloud or comes from a different vendor, so long as it is compliant.

3. The Infotainment Domain

In-vehicle infotainment (IVI) market demands and corresponding solutions demonstrate the need for rapid evolution and architecture adaptation. Planning IVI solutions involves feature compliance and enhancement, robustness, time to market, and cost.

Standard and Agnostic

Infotainment modules may range from a single entertainment display to an expanded cockpit instrument cluster. These modules may connect to the cloud via smartphone, Wi-Fi, or over a network connection (such as Ethernet or CAN) to a telematics unit. This topological complexity, in combination with meeting customer-specific interaction requirements, leads to a portfolio of

infotainment platforms that may use a variety of hardware, operating systems and software architectures.
Development within this framework to meet the demands of multiple OEMs increases costs. Standardization on a single OTA approach can help automate integration and reduce costs.

Enhancing User Experiences

User experience is a critical selling feature for infotainment, and an area of rapid improvement. This is one reason that IVI head units were among the first automotive devices to implement OTA. An industry-wide OTA standard will simplify tracking new network developments and security practices, and support continuous IVI improvement.
The first generation of automotive OTA solutions showed the importance of download flexibility and robustness, but infotainment system and OTA capabilities were usually isolated from other in-car technical domains. Now in newer cars the in-vehicle display becomes an important mechanism to involve the user in the update process for all devices.
In the presence of increasingly complex network architectures, we need to minimize the disruption caused by delivering continuous improvements to dozens of ECUs. A single standard OTA pipeline can help conformity to regional guidelines, configuring when and how users initiate updates, how they approve OEM-initiated updates, or even how they are notified if the updates are performed automatically.
Payload transport within the IVI domain may differ from other domains. Maps or entertainment content and even new applications can be downloaded frequently while the car is in operation, on an as-needed basis determined by bandwidth (and cost) and content consumption rates. This will often not be true in automotive control devices. This dual characteristic is a defining feature of the automotive OTA environment that should not be assumed in OTA technologies that have evolved through consumer markets (where safety systems may not be at play) or in industrial control (without streaming entertainment).

4. The In-Vehicle Network Domain

The in-vehicle network (IVN) is at the core of the connected car. Stand-alone technology domains will be confined to history – the future is a pathway of communication links to each device. The choice of network technology is evolving, so in today's market we must integrate hybrid IVN environments with multiple network standards. The challenge is to create a single OTA pipeline while we cannot yet agree on a common network or OS.

A Common OTA Messaging Protocol

The answer lies in an architectural approach that abstracts the unique differences of the various networks, busses, OSs and even devices that are present in the cars. To standardize the OTA pipeline, we implement a common message protocol, that can lie above the level of the network or bus interface, and has the necessary components to convey the information necessary to complete transactions between the Server in the cloud, and an In-Vehicle architecture with a Client to master the OTA processes, and Agents which abstract the ECUs, sensors, or other edge devices.

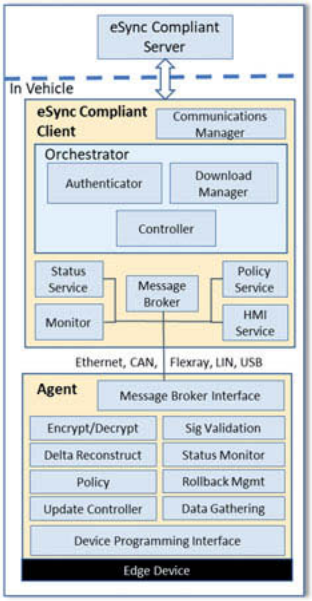


Fig. 3: The eSync OTA data pipeline standardizes the In-Vehicle Client-Agent architecture.

Automating IVN OTA Integration

This standardization bridges the heterogeneous IVN environment, creating a pathway for many cross-platform integration capabilities, such as Device Mapping (or Topology Tree) across all connected devices. Adding and removing features, or even devices, becomes a direct and easily accomplished process of messaging, without significant integration engineering – during development, at deployment, and after delivery. This enables quicker deployments and equipment package feature enhancements, with secure and complete asset inventory management.

Distributing the OTA Workload

Another opportunity from standardizing a Client–Agent architecture is parallel tasking. The bulk of the OTA task work can be done by the Agents, distributed throughout the car, rather than in a single OTA master device. With a Server possessing full information on the OTA capabilities of each Agent, delta files can be constructed for the reconstruction resources reported by each devices. The Client can push the deltas over the IVN to the Agents, which perform their reconstructions and re-flashing in parallel. This shortens vehicle downtime for multi-ECU updates. The same concept of full information in the Server also allows encryption to reach the edge devices for enhanced security.

IVN integration is improved, with greater awareness of vehicle topology and asset revision management, done by a distributed Client-Agent architecture using common messaging protocols. All devices, networks and OSs appear to the Server as a uniform set of Agents.

5. The ADAS, Active- and Passive-Safety Domains

Safety domains, both active and passive, are undergoing faster development than almost any other in the automotive industry, with strong growth of ADAS functionalities, and substantial investments into autonomous driving from all major OEMs.

These new functionalities share a few common characteristics:

- They deal with issues of life and death
- They use multiple sensors of different technologies to recognize the world around the car
- They take control of the vehicle for us (braking, steering, accelerating, etc.)
- They make implement complex algorithms to capture what has been learned in real life

This makes them ideal for OTA, but today the safety domain is usually not OTA-updatable.

Managing Multiple Device Updates

ADAS functionalities usually require complex systems combining sensors, central ECUs to analyze data and make decisions, and intelligent mechatronics to convert these into actions. This complex setup puts a focus on configuration management. OEMs typically spend a lot of time and resource on Quality Assurance and testing to ensure that a given configuration of multiple devices will perform correctly.

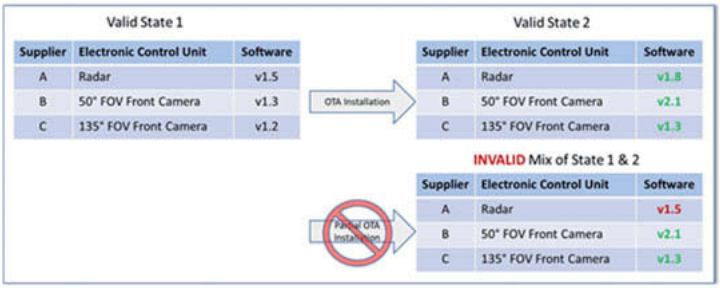


Fig. 4: The integrity of dependent relationships among multiple devices must be ensured before an OTA pipeline can be used to update the ADAS or Safety System domains.

When an OTA update is necessary, they need to maintain the full coherency of the system configuration. This means they need to move from a known, tested and fully verified state to another. It is not acceptable to end with a mixed configuration with some devices in each state, as this would never have been tested. Consider an example with two cameras and one radar (Fig. 4). If the camera updates go well, but the radar update fails, then we must return the complete ADAS sub-system to the last known, tested and verified state.

This is a key requirement for updating the ADAS domain. There must be mechanisms to group together updates of multiple devices and sub-systems, and to treat grouped updates as a single transaction – to succeed, or revert to the previous state if unsuccessful, together.

Update Policy Control

Updating safety devices requires extra attention. For example, while a map could be updated while the vehicle is in motion, a braking ECU should obviously not be. This means OEMs need to define policies that set the conditions to start an update – for the entire vehicle, or for individual devices.

This policy creation can be manual, but when lives are at stake, we also need a mechanism built into the device to establish policy for updating of that ECU, so that the ECU can ensure that an absolute minimum number of preconditions are met before beginning the update.

This is a clear case for the construct of the Agent, so that safety-related policies can be installed in the ECU and consistently followed for updates to that ECU regardless of vehicle model, or the levels of training or attention of the OTA administrators.

Finally, individual ECUs in the safety domain are probably not aware of the interaction of the driver with the user interface in the cockpit. For example, has the driver given permission for the update to start? Has the driver been notified that the brakes are in mid-update, and the

driver cannot start the car and drive at this moment? Standardization helps on that matter by providing standard APIs that can be the used by the vehicle HMI.

Compatibility

Various devices of an automotive sub-system are normally sourced from multiple suppliers, and must work together. Standardization helps ensure the devices all implement the same OTA approach, so they can be reached, updated, or rolled-back smoothly and efficiently.

Delta Compression

Items transmitted over the air could be a simple configuration file, a multi-megabyte binary image of ADAS software, or a large high-resolution map file. Compression reduces transmission costs, and ensure faster updates, or more comprehensive updates in the same time. To enhance the integrity and reliability of updates, a server providing delta files should have complete, up-to-date information on the software installed in every ECU in every car.

Bi-Directional Pipelines

Most of the ADAS functionalities rely on interpreting the world around the car in real time. This interpretation is done by algorithms trained to recognize and process specific conditions.
To help improve these algorithms, Tier-1s and OEMs need data from sensors or ECUs.
An efficient data pipeline, extracting key parameters from development vehicles, not only shortens the development time, but also creates a faster feedback loop. Increasing the value to the customer, and society, by improving safety, reducing accidents and saving lives. Of course, this data capture must comply with applicable privacy laws, which may differ from region to region. So data gathering should also be updateable by the OTA pipeline.

6. The Comfort, Body and Lighting Domains

Comfort, body and lighting substantially influence consumer satisfaction, but there are intense competitive cost pressures. More than many other in-car technology, products within these domains cover a broad spectrum of technologies and performance levels. For example actuators with small microcontrollers, minimal memory and no OS, on a LIN bus, might be mixed with devices using RISC processors, running AUTOSAR on a CAN bus.

Proprietary Approaches Increase Costs

With the proliferation of different approaches, suppliers face unfamiliar OTA technology for each OEM project. With proprietary approaches, ECUs need custom engineering to provide extra resources, with increased testing costs to validate the ECU functions within an OTA data pipeline. Proprietary pipelines mature slower than standard, widely adopted pipelines. On the industry's current path, ECU providers who do business with multiple OEMs will need to implement multiple OTA techniques, requiring slow, costly custom development.

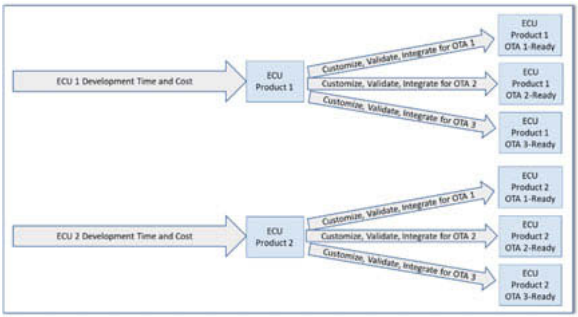


Fig. 5: In a proprietary OTA environment, there is a separate engineering task to integrate each ECU into each OTA solution, including within Tier-1s with multiple teams.

Standardization Reduces Costs

Design once, use many times is a well-proven cost saving concept. With the Server-Client-Agent architecture promoted by the eSync Alliance, the OTA requirements for the ECU are encapsulated in the Agent. Agents can become a part of the shared technology base of automotive business units. So Tier-1 OTA development efforts can be common across multiple products or product lines. New ECUs will then face only the costs of customization of the Agent and will then be available for any eSync pipeline. Compare this efficiency with the current environment, where each ECU product team must invest time and cost to adapt their product to each of the proprietary OTA platforms chosen by their OEM customers.

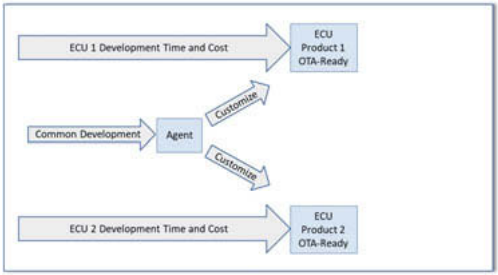


Fig. 6: A standard OTA pipeline means a Tier-1 can develop the core of an Agent once, and then customize it for each ECU product.

A standard pipeline will mature faster, with experience shared across multiple projects. Any weaknesses in the standard will be more quickly recognized and addressed. Pipeline standardization will encourage development tools, validation tools and test suites.
On the OEM side, costs are reduced by enabling Tier-1s to pre-provision their devices for OTA. They can be validated for OTA performance and reliability against standard platforms, reducing the costs and time for testing during vehicle integration.
As we trend towards higher individualization of the vehicle, OTA provides valuable opportunities for customization and personalization. Widespread adoption of a standard OTA pipeline will encourage innovation by focusing teams on developing new features, not integrating and validating multiple OTA mechanisms. There may be any number of value-added features-on-demand to be offered for customer OTA installation, ranging from custom lighting to advanced human-machine interfaces.

7. Cyber-Security

As OTA means connectivity between the vehicle and the outside world, interference by outsiders is rightly seen as a threat.[3] Every OEM and Tier-1 cannot be expected to put in place the resources to ensure cyber-security at the highest state-of-the-art levels. An industry where each participant sets their own approach will suffer greater vulnerability to attack.
A common approach on the other hand, enables collaboration. An open consortium draws many interested parties together – OEMs, suppliers, technology sources – to speed up recognition of vulnerabilities, and implementation of solutions. It also means proposals to improve security are reviewed by multiple parties, and industry experts.
Building a standard pipeline ensures the weakest link is at least as strong as the standard.

8. Conclusions

There is growing recognition that OTA must reach all programmable devices in the car. But today there is a bewildering proliferation of proprietary OTA approaches. It is time now for the industry to explore how a common data pipeline can be built from the cloud to the end devices in the car. This is the founding principle of the eSync Alliance, an open trade association formed by companies rooted in the automotive industry. The first generation (v1.0) of the eSync Alliance specifications were adopted in April of 2019.[3]

A working eSync Compliant platform is available for examination at the conference.

The eSync Alliance is an open and non-discriminatory association. All automotive companies are invited to join, and to participate in setting the standard for a common data pipeline.

[1] eSync Alliance website (www.esyncalliance.org)

[2] “v1.0 of the eSync Compliance Specs Released”, G. Shar, Auto Connected Car News (https://www.autoconnectedcar.com/2019/04/v-1-0-of-esync-compliance-specs-released/), April 16. 2019. Accessed July 18, 2019.

[3] “OTA Specifications Unify Software Updates for Cars”, C. Hammerschmidt, EE News Automotive, April 30, 2019 (https://www.eenewsautomotive.com/news/ota-specifications-unify-software-updates-cars). Accessed July 10, 2019.

Data Structures and Interfaces for High-resolution Maps in Rapid Prototyping Applications of Highly Automated Driving

Dipl.-Ing.(FH) **Marc Giertzsch**, Opel Automobile GmbH, Rüsselsheim

Introduction and scope

To enable highly automated driving a broad set of sensors is required, not only because the underlying technical principles imply different conditions (weather, day/night, etc.) in which the sensors can be used, but also due to the reason that the sensors are used for decisions at different distance levels – so-called micro- and macro-level decisions. Most sensors do not cover large distances, e.g. automotive radar only provides information up to a few hundred meters and, consequently, can only be used for micro-level decisions. Utilizing digitalized road maps enables macro-level decision making, route planning for example. In addition, high-resolution maps support micro-level decisions as well.

In Ko-HAF[1] [1] (Cooperative Highly Automated Driving), Opel used high-resolution maps for both micro- and macro-level decisions. The map was created by a project partner and was given in OpenDRIVE® format, which is a vendor-independent open format for digitalized road maps. This format was not suitable for the usage in the rapid prototyping environment of our demonstrator vehicle. Therefore, we converted the map into a set of data structures tailored to our needs, which enabled interacting with the map in multiple ways, e.g. for route planning as well as for in-lane positioning and trajectory planning.

Scope of this paper is to briefly describe the OpenDRIVE® format, the data structures we designed in order to represent the map in our rapid prototyping environment, the conversion from OpenDRIVE® format to our representation, and to exemplarily describe interfaces against the data structures used for map representation in our rapid prototyping system.

[1] The project Ko-HAF was funded by the German Federal Ministry for Economic Affairs and Energy based on a resolution of the German Bundestag.

OpenDRIVE® format

In order to describe digitalized road maps, an abstract representation of the road network is required. OpenDRIVE® (refer to [2]), which is managed by ASAM e.V. (Association for Standardization of Automation and Measuring Systems), provides a standard for such a representation.

OpenDRIVE® files are xml files, which are organized in the following five sections:

- Header
- Roads
- Signals and signs
- Junctions
- Controllers (representing dynamic signals)

We will now briefly discuss the sections “Header”, “Roads”, and “Junctions” based on the OpenDRIVE® format specification in revision H of version v1.4. For a detailed description of these sections and for the uncovered sections, please refer to the official OpenDRIVE® documentation [3]. We start with a chapter on co-ordinate systems used in OpenDRIVE® as a preparation for the following chapters.

OpenDRIVE® co-ordinate systems

There are three co-ordinate systems used in OpenDRIVE® files (Fig. 1). First, there is the global co-ordinate system[2] $(o_{\mathrm{geoRef}}, x, y)$ with x pointing forward and y pointing left. The origin of the inertial system is given by the geo reference. Note, that the inertial co-ordinate system can be rotated against UTM. For each road, there is a local co-ordinate system $(o_{\mathrm{road}}, u, v)$, with u pointing towards the road reference line (in the case of highways this is the leftmost lane marking) at the beginning of the road and v to the left. Finally, for the course of each road, there is a track co-ordinate system $(o_{\mathrm{refLine}}, s, t)$, with s pointing along the reference line and t pointing to the left. All three co-ordinate systems are right-handed and orthonormal.

Note: In this chapter, x, y, u, v, s, t denote unit vectors to define the co-ordinate systems used. In the further chapters with the same symbols we identify the co-ordinates of the respective co-ordinate systems.

[2] In the OpenDRIVE® specification, the global co-ordinate system is called “inertial system”.

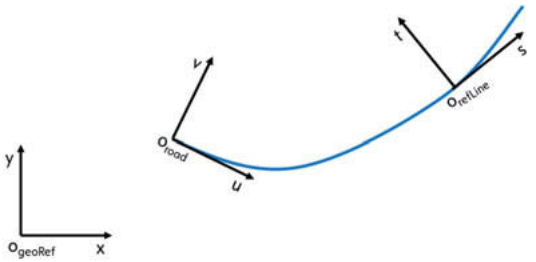


Fig. 1: OpenDRIVE® co-ordinate systems

OpenDRIVE® header section

In the header section, the most important information is the geo reference, typically given in UTM coordinates. All other location information is related to the geo reference via an offset (x, y) in meters. For example in Fig. 2, the point on the circular track marked with the blue x has the offset (x, y) = (951, 631) meters and, therefore, is located at (32U, 495219, 5538238) in UTM coordinates if the global co-ordinate system is not rotated against the UTM co-ordinate system.

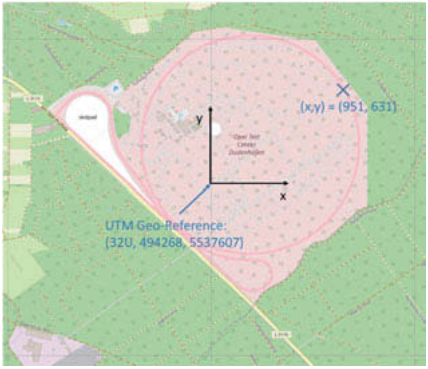


Fig. 2: Example of geo reference and (x, y) offset

Representation of roads in OpenDRIVE®

The basic geometry of a road is given by so-called reference lines. In the case of highways they represent the left border of the leftmost lane and are typically given as parameterized curves of third order (for other representations refer to section 3.2 of [3]). With increasing curve length the representation will be less accurate and, therefore, the road is represented using a

set of curves which are interconnected one after another instead of a single reference line. Formally:

Let $I_R \coloneqq [1, R] \subset \mathbb{N}$ be an index set and $S \coloneqq [0 = s_{\text{startRoad}}, s_{\text{endRoad}}[$, with s being the curve parameter (≡ road length). With $s_{\text{startRoad}} = s_0 < s_1 < s_2 < \cdots < s_{R-1} < s_R = s_{\text{endRoad}}$ we define the partition $P_{\text{refLine}} \coloneqq \{[s_{i-1}, s_i[=: S_i | i \in I_R\}$ and obtain

$$S = \bigcup_{i \in I_R} S_i.$$

With this partition we define the reference lines for one road as follows:

$$\gamma_i : S_i \in P_{\text{refLine}} \to \mathbb{R}^2 : s \in S_i \mapsto \begin{bmatrix} a_u \\ a_v \end{bmatrix} s^0 + \begin{bmatrix} b_u \\ b_v \end{bmatrix} s^1 + \begin{bmatrix} c_u \\ c_v \end{bmatrix} s^2 + \begin{bmatrix} d_u \\ d_v \end{bmatrix} s^3 = \begin{bmatrix} u \\ v \end{bmatrix}_{\text{ref}_i}.$$

Furthermore, each road is partitioned into a sequence of so-called lane sections ($P_{\text{laneSections}}$). In a given lane section the number of lanes is constant. Note, that in general $P_{\text{refLine}} \neq P_{\text{laneSections}}$.

In each lane section, the lane widths are described by polynomials of third order with the domain being the curve parameter of the reference lines. Again, it is required to partition the interval $\tilde{S}_j \in P_{\text{laneSections}}$ in order to ensure the desired accuracy of the lane width polynomials. Due to the fact, that the lane width polynomials may differ strongly per lane in a given section, that partitioning is done separately per lane. Hence, $\tilde{S}_j$ is partitioned by l partitions ${}_l\tilde{P}_j \coloneqq \left\{ {}_l\tilde{S}_{j,1}, \dots, {}_l\tilde{S}_{j,{}_lR_j} \right\}, 1 \leq l \leq \text{number of Lanes in } \tilde{S}_j$, which are not necessarily identical (${}_{\lambda_1}\tilde{P}_j \neq {}_{\lambda_2}\tilde{P}_j, 1 \leq \lambda_1, \lambda_2 \leq l$). In each interval ${}_l\tilde{S}_{j,k} \in {}_l\tilde{P}_j, 1 \leq k \leq R_j$, the lane width polynomial is given by corresponding coefficients ${}_l\Gamma_{j,k} = [{}_la_{j,k} \quad {}_lb_{j,k} \quad {}_lc_{j,k} \quad {}_ld_{j,k}]$.

At a given $s \in S$ the lane widths for the l lanes which exist there, are computed by the following formula:

$$\omega : S \to \mathbb{R}^l_{\geq 0} : s \in S \mapsto \Gamma_s \begin{bmatrix} s^0 \\ s^1 \\ s^2 \\ s^3 \end{bmatrix} \coloneqq \begin{bmatrix} {}_1\Gamma_{j_1,k_1} \\ {}_2\Gamma_{j_2,k_2} \\ \cdots \\ {}_{l-1}\Gamma_{j_{l-1},k_{l-1}} \\ {}_l\Gamma_{j_l,k_l} \end{bmatrix} \begin{bmatrix} s^0 \\ s^1 \\ s^2 \\ s^3 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_{l-1} \\ w_l \end{bmatrix}.$$

In Fig. 3 an exemplary partition of the reference line is shown, as well as exemplary partitions for the lane width polynomials for two lane sections. At $s = \sigma$ the lane widths are computed as

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} {}_1a_{2,2} & {}_1b_{2,2} & {}_1c_{2,2} & {}_1d_{2,2} \\ {}_2a_{2,1} & {}_2b_{2,1} & {}_2c_{2,1} & {}_2d_{2,1} \end{bmatrix} \begin{bmatrix} s^0 \\ s^1 \\ s^2 \\ s^3 \end{bmatrix}.$$

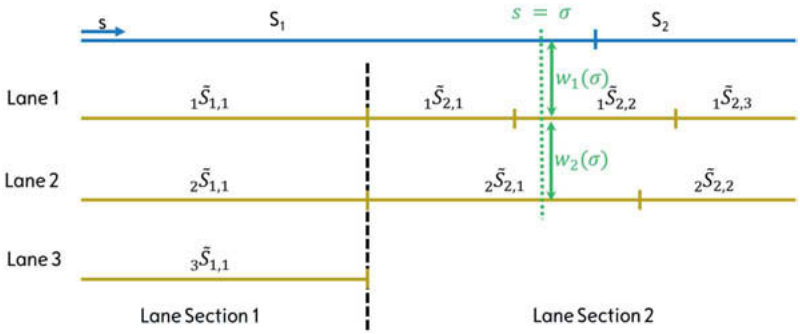


Fig. 3: Relationship between reference line partition, lane sections and lane width partitions. The reference line is shown in blue, lane markings in yellow, and the lane section boundary in dashed black.

Representation of junctions in OpenDRIVE®

Junctions are modelled by roads as well, these so-called connecting roads connect an incoming road to an outgoing road. Every non-connecting road, which lies in the interior of the map represented by the OpenDRIVE® file, has at least two connecting roads – one at the “start” of the road and one at the “end”. An incoming road can have more than one connecting road, e.g. if both a right turn and straight exist (refer to Fig. 4), and an outgoing road can have more than one connecting road.

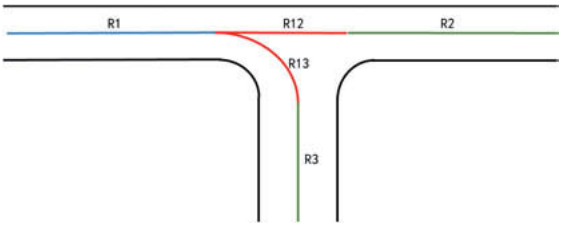


Fig. 4: Junction example. Incoming road in blue, connecting roads in red, and outgoing roads in green.

The road linkage for roads inherent to a junction J can be described by an adjacency matrix

$$A = (a_{ij})_{1 \le i,j \le |J|} = \begin{cases} 1, if\ (R_i, R_j) \in J \\ 0, else \end{cases}, |J| \coloneqq \text{number of roads inherent to junction } J,$$

with the roads being the nodes of the graph.

Data structures for lane center representation

The data structures Opel utilized in Ko-HAF to represent digitized road maps have been chosen based on three design goals. First, the data structures shall enable fast computation of operations often used, for example matching of the current vehicle location to the map (referred to as map matching), or computation of the distance from the current vehicle location to a desired path. Second, a representation of the digitized road map based on the lane centers was best suited for the tasks performed. As the demonstrator vehicle was equipped with a GPS that provides the ego position in UTM co-ordinates, storing the lane center co-ordinates in UTM co-ordinates was desired. Finally, for the rapid prototyping environment used in our test vehicle, the most suitable basic data structure for the storage of larger datasets is an array. To summarize, we represent the digitized road map by storing UTM co-ordinates of the lane centers of the roads in an array-shaped data structure. We will now discuss in detail how this is done.

With a given step size Δ, we first sample the reference lines of a road with the sequence

$$\mathcal{S} = (\mathrm{n}\Delta)_{0 \le \mathrm{n} \le \left\lfloor \frac{\mathcal{L}}{\Delta} \right\rfloor}, \mathcal{L}\colon \text{Length of road}$$

and compute the local (u, v) co-ordinates by applying the formula for $\gamma_i(s), s \in \mathcal{S}$ already discussed in the section “Representation of roads in OpenDRIVE®”. In order to compute the corresponding lane widths, we now need to figure out which lane width polynomials to use. This is done by matching the curve parameter $s \in \mathcal{S}$ to the lane width partitions (refer to section “Representation of roads in OpenDRIVE®”). With the matched lane width polynomials the distances from the origin point at the reference line to the lane centers compute as follows:

$$[s^0 \quad s^1 \quad s^2 \quad s^3]\Gamma_s^{\mathrm{T}} \begin{bmatrix} \frac{1}{2} & 1 & 1 & 1 & 1 \\ 0 & \frac{1}{2} & 1 & 1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} = [d_1 \quad d_2 \quad \ldots \quad d_{l-1} \quad d_l] =: D$$

Using the normal vector of the reference curve in the origin point, the lane center co-ordinates can be computed from the origin point $\gamma_i(s), s \in \mathcal{S}$ and the distance vector D by vector addition.

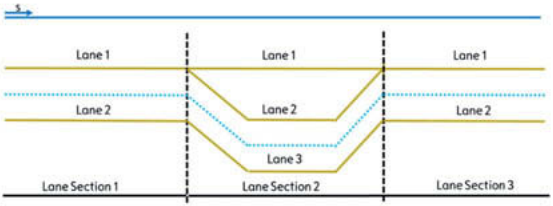
Finally, a change of basis is done[3] by multiplying the local (u, v) co-ordinates of the lane centers with a combined translation and rotation matrix in order to retrieve the global (x, y) co-ordinates of the lane centers in UTM.

So far we achieved a matrix $\begin{bmatrix} x_1 & x_2 & \cdots & x_l \\ y_1 & y_2 & \cdots & y_l \end{bmatrix}$ of global co-ordinates in UTM for all lane centers corresponding to one origin point $s \in \mathcal{S}$ for a given road. For that road, the computations described above are repeated for all $s \in \mathcal{S}$ and the co-ordinates are stored in one lane center matrix as follows (for non-existing lanes we will set $x = y = 0$):

$$LCM_{\text{Road}} = \begin{bmatrix} x_{1,\mathcal{S}(1)} & x_{1,\mathcal{S}(2)} & \cdots & x_{1,\mathcal{S}(|\mathcal{S}|)} \\ y_{1,\mathcal{S}(1)} & y_{1,\mathcal{S}(2)} & \cdots & y_{1,\mathcal{S}(|\mathcal{S}|)} \\ x_{2,\mathcal{S}(1)} & x_{2,\mathcal{S}(2)} & \cdots & x_{2,\mathcal{S}(|\mathcal{S}|)} \\ y_{2,\mathcal{S}(1)} & y_{2,\mathcal{S}(2)} & \cdots & y_{2,\mathcal{S}(|\mathcal{S}|)} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ x_{L,\mathcal{S}(1)} & x_{L,\mathcal{S}(2)} & \cdots & x_{L,\mathcal{S}(|\mathcal{S}|)} \\ y_{L,\mathcal{S}(1)} & y_{L,\mathcal{S}(2)} & \cdots & y_{L,\mathcal{S}(|\mathcal{S}|)} \end{bmatrix}, L \coloneqq \text{ maximum number of lanes in a lane section}$$

As discussed in the section “Representation of roads in OpenDRIVE®”, a road is partitioned into lane sections. Between two adjacent lane sections, the number of lanes may change. For example in lane section 1 we might have ℓ lanes, whereas in lane section 2 we might have ℓ-2 lanes, and in lane section 3 we might have ℓ +1 lanes. New lanes can even emerge between existing lanes in a new lane section and lanes between others can end. Nevertheless, in every lane section lanes are numbered consecutively from 1 to L, which means, that lane number m in lane section k is not necessarily the same “physical” lane as lane number m in lane section $k + 1$. This means that we need to jump between multiple rows in the lane center matrix to retrieve one physical lane center. In Fig. 5, this situation is illustrated by a relatively simple example. In lane section 2 a new lane emerges between lanes 1 and 2 of lane section 1. The new lane ends in lane section 2, as well. To retrieve the lane center marked in light blue, the light blue co-ordinates of the lane center matrix needs to be used.

[3] Strictly speaking, there are two base changes. One from the local co-ordinate system to the global co-ordinate system and one from the global co-ordinate system to the UTM co-ordinate system.



$$LCM_{\text{Road}} = \left[\begin{array}{ccc|ccc|ccc} x_{1,\mathcal{S}(1)} & \cdots & x_{1,\mathcal{S}(j)} & x_{1,\mathcal{S}(j+1)} & \cdots & x_{1,\mathcal{S}(k)} & x_{1,\mathcal{S}(k+1)} & \cdots & x_{1,\mathcal{S}(l)} \\ y_{1,\mathcal{S}(1)} & \cdots & y_{1,\mathcal{S}(j)} & y_{1,\mathcal{S}(j+1)} & \cdots & y_{1,\mathcal{S}(k)} & y_{1,\mathcal{S}(k+1)} & \cdots & y_{1,\mathcal{S}(l)} \\ x_{2,\mathcal{S}(1)} & \cdots & x_{2,\mathcal{S}(j)} & x_{2,\mathcal{S}(j+1)} & \cdots & x_{2,\mathcal{S}(k)} & x_{2,\mathcal{S}(k+1)} & \cdots & x_{2,\mathcal{S}(l)} \\ y_{2,\mathcal{S}(1)} & \cdots & y_{2,\mathcal{S}(j)} & y_{2,\mathcal{S}(j+1)} & \cdots & y_{2,\mathcal{S}(k)} & y_{2,\mathcal{S}(k+1)} & \cdots & y_{2,\mathcal{S}(l)} \\ 0 & \cdots & 0 & x_{3,\mathcal{S}(j+1)} & \cdots & x_{3,\mathcal{S}(k)} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & y_{3,\mathcal{S}(j+1)} & \cdots & y_{3,\mathcal{S}(k)} & 0 & \cdots & 0 \end{array}\right]$$

Fig. 5: Example of a new lane emerging between two others in road topology and lane center matrix

In order to avoid the need of considering multiple rows in the lane center matrix for one physical lane, we instead store the corresponding co-ordinates in two consecutive rows. Therefore, we first create and fill a matrix with $2l_\rho, l_\rho \coloneqq \text{num physical lanes}$ rows, where a physical lane is stored in 2 consecutive rows. Next, in order to dense the lane center matrix, we rise the co-ordinates for complete physical lanes as far as possible and the rows at the bottom of the matrix without any lane information are removed. This two-step process is in Fig. 6.

Now, the (x, y) co-ordinates of a physical lane are represented in the lane center matrix contiguously in two consecutive rows, but one row of the lane center matrix can have co-ordinates of more than one physical lane. To distinguish the physical lanes, a second matrix is built with dimension $\#(\text{rows of lane center matrix}) * \#(\text{lane sections})$, which defines the lane center matrix topology. Each entry of that matrix shows which physical lane resides in the corresponding area of the lane center matrix. The topology matrix is built during the generation of the dense lane center matrix in a similar way.

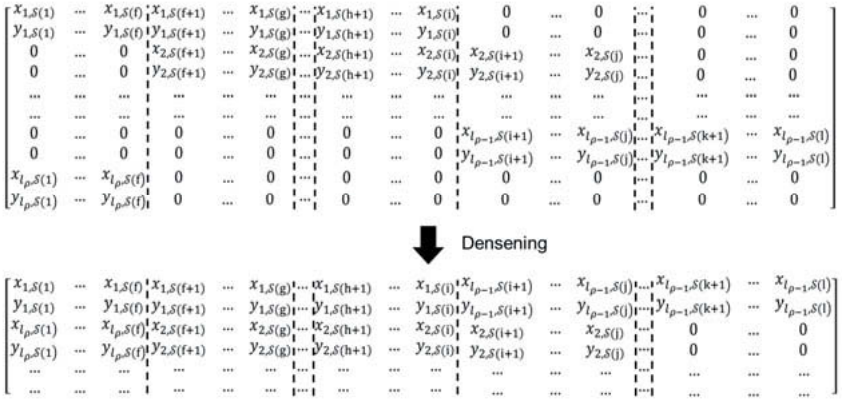


Fig. 6: Creation of the lane center matrix

The lane center matrix and the corresponding topology matrix is computed for each road existing in the OpenDRIVE® file and the resulting matrices are stacked together in a structure as shown in Fig. 7.

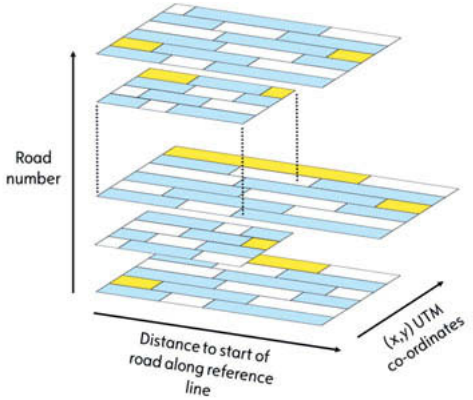


Fig. 7: Lane center structure for all roads. Blue/White areas are co-ordinate pairs for lanes, orange areas are zero.

The two data structures we have discussed so far already enable map matching of the vehicle, given an exact GPS position in UTM co-ordinates, by searching the UTM co-ordinate pair min-

imizing the distance to the vehicle position. Note, that this search does not need to be exhaustive, because most roads of the digitized map are too far away from the current vehicle location. We start by computing the distances to the first and last UTM co-ordinates per road. Next, for the nearest road (or nearest roads in case of similar distances), we perform the same step per lane section retrieving the lane section which is closest to the current vehicle location. Finally, we search within that lane section for the co-ordinate pair with minimal distance to the current location of the vehicle. Even in that lane section we can reduce the search effort by simply restricting our search space to co-ordinate pairs with $\left\| \binom{x}{y}_{\mathrm{loc}} - \binom{x}{y}_{\mathrm{laneCenter}} \right\|_2 \leq r_{\mathrm{search}}$ with r_{search} being the search radius. Map matching is an operation which is used very often, and with the data structures we discussed it is very efficient. Instead, if we want to perform map matching directly using the OpenDRIVE® file, this will be much harder and will take more time. Hence, map matching is one of the operations where we have big advantages with the data structures discussed.

Next, we need a data structure to interconnect roads and to interconnect lanes which start in one road and end in another. As mentioned in the chapter “OpenDRIVE® format” a digitized road map is represented as a set of roads and junctions in OpenDRIVE®. Each junction connects a set of incoming roads to a set of outgoing roads via so-called connecting roads. Furthermore, every non-connecting road has, at least, one connecting road as predecessor and one connecting road as successor. In other words, the road network is represented as a directed graph in OpenDRIVE®. Hence, we represent the linkage between roads and their corresponding lanes using an adjacency list $\mathrm{A}\colon (R_{\mathrm{in}}, L_{\mathrm{in}}) \rightarrow (R_{\mathrm{out}}, L_{\mathrm{out}}), R\colon \mathrm{Road}, L\colon \mathrm{Lane}$.

To be able to implement all required interfaces, more data structures than those discussed here are necessary, e.g. a data structure to represent whether or not lane changes are permitted at a given location. In this paper we will not cover these data structures, because they are not required to understand the basic concept of the chosen representation.

Offline and online conversion

Depending on the size represented by the digitized road map, the lane center structure together with the accompanying data structures may require a lot of memory. Therefore, in a first project phase, a smaller OpenDRIVE® map has been converted offline and the generated data structures were then used in the rapid prototyping environment.

To work with a larger map, that map is partitioned in so-called tiles – i.e. multiple OpenDRIVE® files – which are small enough to be processed online. The relevant next tiles, depending on the driving direction and the tile the vehicle is currently located in, are uploaded to the vehicle

if it gets close to the boundary of the current tile using an over-the-air data connection. Using the same code as for the offline conversion, the received tiles are converted online and can be used when required.

Exemplary interface: Distance to next highway exit

One common scenario in Ko-HAF is an automated ride on a highway with automatically taking a pre-defined highway exit. In such a scenario, the driver will take over vehicle control once the exit maneuver has been performed. As shown in Fig. 8, the scenario may include the automated change between several highways (from highway $\mathrm{H1} = \{\mathrm{R1}, \mathrm{R12}, \mathrm{R2}\}$ to highway $\mathrm{H2} = \{\mathrm{R3}, \mathrm{R34}, \mathrm{R4}, \mathrm{R45}, \mathrm{R5}\}$ in the example).

In order to enable the driver to safely takeover vehicle control, a distance bar is shown, indicating the distance to the point where the driver is expected to resume control. The maximum distance to be shown should be half the length of one map tile. Given the current location of the vehicle, we can compute the distance to the takeover point with the data structures defined in the chapter “Data structures for lane center representation”.

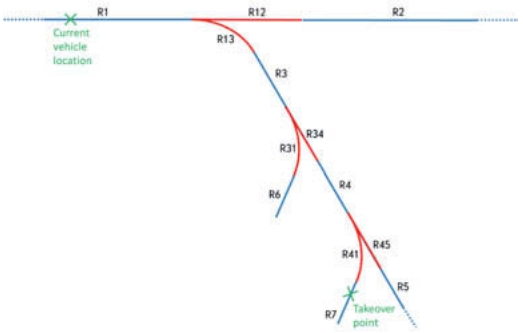


Fig. 8: Exemplary driving scenario. Normal roads are shown in blue, connecting roads in red.

First, we need to compute a path σ_m from the current vehicle location to the takeover point. As the map is represented as a directed graph and the number of roads is small enough due to the map tile size, it is possible to compute the path if the graph is either not multi-partite or both current location and takeover point are within the same partition of the graph. For such cases there exist multiple algorithms in the literature (e.g. [4]) and we will not cover the routing algorithm here, because it is out of the scope of this work. Given the path σ_m we can now compute the distance to the takeover point. The length of the roads in-between the current location and the takeover point can simply be estimated by

$$\sum_{r \in \sigma_m \setminus (\sigma_m(1) \cup \sigma_m(|\sigma_m|))} \Delta |\mathcal{S}_r|.$$

Additionally, we need to consider the distance to the end of the current road and the distance from the start of the last road to the takeover point. Hence, we need to match the current vehicle location and the takeover point, as described in “Data structures for lane center representation” in order to retrieve the corresponding column indices i_c for the current vehicle location and i_t for the takeover point. With those indices we compute

$$S_\mathrm{first} = |\mathcal{S}_{\sigma_m(1)}| - i_\mathrm{c};\ S_\mathrm{last} = i_\mathrm{t}$$

and, therefore, the complete distance from the current vehicle position to the takeover point is computed as follows:

$$D_{\mathrm{c}\to\mathrm{t}} = \Delta \left(S_\mathrm{first} + S_\mathrm{last} + \sum_{r \in \sigma_m \setminus (\sigma_m(1) \cup \sigma_m(|\sigma_m|))} |\mathcal{S}_r| \right)$$

This formula of course underestimates the length to the takeover point, because the reference line is sampled and, hence, the distance along the reference line from sample to sample is greater or equal to the straight line between the two samples. But, as most parts of the path will be highways or at least roads with low curvature, the error is not that large. Required lane changes can be respected in the above formula as well. Typically, the error we induce by neglecting required lane changes and by underestimating the road lengths is very small, because the sample size is chosen rather small. Additionally, unforeseen maneuvers like overtaking will have more impact on the total distance to the takeover point and the distance bar is updated frequently.

Exemplary interface: Distance to the lane center

Another common scenario in highly automated driving is that we need to compute the distance $d_{v\to\mathrm{LC}}$ between the vehicle center and the desired vehicle path (e.g. the lane center of the current lane).

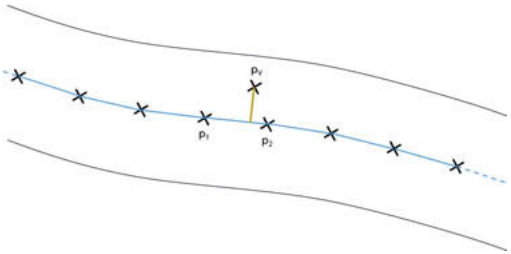


Fig. 9: Example of current vehicle location in a sampled lane

For example consider the situation shown in Fig. 9, where the vehicle center is currently located at p_v. First, we search for the nearest two lane center samples p_1 and p_2 as described in the chapter “Data structures for lane center representation”. To compute the distance from p_V to the connection line between p_1 and p_2 we use the standard scalar product $\langle a, b\rangle := a^T b = |a||b|\cos\sphericalangle(a, b)$ which is zero if and only if the vectors a and b are orthogonal to each other. We define the vector $s := p_1 + \lambda(p_2 - p_1), \lambda \in \mathbb{R}$ pointing from p_1 into the direction of p_2 and compute λ_s, such that $\langle p_v - s, p_2 - p_1\rangle = 0$, which yields $d_{v\to\mathrm{LC}} = |p_v - s|$.

Summary

In this work we discussed the basics of the OpenDRIVE® format. Furthermore, we discussed the basic data structures used to represent the high-resolution map, as given by our project partner, in our rapid prototyping environment. Using two exemplary interfaces we have shown the advantages of our data structures in comparison to use the OpenDRIVE® representation directly. The main advantage is, that for computations which need to be performed very often we can use basic algebraic operations and, hence, achieve a good runtime performance. Instead, performing those computations directly on the OpenDRIVE® map file would be more complex and would require higher effort.

The major drawbacks of this approach are that it is not as generic as the OpenDRIVE® representation and the risk that we need to maintain or change our implementation if we retrieve map files based on newer OpenDRIVE® versions which are partially incompatible with older versions.

Another downside of the data structures we discussed is that the memory consumption for the lane center matrices will be very high especially if we sample the road with a small step size. This downside can be overcome by adaptive sampling, meaning that we will increase the sample step size when the curvature of the reference line at a given location is small and when the curvature of the corresponding lane width polynomials is small as well. This would require one additional vector, containing the cumulated sample sizes per road, and would not increase the complexity of the interfaces drastically.

References

[1] "Ko-HAF," [Online]. Available: https://www.ko-haf.de.

[2] "OpenDRIVE," [Online]. Available: http://www.opendrive.org.

[3] "OpenDRIVE Format Specification v 1.4," [Online]. Available: http://www.opendrive.org/docs/OpenDRIVEFormatSpecRev1.4H.pdf.

[4] W. Zeng and R. L. Church, "Finding shortest paths on real road networks: the case for A*," *International Journal of Geographical Information Science, Vol. 23, No. 4,* pp. 531-543, April 2009.

Multilateralism at its best: A blockchain-based platform enabling data sharing, monetization and service differentiation in the automotive industry

Dr. **Karoline Bader**, **Volkmar Knaup**, **Stefan Schneider**,
Continental Secure Data Germany GmbH, Aschheim

Abstract

The sharing and monetization of vehicle sensor data represents both a challenge and an opportunity for the automotive industry. The increasing number of connected vehicles generating data, promises a high potential to establish new business models and revenues. Vehicle manufacturers need to define clear data monetization strategies and can choose from different approaches to unlock the full potential of that data. The use of data platforms as sales channel is currently a popular approach in the industry. However, before data can be monetized, there are several legal, technological and commercial requirements which need to be fulfilled. In case manufacturers aim to sell data with the support of an external platform provider, two major technological trends can be observed: On the one hand, centralized platforms and marketplaces which act as brokers and build their platform on a centralized architecture. On the other hand, there are platforms relying on a distributed peer-to-peer (P2P) network which are supported by distributed ledger technology like blockchain. This paper puts an emphasis on the latter approach and outlines how the distributed approach enables new abilities to establish successful business models. Considering a B2B-related data monetization, the distributed architecture accelerates the creation and improvement of vehicle services without the need of a central data broker. Additionally, the used technology stack for the distributed approach, enables C2B business models showing that control and power of data monetization could even be spread to each car driver himself. With such an empowerment of drivers, vehicle manufacturers can further differentiate from their competitors by delivering them a new experience.

1. Current market trends

Nowadays, the automotive industry is facing a time of dramatic change. The emergence of new digital technologies does not only transform the industry setting, but it disrupts and even attacks the traditional business models of vehicle manufacturers and automotive suppliers. However, such a change always comes with new business opportunities for the whole ecosystem. In this regard, automotive players are required to possess the right (dynamic) capabilities and understand how to pull the strings (Bader & Enkel 2014a, KPMG 2019). Especially when markets develop, separate, crash or even die over time, dynamic capabilities help firms to successfully foster such market changes and create novel resource configurations. Sensing, seizing and reconfiguration are the three capability levels of relevance. They describe a firm's ability to integrate, build and reconFig. internal and external competences in rapidly changing environments (Teece 2018).

During recent years, connectivity of vehicles has increased step by step. First, hardware components in vehicles have become more interconnected communicating with each other. In addition, manufactures have enabled selected vehicle models of their own brands to exchange data within their fleet. Second, vehicle manufactures have gone one step further enabling vehicles across their own brands to exchange relevant data with each other relating to a system of connected vehicles. Third, industry boundaries have become more and more semi-permeable leading to a system of systems (Dagnino 2004, Porter & Heppelmann 2014).

Analysts expect that in 2020, 30 million cars are sold with embedded connectivity. Taking these market developments into account, the relevance of data platforms which operate across industry boundaries is uncontroversial (SBD 2019). Data platforms are defined as products and services which bring together user groups in a two-sided network. They provide guidance, infrastructure, rules and restrictions facilitating the transactions and trading of data between different user groups (Eisenmann et al. 2006).

The automotive industry considers data platforms as one lucrative possibility to generate new business and benefit from the possibility to sell data to various firms via one platform. There are different roles around this data platform business. Some firms position themselves as (neutral) platform providers benefiting from transaction fees, membership fees and/or licensing fees. Other firms take the role of a data provider. They monetize data via a multi-channel strategy with the support of different data platform providers. Finally, some firms take the role of a data consumer and position themselves, for example, as service providers. They buy data to better train their service algorithms and offer the best possible services in the industry. Thus, they generate revenue through service licensing fees (SBD 2019).

2. Data platform business

2.1. Market requirements

There are several legal, technical and commercial requirements connected to data sharing and monetization across companies and industries. Connected vehicles generate an increasing amount of data with integrated hardware like sensors and electronic control units (Han et al. 2018; Pillmann et al. 2017). Before data can be processed, there are various legal requirements which need to be fulfilled. In the European Union, the GDPR (General Data Protection Regulation) imposes the requirement to have a basis to lawfully process data from a data subject, which relates to the vehicle driver in this case. One of the most prominent legal bases is to get the consent of the vehicle driver in order to process his personal data (EU GDPR 2016, Art. 6). As soon as the driver has given his consent, the vehicle manufacturer is allowed to collect, use and trade this data with respect to a specific purpose. This procedure is compliant with the European data protection law. (EU GDPR 2016, Art. 7)

Vehicle drivers and end users in general get more and more conscious about the value that they contribute with their data (McKinsey & Company 2016). Therefore, vehicle manufacturers need to find ways to encourage vehicle drivers to share their (personal) data with them. Incentive systems shall motivate vehicle drivers allowing manufacturers to access their data and monetize it. Before data can be monetized, it must be processed to make it readable and usable. There are various challenges regarding the collection and processing of data on the individual vehicle manufacturer side. These challenges are not further discussed in this paper, as the focus lies on the monetization and sharing of data across firms and respective opportunities for the automotive industry.

Currently, vehicle manufacturers put two approaches of data monetization into practice: First, some of them sell data by finding and addressing interested data consumers on their own. Second, others outsource this activity and use various data platforms which manage the data monetization for them. Vehicle manufacturers look into ways to monetize vehicle data, while at the same time staying in control of the data and the price determination. Several vehicle manufacturers feel confronted with the risk of not gaining the full value of the vehicle data. Instead, they are afraid that some platform providers get too much of the cake. Accordingly, trust and transparency are key aspects when data is monetized via a platform (McKinsey 2018).

On the other side of the equation, data consumers like service providers, cities and insurances aim to achieve a better coverage of data without addressing each single provider of data separately. Especially service providers face the challenge to develop and improve comprehensive services which rely on a solid coverage and penetration of data. In-vehicle service providers which offer services related to safety (e.g. road weather service) and comfort (e.g. parking service) strongly build on real-time data. Therefore, near real-time data exchange along the value chain is essential. Additionally, these in-vehicle services only achieve a high quality and confidence level if both the coverage and penetration of certain data attributes for a specific region are warranted. Often, one single vehicle manufacturer is not able to provide all the data required for such in-vehicle services. This fact shows that there are several requirements imposed from a data consumer perspective which definitely needs to be considered in the value chain.

2.2. Technological trends

From a technological point of view, two trends can be monitored in the market when it comes to data sharing and monetization via platforms. These two technological trends are illustrated in Fig. 1. At the same time, they also reflect two different commercial approaches of data sharing and monetization.

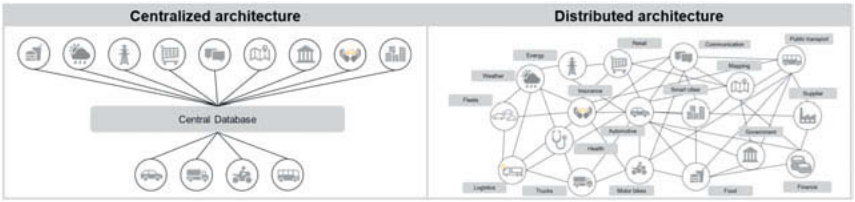


Fig. 1: Centralized and distributed platform architectures

The centralized architecture uses a server-client approach where the platform provider positions itself in the middle between the data providers and data consumers. The platform centrally coordinates the monetization of data when bringing data providers and data consumers together. Thus, the data is sent from the data provider to the platform who accesses this data and most frequently centrally stores it. The centralized data platform commonly also covers the tasks related to the standardization, normalization and quality assessment of data. Finally,

the data is offered by the data platform provider and sold to consumers like insurances, cities and service providers (SBD 2019).

In contrast, the distributed architecture brings data providers and data consumers directly together and does not require a central server in the middle. By using this architecture, a data platform provider enables a direct data exchange and bilateral transactions between data providers and consumers in a distributed peer-to-peer network (Schollmeier 2001). The data platform operates as an enabler and facilitator of the ecosystem and allows multiple parties to directly share data with each other. Since the technology rests in a distributed architecture, data consumers could either connect to a vehicle manufacturer's central data lake or even to multiple connected vehicles participating in the ecosystem. Data is not stored on a central server of the platform provider. Instead, it lies in the nature of the architecture that the control and administration is spread among the participants in the network meaning that data stays on the edge. The use of distributed ledger technology enables such a distributed architecture. There are multiple distributed ledger technologies with different maturity levels in the market. Blockchain technologies are a subcategory of distributed ledger technologies and this paper concretely focuses on the Ethereum blockchain (Wood 2017).

3. Blockchain-based data monetization

3.1. Benefits of using blockchain technology

Blockchain technology brings several benefits concerning the monetization of data. Due to the nature of the technology, it addresses several of the above-mentioned requirements. By looking at these requirements, the suitability of the blockchain technology, e.g. Ethereum, becomes obvious. In the following section, the value of the distributed architecture and the use of blockchain technology will be assessed for the above-mentioned requirements and features: consent management, incentive system, control over data monetization and service development.

Consent management: Information which is documented on the blockchain cannot be tampered by any party. This characteristic of the technology is especially beneficial by connecting it to the consent management system. By documenting the status of a driver's consent on the blockchain, it can always be tracked securely, transparently and efficiently by the involved parties. Thus, the status of a driver's consent is always shown correctly without the risk of having false or corrupt information. In this regard, every party of the value chain stays compliant with data protection laws like GDPR. Additionally, the blockchain enables traceability about

the access to data which further increases the transparency for all participants along the value chain (Miraz & Maaruf 2018).

***Incentive system for vehicle drivers*:** Vehicle drivers are increasingly conscious about their generated data and their contribution to the value chain. They expect to receive a value in return when sharing their data with the ecosystem. As a distributed data platform uses cryptocurrency to securely and efficiently manage bilateral transactions between data providers and consumers, this integrated payment system can be used to directly and seamlessly reward drivers with so-called tokens. By so doing, vehicle drivers have an own cryptocurrency wallet in their vehicle and connected car application and earn their piece of each data-related transaction. This ability represents a major differentiator for vehicle manufacturers and enhances their market intelligence by, e.g. directly getting the information for which rewards car drivers would actually share personal data. Furthermore, it provides more flexibility and room for vehicle drivers to also benefit from their contribution to the value chain.

***Control over data monetization*:** Most vehicle manufacturers are concerned about not getting the full value of the data in case it is monetized via an external data platform. Consequently, trust and transparency are obligatory to overcome these concerns. A distributed platform architecture enables the data provider to stay in control of the data monetization and pricing, as such activities stay with each participant in the peer-to-peer network. Thanks to the blockchain, the transactions directly take place between a data provider and consumer. In comparison, centralized architectures use a middle man as intermediate managing the data monetization and transactions in the ecosystem. Considering specific activities such as data normalization and standardization, such centralized systems usually operate as data controllers (EU GDPR 2016, Art. 24). Based on their positioning, they might benefit from information asymmetry and power control.

In contrast, distributed systems foster transparency and agility. Data is kept distributed in the network and stays on the edge with every data provider. This means that data is not centrally stored on a third-party's server which could somewhat lead to a loss of control. Instead, data remains solely with each data provider until a specific data consumer purchases the access to the data. The distributed data platform enables ad-hoc contracting and an end-to-end encrypted data exchange between providers and consumers of data. The platform operates as data processor (EU GDPR 2016, Art. 28). It can only access the meta data of a data stream, but not the content of the data stream as such. Only the data provider and the data consumer own this privilege after the purchase of a respective key. Such security and transaction mechanisms increase trust in the distributed ecosystem (Wood 2017, Miraz & Maaruf 2018).

Development of services: Especially service providers developing in-vehicle services to increase driver safety and comfort, heavily rely on the access to vehicle data. Besides other data sources, a road weather service can be fed with the road surface condition measured by the vehicle itself. To enable such a service, near real-time data exchange is required. By using a distributed architecture where data is directly streamed from a data provider's backend to a data consumer's backend, no middle man in-between touches the data and further reduces data stream velocity in the value chain. This enables a fast and direct exchange of data which is used to improve the quality and confidence level of in-car services. The same argumentation holds for a comfort-related service like an on-street parking service, as data exchange in network speed is key for parking on highly frequented urban streets where vacant spots are often already taken on short notice.

Because of the suitability of the blockchain technology to address the different market requirements, Continental has decided to leverage this technology for data sharing and monetization. Putting a clear emphasis on open innovation (Bader & Enkel 2014b) and cross-industry innovation with regard to radical innovations (Enkel & Heil 2014), Continental collaborates with both Hewlett Packard Enterprise and Crossbar.io on the blockchain-based data monetization platform. Illustrated in Fig. 2, the applicability and implementation of the technology stack for data sharing and monetization is shown based on two major use cases.

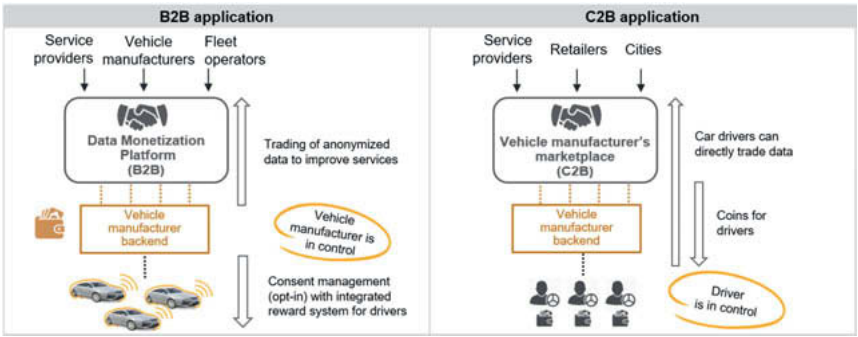


Fig. 2: B2B and B2C applications of the distributed architecture

First, companies are enabled to monetize and share data via the B2B platform. The platform brings data providers such as vehicle manufacturers and data consumers such as service providers or other vehicle manufacturers together. The platform uses the smart contracts of

the Ethereum blockchain for the registration of both data providers and consumers and for the overall payment. The focus of the B2B platform is to enable the improvement of in-vehicle services which heavily rely on both coverage and penetration of data.

Second, the usage and implementation of the technology stack is possible enabling a vehicle manufacturer to launch its own marketplace and even connect it to an overall distributed ecosystem. In this C2B application, a vehicle manufacturer can directly equip its vehicle drivers with own cryptocurrency wallets and can enable them to directly trade and monetize their data out of the individual vehicles via “Earn-as-you-Drive”. Such an application needs to be reflected regarding the architecture and computing power of the manufacturer’s connected vehicles. This C2B application refers to a second business opportunity which is not further elaborated in this paper.

3.2. Data monetization accelerating service improvement and differentiation (B2B)

Vehicle drivers need to give their consent to permit manufacturers the sharing and trading of their personal data. According to GDPR, GPS (Global Positioning System) as location data is considered to be personal data, when third parties can exactly follow a vehicle driver’s journey (EU GDPR 2016, Art. 4). This data attribute is obligatory when it comes to the improvement of both safety and comfort-related vehicle services. However, only with the consent of the driver and a clearly defined and articulated purpose, the respective vehicle manufacturer can utilize and monetize such data. An integrated consent management system where vehicle drivers are in full control of their data shall allow for an easy opt-in and opt-out via a connected mobile application or the HMI (Human Machine Interface) in the vehicle. The status, whether an individual driver has given his consent or not, can always be tracked securely via the blockchain.

For distributed data platforms, this is critical and a prerequisite to stay compliant with data protection laws like GDPR. The blockchain adds security, efficiency and transparency along the whole value chain, since the status of the consent cannot be tampered by any party in this chain. Besides, the consent management is seamlessly connected with an incentive system meaning that vehicle drivers can be rewarded by either incentives or tokens for sharing their data. After having received the drivers’ consent, the vehicle manufacturer is able to use, provide and monetize the data via the distributed platform. Data which is offered by multiple vehicle manufacturers on the platform is not centrally stored, but stays distributed on the data providers’ servers. Potential data consumers such as service providers are able to investigate the data offers via the distributed platform. Besides the investigation of sample data, further

information with respect to coverage, fleet penetration, variety of data attributes, sampling latency and other characteristics help them better assess the potential purchase.

Besides the distributed data storage, the monetization of data also remains in control of the data providers. They benefit from sovereignty and can set the prices for their data themselves. Whenever data is sold, the cleared transactions are documented on the blockchain-based data monetization platform. Nevertheless, individual transactions between the parties are handled off-chain. Furthermore, data is streamed off-chain from backend to backend. The data is fully end-to-end encrypted and is only readable when the data consumer has purchased the respective encryption key (Wood 2017). This distributed ecosystem favors the improvement and acceleration of service creation and development. An on-street parking service, for instance, showing where parking spots along the street are vacant, heavily relies on real-time vehicle sensor data. However, the vehicle sensor data from just one vehicle manufacturer often is not enough to provide a high-quality service with reliable predictions about the availability of parking spots. Since the coverage of connected vehicles of one manufacturer differs by regions, there are gaps and white spots with respect to available data. Consequently, the individual fleet penetration would not be sufficient.

By accessing data from other vehicle manufacturers and data sources like intelligent street lamps or ground sensors, an on-street parking service provider is able to buy data for regions where the current data coverage and penetration is not good and reliable enough. Such in-vehicle services are offered to vehicle manufacturers which enable their drivers to subscribe to different service packages related to safety and comfort. The applicability of the blockchain-based data monetization platform can foster service differentiation of vehicle manufacturers related to real-time service use cases. With services fed by multiple data sources, vehicle manufacturers can differentiate themselves from the competition by offering services with a high quality, coverage, reliability and innovativeness. Finally, Fig. 3 summarizes the business flows along the value chain with regard to data monetization in a distributed ecosystem.

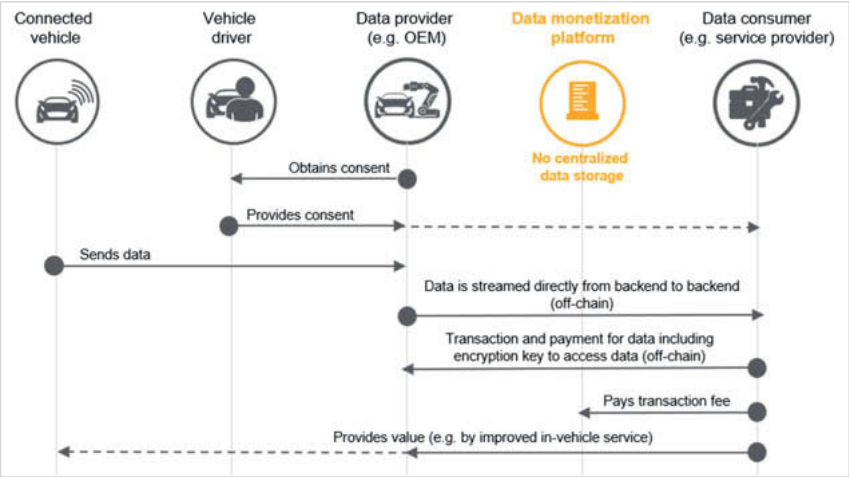


Fig. 3: Business flows for data monetization via a distributed platform approach

4. Implications and outlook

Although vehicle manufacturers might have different motivations regarding data sharing and monetization, all of them have one common goal: Fostering a driver-centric approach including the provision of the best possible service experience for their own drivers. Considering data monetization strategies, vehicle manufacturers currently have great opportunities to capitalize on their assets. They are the ones with full access to the comprehensive sets of vehicle data and with direct access to the vehicle drivers themselves (KPMG 2019, SBD 2019).

Some vehicle manufacturers put a strong emphasis on a multi-channel strategy for data monetization to generate additional revenues and cross-finance expensive vehicle sensors. Specific vehicle manufacturers have started to already apply the “NEVADA-Share & Secure” concept initiated by the VDA (German Association of the Automotive Industry) to accelerate vehicle data access. With the support of multiple neutral servers, these vehicle manufacturers offer selected data packages on the market (Reich et al. 2018, VDA 2017). Other vehicle manufacturers seem to rather strive for the implementation of an “Earn-as-you-Drive” model to differentiate themselves from competitors with such a service feature. Unequivocally, the market senses that increasing connectivity enables new business models regarding both data and service business.

Despite bright business opportunities with data and a good position regarding the vehicle data access, multiple vehicle manufacturers still face the challenge to align their backend architectures. Furthermore, it is difficult to identify which data is of value and to define the right price. Finally, data extraction from the connected vehicles demands manufacturers to take upfront investments. Currently, the automotive industry cannot tell with certainty which data monetization strategies, data models and service businesses make these upfront investments justifiable and reasonable. These days, automotive players face the challenge to balance both their traditional business and prepare for service and data business. They require new capabilities related to services, data and analytics. Furthermore, they need to take risks, be willing to cannibalize existing solutions and willing to focus on future markets. Thus, it is important to understand and internalize as quickly as possible that failing to prepare means preparing to fail (Bader & Enkel 2014a, Teece 2018).

Assuredly, one aspect of preparation considers a deeper understanding of data platform business and related market mechanisms. Current market activities highlight the emergence of an increasing number of data platforms enabling automotive players to share and monetize data and accelerate service and analytics business. Connected vehicles have already become more than just connected and smart products. By linking their own vehicle data lakes to third party servers and marketplaces, vehicle manufacturers advance the development of a product system. Related to the formation of product systems, comparable trends can be observed in other fields and industries such as logistics, production, energy, living and beyond.

However, industry boundaries have already started to expand beyond product systems to system of systems (Porter & Heppelmann 2014, 2015). The future demands further approximation and merging of currently scattered product system landscapes. Accelerating a system of systems in a distributed way is recommendable, as markets should not be dominated by monopolies. Instead, they should be characterized as autonomous, multilateral, transparent and efficient systems. Considering both a systemic and distributed view, one can take a different look on markets and ecosystems being able to move thinking from “parts to wholes, from objects to relationships, from structures to processes and from measuring to mapping” (Vargo et al. 2017).

5. Literature

Bader K. and Enkel E. (2014a). Towards service-based business models in product-centric firms: A capability approach, Conference Proceedings, International Society for Professional Innovation Management Conference, Dublin, Ireland.

Bader K. and Enkel E. (2014b): Understanding a firm's choice for openness: Strategy as determinant, International Journal of Technology Management, Vol. 66, No. 2/3, pp. 156-181.

Dagnino G. (2004). Complex systems as key drivers for the emergence of a resource- and capability-based interorganizational network, Emergence: Complexity and Organization, Vol. 6, No. 1/2, pp. 61-69.

Eisenmann T. R., Parker G. and van Alstyne M. (2006). Strategies for two-sided markets, Harvard Business Review, Vol. 84, No. 10, pp. 1-11.

Enkel E. and Heil S. (2014). Preparing for distant collaboration: Antecedents to potential absorptive capacity in cross-industry innovation. Technovation, Vol. 34, No. 4, pp. 242-260.

EU GDPR Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

Han J., Kim H., Heo S., Lee N., Kang D., Kim K., Yoon W., Byun J. and Kim D. (2018). G1 connected vehicle: An integrated vehicle information platform and its ecosystem for connected vehicle services based on GS1 standards. Proceedings of the IEEE Intelligent Vehicles Symposium (IV) 2018. Changshu, China.

KPMG (2019). Global Automotive Executive Survey 2019, released: March 2019, pp. 1-48.

McKinsey & Company (2016). Monetizing car data: New service business opportunities to create new customer benefits, released: September 2016, pp. 1-60.

McKinsey & Company (2018). From buzz to bucks: Automotive players on the highway to data monetization, released: March 2018, pp. 1-48.

Miraz M. and Maaruf A. (2018). Applications of Blockchain Technology beyond Cryptocurrency. Annals of Emerging Technologies in Computing (AETiC), Vol. 2, No. 1, pp. 1-6.

Pillmann J., Wietfeld C., Zarcula A., Raugust T. and Alonso D. C. (2017). Novel Common Vehicle Information Model (CVIM) for future automotive vehicle big data marketplaces. Proceedings of the IEEE Intelligent Vehicles Symposium (IV) 2017. Redondo Beach, California, USA.

Porter M. and Heppelmann J. (2014). How smart, connected products are transforming competition, Harvard Business Review, Vol. 92, No. 11, pp. 64-88.

Porter M. and Heppelmann J. (2015). How smart, connected products are transforming companies, Harvard Business Review, Vol. 93, No. 10, pp. 96-114.

Reich A., Krämer N. A. and Lenninger R. (2018). Fahrzeugdaten-Management. Standardisierter Zugang als Basis für neue Geschäftsmodelle, ATZelectronik, Vol. 13, 02/2018, pp. 42-47.

SBD (2019). Data monetization: Strategies for the connected car, released: June 2019, pp. 1-126.

Schollmeier, R. (2001). A Definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. Proceedings of the First International Conference on peer-to-peer computing 2001, pp. 101-102, Linköping, Sweden.

Teece D. J. (2018). Business models and dynamic capabilities, Long Range Planning, Vol. 51, No. 1, pp. 40-49.

Vargo S., Koskela-Huotari K., Baron S., Edvardsson B., Reynoso J. and Colurcio M. (2017). A systems perspective on markets - Toward a research agenda, Journal of Business Research, Vol. 79, pp. 260-268.

VDA (2017). Positionspapier - Zugang zum Fahrzeug und zu im Fahrzeug generierten Daten, accessed via: https://www.vda.de/de/themen/innovation-und-technik/datensicherheit/was-ist.html.

Wood, G. (2017). Ethereum: A secure decentralized generalized transaction ledger, released: April 2017, pp. 1-32.

AI and the Evolution of Model-Based Design

Jim Tung, MathWorks, Natick, Massachusetts, USA

Abstract

The growth of AI functionality in automotive creates a challenge in defining best practices for developing those features and integrating them in an automotive system-level development and verification workflow. There are skills gaps, differences in culture, staff preferences, and other aspects to address. To help in that transition, the widely used Model-Based Design approach is evolving in ways that accommodate and support today's AI development approaches as well as the established workflows that most organizations have.

The Value of Model-Based Design in Automotive Electrification

Over the last two decades, Model-Based Design has been a primary approach used by automotive OEMs and Tier 1/2 Suppliers worldwide to apply electrification in many automotive systems, including engine management[1], suspensions for heavy-duty trucks[2], hybrid powertrain control[3], body control[4], braking[5], and HVAC[6], as well as smaller features such as wiper systems[7].

In each of these domains, Model-Based Design involves the systematic use of models: designing systems using simulation, automatically generating AUTOSAR[8]-compliant production ECU code from models, and testing and verifying the embedded systems with ISO-26262 certification[9] (including the first Chinese company to obtain ISO-26262 ASIL D certification for a locally developed product[10]). The approach can integrate models created in different tools from multiple vendors, often created by different teams. In that time, automotive companies have developed the skills, processes, and supporting toolchains to support Model-Based Design approaches, not only within their engineering organizations, but also to clarify and streamline the critical interactions between OEM and Supplier.

Organizations that use Model-Based Design consider it an agile development workflow[11] [12] – for systems development, not just software development. It has enabled automotive organizations to jettison the traditional waterfall approach for ECU development, using simulations to clarify and refine requirements, doing rapid prototyping to quickly experiment with functions on the production hardware, and using "shift left" approaches (e.g., software-in-loop,

processor-in-loop, and hardware-in-loop) to find and resolve design and implementation errors early in the development process. Engineering groups cited in the references have been able to identify more than 95% of requirements issues before implementation, compared to 30% previously, with issue resolution as soon as six weeks, instead of a year or more [13]; reduced development time by 40-80%[14] [15]; and cut verification time by 50%[16], with certification process time reduced by 20-30%[17] [18]. Of course, the degree of benefits differs case-to-case. However, the advantages generally have been sufficiently compelling that Model-Based Design approaches are used in perhaps every automotive OEM and Tier 1, as well as a significant number of Tier 2 suppliers.

AI Moves into Automotive Systems

Today, AI systems have been introduced incrementally into the vehicle, for ADAS, prognostics, and more. That has meant new types of perception, classification, and other algorithms. Model-Based Design has evolved to support the new technologies, and it has been used effectively by engineering groups who use Model-Based Design approaches to introduce perception[19] and machine-learning based systems into automotive. Some forward-looking organizations are exploring ways to implement AI and DNN-based systems for engine control, directly using Model-Based Design[20].

However, in many situations, this is not the case. Often, the technical personnel that are tasked with creating perception and other AI algorithms don't have a background in Model-Based Design. Instead of engineering backgrounds, they often have computer science backgrounds. Instead of modelling and simulating to understand the system's behavior, they often prefer code-first approaches, supplemented by agile software methods and continuous-integration processes. This seems significantly different from the model-based approach that is now established in most automotive organizations.

At present, L2 and L2+ features can be added in a relatively isolated manner, as modules that are largely separable from other automotive functions. However, as automotive companies look forward to integrating higher levels of autonomous functionality in their vehicles, they are starting to see the looming challenges of the development workflows. In doing so, they sometimes start with a perspective that code-based approaches for AI development and the Model-Based Design approach for automotive systems development are at odds with each other.

A Path Forward

However, the Model-Based Design approach has supported code-first workflows from the beginning, over 20 years ago. That includes the ability to integrate handwritten code as a subsystem that can be simulated as part of a system model, and emitted as part of the automatically generated code. That was a requirement from the beginning of Model-Based Design adoption, since all microcontroller code was handwritten at that time. Today, most of the community isn't aware of this, because the ability to automatically generate the production code from models has provided such compelling value and because of the Model-Based Design process maturity that most companies now have.

However, the ability to integrate code in models (especially when combined with the ability to use hand-optimized libraries with automatically generated code) provides an opportunity to integrate the system development workflow that leverages the established model-based workflows of the engineering community while respecting the code-first practices of the AI algorithm developers. We are starting to work with some companies on an even more compelling vision: to align the agile-system and agile-software practices of their communities, while shifting those approaches to take advantage of hyperscale cloud resources[21].

This must be done at an organizational level, since it must account for the team skills, workflows, and types of culture in the different teams, while supporting the strategy, make-up, and personality of the organization.

References

[1][1] “Chery Enables In-House Development of Engine Management System Software with Model-Based Design”, https://www.mathworks.com/company/user_stories/chery-enables-in-house-development-of-engine-management-system-software-with-model-based-design.html

[2][2] “Continental Develops Electronically Controlled Air Suspension for Heavy-Duty Trucks”. https://www.mathworks.com/company/user_stories/continental-develops-electronically-controlled-air-suspension-for-heavy-duty-trucks.html

[3][3] “GM Standardizes on Model-Based Design for Hybrid Powertrain Development”, https://www.mathworks.com/company/user_stories/gm-standardizes-on-model-based-design-for-hybrid-powertrain-development.html

[4][4] “Lear Delivers Quality Body Control Electronics Faster Using Model-Based Design”, https://www.mathworks.com/company/user_stories/lear-delivers-quality-body-control-electronics-faster-using-model-based-design.html

[5][5] “TRW Automotive Develops and Tests Electric Parking Brake”, https://www.mathworks.com/company/user_stories/trw-automotive-develops-and-tests-electric-parking-brake.html

[6][6] “GM Engineering Europe Develops HVAC Controller for GM Vehicles Using Model-Based Design”, https://www.mathworks.com/company/user_stories/gm-engineering-europe-develops-hvac-controller-for-gm-vehicles-using-model-based-design.html

[7][7] “Mitsuba Accelerates Development of Reversing Wiper System”, https://www.mathworks.com/company/user_stories/mitsuba-accelerates-development-of-reversing-wiper-system.html

[8][8] “LG Electronics Develops ISO 26262–Compliant Power Inverter Control Software with Model-Based Design”. https://www.mathworks.com/company/user_stories/lg-electronics-develops-iso-26262-compliant-power-inverter-control-software-with-model-based-design.html

[9][9] Ibid.

[10][10] “KOSTAL Asia R&D Center Receives ISO 26262 ASIL D Certification for Automotive Software Developed with Model-Based Design”. https://www.mathworks.com/company/user_stories/kostal-asia-r-d-center-receives-iso-26262-asil-d-certification-for-automotive-software-developed-with-model-based-design.html

[11][11] “Multidomain Model-Driven Software Development at Volvo Car Group”, Jonn Lantz (Volvo Car), presented at MathWorks Automotive Conference, September 24, 2015. https://www.mathworks.com/videos/multidomain-model-driven-software-development-at-volvo-car-group-108100.html

[12][12] "Model-Based Software Development: An OEM’s Perspective“, Simon Fürst (BMW), presented at MathWorks Automotive Conference, September 24, 2015. https://www.mathworks.com/videos/model-based-software-development-an-oems-perspective-108099.html

[13][13] See Lear reference
[14][14] See Chery reference
[15][15] See Lear reference
[16][16] See Continental reference
[17][17] See KOSTAL reference
[18][18] See LG Electronics reference
[19][19] “Traffic Sign Recognition for Driver. Assistance Systems “, Dr. Alexander Behrens (Continental AG), presented at MATLAB Expo 2014. www.mathworks.com//content/dam/mathworks/mathworks-dot-com/solutions/automotive/files/de-expo-2014/traffic-sign-recognition-in-driver-assistance-systemsmatlab-at-continental.pdf
[20][20] ”Implementation development process in application of AI to in-vehicle control ECU” (translated), 横山夏軌 (DENSO TEN Co, Ltd.), presented at MATLAB Expo (Tokyo, Japan), May 28, 2019, www.matlabexpo.com/content/dam/mathworks/mathworks-dot-com/images/events/matlabexpo/jp/2019/b3-ecu-ai-implementation-densoten.pdf
[21][21] “Big Data, Data Analytics, and Machine/Deep Learning Infrastructure at Caterpillar”, Larry Mianzo (Caterpillar), presented at MathWorks Automotive Conference, May 9, 2017. www.mathworks.com/content/dam/mathworks/mathworks-dot-com/company/events/conferences/automotive-conference-michigan/2017/proceedings/big-data-data-analytics-machine-deep-learning-infrastructure-at-caterpillar.pdf

On modern automotive software development

Forever stuck in the middle?

Dr. **Riclef Schmidt-Clausen**, **Uwe Reder**, **Rainer Lange**,
e.solutions GmbH, Ingolstadt

Abstract

The field of software development is gaining importance exponentially in the automotive sector. At the same time, we see a similar growth of complexity as well as great opportunities in terms of technology and cost.

This is the advent of the hypercomplex and for the outsider somehow hard to understand brave new world of software opportunities. This triggers the desire to be able to somehow harness and control this technological paradigm change, that cannot be ignored any more. We take a look at and assess some of the currently seemingly popular approaches, which are pursued to tackle this challenge. Also, we present some successful players and their way to cope with these new challenges. Our focus is on infotainment development, since here we not only have a fast pace of innovation, a rapidly changing playing ground of technology partners and a perceived direct competition and comparison with the smartphone industry.

Introduction

A lot has been said and written about the current change in the Automotive Industry, one major driver of which are the changes in and the increasing importance of software development. The change we see in automotive software is quite different from e.g. the transformation from combustion engines to e-mobility. In the latter, one technology is replaced by the other, which in itself is a manageable process, since we roughly know, where things are going. That path is far from smooth, and the collateral effects on decade old and hardened organizations, competences and after all the employment situation are dramatic. Still, the direction is clear, measures can be taken and the whole transformation appears somewhat controllable with the tools we know from the past. On the other hand, software is less easily understood, the changes here are much more dramatic, since what is new here is growing in complexity in almost every aspect, be it technically, in terms of security, legally, financially, and recently even more politically.

So, what are the challenges, we are facing in the field of automotive software development?

On the technological side we see a trend high performance computers in the car, integrating functions that had been previously located at independent ECUs. Autonomous driving leads to unprecedented expectations in terms of safety, reliability and security. Hence, to keep up to date, not only in that field, a proper Update OTA becomes imperative. This new level of technical interaction into the driving process as well as the rising awareness and expectations on data security has led to the UNECE Task Force on Cyber Security and (OTA) software updates (CS/OTA), the recommendation of which will have a great impact on the development and maintenance processes of automotive software. To make things even more challenging, we see a dramatic increase in cross domain functions, which not only require an efficient cooperation across different car domains but as well across formerly rather independent departments of the involved players. Nevertheless, the end customer expects a high level of quality, we have to keep in mind, we are talking about a mobile device costing several 10.000s of Euros and not a smartphone, which is –despite more and more price tags approaching the 4 digits – much cheaper and typically gets replaced every two to four years.
Probably due to the rather recent rise of the importance of automotive software, the understanding of which seems to decrease on the way up on the hierarchical ladder in an industry, where mechanical engineering has been dominant since its beginning. Simple solutions, simple explanations are called for, neglecting the complexity of the issue, leading to the potentially wrong mindset and questionable guidelines and decisions.

Tradition meets Progress – and fails

The automotive industry has been growing consistently during the last 100+ years. The initial multitude of 100s of little startups as well as bigger car makers has been consolidated to a handful of big players. Until some years ago this growth has been rather evolutionary, even the advent of electronics having caused only a small ripple compared to what we see now. The challenges of the past have been met with traditional measures, mainly an increase in headcount and a broadening of the skill base necessary to develop a car. Software only played a minor role. It was specified by the OEM, developed by the supplier and typically bundled with the corresponding hardware, i.e. an ECU. Therefore, the necessary effort, techniques and challenges of the required software development processes were rather intransparent to the OEM. This shortcoming on the OEM side was understood, and met by acquisition of the necessary know-how by hiring software specialists. Software subsidiaries were widely founded, fully owned or as joint ventures, in order not only to understand, what the supplier was actually doing, but also to increase the own share in the software value chain. There are plenty of good examples, where this worked out. Almost every major carmaker today has directly or indirectly

a significant share in the software development process. So far so good, but what is going wrong right now?

Initially software in the car has been largely driven by simple and stable products, where a rather simple functionality, reliability and robustness was in focus, such as engine, body electronic ECUs or a radio. Software was a means to replace or reduce electronic hardware necessary for a functionality. Subsequently the amount of Software driven functions increased at a growing pace, and the problems began. Around this time, triggered by some severe problems in the aerospace industry, measures were called for, to control this ever growing technology. If you can't control it, or don't understand it, regulate it. So standards were derived following the paradigm of the past, by creating stable processes, proper documentation and a multitude of requirements with regards to the what, how and when of the software development process. Other fields, like the physical production of cars had shown, that this approach can be highly successful and efficient.

Things however are not working out as smooth as expected. The complexity of software functionality has grown so much, new requirements from every direction such as legal, financial, processes and last but not least new technologies, that the traditional way just doesn't fit any more. The resulting problems can be seen everywhere in the automotive industry. In the past, a car launch was dominated by mechanical issues, tolerances, reliability, surface quality and appearance. Today, software issues are a major challenge for every car launch, despite thousands of pages of process and documentation requirements and development standards. Despite all the promises at the beginning of each project, finally the car gets launched through a lot of blood, sweat and tears in the involved teams, aka overtime including weekends, budgets not met and the solemn promise to make it better the next time. With another supplier or OEM, depending on whom, you ask.

Ignoring complexity doesn't solve the problem

So what has happened? Let us take a look at some actual requirements we are faced with in the infotainment development, which were derived from other, by far less complex fields of software. We are certain that every carmaker has requirements like these or similar, since every carmaker is trying to tackle the same issues, has the same problems, as described above, so the approaches are quite similar. In a classical project, as a supplier you get a specification, ideally complete and precise without room for interpretation. You develop according to the specification, and in the end you get a result which roughly resembles what the customer had in mind.

In infotainment, you typically have an incomplete specification, about one quarter of which will have been changed through the progress of the development project. Just as well a lot of requirements refer to cross domain functionalities, which often are not really aligned across the involved departments. In an attempt to tackle this complexity, every party in the project tries to fulfil its part of the specification in order not to be held responsible. Experience shows, that about ¾ of the major problems in a project arise from cross domain requirements, so even if everyone does a perfect job in isolation, not even half the work is done once the feature complete milestone seems to have been met. This in itself is not new, everyone is trying to find a solution for this ever repeating problem. There are plenty of solutions, each of which promises to be the holy grail of automotive software development. Each of which promises to be the remedy for what we have learned to be the normal during the last decade: the infotainment industry being somewhat of a modern elite and meat grinder at the same time.
So we do a reality check on some of the requirements, a software supplier typically gets confronted with these days. Out of the described challenges and uncertainties a lot of OEMs have taken an initiative to finally control software development projects from tier1 supplier.

Automotive SPICE – the gold standard?

This process assessment model (aSPICE) is an accepted standard to rate the quality of a software development process. The brand aSPICE is owned by the VDA – the german association of the automotive industry – and accepted by many international carmakers as well. OEMs use it widely as a means of qualification of their suppliers, so subsequently a unique ecosystem of consultants has grown around it. To clarify some of the shortcomings of this idea, you may excuse the exaggeration as a means to make a point.
It is comfortable indeed for the OEM, since the perceived quality of a supplier is reduced to the aSPICE level, a simple grade replaces the need to understand the full process. The SPICE-Level tends to degenerate to something like a school degree, where no one questions the method. All other achievements of the supplier are only rated in connection with the corresponding SPICE rating. If you don't comply, everything you actually do to improve the quality of your product is considered an excuse. Even worse, your customer not only wants you to comply to a certain aSPICE Level, it is expected, that your whole software value chain complies. This in itself is a good idea, as long as you as a Software Tier1 have an actual subsupplier market with proper competition and an attractive business. The problems begin, if you have suppliers or technology partners, that do not really care about aSPICE. This is for example the case with open source software. In the past for example QNX was widely used as an operating system in infotainment headunits. However it gets more and more replaced by Linux

and Android. Unquestionably you won't enforce aSPICE here. Just as well, in infotainment the chipmakers typically derive their SOCs from Consumer Chipsets, including the necessary Software to run it, the Board Support Package (BSP). Will they adapt the processes of their main business – SOCs for Smartphones – in order to comply to a requirement of a – in terms of numbers – miniscule business? Certainly not, and this is widely accepted. There are plenty of other examples, where the aSPICE requirements or OEM specific software process requirements cannot be transferred 1:1 to infotainment projects. Metrics, which work fine in other, less complex software may be impossible to fulfil. The mandatory documentation is referring to the definition of a software unit. However there is room for interpretation, as to at which level of your software architecture a unit is defined. The smaller a unit, the greater the documentation effort in total. This however may contradict an agile development process, which is becoming quite popular in infotainment, to meet the challenge of ever changing functional requirements. It would slow down your development process considerably, if you aim too low in your unit definition. However, we have come across OEM quality requirements where exactly this happens.

A call for action

So OEM requirements often overlap or contradict aSPICE imperatives. Both of which, neither the OEM requirements nor the aSPICE standard are going in the wrong direction, quite the contrary. But in full they are not applicable for infotainment software development projects. Enforcing these onto software suppliers may have adverse effects. The supplier market could diminish, suppliers could try to achieve their aSPICE level on an isolated, smaller project or a fake process framework is set up, in order to seemingly comply. All of these resulting actions do not improve software quality at all.

At the same time we see a growing trend towards partners from the consumer electronic industry, currently popular is Google with the Android OS. This could lead to the paradox situation, that OEMs, frustrated by the experiences from the past on the one hand hinder their classical suppliers by forcing a multitude of process regulation upon them, and on the other hand surrender to aSPICE-free Android.

As a finding, we conclude that some elements of aSPICE and OEM specific software process requirements do not fit the infotainment field. It is in our strong interest however, to create a framework which does fit the infotainment world. We have to tune our processes to be fast, flexible and deliver a proper quality on time, or update timely.

Process assessment models like aSPICE mainly focus on the way, how software is built. These process requirements do make a lot of sense in the safety or homologation areas. IN the infotainment we see some newcomers like Google or Apple, which are very successful without these process requirements. In every area, where strict process models are enforced, like in the aerospace or military industry, typically budgets and timelines are not met. Think of a restaurant, where you try to assess the quality of food by auditing the process adherence of the procurement, the cook and the cleaning personnel. Process adherence is quite important, however it does not guarantee, that your food is good, on time and not overpriced.

In infotainment we feel, that we are at a crossroads: shall we keep on trying to adhere to unfitting standards and run danger of being replaced by other due to lack of speed and flexibility? Or should we try to learn from the new players in town and generate a new gold standard for infotainment development? We should use our experience across the industry in order to lead infotainment development into a new era.

An era where we are able to meet the pace of consumer electronics – in the automotive world!

The Future of Digital Car Access

Service Potentials and Ecosystem Challenges

Kai Lars Barbehön, Dr. **Olaf Müller**, **Daniel Knobloch**,
BMW AG, München

Abstract

The digitization of vehicle access represents a significant capability to transform the automotive industry from a product supplier to a sustainable mobility service provider. Access via smartphones is already an essential part of carsharing offerings to reflect the flexibility in the client / vehicle relationship. Retail vehicle equipment is also increasingly showing digital solutions based on different radio and security technologies (e.g. NFC / Bluetooth). These are all proprietary product offerings that currently do not reach the goal of high ease of use with maximum security at the same time. In addition, the penetration rates of the customer offers are limited by the lack of compatibility, as well as the necessary scalability of the architecture for mapping an overall ecosystem.

The paper describes the resulting need for standardization of digital access architectures between vehicle and smartphone manufacturers. The Car Connectivity Consortium (CCC) forms a suitable platform for this, after all well-known representatives of vehicle and mobile phone manufacturers as well as mobile service providers are represented there. The article describes a BMW standardization initiative in this field and discusses the challenges and the current state of such a standardization, taking into account technological aspects as well as issues of certification covering anti-theft requirements and the insurance industry. In addition, an outlook is given on the significant potential of such a digital platform for presenting a variety of mobility services and for further increasing the comfort of the phone / car interface through the use of innovative wireless technologies.

Digitalization of vehicle Access

Smart devices are becoming more and more a central role in our lives. Many every-day activities get consolidated into smart device functions, increasing its capabilities constantly. This development has been seen in many industries over the last decades.

Fig. 1: Digitalization of vehicle Access – Analogy to other industries

Where mechanical typewriters were used before, tablets are used today. Telephones have almost fully been replaced by wireless smartphones. Vehicles which used to be purely mechanical are being enhanced with more and more with complex electronics.
Vehicle keys have evolved from the mechanical key to electronic identification devices, adding security and comfort to the system. In 2015 BMW has introduced a Display Key, the first time adding a user interface to the key fob. The integration of a vehicle key into smartphones is the next step on this journey of digitalization of vehicle access.

Fig. 2: Digital Key Service Potentials to the customer

Development of a Digital Key

Integrating the vehicle key functionality into smartphones enables many use cases. The benefit of not having to carry a second device beside your smartphone is obvious. However Digital keys can also easily be shared between different users. It is possible to send a shared key to a friend who lives on the other side of the world as well as to a family member who sits on the other side of the kitchen table. Shared keys can be equipped with individual driving rights. The ability to limit speed or power with the assignment of a key opens up a variety of completely new business models. In one smartphone one or more digital keys may be loaded. Whether it requires to having access to all the family cars or the complete fleet of a business, it requires only one device to be carried around in order to have access to all vehicles.
A Digital Key can be integrated into smart devices like smartphones. However its digital nature offers the possibility of integrating it in many more form factors like NFC-cards, rings or watches. Together with the Digital Key comes a digital identity. The identity of a vehicle owner is clearly different from the identity of a friend who has received a shared key. Personalization to the actual driver is a key benefit of Digital Keys. BMW recognized the potential of such a Digital Key Platform, introduced a solution in the market and follows an innovative roadmap for future enhancements.

Defining a technology, enabling the vehicle access with smart devices is a challenge. Ordinary car keys are using different radio technologies as well as different protocols. It would be the straight forward solution to integrate such existing technologies into smartphones, however that would require the integration of multiple new radio technologies and protocols into smart devices. The biggest technological hurdle of this approach is, that existing key fobs use radio frequencies in the range of 100 kHz which requires large loop-antennas. Mainly because the required real estate inside a device is such an integration into smartphone complex and costly. Using existing smartphone technology would be the second straight forward way. There are various technologies available. In the market 3G/4G, Bluetooth, WiFi, are widely adopted in smartphones. Unfortunately none of those existing technologies provides the necessary security and user experience. To build a sustainable system, which fulfils the requirements of modern car makers, security needs to be at a level, that the system can demonstrate resistance to penetration attackers with high attack potential [1]. Building a system with a lower security may easily harm the whole industry. As soon as a certain market penetration is reached, the development of attacks would become profitable and seen in the field. Even a theoretical possible attack well demonstrated in a YouTube movie can create a customer reluctance, damaging the perceived security of a Digital Key solution.

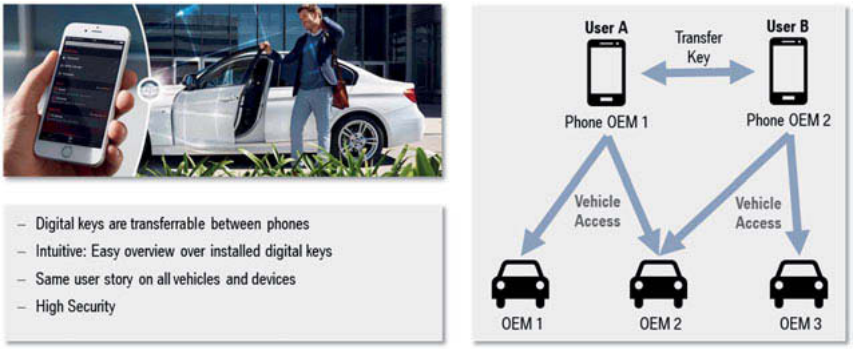


Fig. 3: Digital Key Ecosystem – the ideal world of Digital Vehicle Access

To solve this task, a brand new technology must be developed. Its design has to fit vehicle requirements in terms of security and user experience as well as smartphone requirements to allow high market adoption.

Accomplishing a Digital Key technology to be a scalable solution, it requires smartphones from different vendors to be interoperable, as well as vehicle manufacturers to agree on a common way for vehicles to communicate with smartphones. This requires the standardization of a protocol between smartphone and vehicle to access the vehicle, as well as a standardized way to transfer a key between smartphones.

The only long term sustainable solution is as an industry-wide standard.

Creating a Standard

In 2016, BMW started to develop a technology for Digital Vehicle Access by means of industry partnerships with the goal to create a global standard for Digital Keys.

Fig. 4: Development of an Ecosystem

In particular, BMW contributed the resulting cross-industry solution to the global standardization body Car Connectivity Consortium (CCC) [2]. The Car Connectivity Consortium, supported by the major smartphone and automotive manufacturers, is a cross-industry standard organization, developing standardized technologies for interfaces between vehicles and smart devices to create sustainable and flexible ecosystems for superior user experience over vehicles devices in the market.

Besides the development of a technical specification, agreed amongst all industry, the CCC is also defining a certification program, under which it is ensured that implementations fulfil all interoperability and security requirements necessary to implement this standard.

The CCC is expected to publicize this standard as its second release of the Digital Key specification towards the end of 2019.

Standardized Digital Key

The CCC Digital Key in its second release is a standardized vehicle access ecosystem, enabling to use Smart device as a car key. Upon a successful pairing with a vehicle, a Digital Key is created in the device and will allow the smart device to be used as a car key.

Between certified devices easy and intuitive sharing of keys is enabled. Every vehicle, supporting CCC Digital Key technology is interoperable with certified devices.

The essence of the standardized approach is a flexible Key management logic, enabling all relevant use cases, such as

- Access vehicle using a smart device
- Start Engine using a smart device
- Share keys with friends
- Specify access rights and driving entitlements

The key management logic is a secure software architecture. Its core is a Digital Key Applet component, designed to run in a secure element. It comprises all security relevant functionality like storage of keys, authentication algorithm and crypto functions. The Digital key Applet has direct connectivity to the NFC interface to communicate to the vehicle. A framework component is managing all supporting operations and managing the connection to the device backend. Trust is established by using a Public-Key-Infrastructure (PKI).

High scalability and interoperability is established by standardizing the two key interfaces.

Vehicle to Device Interface: This interface is used to authenticate the device towards the vehicle. A protocol is designed to provide mutual authentication with integrity, confidentiality, forward secrecy as well as tracking resilience.

Vehicle Server to Device Server Interface: To allow standardized communication between the vehicle servers and the device servers, a communication interface is defined for pairing a Digital Key with a vehicle as well as sharing a Digital Key.

The Digital Key as it is being currently standardized in CCC as a highly secure and flexible architecture. It is designed to be future proof and building on the highest available state-of-the-art security hardware found in modern devices. By its structure, the protocol separates the authentication from the underlying radio technology providing the distance commitment. The currently chosen NFC technology may be replaced by other appropriate technologies in newer releases of the Digital Key.
In future developments BMW intends to extend this architecture to support passive entry use cases, where a user can keep its key in his pocket just as in current comfort access with key fobs. For this purpose, BMW follows the line of industry partnerships with major smartphone manufacturers. As in the NFC case, the overall aim is to establish an industry-wide standardized solution in the future.

Conclusion

In this paper the developments of the digitalization of vehicle access is presented. It is shown how a standardized technology for vehicle access is inevitable for building a long term sustainable ecosystem for smartphone based vehicle access. The active role of BMW in the context of the developments in the Car Connectivity Consortium are presented and with the upcoming publication of the Digital Key specification release 2 it is shown, how this top-class goal will be achieved in the industry. The development of a flexible, NFC based Digital Key architecture is the basis of future Digital Key developments. Comfortable use cases like passive entry / passive start will be supported, as appropriate technologies become available in the market.

References

[1] ISO/IEC 15408-3 Information technology Security techniques — Evaluation criteria for IT security

[2] Car Connectivity Consortium (CCC) https://carconnectivity.org/

Potential of Training Neural Networks Using Virtual Environments

Raphael Pfeffer, **Natalya Ahn**,
IPG Automotive GmbH, Karlsruhe

Abstract

Artificial intelligence (AI) has increasingly been brought into the focus of public attention in recent years. The automotive industry can significantly benefit from this – especially in the development of autonomous driving functions and advanced driver assistance systems (ADAS). In addition to established OEMs, industry startups have been investing large sums of money and enormous efforts in the further development of this technology. Most of the major trends in the automotive industry can be supported by AI-based algorithms.

Reliable performance of autonomous driving functions requires fast and error-free acquisition of traffic and environmental situations, for instance by image processing systems. Training the algorithms entails an extremely high investment of time and costs. These algorithms acquire their capabilities through a learning process that is based on training data. In addition to the actual data, this data – at least in supervised learning – contains meta information such as the position and classification of all relevant objects. Today, manual labeling of the training data is a common practice and the reason for the very high expenditure of time and money. Moreover, a large amount of the training data is required with high variance to ensure that all conceivable environmental situations are covered.

The quality of the trained neural network depends on the quality of the training data used. Neural networks will only achieve high performance if the training data has been thoroughly annotated with utmost accuracy. Especially the annotation of data from cameras, but also other sensor data requires extremely high manpower investment. High error-proneness in the labeling process is another disadvantage of manual annotation. When algorithms for autonomous driving functions are trained based on such faulty data, it is only logical that they will not be able to accurately function in specific situations.

The annotation quality of training data can be clearly enhanced when the annotation of the (image) data, rather than being manually performed by humans, is automated, which is possible for instance with synthetically generated data. The CarMaker open integration and test platform from IPG Automotive provides error-free annotations and significantly reduce

manpower requirements. The simulation environment is able to generate synthetic data such as images, which are automatically provided with the corresponding meta information, i.e. automatically labeled.

Large training data sets can be efficiently created in this way. If necessary, the training can subsequently be fine-tuned with a reduced amount of real-world data.

Furthermore, CarMaker offers the opportunity to integrate the trained deep learning algorithms into a closed-loop environment and to test them in realistic driving situations by means of simulation. It is also possible to test and validate AI-based algorithms using ROS (Robot Operating System) middleware and to thus utilize ROS capabilities for the development of automated driving functions. This complements the simulation with CarMaker by the possibility of using existing ROS models and libraries, for instance for navigation, perception, motion planning and sensor data processing.

This paper presents ways in which the effort required for training neural networks can be significantly reduced by means of CarMaker. Based on examples, the authors demonstrate how training data sets that are suitable for training artificial neural networks can be created in simulation. Subsequently, an integration of CarMaker and ROS middleware is described as an approach to testing and validating trained algorithms in a closed-loop environment.

Artificial Intelligence and its Role in the Development of Automated Driving

In the automotive sector, environment perception, decision-making and route planning are typical areas of application for artificial intelligence algorithms. Recognition of the driver's intentions, by means of voice or gesture detection, are also typical applications for artificial intelligence. In general, integration of artificial intelligence methods provides solutions to diverse challenges that are difficult to tackle with a logic-based algorithm.

In science, artificial intelligence (AI) refers to machines trying to emulate human thought processes. Machine learning is a subcategory of artificial intelligence and describes any type of system that is able to build knowledge of its own based on its prior experience. A neural network, also known as an artificial neural network, is a type of machine learning system that models biological brain and its neurons. Deep learning refers to a neural network consisting of a hierarchy of hidden layers. The utilization of deep learning algorithms encompasses two phases: the training phase and the inference phase.

The virtual environment can be applied for both phases of the deep learning algorithms. During the training phase, the network is trained with training data in order to enable it to solve the relevant problem. A large volume of high-quality, labeled data is indispensable in this phase (as it is the case for supervised learning). Synthetic data generated in a virtual environment

has the advantage of allowing for precise labeling and high scenario coverage in a very short period of time. During the inference phase, the trained network is tested with a new data set. In the early phase, this is also done in simulations.
The following two sections explain the training phase with synthetic data, on the example of camera image data and the inference phase through the integration of a middleware in further detail.

Training of AI-Based Algorithms Using Synthetic Image Data

The emergence of more and more complex and powerful algorithms requires increasingly large data volumes that also have to exhibit specific attributes.
The so-called Big Data Challenge summarizes the requirements for data in the 5V Model [1]: “5V” stands for “volume,” “velocity,” “variety,” “veracity” and “value.” They describe the characteristics that data must have in order to be of appropriate value, also for training respective networks. Accordingly, the required volume, efficiency, suitable variety, veracity and a complete set of corresponding meta information are important in the context of this data, which entails a huge effort. Particularly the annotation of meta information required for supervised learning processes may generate a considerable manual effort, depending on the method used and the required quality.
Even with optimal tool support, the so-called labeling of data frequently entails a manual time expenditure of several minutes per image, at least when it is necessary to classify the data on pixel level (semantic segmentation). For instance, for training data sets with a size of 100,000 images, about 6,000 to 80,000 man-hours are necessary, depending on the quality of labeling and the labeling task itsself. Aside from the expenditure of time, manual annotation leads to errors. Complex annotation tasks in particular may produce an error rate in the meta data with a potentially adverse effect on the training of the AI algorithms.
The utilization of synthetic data versus real-world data for training AI-based algorithms provides various advantages. Synthetic data can be generated in large volumes, in an automated process and with a comparably minor effort. In addition, it is possible to automatically generate meta information from the ground truth boundary conditions so that, in contrast to manual annotation, no errors occur in the labeling process.

Existing Approaches of the Utilization of Synthetic Data

Various approaches already exist that demonstrate the feasibility of training a neural network with a sufficient volume of purely synthetic image data so that it will perform at least as well as a neural network that was exclusively trained using real-world data [2].

In the area of pedestrian recognition by advanced driver assistance systems (ADAS), Marin et al. showed how a virtual environment can assist in training of appearance-based pedestrian models [3]. Authors investigated the utilization of histograms of oriented gradients features and so-called linear support vector machines in order to train a classifier for the recognition of pedestrians. The study revealed that only very small differences exist between the utilization of a detector that is completely based on virtual data and was trained using the Half-Life 2 engine and a detector based on real-world data. Similar techniques were employed with a different data set in the example described in [4]. In this configuration, augmented reality training sessions were used instead of completely virtual or completely real data in order to train a classifier for the recognition of pedestrians. For this purpose, virtual objects were rendered and placed in front of a real-world image background in order to enhance recognition performance. This study showed that in this case the classifier that was exclusively trained with augmented reality data does not achieve the same performance level as the classifier that was trained with large volumes of real-world data.

Even so, complementing real-world data by augmented reality can enhance the recognition performance of the classifier. Rajpura et al. showed in [5] that the addition of synthetically generated images to the real-world data set significantly enhanced recognition performance. Remarkably, this effect was achieved with a relatively small volume of real-world data. Only 500 real-world images, complemented by 4,000 synthetically generated ones, were used. With larger data sets, reduced accuracy was observed, which suggests overfitting to synthetic data.

Labeling in a Virtual Environment

This work shows how images resulting from a simulation can be used for training of AI-based algorithms. For this purpose, the CarMaker open integration and test platform is employed. With the simulation solutions of the CarMaker product family, it is possible to run virtual test drives representing real-world driving scenarios. The environment – including the road, road users and dynamic objects, the surroundings, etc. – can be configured as needed. The fully parameterizable vehicle model, the so-called virtual prototype, is a virtual model of a real-world prototype. All real-world components of the vehicle are represented by a corresponding model in the virtual world.

While the simulation is running, a visualization integrated into the simulation environment generates a 3D animation of the environment and its output can be used as the basis for the training. In addition, it is possible to automatically co-generate other training data during the simulation such as Lidar point clouds or time series signals. Complete information about all included objects (ground truth) can automatically be included in the data as well.

Figure 1 shows examples of two typical annotation methods that can be automatically generated from the visualization. In the bounding box annotation (Figure 1, right), the objects contained in the image that are relevant for the classifier of the AI-based algorithm are marked by a rectangular shape. The labeled training images automatically generated in this way can subsequently be exported frame-wise and with additionally required meta information such as the position of the bounding boxes in the image.

Thus, all marked objects can subsequently be distinguished from each other such as passenger cars, pedestrians, bicycle riders, traffic signs, etc.

The pixel-accurate semantic segmentation (Figure 1, center) follows the same basic principle. Here, based on the previous freely defined object classes of the simulation, a color code is allocated to all objects contained in the image. In this way, all object classes are depicted according to the respectively defined color scheme as early as in the rendering process of the simulation. Together with the “original” synthetic image (Figure 1, left) the meta information that has automatically been gained in this way can be used for the training as well.

Thus, the generation of the image data and annotation of the required meta information can be fully automated and result in major time savings, as tens of thousands of annotated images from synthetic scenarios can be generated overnight.

Fig. 1: Synthetically generated image (left), pixel-accurate semantic segmentation (center) bounding box annotation (right)

In a study performed by Pfeffer et al. [6] the performance of the bounding box annotation in CarMaker, for one, was compared with other synthetic image sources and, for the other, it was investigated how well the synthetically generated images performed in training of a neural network compared with real-world images (see Figure 2, left). The investigation showed that in spite of a rather average performance of the simulation environment in comparison with image sources of the video game engine GTA V and Carla[1] a good level of performance

[1] http://carla.org/

(median average precision in object detection of the neural network) was achieved in the training process. This was the case particularly when a small volume of real-world images (5%) was mixed with the synthetic training data (see Figure 2, right).

The study also showed that due to the use of CarMaker as the synthetic image source it was possible to reduce the effort (costs) of the training by several factors.

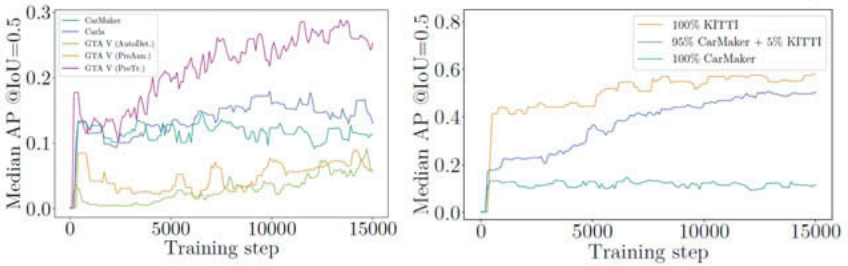


Fig. 2: Results of CarMaker in comparison with other synthetic image sources and annotation methods (left); performance in comparison with real-world data (KITTI[2]) and mixed data (right) [6]

Testing of AI-Based Algorithms

In addition to the possibility of efficiently generating large volumes of annotated training data for AI algorithms, the use of simulation enables the testing of previously trained algorithms. This phase is also referred to as the inference phase. The test and validation can be performed in CarMaker itself as described in [7], where a lane keeping assist system trained on real-world data was successfully tested in the virtual environment of CarMaker.

According to the use case of the developed AI algorithm, diverse scenarios can be generated in simulation and the algorithm tested for correct functional performance. The simulation also allows for the use of variation techniques that make it possible to achieve a higher degree of test coverage. For instance, variations of dynamic and static objects in terms of color, maneuvers or size within the basic scenario are possible by modifying just a few parameters in CarMaker. In addition, it is possible to change both the layout of the route and the prevailing weather conditions, for instance in the form of the illumination of the scenes. In combination, this results in a large number of scenarios for automated functional testing of AI-based algorithms.

[2] http://www.cvlibs.net/datasets/kitti/

Furthermore, testing and validation can be performed by means of a middleware framework, the so-called robot operating system (ROS). ROS and its integration in CarMaker will be explained in greater detail below.

Overview of ROS

ROS refers to an open-source framework [8] that has come to be regarded as a standard platform for the development of robotic applications and whose scope of application has recently expanded significantly. Developers of advanced driver assistance systems and automated driving functions in particular benefit from ample opportunities to support their developments by means of ROS. Algorithms, supported by existing ROS models, software packages and libraries may be directly integrated into ROS. In addition, many other ROS packages and predefined messages (topics or services) are included. Aside from hardware abstraction and package administration, another component of ROS lies in the communication between processes (nodes) – typically operating in parallel – that are executed either on the same system or on different systems (connected to each other via Ethernet for example).

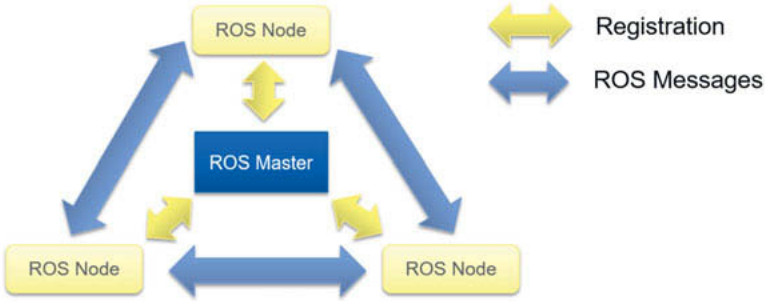


Fig. 3: Generic ROS network composed of three ROS nodes

Figure 3 shows a generic ROS network with three ROS nodes. The ROS nodes register with the ROS master and communicate with each other via topics or services. The ROS services can be used for synchronous or asynchronous bi-directional communication between two nodes (e.g. status query or reset), while the ROS topics offer the opportunity to send data from one ROS node (publisher) to other ROS nodes (subscribers). These are therefore used as the central means of communication within ROS.

CarMaker Interface with ROS

In CarMaker, an interface (CarMaker ROS Interface) is available for interacting with a ROS network. It represents a direct interface between CarMaker and ROS and is based on a ROS node that has been integrated into the CarMaker application. This simulation-internal ROS node is programmed directly as a ROS package (using basic ROS tools) and started together with CarMaker. Thus, communication within the ROS network is achieved via standard ROS mechanisms. The integrated ROS node acquires all data generated for the virtual vehicle, such as data from sensors and of the driving condition, and transmits it to an external ROS node via ROS topics. Such a ROS node could contain an ADAS/AD function or a trained AI-based algorithm to be tested which, in turn, transmits control commands to the virtual vehicle. In addition, a complete autonomous software stack can be developed in ROS and linked with a virtual prototype via the CarMaker ROS interface, see Figure 4.

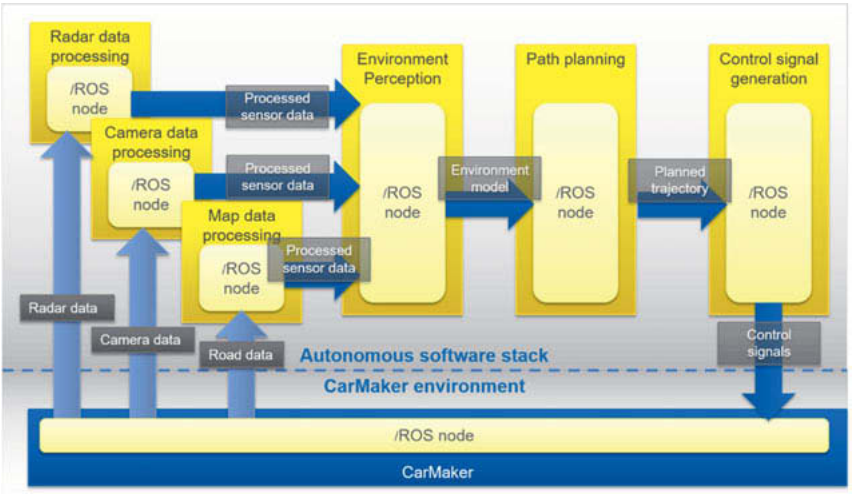


Fig. 4: Testing and validation of AI-based algorithms by means of CarMaker and the ROS environment

The CarMaker simulation environment in which CarMaker serves as the simulation platform and a ROS node which, as described above, is directly integrated into CarMaker as a shared library are depicted at the bottom of the figure. Above CarMaker an exemplary autonomous software stack is depicted that contains all algorithms enabling automated/autonomous driving.

CarMaker provides the input data for the individual sensor data processing algorithms such as Radar, camera and map data. Other sensor data such as ultrasonic and Lidar are conceivable as well. The data is processed by the individual algorithms with object lists being created from the raw data in most cases. As an example, an AI-based object detection algorithm, which is trained based on synthetic image data (see previous section), can be integrated in the camera data processing ROS node for testing. The environment perception block could then perform the sensor data fusion, localization and object tracking. The sensor data fusion, which receives the individual processed sensor signals as input, merges individual object lists and map data to create a unified environment model. With its help an autonomous vehicle can identify its location, the locations of other road users and, accordingly, the areas in which it can travel without risk.

Based on this environment model, the next possible maneuver can be determined in the path planning process. Based on this maneuver and the environment model, the final trajectory including the speed curve that the vehicle is supposed to follow is planned. In order to be able to follow this trajectory, the control signals for the vehicle have to be determined. They are calculated by the algorithm for the control signal generation and subsequently sent to the ROS node inside CarMaker. CarMaker then converts these control signals into a vehicle movement based on the respective vehicle model.

The fact that also reactions of road users who are being simulated in the CarMaker simulation environment can be observed is a major advantage of such closed-loop approach. Furthermore, the sensor signals directly depend on the vehicle movement and therefore always match the environment. This is not the case with an open-loop simulation that is only based on measured data.

Additionally, with this type of simulation environment it is possible to separately look at and test individual subareas of the algorithms from the autonomous software stack. In this case the processing of the sensor data and the creation of the internal environment model can be skipped. The algorithms for the path planning and control signal generation are simulated and tested based on the environment model provided by CarMaker.

All blocks of the presented exemplary autonomous software stack can be supported by AI-based algorithms, which can be developed and tested separately and in parallel in a distributed ROS network. To analyze the simulation, both the mechanisms of the CarMaker simulation environment (e.g. 3D visualization and signal recording) and ROS (e.g. rqt Topic Monitor, rosbag for data logging) can be used.

Summary

AI-based algorithms are increasingly being integrated in the automotive sector, particularly in the development of automated and autonomous vehicles. A virtual test environment can support the development of AI-based algorithms both in the training and in the inference phase. Thus, simulation has become a major and indispensable element of the development of autonomous vehicles.

This paper has illustrated how CarMaker can be used for training of deep learning algorithms with accurately labeled synthetic image data. Subsequently, it showed how CarMaker and ROS can be coupled together to test and validate trained AI-based algorithms.

References

[1] Mrozek, D.: Scalable Big Data Analytics for Protein Bioinformatics, Efficient Computational Solutions for Protein Structures, Springer Nature Switzerland, 2018

[2] Johnson-Roberson, M., Barto, C., Mehta, R., Sridhar, S., Rosaen, K., Vasudevan, R.: Driving in the matrix: can virtual worlds replace human-generated annotations for real world tasks? In: Proceedings of International Conference on Robotics and Automation (ICRA), 2017

[3] Marin, J., Vazquez, D., Geronimo, D., Lopez, A.M.: Learning appearance in virtual scenarios for pedestrian detection. In: IEEE Computer Vision and Pattern Recognition (CVPR), 2010

[4] Nilsson, J., Fredriksson, J., Gu, I.Y.-H., Andersson, P.: Pedestrian detection using augmented training data. In: 22nd International Conference on Pattern Recognition (ICPR), 2014

[5] Rajpura, P.S., Bojinov, H., Hegde, R.S.: Object detection using deep CNNs trained on synthetic images. In: Computer Vision and Pattern Recognition, 2017

[6] Pfeffer, R., Bredow, K., Sax, E.: Trade-off analysis using synthetic training data for neural networks in the automotive development process, Accepted paper at IEEE Intelligent Transportation Systems Conference (ITSC), 2019

[7] Innocenti, C., Linden, H., Panahandeh, G., Svensson, L. and Mohammadiha, N.: Imitation Learning for Vision-based Lane Keeping Assistance, in Proc. of the International Conference on Intelligent Transportation Systems (ITSC), 2017

[8] The Robot Operating System (ROS)

Mission AI in Automotive

Collaboration Models and Functional Safety

Dr. **Ulrich Bodenhausen**, Vector Consulting Services GmbH,
Ulrich Bodenhausen AI Coaching, Stuttgart

Abstract

AI offers great potential for many automotive applications. Collaboration models of several partners will be an important way for the realization of automotive functionality with AI. The big challenge in the collaborative development of AI based systems is to share resources on the one hand and on the other hand reduce the risk from the fact that the AI partner will gain domain knowledge right from the heart of the business case. We propose three solutions to this, which should be applied in parallel.

1. AI in the Automotive Industry

Artificial Intelligence (AI) offers great potential not only for autonomous vehicles, but also for many other automotive applications, where an AI-based algorithm can be used to augment a car functionality by usage of available or new sensor data. Voice assistants are one example of an already common application of AI in cars. According to a recent report about the market in the U.S. [1], voice assistants are already used at least daily by approx. 27 million users. Concerning the usage of devices, consumers say that they are equally likely to have used voice through the voice assistant native to the car and a smartphone connected to the car via Bluetooth.

How can AI be used to augment automotive functionality? Existing ECUs can be extended by additional sensors combined with AI to enhance the functionality or diagnostic functions can be extended by AI to improve diagnosis. Figure 1 shows both principles.

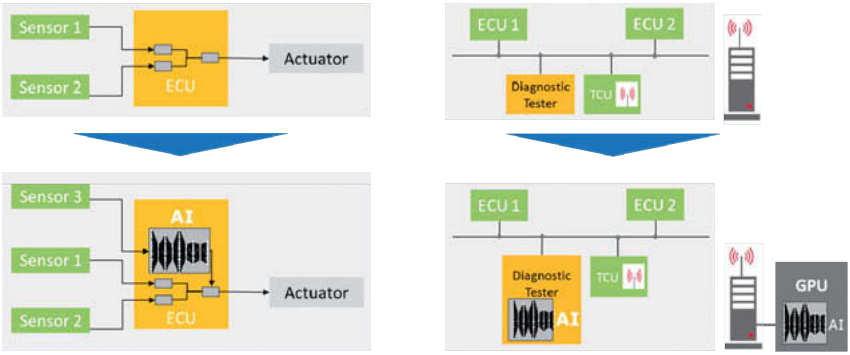


Fig. 1: AI offers great potential to enhance automotive functionality. Left side: Existing ECUs can be extended by additional sensors combined with AI to enhance the functionality. Right side: Diagnostic functions can be extended by AI, both in the diagnostic tester as well over the air.

In practice of real-world AI applications, the optimization of the performance is very important to achieve a functionality that meets the expectations of the users. Example: In case of a voice assistant, the number of languages that are accepted, the size of the vocabulary, the recognition accuracy in various noisy environments are important criteria for user acceptance. To meet the expectations of the users, functional improvements of the AI module by optimization of internal architectural parameters (such as number of hidden layers and detailed connectivity of Deep Learning Neural Networks) and big data for the training are key success factors for excellent performance. This results in a big increase in the required resources [2]. The steps for the successful development of AI-driven applications and the implications on resources and time are shown in Figure 2.

Fig. 2: Seven Steps for Your Success with Development of AI Applications. Note the high increase of resources shortly before market entry.

In case of safety critical systems, one big challenge is to provide the safety argument for systems using AI/ Machine Learning. The recently published standard ISO/PAS 21448 “Road vehicles - Safety of the intended functionality” (SOTIF) [3] was developed as a standard for road vehicles that - besides other techniques - use functions that use AI/Machine Learning algorithms. However, there is not yet much practical experience with this standard, and it is clear that it will require expertise and resources to provide the safety argument according to the standard.

In summary, AI provides great potential, but it will require specific AI-know how, intensive data collection and functional improvement as well as expertise and resources for safety assurance. It is likely that a collaboration of several partners is needed to bring together the required expertise and resources.

2. Collaboration Models and the “AI Collaboration Dilemma”

Building up all required competencies and resources in the own company will likely take too much time. Therefore collaboration with partners will be a promising approach. The automotive industry has several decades of experience with collaboration models. Approximately 25 years ago many suppliers faced the challenge to develop from a component supplier to a system supplier. Collaborations of several partners were established to bring together the required competencies. Long-term experiences are that these collaborations can be very stable if there is a real technical benefit for product, a real value contribution for all partners (and not only for some) and that the culture of the participating parties fits together.

For the collaboration aiming to bring in AI competencies, potential partners can come from the whole spectrum from start-up companies to established big AI-driven companies. There is not a long history of experiences concerning this, but in some cases it has been clearly observed that business value indicators of the collaboration partners have changed significantly over the duration of the collaboration, see Figure 3.

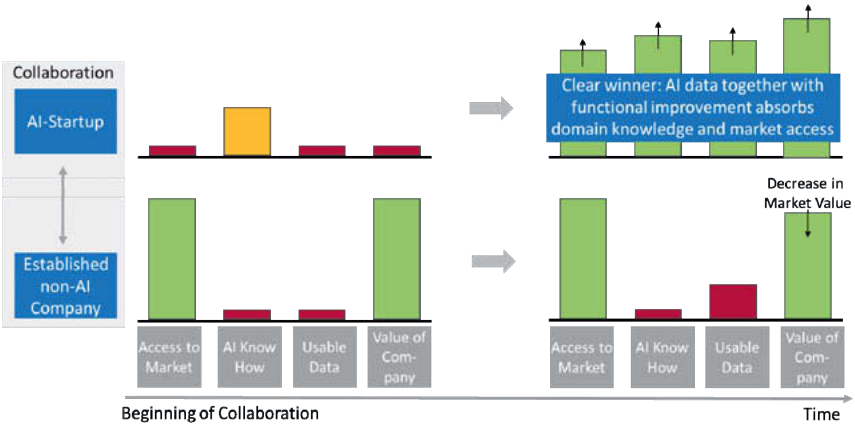


Fig. 3: Emergence of business value indicators in a collaboration involving AI. At the end of the collaboration the AI-Startup is the clear winner, because important domain knowledge was absorbed and good access to the automotive market was achieved. The effect dan be called “AI Collaboration Dilemma”.

There is a significant probability that the AI-competent partner in the collaboration will absorb important domain knowledge and market access and therefore will clearly be the winner. The effect can be characterized as the “AI collaboration dilemma”: The AI partner gains domain knowledge right from the heart of the business case. The reason is that the AI partner will work intensively with the data and AI algorithms to make the algorithm extract valuable information. By doing this the partner will learn a lot about the business, in a similar way the algorithms learn from the data. This cannot be well separated.

3. One Challenge and Three Solutions

The big challenge in the collaborative development of AI based systems is to share resources on the one hand and on the other hand reduce the risk from the AI Collaboration Dilemma. We propose three solutions to this, which should be applied in parallel:

3.1 Distinguish Clearly Between Competitive Activities and Sharing Activities

All AI development and SOTIF activities should be assignment to either sharing activities (activities that should be shared in the collaboration) or competing activities (activities where there should be a beneficial level of competition between collaboration partners), see Fig. 3.

The collection of data for training and testing can be well shared among partners to increase efficiency of the overall team, without the risk of one-directional absorption of domain knowledge by one collaboration partner. All SOTIF evaluation activities must be done as sharing activities, because the whole system must be considered as one safety item.

The conclusion, if the residual risk is acceptable, must be done by every participating partner separately. All functional improvements are the core of the competitive activities. Those partners, who do not already have the competencies for functional improvement, should extend them to become a competitive partner in the collaboration.

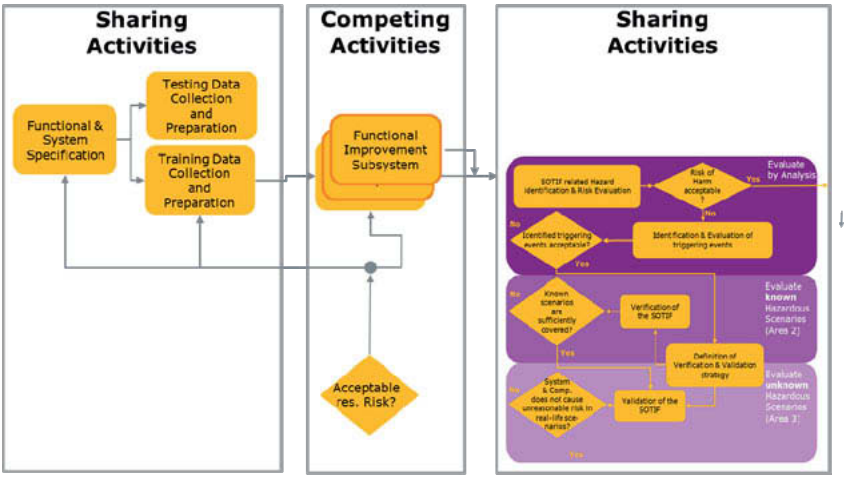


Fig. 3: SOTIF activities extended with activities for data collection and preparation and assignment to sharing activities (activities that should be shared in the collaboration) and competitive activities (activities where there should be a beneficial level of competition between collaboration partners).

3.2 Achieve Competitiveness of Parties and Competition Between Parties

AI competencies should be enhanced at all participating parties, even if there are dedicated AI specialists in the collaboration. This will reduce the risk of the AI collaboration dilemma. Important AI developments in the U.S. have in many cases been carried out as competition of teams (examples: DARPA Grand Challenge and Urban Challenge, Spoken Language Programs). It is also a frequent method to have multiple teams working on same topics at top universities.

3.3 Ensure Technically Diverse Redundant Submodules to Support Safety Argument

If several submodules are already available from the competitive approach described in paragraph 3.2, then they should be implemented as diverse redundant submodules to improve the performance and support the safety argument. The background is that redundancy is a common method to enhance safety. Example: ISO 26262 considers ASIL decomposition as a concept to provide the functionality of one system with high ASIL level by two subsystems with lower ASIL level under the condition, that sufficient independence of the elements is assured. However, independence of two learning subsystems using the same training data needs tailored algorithms to really assure sufficient independence, because the data is shared. Independence of a rule-based subsystem and a learning subsystem also needs consideration, if both are optimized on the same set of known data/scenarios. Therefore appropriate tools for the measurement and control of independence are needed.

4. Conclusions

AI offers great potential, not only for autonomous vehicles, but also for many other automotive applications. Recent numbers from a survey on the usage of voice assistants in cars shows the dimension of AI-driven applications for the automotive industry (27 million “at least daily” users in the U.S.). Beneficial and safe application of AI is an important mission for the automotive industry.

Collaboration models of several partners will be an important way to augment automotive functionality with AI. In order to make the collaboration beneficial for all participants and reduce the risk of the AI Collaboration Dilemma, resources should be especially shared for data collection and all SOTIF related evaluations. Functional improvements should be organized to keep a beneficial level of competition in the collaboration. If several submodules are already available from the competitive approach, then they should be implemented as diverse redundant submodules to improve the performance and support the safety argument.

[1] Kinsella, B., Mutchler, A., “In-Car Voice Assistant Consumer Adoption Report”, January 2019, voicebot.ai

[2] Bodenhausen, U., “Quick Start with AI for Businesses”, MLConference, June 2018

[3] ISO/PAS 21448:2019, Road vehicles - Safety of the intended functionality (SOTIF)

Engineering and Hardening of Functional Fail-Operational Architectures for Highly Automated Driving

Identifying and shaping the operational design domain

Dr. **Rasmus Adler**, Dr. **Daniel Schneider**,
Fraunhofer IESE, Kaiserslautern;
Takeshi Fukuda, Hitachi Automotive System Europe GmbH

Abstract

Rising automation levels in the automotive domain call for a shift from the fail-safe to the fail-operational paradigm. Fail-operational architectures and behaviors are, however, inherently more complex and thus require special diligence in order to assure safety. To this end, we present a methodology that facilitates the design of fail-operational architectures from early design stages on, by enabling informed assessment regarding the fitness for purpose of gradually evolved architectures. The method specifically considers resilience regarding dynamic changes in environmental conditions (including V2X aspects) as well as in internal capabilities.

Introduction

Engineering adequate fail-operational architectures and behaviors constitutes a key challenge in automotive safety engineering for highly automated driving. In contrast to the fail-safe paradigm, which is the standard today and which is usually adequate for human-operated systems, fail-operational architectures are typically far more complex and demanding to engineer. This is due to the fact that a fail-operational architecture must actually realize a higher level of automation itself. To be fit for any conceivable contextual situation, it needs to possess adequate perception capabilities and be able to plan and determine its exhibited behaviors. In other words, an appropriate fail-operational architecture needs to be resilient with respect to any contextual situation, as it must always guarantee safe operation.

Adaptation with respect to the current driving situation is not a new topic. There are different calculation variants to determine a physical state variable like the speed of a car. Typically, a variant is valid only in some driving situations, which is due to underlying calculation assumptions like “wheels are not spinning”. For this reason, even in the case of non-automated driving we have to select several variants at design time and implement an adaptation behavior that activates always those variants that fit the current situation best. In some rare situations,

we may have to choose variants that result in limited perception capabilities of the vehicle. This has consequences for trajectory and maneuver planning, and we have to choose degraded modes of operation, such as a mode where passing is not allowed. Consequently, the adaptation behavior defines under which conditions the vehicle is normally operational, degraded operational, and non-operational. This is therefore related to the operational design domain (ODD), which is defined in the standard J3016 [9] as “the specific conditions under which a given driving automation system or feature thereof is designed to function, including, but not limited to, driving modes.” More precisely, J3016 stipulates that an ADS may have one or more features, with each feature having exactly one ODD. It recommends specifying the ODD but provides limited guidance for identifying the specific conditions and thus for assuring their completeness. Our approach addresses this limitation and aims at systematic engineering of an adaptation behavior that maximizes the ODD on the one hand and minimizes the costs for sensors and related perception algorithms on the other hand. The fundamental idea for optimizing adaptation behavior with respect to availability of features and costs is described in detail in [2]. In a recent project [3], we enhanced this approach and demonstrated its practical suitability for identifying the ODD.

In the following, we will present the most important aspects of our methodology and summarize our experiences in applying it to a case study. First, we will introduce an architecture modeling approach and elaborate on the challenge of adapting architectural elements. Next, we will discuss adaptation behavior modeling and present enhancements for modeling the operational design domain. Then we will present automated analyses for optimizing the adaptation behavior. Finally, we will present related work before concluding this paper.

Architecture Modeling

We focus on the adaptation of calculation variants in order to maximize the availability of features in the case of critical events, including environmental conditions such asa smudged camera lens, internal failures, or V2X problems. In the following, we will first propose an approach for modeling calculation variants and introduce related terms. Afterwards, we will discuss its application in our case study and the problem that motivates the modeling of adaptation behavior, which we will then present in the next section.

Fundamental modeling approach and terms

Our approach builds on a common actor-oriented modeling approach[4]. In actor-oriented modeling, “boxes” are connected via ports. UML[5], SysML [6], Simulink [7], and many other languages support this kind of modeling. For instance, SysML has “internal block diagrams”, while UML has “component diagrams”. SysML uses the term “block”, whereas UML uses the term “component”. In the following, we will call them **functions** as this fits our scope best. A function receives input information via **input ports** and computes output information, which it provides via **output ports** to other functions. Functions can be hierarchical, that is, composed of other functions. For modeling a composition, we connect output ports with input ports. We call functions that are not refined **basic functions**. Basic functions comprise algorithms (calculation variants) and switch between them to compensate for the lack of some inputs or insufficient input quality. A hierarchical function thus has different states depending on its directly and indirectly contained basic functions. We call these states configurations. The hierarchical function at the top level is the system and its configurations are the **system configurations**.

Example

Figure 1 exemplifies the modeling of a basic function “ego speed” using the tool Enterprise Architect [8]. The function shall calculate the speed of the vehicle from the GPS, camera images, or wheel speeds. To this end, it has five algorithms. The first one uses only the GPS. The second one uses only information from the camera. The third one uses the speed of all four wheels. The fourth one uses only the speed of the rear wheels. The fifth one uses only the speed of the front wheels. We assume that we know that these are all sensors and algorithms that may contribute to the availability of user-perceivable features including safety-critical ones. Otherwise, we would have modeled further sensors and algorithms. In this way, we define the design space for our fail-operational architecture.

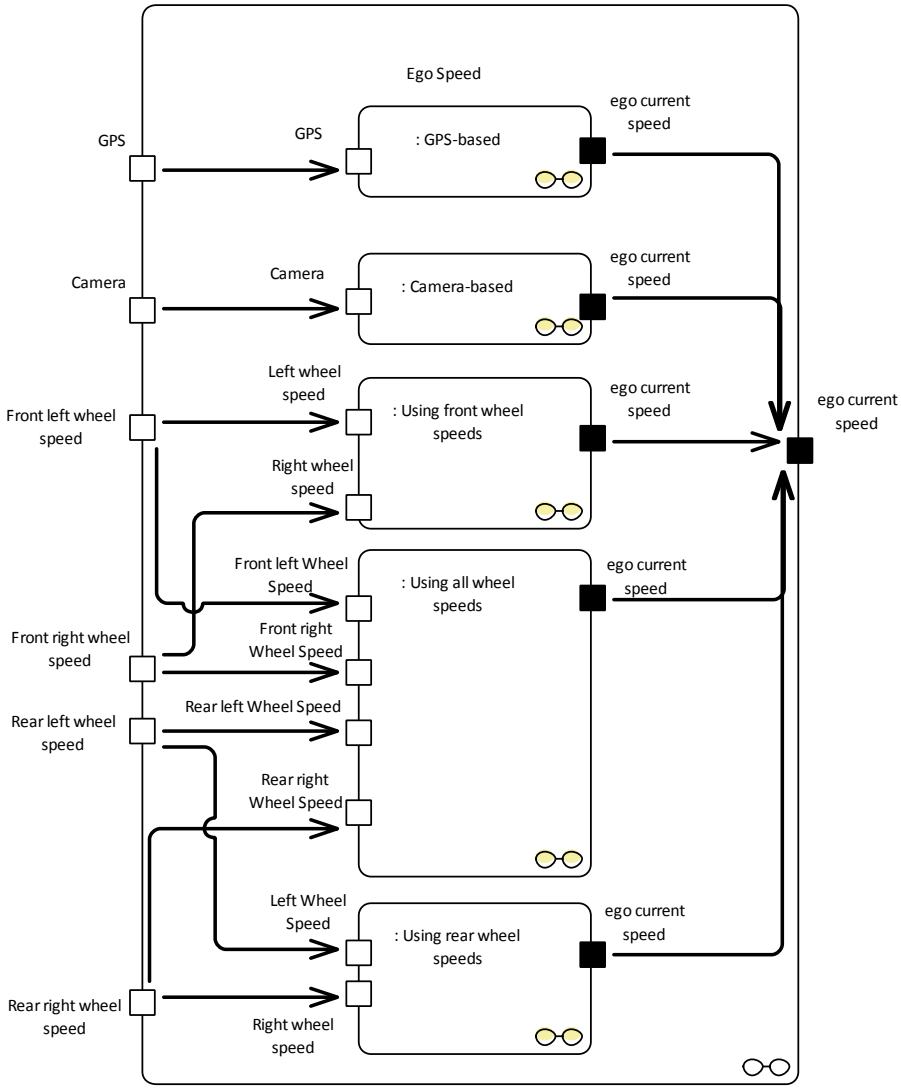


Fig. 1: Example of a basic function with five algorithms

As shown in the figure, an algorithm may use only a subset of the inputs. Furthermore, we allow that it provides only a subset of outputs because the functions requiring the missing

outputs might compensate for such unavailability by means of adaptation. If no algorithm is applicable, we have to shut down the function. In many cases, the adaptation behavior of the other functions can completely compensate for such a shutdown. Thus, it is a reasonable default state, which we will consider in the following as an algorithm requiring no inputs and providing no outputs. By default, a function therefore has at least two algorithms.

Problem

The adaptation behavior has to define (1) which system configuration should become active depending on the currently available system inputs and their qualities, and (2) how to switch between system configurations at runtime. The first task is complex due to an explosion of possible system configurations. In our last case study, for instance, we had 34,560,000 system configurations; in previous evaluations [1]of the approach, we had 10^{72} system configurations. The second task is complex because complex reconfiguration chains are necessary to avoid invalid states. If one system input becomes unavailable, then first the functions that use this input adapt their algorithms to compensate for this loss as best they can. If they cannot compensate for it completely, then the connected functions have to adapt, and so on. In many cases, some functions need to be adapted several times due to feedback loops. Without modeling and analysis support, it is hardly possible to cope with this complexity.

Adaptation Modeling

In order to cope with the explosion of system configurations, our approach uses modularity and hierarchical abstraction. The idea behind modularity is to constrain and define the adaptation behavior at the lowest hierarchy level and completely avoid the complexity of combinations. The idea behind hierarchical abstraction is to reduce the number of possible combinations at every hierarchy level so that an explosion of system configurations is avoided. However, in the following we will concentrate on modularity because it is a prerequisite for implementing hierarchical abstraction.

First, we will introduce an adaptation type system that allows modular specification of pre- and post-conditions for each algorithm. Second, we will explain the modeling of pre- and post-conditions. Third, we will discuss the prioritization of algorithms in order to maintain a deterministic adaption behavior if several preconditions are fulfilled. Finally, we will reflect on the trigger for adaptation and introduce further triggers.

Adaptation type system

The adaptation type system shall extend the normal interface of the function so that every algorithm can specify its preconditions for becoming active and its post-conditions that need to be considered by other algorithms.

To this end, it introduces a **semantic type** for every port. The semantic type defines the semantics of the values provided by the port. This is in analogy to a data type or a unit type that defines data or the unit of a value. For perception functions, this is generally a physical state variable like the “ego speed”. Furthermore, it introduces **quality properties** and related **quality property values** for each semantic type; for instance, “ego speed precision” and the values “low precision”, “medium precision”, and “high precision“ together with a definition for the meaning of low, medium, and high.

Last but not least, it introduces **algorithm types** to classify algorithms with respect to their assumptions for providing correct output. If the output is a physical state variable like “ego speed”, then an algorithm typically assumes a physical model like the single-track model. This model approximates the real world. Due to this approximation, it is only valid under certain assumptions. If we calculate the vehicle speed from the wheel speeds, then we assume, for instance, that the wheels are not in the air or spinning on the ground. The algorithm types extend semantic types with such assumptions and are part of the overall adaptation type system. This means that, besides the data value, a port provides the following adaptation values: (1) the type of algorithm that was used for calculating the value and (2) some quality property values.

Adaptation constraints

Based on the adaptation type system, we introduce **pre- and post-conditions** for formulating adaptation constraints. A precondition of an algorithm defines for each of its inputs which adaptation values are acceptable to activate the algorithm. For instance, an algorithm that uses the “ego speed” may constrain usage to three wheel-based calculation variants by referring to a related adaptation type. Furthermore, it may require a certain precision of the ego speed by referring to the quality property values “medium precision” and “high precision”. A post-condition could then define how the precision of the output varies with the current precision of the inputs.

Adaptation goal

The goal of adaptation is to keep all user-perceivable vehicle functionalities alive in the best possible way. This means that if the adaptation values at the local interface of the function satisfy the preconditions of several algorithms, then the function should choose the algorithm that achieves this goal best.

In the best case, this goal can be achieved by means of fixed priorities; function experts easily come up with such a prioritization. Accordingly, we handle this case by **prioritizing** the algorithms and defining that it is always the highest-priority algorithm with currently fulfilled preconditions that becomes active.

In general, local and static prioritization cannot achieve the global optimum. Also, static prioritization might be optimal, but hard to identify. In order to cope with these problems, we recommend applying hierarchical abstraction, as described in [2].

For this work, we assume the best case and focus on the analysis of the adaptation behavior resulting from static priorities. Even though we know that we achieve the best result with static priorities, we do not know how good they are and what our ODD is. To analyze the availability of features, we need to model how the algorithms relate to user-perceivable **modes of operation**. For instance, in some degraded modes, a vehicle's highway pilot system could continue lane keeping, but passing might not be possible. In this context, we differentiate between **fully operational modes**, **degraded but operational modes**, and **non-operational modes** that will end up in a non-operational state of the vehicle, like parking on the hard shoulder of the road. In our case study, we were able to define these modes by considering only the algorithms in the functions Maneuver Planning and Trajectory Planning. The pre- and post-conditions implicitly define the relations to the other algorithms and the system input. Consequently, the modeling of pre- and post-conditions can also help us to specify which system configurations provide which features, because we need not consider each system configuration.

Adaptation trigger

So far, we have specified the triggers for adaptation in terms of adaptation values. In the following, we will discuss how these adaptation values relate to the ODD, including special environmental conditions, internal failures, and V2X problems.

From the definition of the ODD in [9], we conclude that the ODD should rather constrain real-world aspects such as "tunnel" than technical aspects such as "no GPS", but with our adap-

tation type system, we would rather capture technical limitations. Consequently, we need to bridge the gap between the technical limitations described by our adaptation model and a description of the ODD. To this end, we introduce **real-world attributes** like “tunnel” and link them to our technical constraints, e.g., “no GPS”. Furthermore, we cluster the attributes into **real-world attribute clusters** such as weather, static road conditions, and so on in order to reason about their completeness by asking, “Do we have all attributes of this cluster?” This idea of structuring real-world attributes has similarities to an approach pursued in the Pegasus project [10] that describes the real world in different layers and adds elements to these layers. In the future, we might align our model with the results of the Pegasus project.

Another trigger for adaptation that we have not discussed so far are **internal failures**. The technical components that realize a function may fail; as a result, the function would no longer be able to provide the semantic type. The adaptation behavior of the other functions may compensate for such unavailability. We only need to trigger a shutdown of the affected functions. To capture this aspect, we enhanced our adaptation model with a deployment view and a failure model. The deployment view assigns functions to **technical components**. The failure model defines possible detectable failures for the technical component. Based on this additional information, we are able to simulate how the adaptation behavior handles failures in specific situations.

Our approach can deal with unavailable or low-quality**V2X** data by means of semantic types and quality properties. The interesting part comes with the algorithm type. If we have V2X information from an algorithm in the cloud or somewhere else, then this algorithm is based on some assumptions, which might be in conflict with our usage of the V2X data. Our approach maybe suitable for solving this issue if it is also applied beyond the physical boundaries of the vehicle.

Analysis

The adaptation model enables various ways of analyzing fail-operational behavior. It enables, for instance, verifying the reconfiguration sequences, as we did in [1]. In previous work, we also analyzed and optimized the fail-operational behavior. However, as we had no link to the real world, we only optimized the trade-off between technical limitations and costs. We evaluated whether it is worth realizing the algorithms and sensors to be operational with or without a technical limitation such as “no GPS”. This was sufficient for non-automated driving, as a shutdown is generally acceptable. From an industrial perspective, the need was apparently not great enough to adopt and rollout the approach. The transition to automated

driving has changed this situation, as it has increased the need. Standards like ISO 21448 [17] demand specifying the ODD. Research in this direction is still increasing and terminology for raising research questions has come up. For instance, the work in [14] differentiates between the Restricted Operational Domain (ROD) and the ODD in order to highlight that a degraded mode can handle fewer real-world situations. Figure 2 illustrates the reduction in the case of functional degradation.

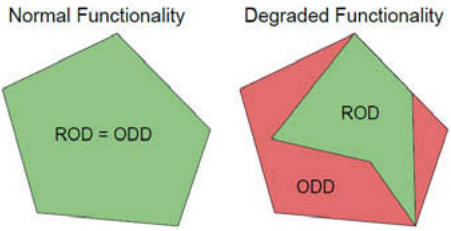


Fig. 2. Operational Design Domain (ODD) and Reduced Operational Domain (ROD)

Our enhancement of the modeling approach leads to a definition of the ODD and the ROD. For any combination of real-world attributes, we can determine the related operating mode and thus know whether it is operational or not. We also know how this is reduced if we have an internal failure. As we already know the ODD and ROD, our analyses are not targeted at the identification of these domains. In our case study, we implemented a reachability analysis [13]to find out whether there are some algorithms that would never activate in our ODD. Furthermore, we implemented an analysis to find the likelihoods of the operating modes.

Related work

Most related work is our previous work. We adopted the adaptation modeling approaches from Chameleon [14] and its successor Mars [1][2]. Chameleon introduced modular adaptation of behavior modeling using preconditions, post-conditions, and priorities. It focused on the analysis of complex reconfiguration sequences and the problem that switching the algorithm in function A may require adaptation of all components that receive values from this function A. In contrast to Chameleon, we focus in this work on the trade-off between maximizing the ODD and ROD versus minimizing costs.

MARS enhanced Chameleon with hierarchical abstraction. However, we did not take advantage of this enhancement, as it was not necessary in our case study. MARS further introduced the modeling of user-perceivable operating modes, which we adopted in our case

study. MARS also introduced an environment model for running Monte Carlo simulations. This environment model generated the system input, that is, semantic types and qualities. We enhanced this approach with real-world aspects in order to get the ODD and ROD. Furthermore, we introduced component failures and V2X failures as triggers for adaptation.

Related work in the context of V2X and safe runtime collaboration of systems is provided by ConSerts [15] and the DEIS project. Both approaches provide solutions for modeling the trustworthiness of provided services or information. The idea is to formalize assurance cases so that they become machine-readable. Systems can then exchange their assurance cases and find solutions how they can work together safely. Assurance cases provide argumentation for a claim or goal. The argumentation comprises assumptions that the receiver of the case has to check. We could interpret our adaptation information as an assurance case. The semantics define the goal, as it refers to what we want to have, e.g., the physical state variable “speed”. The algorithm refers to the assumptions under which we achieve this goal. This analogy is in harmony with the idea we proposed for V2X scenarios.

A related research field concerns the identification of the ODD. Most research in this direction has a strong focus on sensors and their mounting. This work complements our work. We make assumptions about the places where the sensors are mounted. Approaches that derive the ODD from sensor sets make assumptions about the algorithms and their adaptation behavior.

The research around fail-operational architectures has a strong focus on the handling of internal failures, particularly hardware failures. A typical example is the work by Ishigooka et al. [16], which focuses on technical aspects, such as CPU load and memory consumption. We created a first link to this research area by introducing component failures as triggers for adaptation.

Summary, Conclusion and Future Work

In this paper, we presented a methodology aimed at facilitating the systematic engineering of fail-operational architectures and behaviors for highly automated vehicles. The method is already applicable in early development phases and enables evolutionary refinement driven by revolving appropriateness analyses.

Our experience gained during industrial case studies [3] suggests that the proposed methodology operates at the right level of detail. It is sufficient for responsible engineers to make a decision on whether the defined fail-operational behavior of an architecture is appropriate or not in the intended ODD.

We have reached the point where systematic engineering of fail-operational behavior has become inevitable. As long as shutdown was an acceptable solution and being operational was rather a matter of availability than of safety, industry was free to lag behind the state of the art in engineering adaptation behavior. The safety criticality in the context of automated driving has changed this situation. Standards define terms like ODD and demand its specification. In many cases, the term “minimum risk state” fits much better than the term “safe state”, and the term “resilience” is experiencing a renaissance. Furthermore, resilience is increasingly focusing on handling critical context situations related to functional insufficiencies, as discussed in SOTIF [17]. All these trends and issues perfectly match with our mature methodology, which has a long history going back to 2005. We believe that the break-even has been reached and that our success story with Hitachi was the beginning of a stepwise adoption of this method in industrial practice.

For future work, we plan to develop tool support for the methodology based on our safeTbox tool [19]. This shall lower the modeling effort and maybe provide semi-automation for some aspects; another aim is to allow execution of the presented analysis without any implementation overhead. We also plan to integrate more of the concepts from the MARS approach into the presented methodology. For instance, we omitted optimization of the adaptation behavior described in the previous work section. Last but not least, we plan to link our research to research for defining test scenarios [10] and to research about sensor set evaluations.

[1] R. Adler, I. Schaefer, M. Trapp und A. Poetzsch-Heffter, „Component-based modeling and verification of dynamic adaptation in safety-critical embedded systems,“ ACM Transactions on Embedded Computing Systems, Bd. 10, 2010.

[2] R. Adler, A Model-based Approach for Exploring the Space of Adaptation Behaviors of Safety-related Embedded Systems, Kaiserslautern: Fraunhofer, 2013.

[3] „Hitachi Ltd. Success Story: Safety Engineering for vehicles of higher automation levels - Fraunhofer IESE,“ [Online]. Available: https://www.iese.fraunhofer.de/en/customers_industries/automotive/hitachi-success-story-safety-engineering-for-vehicles-of-higher-automation-levels.html. [Zugriff am 21 07 2019].

[4] E. A. Lee, S. Neuendorffer und M. J. Wirthlin, „Actor-oriented design of embedded hardware and software systems,“ Journal of circuits, systems, and computers, Bd. 12, Nr. 03, pp. 231-260, 2003.

[5] „Welcome To UML Web Site!," [Online]. Available: https://www.uml.org/. [Zugriff am 21 07 2019].

[6] „SysML Open Source Project - What is SysML? Who created SysML?," [Online]. Available: https://sysml.org/. [Zugriff am 21 07 2019].

[7] „Simulink - Simulation and Model-Based Design - MATLAB & Simulink," [Online]. Available: https://www.mathworks.com/products/simulink.html. [Zugriff am 21 07 2019].

[8] „Enterprise Architect 14.1," [Online]. Available: https://www.sparxsystems.de/uml/neweditions/enterprisearchitectbeta00/. [Zugriff am 21 07 2019].

[9] SAE, J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, 2016.

[10] „About PEGASUS - pegasus-EN," [Online]. Available: https://www.pegasusprojekt.de/en/about-PEGASUS. [Zugriff am 21 07 2019].

[11] ISO, Road vehicles -- Functional safety, 2018.

[12] I. Colwell, B. Phan, S. Saleem, R. Salay und K. Czarnecki, „An automated vehicle safety concept based on runtime restriction of the operational design domain," in 2018 IEEE Intelligent Vehicles Symposium (IV), 2018.

[13] K. Schneider, T. Schuele und M. Trapp, „Verifying the adaptation behavior of embedded systems," in Proceedings of the 2006 international workshop on Self-adaptation and self-managing systems, 2006.

[14] M. Trapp, „Modeling the Adaptation Behavior of Adaptive Embedded Systems," Kaiserslautern, 2005.

[15] D. Schneider und M. Trapp, „Conditional Safety Certification of Open Adaptive Systems," ACM Transactions on Autonomous and Adaptive Systems, Bd. 8, 2013.

[16] T. Ishigooka, S. Honda und H. Takada, „Cost-Effective Redundancy Approach for Fail-Operational Autonomous Driving System," in 2018 IEEE 21st International Symposium on Real-Time Distributed Computing (ISORC), 2018.

[17] ISO, Road vehicles -- Safety of the intended functionality, 2019.

[18] I. Colwell, B. Phan, S. Saleem, R. Salay und K. Czarnecki, „An automated vehicle safety concept based on runtime restriction of the operational design domain," in 2018 IEEE Intelligent Vehicles Symposium (IV), 2018.

[19] safeTbox, Available: https://www.iese.fraunhofer.de/en/competencies/safety_engineering/tools_safety/safetbox.html. [Zugriff am 06 08 2019]

Safety for Automated Driving with High Performance ECUs

Dr. **M. Oertel**, **J. Wolf**, Vector Informatik GmbH, Stuttgart

Abstract

The competition to provide advanced automated driving experience dominates the development of electronic automotive systems today. To meet the requirements on computational power, flexibility in development processes, and updateability of embedded software, E/E architectures are changing. Centralized high-performance ECUs are the result of this endeavor. But how to achieve functional safety in this scenario?

Not only the hardware and software architecture of these microprocessor-based systems is different from the traditional microcontroller-based automotive systems, also safety goals have changed. Fail-safe systems have a well-defined safe state in case of failure. This will be supplemented or replaced by fail-operational systems, which require a defined reliability for the provided service, even in case of failure. As of today, these systems require a combination of microprocessor and microcontroller, raising the question of how to distribute safety requirements.

Looking deeper into the software on the high-performance system, challenges are introduced by aspects like the usage of object-oriented programming languages (e.g. C++), more frequent and modular software updates over-the-air, as well as high level operating systems. In this article, hands on experience of automated driving projects is given, starting from an introduction to terminology and concepts, going over to challenges on the level of the ECU design, and finally looking at the application development itself.

Introduction and Terminology

Today, almost all systems in a car are designed to work with a single channel. Single channel means that if an element in the chain from sensor via logic to the actuator fails, then the entire system fails. Redundancy through a second sensor or microcontroller, for example, are used solely to reliably detect a fault in this chain. The usual reaction in case of a detected fault is to reset the control unit or to switch of the complete system and inform the driver, leading to a safe state (usually supported by HW measures). The driver then needs to adapt to the new

situation and stop the car in case of an emergency. A reset or complete deactivation of a system are considered safe states. Such a system is then called a fail-safe system.
The software of fail-safe systems must detect faults in hardware and software within a specified tolerance time interval. Often, mechanisms like deadline monitoring for certain software functions or additional checks on the integrity and liveliness of the received data (see also E2E) are used.
If the driver should not be the fallback solution in the safety concept, there must be an electronic fallback. A prerequisite for such a system is, that each channel can detect its faults on its own. In case of a fault the active channel is deactivated, and a second channel takes over control of the vehicle. Such a system can be considered a fail-operational system.
In hardware development it is well established that each part of a system can fail at any point in time with a certain probability. There are different mathematical models for this probability. With these methods it is possible to evaluate and argue why a system can be assumed to be safe. The crucial point here is that *there are* mathematical models.

In contrast, faults in software are systematic faults. Attempts to create a generally accepted model for probabilities of systematic faults have not succeeded to date. Thus, fault avoidance must be implemented to avoid invalidating the probability-based arguments of the analysis on a system level. Fault detection is no longer sufficient, since it does not help the driver if the system detects that it no longer can control the vehicle. The driver relies on the system controlling the vehicle even in case of a fault.
How can fault avoidance for complex software be shown? Systematic faults can only be prevented by an adequate development process. The safety standard ISO 26262 defines a lower bound for the activities and methods. In contrast to fail-safe systems, many more functions become safety-related in fail-operational systems, not only the functions that detect and mitigate faults. For AUTOSAR basic software there will be safety requirements for the communication and other basic functionality.

ISO 26262 defines these types of requirements as safety-related availability requirements. They lead to a typical design of today's highly automated driving ECUs: a microprocessor-based part that performs the nominal function and a microcontroller-based part that acts as a monitor to the microprocessors and as a fallback to continue degraded automated driving in case of a failure of the microprocessor-based parts. An example system architecture is depicted in Fig. 1. There are multiple reasons for these kinds of systems architectures. On the one hand, the high-performance microprocessors are newly introduced into the automotive

domain. Their development process to mitigate systematic faults in hardware, their diagnostic coverage capabilities to detect random faults do not meet current automotive standards. This is natural due to the increased complexity of such microprocessors. On the other hand, also crucial software, such as reliable, safety-qualified communication stacks, are currently only available for microcontrollers. This lack of mechanisms and confidence leads to supplementary use of microcontrollers with well-known software that is used for many years now in the automotive domain.

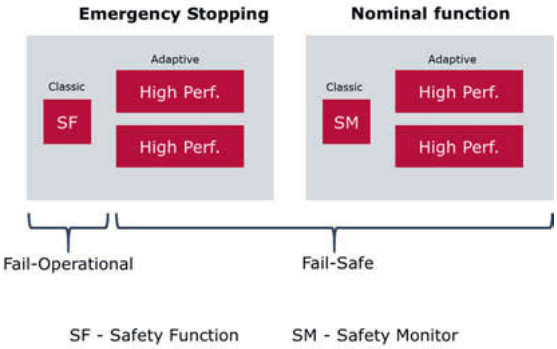


Fig. 1: Typical Automated Driving Architecture

High Performance ECUs

In this section the challenges of the design of the fail-safe, high-performance part of an automated driving system is discussed.

For use cases like automated driving, infotainment or central applications servers a different dimension of computational power is needed from what is available in today's automotive microcontrollers. The microprocessors used in this high-performance segment feature a Memory Management Unit (MMU) providing virtual memory. In contrast, microcontrollers use a Memory Protection Unit (MPU), which also provides a separation of memory, but all code is running in the same address space. This hardware difference leads to a different class of operating systems used on these ECUs. While the microcontroller-based ECUs typically run OSEK-based operating systems, such as the AUTOSAR Classic operating system, the high-performance ECUs make use of operating systems known from the IT domain, such as Linux, QNX or PikeOS.

These operating systems provide filesystems and a binary loader, enabling the installation of additional software on the ECU in the field, without the need to reconFig. or recompile the operating system or related system services. This enables an independent development of the different software components.
Delivering an “empty” ECU to the end customer and installing software afterwards sounds tempting. But is this approach also applicable for safety-related software such as Lane Keep Assists or Highway Pilots?
With the number of downloadable, safety-related content the testing effort for the ECU increases, since every combination of software on the ECU would need to be tested (cmp. ISO 26262-6:2018 Annex C). To avoid this, the operating system needs to guarantee freedom from interference for the installed applications. This means, that it is not possible for a newly installed application to influence the behavior of the already running applications. The influence can be caused by multiple aspects:

- an application could accidently write to the memory of another applications, causing a data or program corruption, in worst-case even an undetected wrong behavior, or
- the new application could influence the timing behavior of an applications, by consuming such a level of CPU time, that other applications are delayed, or
- a newly installed applications could interfere with existing communication channels .

For safety-related software, the mechanisms ensuring this freedom from interference must to be developed with the same ASIL as the software that can be influenced.

While all operating systems provide a mechanism using the MMU to separate processes from memory point of view, this mechanism is in many cases not developed according to safety standards. Typically, a special “safety variant” of the operating system needs to be used. Especially for Linux this is not available, creating the need to qualify needed parts of the Linux kernel project specific (cmp. ISO 26262-8:2018 Clause 12). There have been projects started to give guidance, how this qualification could be done [1], but it will take a while until results can be used in series production projects. Commercial off-the-shelf operating systems explicitly developed for safety-related systems provide a certain benefit here.

Compared to ensuring freedom from interference with respect to memory, showing the freedom from interference with respect to timing is even harder. Operating systems designed for the IT domain typically lack proper hard real-time features. Linux can e.g. be extended with the

preempt_rt patch [2], providing the “earliest deadline first” (EDF) scheduling algorithm and optimized kernel locking, but even then, newly installed tasks may impact other applications if too many threads are created. Taking PikeOS as a counter example, separate time domains are created for different applications, bounding the execution time and in turn ensure freedom from interference [3]. QM applications, which might need a more dynamic usage of the CPU can be put into one specific time domain.

FFI for communication can be established in fail-safe systems using the E2E mechanism, to detect failures.

If these scheduling and protection mechanisms are available with the highest ASIL of the applications that should be executed, the idea of installable applications without integration testing all combinations also for safety-related software might be feasible.
Still, high-level operating systems do not only consist of a scheduler, the whole package, like libraries, system calls, filesystem, might affect functional safety. But what are the critical points?
Since the spatial separation is provided between processes , all software running in the same process, needs to comply to the same ASIL. While in AUTOSAR Classic this is a requirement easy to fulfill, when using high-level OS this results in additional effort. Standard libraries used on these types of systems, such as the C or C++ Standard Library are linked to the applications software and are executed in the same context. Hence, ASIL versions of this libraries must be used. The same applies for applications using automated driving libraries.
Thus, open source software cannot be used in safety-related applications, unless qualification measures according ISO 26262 have been performed. Even if a piece of open source software to be used in a safety context is qualified, there are two additional pitfalls to consider. One is the license of the software, the other is the maintenance strategy. The license is typically more a legal topic, than a technical one. Some licenses require publishing parts of the code that uses it, or commercial usage is restricted. The issue with the maintenance is more complex. Some open source software deals with very complicated calculations, such as cryptographic algorithms. In case of a defect in this part of the software, it is not easy to fix them in short time, since this code is not self-explanatory. One viable solution would be to hire maintainers for the software to be on “stand-by”, or directly contract full-time maintainers for such software. Both approaches are not easy to set up and provide no guarantee that the maintainer is familiar

with that part of the codebase which has the defect. This is a risk in particular for huge codebases such as the Linux kernel itself. As a result, the decision for open source software in the safety-related part of an application should be carefully evaluated.

Safe Application Development

After the identification of challenges on ECU level, this chapter has a focus on the development of the software components itself. Important aspects are issues introduced by the programming language C++, the safety features of the AUTOSAR Adaptive Platform, as well as a potential impact of dynamic loadable software components on the exchange of safety artifacts in the supply chain.

Safe Application Development with C++

C++ is an object-oriented programming language allowing a more intuitive way of programming. Unfortunately, many of these advantages contradict with requirements from the ISO 26262.

The first issue is created by the usage of dynamic memory allocations. That is, memory is requested from the operating system during runtime and released when it is not needed any more. This creates a whole new class of potential faults that were not possible in automotive software yet:

- Non-deterministic time required for (de)allocation
- Fragmentation of memory (heap)
- Memory leaks, i.e. memory is not returned to the operating system, but application forgot about that memory
- Exhaustion of available memory

The use of invalid memory, i.e. using memory that does not hold the intended information, has always been a big issue in software development, but with dynamic memory allocation this problem becomes even more complex to solve. Fortunately, for a fail-safe system, only the use of invalid memory (e.g. use-after-free) is critical. All other problems can be detected, and the system can be shut down safely. Memory leaks are not dangerous on their own but lead to a much quicker memory exhaustion than expected. Exhaustion can be easily detected and handled by software of a fail-safe system. Fragmentation of memory leads to non-deterministic time behavior. It is easily to detect if a non-deterministic timing behavior endangers the safety of a system. Still there are options to increase the availability, such as pre-allocation of memory from the OS and using a memory allocation strategy suited for embedded systems (e.g. [4]).

Second, generic programming, in C++ terms called templates, a very powerful feature of modern programming languages, creates issues. It creates new questions how to test functions that are programmed independent of a specific type. One approach would be to create test cases for each instance of such a generic function. If not all instances are known at the time of development, e.g. when providing libraries to another party for integration, a solution could be to provide a test suite which can adapt to the instances that are created when using the library.

The third issue is related to the C++ error handling using exceptions. Their use requires dynamic memory allocations and complex unwinding of the stack. Moreover, guaranteeing exception safety, i.e. the current object is still in a valid state even if an exception is thrown, is hard to implement and show. The most pressing problem, however, is that static code analysis with today's tools is almost impossible to do. The C++ standardization committee has recognized those issues and is working to change the C++ specification in future releases. For now, one could use a different approach that is also taken by programming languages like Rust (i.e. using a Result data type).

Fourth, features like runtime type information (RTTI) increase the cost for compiler qualification, since even in C++17 RTTI does not follow the C++ dogma "you only pay for what you use".

Safety Mechanisms in the AUTOSAR Adaptive Platform

Using the programming language correctly is an important aspect to achieve functional safety for high performance ECUs. Fortunately, there exists support in form of preimplemented safety machanisms inside adaptive AUTOSAR, to realize the safety concept.

The mechanisms provided by the AUTOSAR Adaptive Platform are very similar to the Classic Platform. To ensure communication integrity between applications the end-to-end protection mechanism is used. In this communication pattern, information is sent periodically on the network from the sender to the receiver. This information is enriched with a sequence counter, a data ID and a CRC. Thus, faults in the communication can be detected (see Table 1)

Table 1: Overview Communication Faults and Countermeasures

Which faults are possible?	How do we address them?
> Failure of communication peer (even in lower software layers) > Message masquerading > Message corruption	> **CRC** over data, **Data ID** and Sequence counter > Allows to detect corruption and masquerading of the signal
> Unintended message repetition > Insertion of messages > Re-sequencing	> **Sequence counter** > Allows to detect faults in the order of messages > Allows to detect lost/repeated/inserted messages
> Message loss > Message delay	> **Timer** on receiver side > React on lost messages > React on delayed messaged

For the E2E mechanisms it is important to note, that this is purely a detection mechanism, the reaction on discovered faults needs to be performed by the receiver application. Hence this mechanism is only intended for fail-safe systems. For fail-operational system a redundant channel is needed.

For intra-ECU communication the E2E mechanisms are typically not used because of performance reasons. Hence, for adaptive AUTOSAR systems, a safe channel needs to be provided by the operating system, i.e. the operating system does not introduce systematic faults into the communication. For ECUs using a high-level operating system, this requirement is difficult to achieve. Although, other applications cannot directly compromise the channel (use of MMU), the kernel itself is a threat. Either the kernel is developed completely according to the required safety integrity level with an appropriate safety requirement, or any other argumentation exists, why the kernel cannot compromise the channel. Besides the process memory separation this safe communication channel is the second important safety requirement on the operating system.

AUTOSAR Adaptive provides also mechanisms for Timing monitoring. Although some operating systems (like PikeOS) offer a scheduler developed according to the highest safety integrity

level, a timing monitoring using a Watchdog is advised. In AUTOSAR Adaptive this functionality is realized in the “Platform Health Management” (PHM). Using so called health channels, the application can report its status (checkpoint) to the PHM, which can decide based on a given configuration, if the application still behaves inside the given timing frames. An overview of the possible faults and reaction patterns can be found in Table 2.

Table 2: Overview Timing Faults and Countermeasures

Which faults are possible?	How do we address them?
> Execution of code without request > Code not executed although requested > Execution of code started too early or too late > The execution time of a code is longer or shorter than expected > The program flow of a code differs from the expected behavior	> **Deadline Monitoring** > Applicable for aperiodic entities > Time between two checkpoints is compared to min/max values > **Alive Monitoring** > Applicable for periodic entities > Number of checkpoints in interval is monitored > **Logic Monitoring** > Detect wrong execution order > Validate checkpoint activation sequence against preconFig.d execution graphs

The third aspect, in which AUTOSAR Adaptive can support, is persisting data. Although file systems are typically used in high-performance ECUs, these do not guarantee correctly written data with the required ASIL. Hence, protecting the data in the persistency library is a more convenient and cost saving approach. Again, an overview can be found in Table 3:

Table 3: Overview Persistency Faults and Countermeasures

Which faults are possible?	How do we address them?
> Data loss > Data masquerading of > whole file > key-value pair > Data corruption	> **CRC** over data and key > Allows to detect corruption and masquerading of data > **CRC** over whole file with "magic identifier" to prevent QM applications from creating valid files > Persistency cannot ensure that data is written at all. It can only detect loss.

Software Development Process

One of the main advantages of systems build with the AUTOSAR Adaptive Platform, is the possibility to install additional applications during runtime. For safety-related applications we already identified the issue and a solution ensuring freedom from interference between applications. In addition, the workflow will need additional steps and work products with respect to safety. More concrete: Due to the fact that there is no human integrator anymore, putting all applications to the target, the driver can install these from an "App-Store" in the vehicle. Hence, the aggregation of failure states is not performed by a single person anymore. For the AUTOSAR Adaptive Platform this means, that the integrator is not able to state, which combinations of faults in the system should lead to e.g. a process restart, a partition restart, or even a permanent deactivation of functionality. To overcome this gap the "health channel contribution" has to be used. Each application is therefore able to inform the PHM, what kind of faults have been detected. Even newly installed applications take part in this communication. The degradation concept of the ECU must specify and communicate the types of detected faults to the application developers. Hence, beside the interface, also safety specifications are passed using AUTOSAR XML files. It is quite difficult to agree between the ECU-Architect and the application developers on a set of faults, especially since these roles might be distributes across companies. Hence, multiple iterations over these files are expected. Still, the PHM finally needs to know the types of faults to be able to generically react, if a health channel sends an alarm.

Conclusion

This contribution summarizes the challenges and possible solutions to build a safe and reliable system using high performance ECUs. The focus is on software, leaving hardware metrics aside. Fail-safe systems can be developed today using the strategies described for the ECU and the software level. The limiting factor for fail-operational systems on high performance

ECUs are the availability of important OS features, such as the communications stacks, which need to be ASIL quality or the reliability of the hardware. While Vector is already providing ASIL com-stacks for AUTOSAR Classic, as a basis for future fail-operational systems, no high-level OS is known providing such features. Even though on some operating systems, such as PikeOS, development has been started, it is not expected that such software will be available before 2021/2022, leaving the first generation of high-performance ECUs to be fail-safe only.

References

[1] https://elisa.tech/

[2] https://rt.wiki.kernel.org/index.php/Main_Page

[3] https://www.sysgo.com/products/pikeos-hypervisor

[4] http://www.gii.upv.es/tlsf/

Impact of Cybersecurity and Safety Standards on ADAS Software Development Practices

Obaid Ur-Rehman, **Gerhard Wallraf**, **Bastian Holderbaum**, **Marco Jentges**, FEV Europe GmbH, Aachen

Abstract

During recent years, significant progress has been made by the automotive industry towards achieving autonomous driving functionalities. At the core of an autonomous vehicle is the advanced driver assistant system (ADAS). In order to realize the goal of a fully autonomous vehicle, new functionalities are envisaged and continuously integrated into modern vehicles. The advanced features come with increased security and safety considerations. Traditionally, the achievement of cybersecurity and safety had different requirements. However, a successful security attack on a vehicle or its component(s) can have a negative impact on its operational safety, and therefore a holistic approach is needed in which the two domains complement each other. Recently, it has been argued that there are similarities between the security and functional safety engineering processes. The newly published international functional safety standard, ISO 26262:2018, stops short of describing a security engineering process. However, it recommends that the interactions between functional safety and cybersecurity should be taken into account as early as in the concept phase of a development project. The ISO/SAE 21434 Automotive Cybersecurity Standard, which is under development at the time of this writing, tackles the complete cybersecurity lifecycle of a vehicle right from the concept phase up until its decommissioning.

Whereas the safety lifecycle is dominated by process oriented thinking and a methodological mind-set with the aim of preventing faults, security engineering is shaped by techniques to prevent intentional faults (security attacks) which might harm the system and may have potentially degrading consequences for safety.

In this paper, the challenges of interactions between safety and security engineering for ADAS systems are discussed. It is shown how the communication, interaction and cohesion between the functional safety and security teams are helping in the engineering of ADAS software. Moreover, best practices and findings regarding efficient and integrated engineering processes are highlighted.

Keywords: ADAS; Safety; Cybersecurity; HARA; TARA; ISO 26262; SAE J3061

1. Introduction

Most of the road accidents occur due to errors of human drivers. According to the 2018 annual accident report [1] of the European Road Safety Observatory, approximately 11 million injuries were recorded in 2017 due to road accidents in the countries of the European Union [1], out of which the number of fatalities were 25,651. In comparison, in the United States, a total of 37,133 road accident fatalities were recorded in 2017 according to the National Highway Traffic Safety Administration (NHTSA) [2].

Advanced Driver Assistant Systems (ADAS) are designed to aid a vehicle driver by automating certain driving functionalities. Various degrees of automation are possible, ranging from information or suggestion to complete automation. The aim of ADAS is not only to enhance the user experience, but also help in reducing the number of accidents by assisting the human drivers. According to the European New Car Assessment Programme (Euro NCAP), accidents occurring due to late braking intervention are due to the driver's inattentiveness and human error [3]. Recently, there have been comprehensive research activities in the automotive safety field resulting in the introduction of ADAS systems. Some notable examples of ADAS include adaptive cruise control, electronic stability program, collision and obstacle warning, blind spot monitoring, autonomous emergency braking, lane keep assistance, semi-autonomous driving and semi-automatic parallel parking, to name a few. The future and ultimate goal of ADAS is a fully autonomous vehicle which can drive itself from source to destination without any human support.

Ultimately, functional safety and cybersecurity both address the same target: minimizing the risk for a system´s reaction which can cause harm to people. Functional safety considers malfunctions caused by systematic faults or random hardware faults of the system, whereas cybersecurity takes into account the intentional intrusion into the control of the system from the outside. Automotive safety is a well-studied subject that has matured over the years and there have also been successful standardization attempts in the automotive functional safety domain. The development methods for functional safety of road vehicles are defined in the international standard ISO26262. The basic idea of ISO26262 and other functional safety standards is the identification and assessment of risks in the first step and the development of measures for risk mitigation in the second step. In contrast to functional safety, cybersecurity is a relatively new area for the automotive industry. Nevertheless, there is a common understanding meanwhile that security of a modern vehicle has implications on its safety. Even though functional safety and cybersecurity development are typically carried out by different technical teams, a continuous communication and information exchange between both teams is essential. Fig. 1 shows the main interactions of these teams during the concept phase.

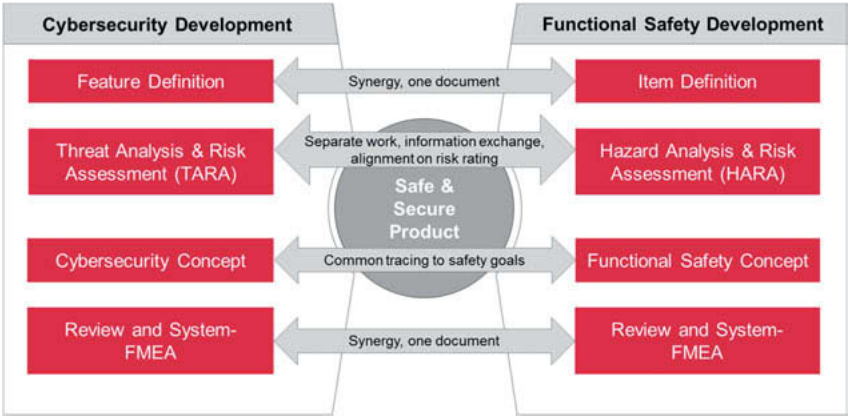


Fig. 1: Interaction paths between functional safety and cybersecurity

Though safety and security have been treated differently so far in the automotive industry, the importance of security in modern vehicles, which are equipped with external communication interfaces to the driver, the infrastructure and the Internet, is ever increasing.

Diving deep, one can notice that there are similarities between the two domains, e.g., in the concept phase and in requirements management. This is also evident from the Hazard Analysis and Risk Assessment (HARA) in safety engineering in contrast to the Threat Analysis and Risk Assessment (TARA) in cybersecurity engineering. They are addressed by ISO 26262 [4] and SAE J3061 [5] respectively.

HARA is a method used to systematically identify all potential hazards caused by malfunctions of a system and to derive the highest level of safety requirements, the so called safety goals. At this level, the root cause for the violation of a safety goal is not yet analysed. Besides a system´s malfunction (functional safety), also an intrusion from outside (cybersecurity) can lead to the violation of the safety goal.

In parallel to the HARA, a TARA is carried out to identify all possible security risks for a system, which, when exploited, may create safety hazards. The security risks would typically not create new unsafe states, but they can definitely impact the likelihood of occurrence of already identified unsafe states. By performing a TARA analysis, the security risks and their likelihood of occurrence are identified. The risks are then rated and prioritized.

In the presented examples, the risk ratings independently derived in HARA and TARA are matching very well. In case of different ratings, an according alignment between the functional safety and cybersecurity experts is important in order to avoid that a risk is underestimated.

2. System Description

In order to demonstrate that a cohesion can be achieved between safety and security engineering, the same use case is considered for HARA and TARA analysis. The ADAS use case is an “Urban Chauffeur”, where the ADAS equipped vehicle is able to perform automated driving manoeuvres inside the city limits at the SAE Level L3 of automation [6]. The HARA analysis is performed based on the method prescribed in ISO 26262. The standard which considers cybersecurity for the whole lifecycle of a vehicle is a joint standard being developed by ISO and SAE, called ISO/SAE 21434 [7]. This standard is under development and in the meantime, the guideline SAE J3061 is being considered for developing the concept phase. However, SAE J3061 does not specify or prefer any particular method for performing TARA analyses and specifies several options.

The HARA method applied at FEV consists of four main steps as shown in Fig. 2.

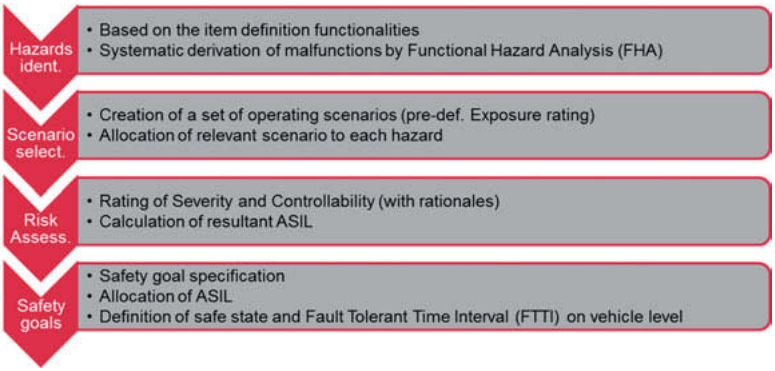


Fig. 2: The four main steps in the Hazard Analysis and Risk Assessment (HARA)

The corresponding TARA method applied at FEV also consists of four main steps, which are analogous to the HARA method, as shown in Fig. 3.

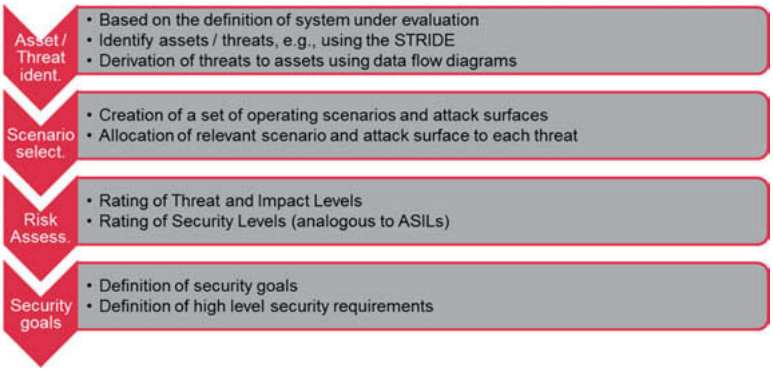


Fig. 3: The four main steps in the Threat Analysis and Risk Assessment (TARA)

It is evident from Fig.s 2 and 3 that the safety and security risk assessments follow analogous steps. Whereas in HARA, the hazards are identified, in TARA the assets and threats are identified. This is followed by creation of operating scenarios and allocation of the scenario to a hazard or threat. This step is followed by risk assessment in both methods. The outcome of this step is an ASIL level in HARA and a security level in TARA (which is analogous to the ASIL level). If the risks are prioritized at more or less the same level, it is a self-check that the analysis is going into the right direction. Finally, the safety and security goals are identified in each of the analysis methods. The HARA of a complex function like the Urban Chauffeur described in this paper, includes more than one hundred hazardous events. One exemplary hazardous event is shown in Fig. 4.

ID	Location	Hazard	Potential Effect	Severity	Severity Comment	Exposure	Exposure Comment	Controllability	Controllability Comment	ASIL	Safety Goal	Safe State
HE030	Driving at constant speed - urban road (50+ kph); Urban Chauffeur (L3 Autonomous) function active	[Hz002] Unintended vehicle acceleration	Front-end collisions with other vehicles	S3	The vehicle can collide with a vehicle/object in front at high speed. This will lead to an S3 injury	E4	defined in accordance with VDA 702	C3	The driver is not monitoring the drive environment. The driver cannot react in time to prevent the hazard.	D	[SG027] Prevent unintended vehicle acceleration	1. Limit the unintended acceleration to a safe value. 2. Inform the driver to take over control of the vehicle. 3. The function shall continue vehicle control until the driver takes over.

Fig. 4: Exemplary hazardous event in HARA for Urban Chauffeur

The example illustrates the increased efforts which are necessary to achieve a sufficient safety for vehicles with higher automation levels. Due to the limited availability of the driver to control a hazardous event, the ASIL is calculated as ASIL D, which represents the maximum rating. Furthermore, the safe state is no longer a simple state like “shut off the function”, but a function on its own, which requires a fail operational behaviour of the system.

In analogy to the HARA performed by a safety engineer, a TARA of the same function is performed by a cybersecurity engineer. Different approaches exist in literature for TARA, some of which are listed in SAE J3061 Annex A. We have adopted multiple approaches, however due to space limitations, only a single approach is being discussed in this paper. It is the TARA method proposed in [8] in the context of the HEAVENS project (see D2 – Security models [9]), referenced in SAE J3061. The TARA results obtained after application of this approach are summarized in Table 1.

Table 1: Exemplary hazardous event in TARA for Urban Chauffeur

ID	Location	Asset	Attacks	Threat	Threat Level	Impact Level	Security Level	Security Goal
T001	Driving at constant speed - highway (100+ kph). Highway Pilot (L3 Autonomous) function active	ADAS ECU, Gateway, Tele-matics, Internet Connectivity Unit (ICU)	[AS001] Packet injection, Spoofing, Tampering, Packet replay	[Th001] Unintended vehicle acceleration resulting in front end collision	$T_{sum} = t_x+t_k+t_w+t_e = 1 + 1 + 1 + 0 = 3$ TL = 3 (High)	$I_{sum} = 10(i_s+i_f)+i_o+t_p = 10(10+1) + 1 + 0 = 111$ IL = 3 (High)	High	[G001] Provide data integrity, authenticity and protection against spoofing

Once again, it is worth mentioning here that in this manuscript, the focus is only on one of the methods proposed in SAE J3061. However, other approaches can also be used, and might be equally sound in obtaining the security level based on the analysis of threats and assets.

In the TARA analysis based on the HEAVENS security model, shown in Table 1, threat analysis refers to the identification of the threats associated with the assets of the Target of Evaluation (TOE) using the Common Criteria approach. It is followed by a mapping of the threats with the security attributes. The threat level is calculated based on four parameters, which are expertise of an attacker, knowledge about the target, window of opportunity and required equipment. Each parameter is assigned a value in the range between 0 and 3, based on Table 2.

After values are determined for each parameter, an addition called sum of threat parameters (T_{sum}) is performed. This sum is then used to assign a threat level.

Table 2: Parameter values for calculating the threat level [8], [9]

Expertise of Attacker	Knowledge of the Target	Window of Opportunity	Required Equipment	Value
Layman	Public	Unlimited	Standard	0
Proficient	Restricted	Large	Specialized	1
Expert	Sensitive	Medium	Bespoke	2
Multiple Expert	Critical	Small	Multiple Bespoke	3

The sum of the values of these parameters (T_{sum}) is assigned a threat level using Table 3.

Table 3: Threat level based on the parameter sum [8], [9]

Sum of threat parameters (T_{sum})	Threat Level
>9	0 – None
7 – 9	1 – Low
4 – 6	2 – Medium
2 – 3	3 – High
0 – 1	4 – Critical

In order to calculate the impact level, another four parameters are used, which are safety, financial, operational and privacy impacts, taking on the values as shown in Table 4. Each of these parameters is grouped into four categories, and the values are summed up to calculate the intermediate impact, which is then assigned an impact level.

Table 4: Parameter values for calculating threat level [8], [9]

Safety - I_s (ISO 26262-3)	Financial - I_f (based on BSI-Standard 100-4)	Operational - I_o (adapted Failure Mode and Effects Analysis (FMEA))	Privacy - I_p	Value
No injury	No impact	No impact	No impact	0
Light and moderate injury	Low	Low	Low	10
Severe and life threatening injuries	Medium	Medium	Medium	100
Life threatening and fatal	High	High	High	1000

Like the threat level mapping, the sum ($I_{sum} = I_s + I_f + I_o + I_p$) is mapped to an impact level here, too, as shown in Table 5.

Table 5: Threat level based on the parameter sum [8], [9]

Impact Sum - I_{sum}	Impact Level
0	0 – None
1 – 19	1 – Low
20 – 99	2 – Medium
100 – 999	3 – High
≥ 1000	4 – Critical

Finally, the security level is calculated based on a matrix of threat and impact level as shown in Table 6. The security levels are analogous to the Automotive Safety Integrity Levels (ASIL) which are assigned based on the HARA analysis. A vulnerability which has the highest threat and impact levels, results in the highest security level “Critical”. Therefore, there is a need to come up with security requirements to deal with this threat. A security level of “QM” implies that there is no urgent need for developing security requirements.

Table 6: Security level based on the threat and impact levels [8], [9]

Security Level	**Impact Level**					
Threat Level		0	1	2	3	4
	0	QM	QM	QM	QM	Low
	1	QM	Low	Low	Low	Medium
	2	QM	Low	Medium	Medium	High
	3	QM	Low	Medium	High	High
	4	Low	Medium	High	High	Critical

It is interesting to note that the outcome of the TARA analysis for the threat presented in Table 1 is a security level “High”, which complements the ASIL level D identified in the HARA shown in Fig. 4. The safety and security goals are stated differently, but the ultimate result is the same safe state. As the two results of the analyses performed by different teams of experts in the safety and security domain complement each other, there is a need to prioritise the hazard / threat and apply even more effort in preventing it. Other TARA methods, such as those shown in [10] and [11] were performed as well and produced a consistent results. However, these are omitted from this paper due to space limitations. They will be discussed in the oral paper presentation.

3. Requirements

Both HARA and TARA are used to identify requirements for safety and security respectively. The systematic derivation of safety requirements from the safety goals is one of the fundamental ideas of ISO26262. Several refinement steps are considered in the safety development, as illustrated in Fig. 5.

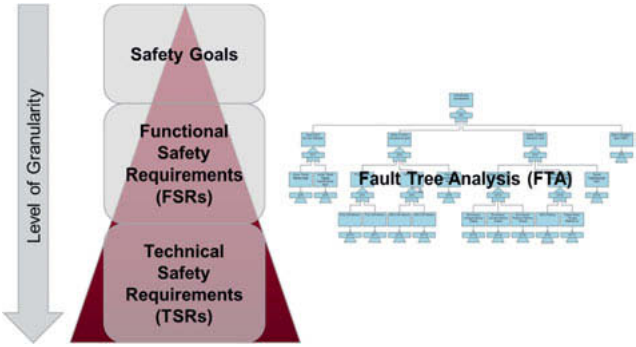


Fig. 5: Derification of safety requirements from safety goals by means of fault tree analysis

The fault tree analysis (FTA) is an established method to support the development of safety requirements, taking into account the respective functional chain which can contribute to the violation of a safety goal. For the exemplary safety goal SG027, Fig. 6 shows in a simplified way how the FTA results in potential root causes for the safety goal violation.

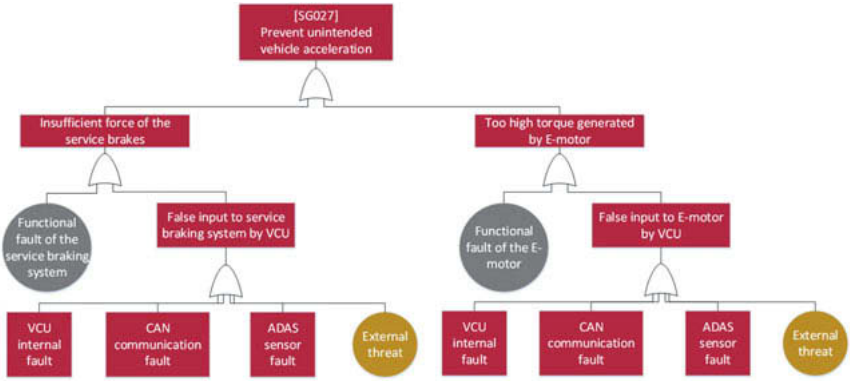


Fig. 6: Simplified extract of fault tree analysis for the violation of safety goal SG027 of the Urban Chauffeur function

The external threat in Fig. 6 could be due to a cybersecurity attack, through which an attacker can provide false inputs such as by packet insertion over the CAN bus. This is extended by the cybersecurity team in consultation with the safety team as shown in Fig. 7. The external

attack is broken down into the possibilities of attacks using an Attack Tree. The Attack Tree is used to identify the high level security requirements through an Attack Tree Analysis (ATA). This is also analogous to the Fault Tree Analysis (FTA) used in the safety domain. The root node (at the top level) represents the ultimate attack goal. Each branch in the tree represents one method of achieving the goal represented by its parent node.

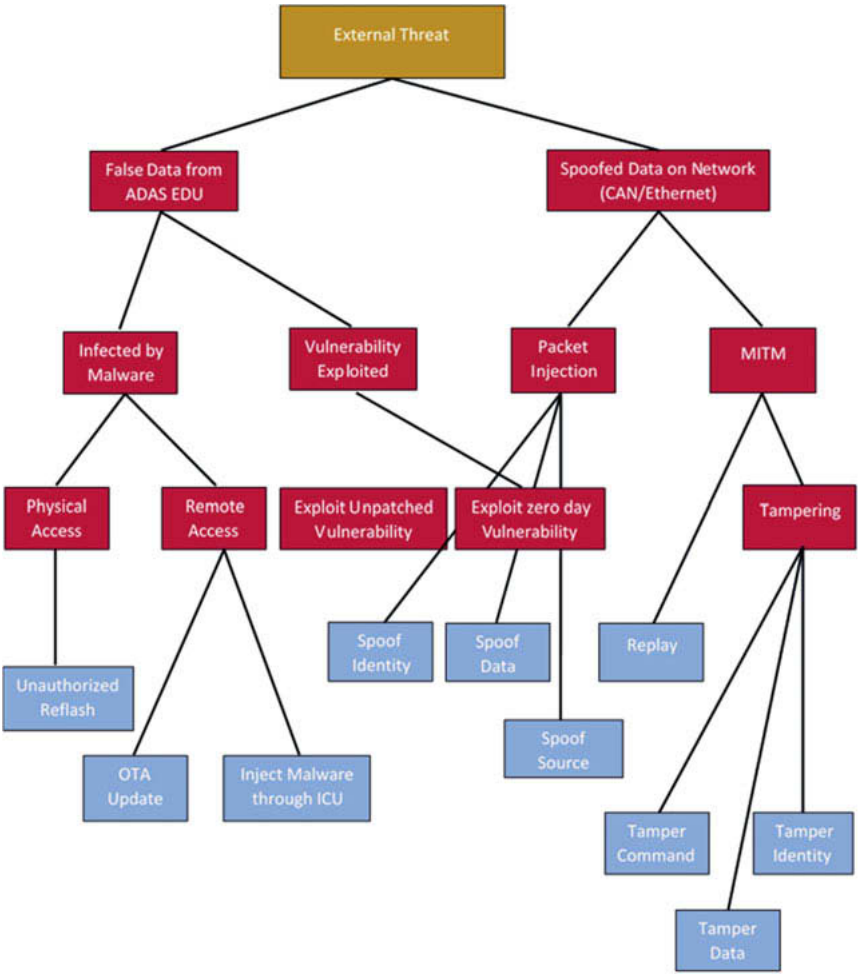


Fig. 7: Simplified extract of an attack tree extending the external threat node of the fault tree

Security requirements are developed using the assets, threats and security levels identified using the HEAVENS project approach (or any other equivalent approach). The attacks are prioritized and listed. The “critical” attacks are then investigated with the attack tree analysis. The leaf nodes in the attack trees identify the most common attacks for which high level security requirements are developed.
As an example, tampering can be prevented using data integrity and authentication. Likewise, the unauthorized re-flashing of firmware can also be prevented by ensuring the data integrity and authenticity.

Conclusion

It has been demonstrated that cybersecurity and functional safety development paths are only partially separated. In a development program, it is important right from the beginning that the technical experts of both domains analyse the system from their respective viewpoints. Nevertheless, a maximum of synergies and interactions between the two development paths shall be aimed at. As a starting point, a comprehensive feature / item definition is needed for both teams. The Hazard Analysis and Risk Assessment (HARA) and the Threat Analysis and Risk Assessment (TARA) are carried out separately, but a joint discussion and alignment on the results is essential. By means of the fault tree analysis, a systematic approach is applied to derive both functional safety and cybersecurity requirements and trace them to the same safety goals. A final system-FMEA shall cover both functional safety and cybersecurity aspects and therefore ensure completeness of the respective requirements. This process ensures the development of a safe and secure system, avoiding unnecessary efforts and delays caused by parallel development paths as much as possible.

References

[1] Annual Accidents Report 2018, European Road Safety Observatory (ERSO), available online at (https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/statistics/dacota/asr2018.pdf)

[2] https://www-fars.nhtsa.dot.gov/Main/index.aspx

[3] M. Paine, D. Paine, C. Newland, and S. Worden, “Encouraging safer vehicles through enhancements to the ncap rating system,” in 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV), no. 11-0107, June 13-16, 2011, Washington DC, USA.

[4] ISO/FDIS 26262 2nd Ed., TC 22/SC32/WG08, Road Vehicles –Functional Safety Parts 1-12”, Mar 2018.

[5] SAE J3061, Cybersecurity Guidebook for Cyber-Physical Vehicle Systems, SAE standard, 2016.

[6] SAE J3016, Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, 2018.

[7] ISO/SAE CD 21434, Road Vehicles - Cybersecurity Engineering. (under development).

[8] M. Islam, A. Lautenbach, C. Sandberg, T. Olovsson, "A risk assessment framework for automotive embedded systems", CPSS 2016 - Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security, pp. 3-14, Xi'an, China, May 30, 2016

[9] D2: Security Models - HEAVENS project (available online at http://autosec.se/wp-content/uploads/2018/03/HEAVENS_D2_v2.0.pdf).

[10] G. Macher, H. Sporer, R. Berlach, E. Armengaud and C. Kreiner, "SAHARA: A security-aware hazard and risk analysis method," Design, Automation & Test in Europe Conference & Exhibition (DATE), March 9-13, 2015, pp. 621-624, Grenoble, France.

[11] A. Bolovinou, U. Atmaca, A. Sheik, O. Ur-Rehman, G. Wallraf, A. Amditis, “TARA+: Controllability-Aware Threat Analysis and Risk Assessment for L3 Automated Driving Systems,” June 9-12, 2019, Paris, France.

Are you Security Compliant?

Current Automotive Security Legislations, Potential Impacts to Automotive OEMs & Suppliers, and First Action Proposals

Moritz Minzlaff, **Marko Wolf**, ESCRYPT GmbH, Munich

1. Abstract

Disruptive changes in the automotive industry like electro mobility and autonomous driving impact not only the technical security realizations at OEMs and suppliers but also the organizational security realizations such as engineering processes or organizational structures.

Our contribution provides specific, effective actions for implementing a security management systems in order to meet and maintain increasingly demanding security requirements at technical and organizational level.

2. Understanding the Challenge

While vehicle safety – for many good reasons – is ensured by detailed, worldwide binding legislations since decades, today's vehicle security legislations are still worryingly vague and mostly non-binding, if existing at all. Considering however that vehicular security vulnerabilities might not only affect proper vehicle operation and vehicular business models, but particularly also safety-critical vehicle functions from today's driving assistants up to future autonomous driving functions (cf. "The Jeep Hack" from Miller & Valasek), this "cyber security legislation gap" needs to and will be closed soon by lawmakers and standardization bodies.

In fact, vehicle security has already become subject of many national and international legislative, regulatory, and standardization activities, for instance UNECE WP.29 Draft Proposal Cyber Security Regulation, the ISO/SAE 21434, and the proposed US Congress SPY CAR Act. These activities cover not only technical security requirements such as vehicular encryption standards (e.g., AUTOSAR SecOC) or vehicular hardware security (e.g., EVITA HSM), but particularly also organizational security requirements like secure vehicle engineering processes (cf. ISO/SAE 21434) or organizational vehicle security functions (e.g., PSIRT). Regulations will soon mandate proper technical and organizational vehicle cyber security measures in order to gain type approval.

Table 1 provides an extract of some of the main, emerging standardization & regulatory activities relevant to automotive security, most of them having a global scope. Thus, all OEMs and suppliers that sell products into one of the automotive markets in EU, US or China, have to address organizational cyber-security regulations now.

Table 1: Selected standardization & regulatory activities relevant to automotive security.

Organization	Formal Reference	Common Title	Type	Regional Scope	Status
ISO/SAE	21434	Cybersecurity Engineering	Standard	Global	Work in Progress
ISO/IEC	27000 series	ISMS Family of Standards	Standard Non-Automotive	Global	Published
SAE	J3061_201601	Cybersecurity Guidebook	Standard	Global	Published
UNECE WP.29	ECE/TRANS/WP.29 /GRVA/2019/2	Draft Proposal Cyber Security Regulation	Homologation	EU, Japan, Korea, others	Draft
China	ICV series	Various	Law	China	Work in Progress
US Congress		SPY CAR Act	Law	US	Work in Progress
EU	Regulation 2017/0225	Cybersecurity Act	Law Non-Automotive	EU	Published
ITU-T SG17	Various	Various	Standard	Global	Work in Progress
AUTO-ISAC	n/a	Various	Best Practices	US	Published (restricted access)

3. Identifying Key Requirements

A central theme to most of the mentioned standardization & regulatory efforts is that of a cyber-security management system (CSMS). It resembles in many ways the information security management systems (ISMS) that are well-known in the information security sector: Both management systems are risk-based approaches that require top management commitment to security (see also action #1) and appropriate organizational measures to achieve and maintain an acceptable risk level. They differ however in their focus: Where the ISMS

Action #1: Align product security strategy with corporate objectives

Implement a product security governance role that directly reports to the board, cf. CISO. Make this role responsible for selecting and prioritizing product security activities in alignment with safety goals, regulatory requirements, and corporate objectives.

has an internal focus on the security of the organization's information assets, the CSMS focuses on the security of the products the organization sells. This focus on the vehicles and their security makes safety impact a crucial dimension for these new types of management systems that is not found in a classical ISMS. Maintaining an acceptable risk level is therefore not a purely business motivated decision, but one of avoiding harm to all road users.

4. Identifying Impacts & Stakeholders

A CSMS as described in the previous section will affect the entire organization. Copying existing ISMS approaches will not suffice as the product security is very different from classical internal IT security (e.g., target access for applying changes). In case of security incidents customer related departments will have to play a crucial role. In fact, the proposed UN regulation requires the CSMS to cover the entire supply chain across the whole life-cycle of vehicle types. Today's tremendously complex automotive supply chains, make involvement of all stakeholder into the security management system a key activity. Internally, engineers have to be trained to develop secure products, marketing departments will have to develop routines for incident handling, security analysts have to be available for ad-hoc, emergency investigations, and legal teams have to draft appropriate security requirements for sourcing suppliers. Once routines are in place, all partners along the supply chain should test and refine their setup through fire drills.

Action #2: Involve all stakeholders in the cyber security management system
For the cyber security management system to be effective and yield the desired results (see action #1), it must involve stakeholders across departments and the supply chain. This requires defined interfaces, a trained workforce, and appropriate contractual agreements when crossing company boundaries.

5. Security Roadmaps & Alignment with Corporate Objectives

The UN draft regulations is currently expected to come into force in the EU in the first half of 2022 and in some markets even sooner than this. Considering development cycles in the automotive domain that often span three or more years, it is clear that the whole automotive industry needs to prepare now. In many cases, the OEMs and suppliers already have a solid foundation upon which to build a cybersecurity management system. Aspects of safety engineering might be reused, a strong and knowledgeable testing team can provide meaningful

input to the risk assessment phase (and of course more easily integrate security testing) and so on. To meet the demanding timelines and avoid engineering process coming to a standstill due to too many new process steps, it is a key activity to assess what process activities are already in place that can serve as cornerstones of the future CSMS. A fit/gap analysis makes these potentially implicit foundations transparent and clearly describes the areas that benefit the most from immediate improvement. Together with a product security governance team, the fit/gap analysis is a crucial tool to become ready for upcoming regulation.

Action #3: Focus on high ROI activities first based on a careful status quo analysis management system

A CSMS implementation will require changes across the organization. Focus on those areas first that yield the highest security benefit or ROI. A gap analysis with respect to relevant standardization and regulatory requirements will provide the necessary foundation to identify those high value improvement measures.

6. Conclusion

Proper automotive security has always been not only a technical, but also an organizational topic. With the deep impact of automotive security especially to driving safety (cf. autonomous driving), automotive security has recently become objective of several international legislations up to type approval. A central theme to most of the standardization & regulatory efforts is that of a cyber-security management system (CSMS) that can be realized by (i) aligning product security strategy with corporate objectives, (ii) involving all stakeholders in the cyber security management system, and (iii) focusing on high ROI activities first based on a careful status quo analysis management system. Considering that the UN draft regulations is expected to come into force in the EU already in 2022, it is clear that the whole automotive industry needs to prepare now.

Integration of Cybersecurity into Development Processes

A Case Study

Florian Stahl, AVL Software and Functions GmbH, Regensburg

1. Introduction / Abstract

„[Cyber] Security is a Process, not a Product.“ says the famous cryptographer and security guru Bruce Schneier and this certainly is more than true for the Automotive Industry where many companies currently still try to secure their components or vehicles by just adding security technologies. But securing such a complex environment like a modern vehicle with more lines of source code than Facebook [1] and wide more than 100 different software suppliers needs more than technologies. It needs processes to control complexity in the first place and to consider cybersecurity in all development phases - from early concept until disposal.

Most conference talks or papers only give a rough outline of necessary process steps or focus on theoretical approaches. However, they often miss or underestimate the challenge to integrate cybersecurity into existing development processes and implement them in practice. For this reason, this paper gives practical insights on how cybersecurity is considered in the development processes of a tier-1 supplier in alignment with current standards like SAE J3061 “Cybersecurity Guidebook for Cyber-Physical Vehicle Systems” and the community draft version of ISO 21434 “Road Vehicles – Cybersecurity Engineering”. In addition, we show examples, challenges, and critical success factors.

2. Cybersecurity in the Development Process

An automotive cyber system can only be secured in a reliable and resilient way if security considerations already start at the very beginning of the development. Still, the security development process does not even end with the start of production. Considering that a vehicle is on the road for many years and the state-of-the-art in security solutions is rapidly evolving, it is vital that the effectiveness of the built-in security solutions is reviewed and maintained constantly over the lifetime.

Fig. 1 gives a rough overview where in the development lifecycle cybersecurity activities take place.

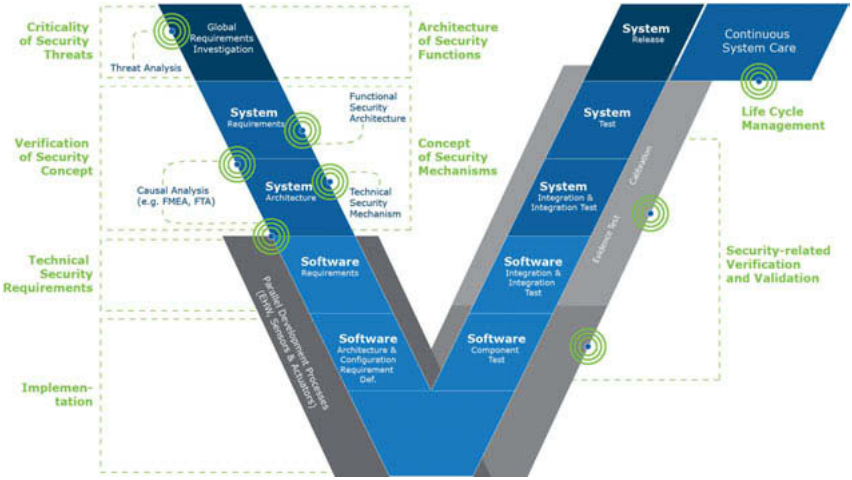


Fig. 1: Cybersecurity Activities in the Development Lifecycle

A cybersecurity relevance assessment is the entry point into cybersecurity engineering when developing a vehicle or some of its components. Once decided that cybersecurity is needed, a Threat Analysis and Risk Assessment (TARA) has to be performed. The essential basis for it is an unambiguous item definition to define the scope of the development. The first part of a TARA is to identify potential threats (Threat Analysis). Usually this is done in a moderated workshop with a cybersecurity expert together with domain experts and developers. In the risk assessment part, they discuss potential threats harming the availability, integrity, confidentiality of the component and evaluate the risk associated with them. Table 1 shows an exemplary RASIC chart with the roles involved in a TARA from a tier 1 perspective.

Table 1: RASIC for TARA – Letters stand for: Responsible, Supporting, Accountable, Informed

Activity	Role			
	Project Security Manager	Requirements Engineer	Project Manager	OEM
Threat Analysis	R	S	S	I
Risk Assessment	R	S	A	S
Risk Report	R	S	S	A

Tools can support TARA and implement different methods for assessing the risk for each threat, like HEAVENS, EVITA, SAHARA or other calculation schemes. The risk can be estimated e.g. by combining the financial, operational, legal and safety impact with the attack potential which considers the time, knowledge and equipment required to perform a successful attack. The output of a TARA is a risk report that proposes how to treat each identified risk e.g. by reducing it with counter-measures (implement security requirements) or accepting it. From a tier 1 perspective, this decision has to be taken by the OEM where medium to upper management should decide on risk treatment depending on the criticality of the risk. TARA is a very important step in cybersecurity engineering because missing threats or false assessment might lead to either critical vulnerabilities or inappropriate costs for security.

After TARA, each threat gets a security goal assigned. These security goals serve as top-level reference to all subsequent security activities like deriving security requirements and describing them in a cybersecurity concept. The following table shows an example for a threat – goal – requirements – combination. One threat can be mitigated by one or more requirements.

Table 2: Threat and Security Goal with Requirements

Threat ID	Threat	Security Level (Risk)	Security Goal ID	Security Goal	Requirement ID	Requirement
T1	Manipulation of ECU firmware	Medium	SG1	Protect the integrity of the firmware	SR1	Ensure that only authentic firmware is flashed onto device.
T1	Manipulation of ECU firmware	Medium	SG1	Protect the integrity of the firmware	SR2	Ensure detection of unauthorized firmware manipulation.

Many security flaws can be addressed by following recommended best practices and highest quality standards in software development. Coding Guidelines like the ones from MISRA or CERT-C are important to avoid weaknesses in the implementation. Since security vulnerabilities can occur in every single line of code (not only code related to security functions like encryption), every developer has to be aware of how to write secure code. Tools for code analysis can help to find issues like buffer overflows or race conditions.

Modern vehicles consist of many interconnected components that have been carefully tested for correct functional behavior to avoid safety problems. On the other hand, testing against potential security gaps is not yet a common procedure within the automotive domain. However, Security Verification should become part of the verification process and depending on the level of criticality and exposure include static code review, dynamic analysis, fuzz testing and penetration testing.

In 2018, the average age of a vehicle in Germany is around 10 years. Due to this fact, it is important to consider a proper lifecycle management for the cybersecurity of vehicles. Therefore, Continuous System Care (also known as Long-Term-Support or Vulnerability Management) considers all topics regarding lifecycle management from the early development phase until end of life. Within Continuous System Care security mechanisms, algorithms and hardware in the ECUs are monitored for security vulnerabilities. If an issue occurs, the variant management database is checked for affected devices. As part of the incident response plan, the vulnerability is assessed with respect to affected security mechanisms, algorithms and hardware. That includes evaluating the vulnerability regarding the affected security goal from TARA, the technical expertise and conditions to exploit the vulnerability. If the vulnerability is not acceptable, corresponding incident response steps like preparing OTA updates or information of the customer will be initiated.
All these processes, activities, and even more is necessary to provide appropriate cybersecurity for a vehicle or its components. Appropriate because 100% security is not possible.

3. Certification of Cybersecurity for Vehicles

That is also one reason why certification of cybersecurity for vehicles cannot be based on technical tests of every version of a car and its components, but rather focus on processes and risk management. UNECE (United Nations Economic and Social Council) proposes such an approach for a Cybersecurity Management System (CSMS) [2]. A CSMS shall cover the following aspects for the development, production and post-production phase:

> *"The vehicle manufacturer shall demonstrate that the processes used within their Cyber Security Management System ensure security is adequately considered. This shall include:*
>
> *a) The processes used within the manufacturer's organization to manage cyber security;*
> *b) The processes used for the identification of risks to vehicle types;*

c) *The processes used for the assessment, categorization and treatment of the risks identified;*
d) *The processes in place to verify that the risks identified are appropriately managed;*
e) *The processes used for testing the security of the system throughout its development and production phases;*
f) *The processes used for ensuring that the risk assessment is kept current;*
g) *The processes used to monitor for, detect and respond to cyber-attacks on vehicle types;*
h) *The processes used to identify new and evolving cyber threats and vulnerabilities to vehicle types;*
i) *The processes used to appropriately react to new and evolving cyber threats and vulnerabilities. […]*
The vehicle manufacturer shall be required to demonstrate how their Cyber Security Management System will manage dependencies that may exist with contracted suppliers and service providers in regards of the requirements [mentioned above].”

In addition, technical tests by the Inspection Authority for each cybersecurity-relevant change would not be feasible due to the sole effort that it would cause.

4. Examples & Critical Success Factors

As of today, most automotive companies have cybersecurity already on their agenda, but with a very different level of maturity. In practice, some do not consider cybersecurity at all yet or skip several steps or processes related to cybersecurity like TARA or risk treatment and come up with a list of security requirements directly. Some OEMs have already quite specific processes and guidelines regarding cybersecurity, but the degree of implementation is still very different. Some project managers are not fully aware of these new requirements and it is not always easy for a supplier to get the required information to take the right decisions regarding cybersecurity. For example, it is hard to perform a TARA without knowing the architecture and security measures of surrounding bus systems.

Thus, the main challenge is not to define and document processes for cybersecurity, but to make them work in practice and to make developers, project managers and other stakeholders aware of cybersecurity as an indispensable requirement in development. Therefore, training and change management is needed and it is rather a question of years than months to make cybersecurity work among all development activities in a huge organization. It is essential to explain everyone why cybersecurity is needed and what are the consequence without it – cars will get hacked, and reputation damaged. Because other than most functional requirements

cybersecurity does not add a value for the customer at first glance and furthermore, cost, time and resource restrictions are additional hurdles.

A widespread approach is to re-use processes and methods from safety for cybersecurity. However, challenges like class breaks (a single attack affects a whole car fleet at once) or targeting the OEM and not the driver do not exist in the safety world. In addition, safety only protects one single objective (driver's health) while cybersecurity has three objectives (confidentiality, integrity, and availability). In short, security can be described as the protection from intentional malicious manipulation, theft, etc. while safety deals with unintentional threats. In the end, aligning cybersecurity too close with safety will fail. But for sure, interconnections have to be considered because cybersecurity issues might cause safety issues and vice versa.

Another success factor is the open communication between suppliers and OEMs to make and keep the whole car secure. However, this openness is still difficult for many companies because they fear to share intellectual property. Other sectors like banking or IT have shown that collaboration is a key success factor for cybersecurity. AUTO-ISAC (an organization where automotive companies share security vulnerabilities and experiences) is a good starting point, but cybersecurity is only as good as the weakest link in the chain and every supplier delivering some security-relevant component has to be considered.

An issue for the growing topic of cybersecurity is the lack of real experts in the automotive domain. They are rare and desperately sought after. Mostly two groups currently jump in to fill this gap:

- People from the safety domain that tend to have a lack of knowledge in cybersecurity
- People from IT / Information Security that tend to have a lack of knowledge in automotive development

5. Summary & Outlook

In conclusion, experts and trainings are needed because automotive companies have recognized the need for cybersecurity in product development and are currently in the transition to define and establish processes. Some have reached quite some maturity, but others have not even started yet and there is hardly an OEM covering its full supply chain yet.

According to the Ponemon Study of Automotive Industry Cybersecurity Practices [3] 30 percent of the automotive companies do not have an established product cybersecurity team or pro-

gram and 23 percent do not have a centralized team yet. Even 84 percent of automotive professionals have concerns that their organizational cybersecurity practices are not keeping pace with evolving technologies.

The expected publication of the ISO/SAE 21434 in late 2020 is expected to further push the topic and will establish cybersecurity in E/E development as a must-have for every automotive company.

Another guiding document from the Alliance of Automobile Manufacturers and the Association of Global Automakers is the Framework for Automotive Cybersecurity Best Practices [4]. It lists five principles for cybersecurity in vehicles that provide a good roundup:

1) Vehicle security by design.

2) Risk assessment and management.

3) Threat detection and protection.

4) Incident Response and Recovery.

5) Collaboration and engagement with appropriate third parties.

Once more, these principles support the statement that processes are more important for cybersecurity than products and technologies. The establishment of such processes will be challenging for the automotive world over the next years. However, there is no alternative if connected and autonomous vehicles shall become and stay reality.

Information sources:

[1] https://informationisbeautiful.net/visualizations/million-lines-of-code/

[2] UNECE Proposal for a Recommendation on Cyber Security, see https://www.unece.org/fileadmin/DAM/trans/doc/2019/wp29grva/ECE-TRANS-WP29-GRVA-2019-02e.pdf

[3] Ponemon Institute (independent study commissioned by SAE and Synopsis): “Securing the Modern Vehicle: A Study of Automotive Industry Cybersecurity Practices”, see https://www.synopsys.com/content/dam/synopsys/sig-assets/reports/securing-the-modern-vehicle.pdf

[4] Framework for Automotive Cybersecurity Best Practices, see https://www.globalautomakers.org/OldSiteContentAssets/press-release/Automakers-Develop-Framework-for-Automotive-Cybersecurity-Best-Practices-assets/framework-autocyberbestpractices-14jan20161-pdf

The transition to HPC-based vehicle architectures

Cyber Security Implications

Andrey Shomer, Argus Cyber Security, Tel-Aviv, Israel

Abstract

As the automotive industry moves towards fully connected cars and autonomous driving, the need for strong computing power inside the vehicle is significantly increasing. Vehicles are gathering, processing, and transmitting increasingly large amounts of data. Sent from multiple sensors, such as lidar, radar, and cameras, the different data streams need to be analyzed simultaneously and in real-time. This challenge is being met with the incorporation of High-Performance Computing (HPC) in-vehicle architectures, empowering the infrastructure with the ability to execute CPU-intensive tasks.

In an HPC-based vehicle architecture, there are several HPCs grouped according to functionality, safety level, and cybersecurity considerations. These centralized high-performance platforms host multiple virtual ECUs, which control the majority of the vehicle's functionalities, on a single processing unit using a hypervisor. The virtual ECUs are based on various operating systems (like Linux, Autosar). To enable the transmission of large amounts of data over higher bandwidths, automotive Ethernet is incorporated as the communication backbone in HPC-based architectures.

This change in onboard infrastructure has introduced new automotive-specific protocols and new attack vectors and has provoked the need for further innovation in protecting vehicles against cyber-attacks. For example, attackers can compromise automotive-specific protocols to launch unprecedented DoS attacks. In these attacks, an enormous amount of data is sent to a victim network node effectively neutralizing it. This was successfully demonstrated by the Argus research team during a next-generation architecture penetration test project: a virtual Adaptive Autosar-based ADAS unit was disabled by a DoS attack, disabling the advanced driver assist functionality in a way that current cyber security solutions could not identify and block.

The new virtualized environment has created a new automotive attack vector in which attackers have proven their ability to break out of a virtual ECU and compromise the host hypervisor - potentially compromising all the virtual guests. This type of attack is known as Guest to Host Escape.

Alternatively, HPC architecture has introduced new security design opportunities, which enable the introduction of dedicated virtual *Security ECUs* to the vehicle architecture, without the added costs of an additional hardware component.

As a global leader in automotive cyber security, Argus' cooperation with multiple OEMs on securing HPC-based vehicle architecture has led to the development of a comprehensive cybersecurity approach that combines host protection and network protection in multiple modular and scalable components: host-level protection on every partition, hardening hypervisor software, advanced heuristics that pinpoint malicious network traffic, a virtual security ECU concept and more.

In this paper, we will elaborate on the security challenges derived from the transition to HPC-based architecture. We will explain how this industry shift will transform existing cybersecurity concepts, and we will explore innovative protection methodologies and solutions that should be introduced into the design and production of vehicles with HPC architecture.

Introduction

Throughout recent years the electric and electronic (E/E) architecture of modern vehicles and cars is going through a significant change. After several decades of stagnation, during which the E/E design changed slowly and incrementally, recent years triggered a revolutionary change with the rapid introduction of software-based systems that require significant computational power and produce large amounts of data. In addition, new automated driver-assist systems (i.e. adaptive cruise control and lane departure warning) heavily rely on numerous sensors producing large amounts of data to be processed and transported through-out the in-vehicle network.

These changes resulted in a need to revise the in-vehicle computation and communication components. A core part of this revision is the adoption of High-Performance Computing (HPC).

While different vehicle and component vendors address the term HPC slightly different, for the purpose of this paper the term HPC relates to In-Vehicle Ethernet communication (802.3bp; 802.3bw; 802.3ch), and/or ECUs which are based on multiple *traditional* automotive MCUs (microcontrollers) or performant computing cores capable of executing POSIX-like operating systems and hypervisors. These two characteristics of HPC-based vehicle architectures introduce new cyber security risks for modern vehicles.

One of the major changes in vehicle E/E architectures is the introduction of new automotive Ethernet protocols, such as the SOME/IP protocol group. Typically, existing Ethernet network protocols cannot be integrated into vehicle platforms “out of the box”. Instead, implementations are tailored to meet the requirements of the automotive world in terms of safety, performance and usability. Decades of experience in network security have shown that new protocols and new implementation for legacy protocols tend to create security gaps that are vulnerable to attacks. In addition, known issues and limitations which are already known in the IT world pose a new threat to automotive in-vehicle networks.

The era of connected cars brought new capabilities to the automotive domain - Over-The-Air updates and continuous remote monitoring and remote diagnostic capabilities. These capabilities are well known from *smart* end-user devices such as personal computers and mobile phones. As opposed to end-user devices, which are based on a single or limited amount of control units, a modern vehicle consists of multiple electronic control units, each relies on its own software package and is usually produced by a different manufacturer. This variance requires vehicle manufacturers to choose one ECU that will centrally manage the aforementioned capabilities.

The gateway is an important member of the E/E vehicle architecture. Traditionally, it was functioning as a bridge between different CAN network segments on the in-vehicle network, but most modern vehicle designs assign new roles to this central unit while also keeping its original purpose. The combination of safety-critical functionalities on the one hand and new connected services on the other introduces new risks both from the functional and cyber-security point of view. Such combination is well illustrated by the gateway example but not limited to it. According to the previously proposed *HPC* definition, a typical gateway is, in fact, a high-performance computing platform, since it is comprised of multiple *traditional* automotive MCUs (microcon-

trollers) or powerful computing cores executing POSIX-based operating systems and hypervisors. In the next section, different HPC architectures are described followed by an analysis of potential cyber-security issues and advantages.

HPC design choices

An HPC based ECU in one vehicle architecture may differ from its counterpart in another vehicle. In this section, several architectures, all of which can be considered *HPC, are* outlined and analyzed from a security standpoint.

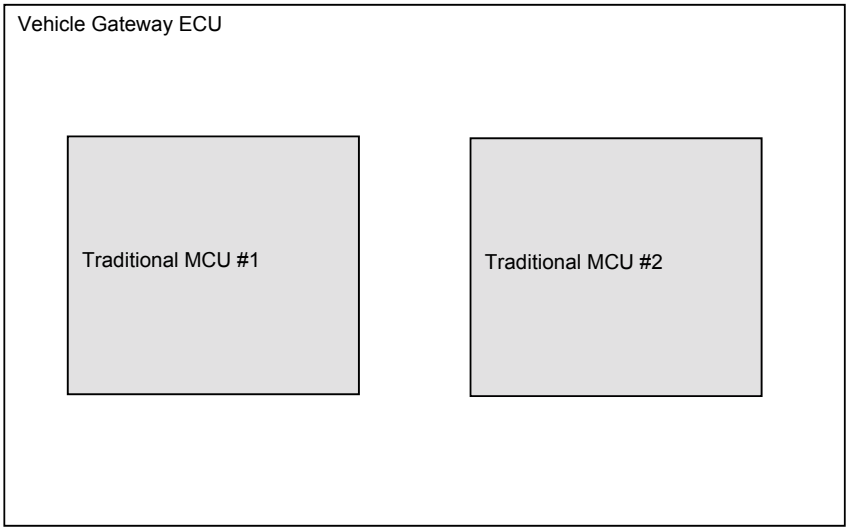


Fig. 1: High-Level architecture of a vehicle gateway incorporating multiple microcontroller units to handle performance-demanding tasks

1. Separate microcontrollers

The demand for additional functionality within central ECUs is rapidly growing, and often exceeds the capabilities of currently available hardware, due to limited hardware resources. Component providers found a way to provide the necessary functionality without increasing the anount of hardware, by designing ECUs which are based on multiple microcontrollers. Less powerful microcontrollers are bundled together on a single printed circuit board (PCB). Features can be then separated between microcontrollers based on safety, functionality or performance.

The separation of functionalities between the different hardware units has a surprising, residual cyber security advantage. Even if an attacker was able to exploit a vulnerability in a software package running on one of the microcontrollers, he might not be able to compromise the second core. The two microcontrollers design provides, de-facto, two separate execution environments that enable the isolation of certain functionalities between the different hardware units. With such architecture in place, a gateway designer may then decide to divide the different functionalities according to the imposed cyber security threat. For example: connected services may reside on one core while safety-critical features may reside on the other. Moreover, such separation introduces an opportunity to divide in-vehicle networks into different segments based on how vulnerable the components within these zones are. An example of such a concept is presented later in this paper in the cyber security solutions section.

2. Performant Processing Unit running a Hypervisor

In certain cases, when an applicable hardware unit is available, manufacturers will choose a Hypervisor (HV)-based architecture. A hypervisor is a software package that creates and runs virtual execution environments. Such execution environments can execute full-blown operating system or a single process. Hypervisors are used to fulfill several goals: limit the amount of resources dedicated to each function or feature, separate software images and distribution, and answer cyber-security needs. When tasks are separated using a hypervisor, the cyber security benefit will be like the advantage achieved by using physically separated computing environments.

At a glance it may appear that the desired separation of execution environments is achieved using a hypervisor, however this may not be the case. In the use case where two or more different microcontroller hardware units provide separate execution environments, the separation is achieved since the microcontroller units are physically distinct from each other. In the

hypervisor use case, the separation is highly dependent on the hypervisor's ability to effectively separate the resources assigned to the different partitions. An illustration of threats which may be caused by flaws in the hypervisor structure is presented in the ARINC 653 Case study section later in this paper.

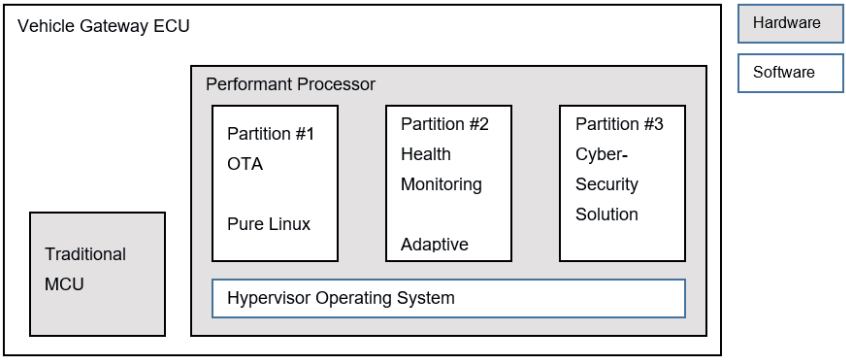


Fig. 2: High-Level architecture example of a vehicle gateway incorporating a powerful core and a hypervisor software solution to separate different system functionalities using different partitions.

3. Hybrid CAN and Ethernet network architectures

The transition from CAN Bus-based automotive networks to fully automotive Ethernet networks is gradual. Most transitional architectures include both Ethernet and CAN network segments. High bandwidth applications such as rear-view cameras, internet connectivity and transmission of radar data are conveyed on Ethernet links while body and safety-critical functions are commanded through CAN or CAN-FD.

The separation into several network types within one vehicle is beneficial from a cyber security perspective. Vulnerabilities which are exploited over the network, usually require specially crafted network frames. This type of vulnerability is further described in later this paper. Having two network types within the E/E architecture effectively forces the disassembly of network frames when forwarded from one network type to the other. Once a malicious frame is disassembled and stripped of its headers, the vulnerability would most likely be eliminated

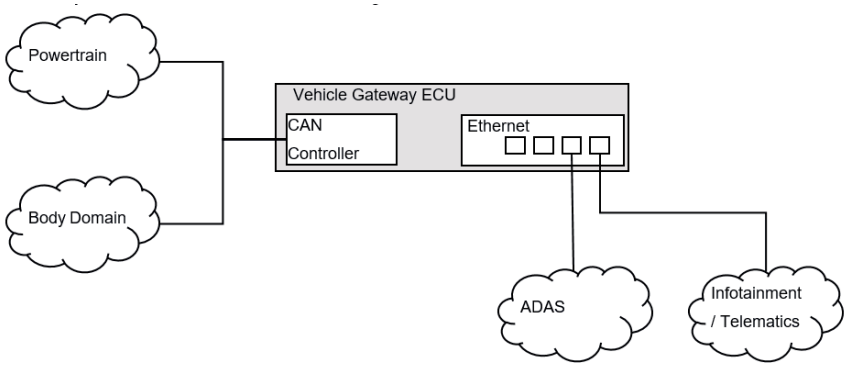


Fig. 3: Illustration of a hybrid (CAN and Ethernet) in-vehicle network

4. All Ethernet in-vehicle network backbone

In-vehicle network design that relies solely on Ethernet communication is cost effective. The cost of maintaining different communication technologies is eliminated and high throughput becomes available to all vehicle domains.

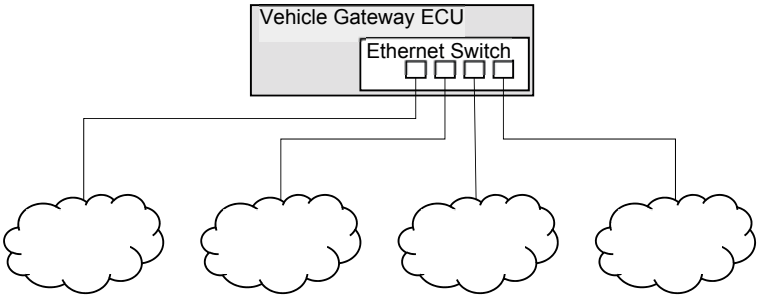


Fig. 3: Illustration of all-Ethernet network backbone

However, from a cyber security perspective the unit is less secure when compared to the hybrid network, as the propagation of an attack based on a flaw in a protocol in the Ethernet stack is now possible throughout the in-vehicle network.

ARINC 653 Case study

The use of hypervisor software to separate functionalities is not limited to the automotive domain. In fact, such architecture is very common in other industries and particularly in avionics systems. The lessons learned in other industries are also relevant for automotive. For avionics systems, a standard was developed to define the requirements of safety-critical systems – ARINC 653

Similar to avionics, the emergence of hypervisor-based automotive systems was driven by the need to separate functions of different safety criticality, but as soon as the first designs were available, it was understood that cyber security should also be taken into consideration. As described in the previous section, regarding several HPC architecture models, the separation of functionalities is effective to isolate a potential attacker from propagating throughout a compromised system.

Some of the hypervisor implementations used in the automotive domain are the same software packages that are certified for ARINC 653 compliance in avionics. Moreover, the security of such implementations is often certified according to the *Common Criteria* specification. Thus, potential security flaws discovered in avionics implementations might already be present or may arise in the future in automotive implementations.

A comprehensive security analysis of hypervisor separation Kernels was conducted by Yongwang Zhao, David Sanan, Fuyuan Zhang, Yang Liu in Refinement-based Specification and Security Analysis of Separation Kernels. The analysis has shown a high risk of an existence of a covert channel in ARINC 653 Separation Kernel implementations. A covert channel refers to the ability to access the data of one hypervisor partition by exploiting another partition. In a system where a covert channel exploitation is possible, an attacker who gained access to a single virtual partition may then use the covert channel to compromise an additional virtual partition within the system.

New threats imposed by incorporation of Ethernet networks

As mentioned in the previous section, the incorporation of Ethernet networks in vehicle network architecture, either as the full backbone or as part of a hybrid network structure, introduces new cyber security threats. The Argus research team demonstrated in two recent penetration testing projects, how attackers can exploit the Ethernet network to maliciously influence in-vehicle ECUs.

1. Implementation vulnerability in a legacy network protocol

During an investigation of a pre-production AUTOSAR Classic Platform, we discovered a high-risk vulnerability in an Ethernet stack implementation which was subsequently fixed. This vulnerability can be exploited by an attacker to perform Remote Code Execution, simply by sending specially crafted IP packets to the Ethernet interface of the ECU.

The vulnerability was discovered in the implementation of IP Fragmentation. The implementation used a control structure to keep track of received fragments and reassemble to the original packet. When a fragmented IP packet is received, it uses a pre-allocated buffer in order to store the packet's fragments and to track the gaps that remain to be filled by future fragments.

The vulnerability results from the assumption that all packet fragments have the same IP-header length. By sending a specially crafted packet with a differently sized IP header, the attacker can cause the payload of previous fragments to be treated as gaps parameters. This could be exploited to achieve arbitrary Remote Code Execution.

Interestingly, a typical AUTOSAR Classic Platform system does not employ any security mitigation techniques. For example, known techniques which are relatively popular on Linux systems such as ASLR, Stack Protection or MAC do not exist on a typical AUTOSAR Classic Platform ECU. This means that once an attacker finds such a vulnerability, there are typically minimal obstacles to prevent him/her from exploiting the vulnerability and relatively quickly, succeed in manipulating the AUTOSAR ECU. Such an attack will have serious ramifications since Classic AUTOSAR based ECUs are the implementation choice for safety-critical components in the vehicle, for example ABS, Steering, and Powertrain. Moreover, depending on the EE architecture, taking over an AUTOSAR Classic Platform ECU may have additional security-related concerns. Compromising a Gateway or a Domain Controller may allow an attacker to bypass network separation and authentication concepts across the EE architecture.

2. Design vulnerability in a new Automotive Ethernet protocol

Another example of a vulnerability in an Automotive Ethernet protocol relates to the introduction of new protocols - SOME/IP-TP and SecOC for Ethernet. A design flaw in AUTOSAR Classic Platform standard was recently identified, which exposes vehicles using SecOC over SOME/IP-TP to Selective Denial-of-Service (DoS) attack.

The SOME/IP-TP segmentation protocol was developed by AUTOSAR, in order to support the transmission of large SOME/IP PDUs (Protocol Data Unit) over UDP. AUTOSAR SecOC (Secure Onboard Communication) is often used to sign SOME/IP PDUs to achieve End-to-End authentication.

The AUTOSAR Classic Platform standard does not specify whether SecOC should sign the complete SOME/IP PDU or each SOME/IP-TP segment individually. Signing the complete PDU achieves end-to-end authentication and is the preferred approach used by known implementations of AUTOSAR Classic Platform. An alternative, which signs segments individually, would not achieve end-to-end authentication, since some network elements need to resegment and resign packets, and therefore would prevent the receiver from authenticating the packet‘s original source. This would be the case when packets are forwarded between different in-vehicle network technologies. For example, when a packet is forwarded from Ethernet to CAN. As part of the conversion process, the forwarding Gateway would need to adapt the segments sizes to CAN MTU (Maximum Transmission Unit) and resign it.

We discovered that signing the complete SOME/IP PDU introduces a vulnerability that enables attackers to prevent specific packets from being received - a.k.a a selective denial of service attack. As a result, an attacker could stop on-going SecOC over SOME/IP-TP datagrams of choice in the network. To exploit the vulnerability, the attacker periodically injects specially crafted packets into existing UDP datagrams with a bogus SOME/IP-TP header, containing an incorrect Offset field. For example, if the attacker injects a ‘0’ offset value after the first legitimate segment was received, the message reassembly process will restart, the received segments will be discarded, and the complete packets will not be accepted.

Interestingly, a traditional DoS condition, for example, one which shuts down all ECU SOME/IP communication, would probably have been taken into account in the safety concept of the system. A selective attack like the one we presented, which is able to prevent an ECU from receiving a specific SOME/IP Event Group or RPC (Remote Procedure Call), may go unnoticed by the victim application (since related errors are handled by lower levels of the platform) - a scenario which may not have been taken into consideration in the safety concept of the system.

Possible cyber security solutions and conclusion

HPC architectures bring new cyber-security opportunities and risks. An HPC-based design may introduce additional security as opposed to a traditional system and can provide opportunities to implement new cyber security concepts. However, if security is not addressed comprehensively new attack vectors will arise.

The optimal cyber security requirements for a particular HPC design should be determined during the design phase of the unit and after a dedicated threat analysis and risk assessment (TARA) is completed. At a minimum the following security should be adopted to help prevent attacks from penetrating the unit:

1. Hardening host hypervisor system

Following the ARINC 653 case-study, it became clear that the hypervisor software itself shall be secured against attacks originating from a virtual partition and targeting another partition or the host system itself. A Hardening solution for the hypervisor host system should include verification of resource allocation within the host system and prevention of malicious code execution attacks.

2. Dedicated security partition

The capability to isolate a specific function from other tasks running on the same unit is useful for implementing a cyber security solution. In a traditional architecture, it would be practically impossible to add a separate unit dedicated to security, due to the high cost of additional hardware. A virtual cyber security appliance, on the contrary, is feasible. Separate hypervisor partition would be a clever choice to gather and analyze data from other partitions and network devices residing on HPC.

3. Network anomaly detection

Introduction of Ethernet communication and new protocols into new vehicle designs require a solution capable of analyzing the traffic on a frame-by-frame basis to detect anomalous frames. Such an analysis would potentially detect and prevent specially crafted frames, which attempt to conduct a code-execution or DoS attack as described in the section referring to new threats imposed by incorporation of Ethernet networks.

References

[1] https://www.elektrobit.com/products/ecu/eb-corbos/

[2] Arun Kumar, Sundar Rajan, ArminFeucht, Lothar Gamer, Idriz Smaili, Nirmala Devi M.: Hypervisor for consolidating real-time automotive control units: Its procedure, implications and hidden pitfalls

[3] Yongwang Zhao, David Sanán, Fuyuan Zhang, Yang Liu:
Refinement-Based Specification and Security Analysis of Separation Kernels. IEEE Trans. Dependable Sec. Comput. 16(1): 127-141 (2019)

[4] Common Criteria for Information Technology Security Evaluation,
3rd ed., National Security Agency, 2012.

[5] Detlef Zerfowski, Andreas Lock: Functional architecture and E/E-Architecture – A challenge for the automotive industry

[6] Paolo Gai ; Massimo Violante: Automotive embedded software architecture in the multi-core age

[7] Why HPC Matters: Autonomous Vehicles: https://www.cio.com/article/3253570/why-hpc-matters-autonomous-vehicles.html

Enhancing In-Vehicle Communication by Authentication and Security

An incremental approach with an example for CAN message authentication

Alexander Hahn,
Automotive Security Group, Microchip Technology Munich GmbH, Heilbronn

Abstract

Modern automotive E/E architectures consist of a huge number of nodes communicating over an openly accessible network inside the car. To ensure authenticity, integrity and protection against external attacks, a secure communication is mandatory. The challenge is to manage the complexity and variety of different communication protocols and nodes, as well as interoperability between different vendors. Rearchitecting a complete E/E architecture to ensure 100% of secured nodes is almost infeasible.

This paper presents a flexible and scalable approach for enhancing security in an in-vehicle network. The solution scales very well across all in-vehicle communication networks, from CAN to Ethernet communication with support for TLS. An example of CAN message authentication is shown in detail. One can selectively decide which ECUs need to be re-architected and which ECUs can remain almost unchanged. This minimally intrusive approach allows the system architect to better adjust the overall network architecture to the security needs with incremental change. Consequently, this reduces effort and risk while significantly reducing re-qualification efforts. Legacy ECUs can be preserved when needed, while feature-rich ECUs may undergo an architecture redesign.

1. Introduction

Security becomes more and more important also inside the automobile. Cars are connected, and we see an increasing number of threats here. Connections are done via telematics services or infotainment, vehicle to vehicle and V2X communication, Cellular phones or Bluetooth® and even through the on-board diagnostics. Todays in-vehicle networks and especially CAN are not architected for security and authentication. What happens if this is not the case?

You can read in the newspapers - there had been several incidents over the last few years. Besides loss of money due to the cost of recalling a fleet of several hundred-thousands of cars, this may also cause a severe damage of brand recognition and finally can lead to lawsuits as well.

The good news is that all this has elevated the awareness to the point where OEMs and consortiums need to take this into account and are updating and developing new cybersecurity specifications. Recently, also governments authorities such as the National Highway Traffic Safety Administration (NHTSA) require to protect integrity of critical ECUs and message communications and recommend to use hardware based security modules.

Another driver is the actual migration from CAN to CAN-FD, as this gives the possibility to introduce security measures in existing CAN networks.

This paper will focus on the following topics:

- Authentic message exchange between nodes with integrity.
- Secure key storage protected against physical attacks – secrets must remain secret.
- Secure boot and secure firmware upgrade possibilities.
- Key management and provisioning infrastructure across different Tier-1 suppliers.
- Minimally intrusive enhancement of existing networks with the example of the CAN network.

2. The Challenge of Automotive Security

Current EE-Architectures in a car are a very complex system, which has been developed and deployed over many years. And they have not been designed with security in mind, such as in the PC domain the introduction of Windows 95 in the mid of the 90s. They consist of multi-nodes, multi-protocols, and multi-vendors (Fig. 1), therefore there is no such an easy way of changing than for an encapsulated system. Adding security just for a handful dedicated ECUs or communications is not enough. One need to take into account that the security concept scales over and compiles with the whole EE-architecture

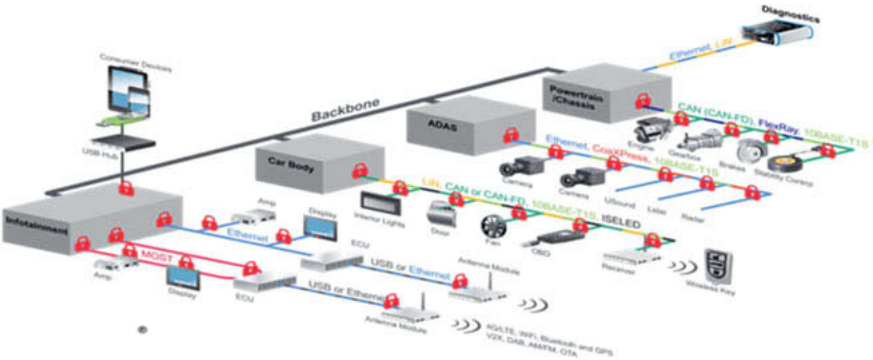


Fig. 1: Network architecture with multiple bus protocols

Today's automotive security model faces the following challenges:

- **Scalability:** The complexity of the automotive network architecture requires a scalable and incremental approach. It is almost infeasible to upgrade all systems at the same time.
- **Interoperability:** With multiple network protocols and multiple Tier-1s supplying ECUs for the same bus, standardization and/or flexibility is needed when enhancing the system for security purposes.
- **Maintain existing performance:** Response times, power-up times and performance calculations, such as bus-load and throughput for the CAN network, all need to be taken into account.
- **Hardware protection:** As cars are being connected, there is a need for dedicated HW-protected security. Adding keys in SW is not enough
- **Infrastructure/Ecosystem/Provisioning:** Provisioning concepts and supply chain are becoming more important for OEMs, while working with different Tier-1s, and applying a common security concept.
- **Preserve legacy where needed:** When the only change in the ECU is the security enhancement, there should be no need to rearchitect SW and ECU – an incremental approach should be preferred. This helps keeping migration effort under control.

3. The Three Pillars for Automotive Security

From a technical implementation perspective, automotive security needs to focus on three main topics:

- **Secrecy of the keys:** Security keys must not be exposed to the outside world – secrets must remain secret.
- **Secure boot and secure firmware upgrade:** Ensure that only authentic and genuine SW will run on the ECU.
- **Authentic communication with integrity:** Message communication between any two nodes must be authentic, the content of the message may not be altered.

Secrecy of keys:

Although more and more of today's system are designed with some awareness of security, they are very often developed by adding security functionality into the car ECUs while keys are stored in SW. But keys in SW are not secure. There are many well-known methods where keys can be accessed in a standard semiconductor device. A probe station where you can access address and data bus lines on the silicon of the microcontroller, like a logic analyser can be bought by anyone for a few thousand Euro. With that an attacker can tear down the ECU of a car and extract secret information, such as keys from the memory of the microcontroller – which then can be used to replicate remote attacks on many cars of the same model all over the world. Never use pure Software to protect private keys!

➔ SW alone is not secure!

How keys are protected matters: So how can you achieve strong security? Looking at a normal ASSP or microcontroller, it is fairly easy to identify memories, CPU modules and busses (Fig. 2a). A secure element does not contain a regular top-level. Instead, there is an active shield across the whole chip (Fig. 2b), which prevents attackers from identifying signal lines and reading content from the device. Various additional techniques are used such as anti-tamper mechanisms, protection against side channel attacks, over-/under voltage detection, scrambling and obfuscation, redundancy, noise generation and many more. And finally, if an attack is detected, the device can erase its own content, so there is no secret information available anymore.

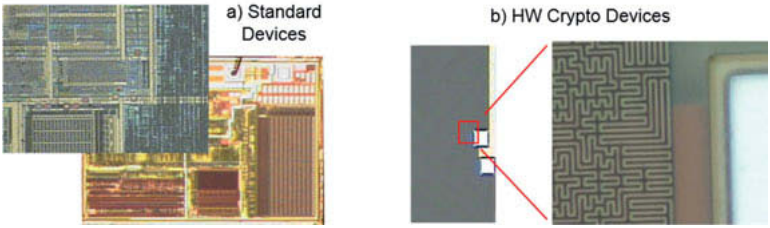


Fig. 2: Top level chip structure of standard cell device vs. crypto device

Such a secure element typically hosts a series of HW crypto modules for algorithms such as AES, SHA or ECC, and handles digital signature analysis, key generation and key exchange. It can store certificates and a number of symmetric and asymmetric keys. And it contains a high-quality random number generator. Finally each device may have a unique number, which can be used when generating keys and calculating signatures. By that, even if an attacker is able – with some significant effort – to retrieve any secret information from the secure element, this is only valid for one device (i.e. one ECU and thereby one car), and cannot be applied to an overall fleet or car model series.

Secure boot and secure firmware update:

To ensure that an ECU is not tampered with malicious (non-authentic) code, secure boot and secure firmware update is required. This allows the ECU to boot only from authentic SW and to only be upgraded with authentic SW, which is proven to be genuine and not altered. Of course keys for this SW builds must be stored secure. For system performance reasons, it is recommended that the secure boot process (and secure FW upgrade process) is configurable, as the boot process has impact e.g. on the system start-up time.

Keys and signature of the boot code are stored in the secure element. Usually the signature of the boot code is generated at OEM side or at manufacturing and is written into the secure element at manufacturing end of line. The secure element may partition this into a series of blocks and calculates a hash value for each, which is then stored safely inside the secure element. When the ECU in the field starts, the SW image is verified prior to the release of the ECU for normal operating function. The hash for each block is recalculated and compared to the stored hash inside the secure element.

For system performance reasons, that can be conFig.d in multiple ways. Hash can be calculated inside the secure element, or inside the MCU, if that is powerful enough and provides some HW crypto accelerator modules. Verification can be done for the whole image, or only for portions, one block, or a couple of blocks, until the functionality can be released in a verified stage. Even during normal operation of the ECU, secure boot verifications may continue on request or on a cyclical base.

During manufacturing end of line, a dedicated IO protection secret can be shared between the MCU and the secure element. Each MCU and secure element pair by that share a unique number, which is used to obfuscate the message. Even if an intruder is able to break into the ECU, he only gets information valid for this individual ECU, and cannot apply this to other ECUs via a remote attack.

➔ Only run genuine and authentic SW!

The same concept is applied when using field updates of the ECU firmware. During download, the new SW build is verified, and the new SW image is only written into internal flash memory, when the code is authentic and it can be ensured that the SW has not been altered.

Protecting communication:

With the increasing number of ECUs and numerous in vehicle communication interfaces, it becomes more and more important, that communication is protected. Message between different ECUs need to be authenticated to ensure that the message originates from a trusted source. And you need to ensure, that the content of the communication is with integrity and genuine, i.e. the message is not altered. Communication inside the car can be everything, from CAN over Ethernet to connectivity to the cloud or telematics system, or even connecting to the charging station and billing infrastructure, when having an electrical vehicle.

➔ Communication needs to be authenticated and with integrity!

4. Example: CAN Message Authentication

One of the challenges inside the car is the huge number of existing nodes, that are already developed, tested and qualified. For example, when looking on the CAN bus, those nodes may

not have powerful MCUs, and may not require a significant functional upgrade, when rearchitecting the network for security. Therefore, an incremental approach can help. The actual migration from CAN 2.0 to CAN-FD gives a good opportunity by using a dedicated device, which can autonomously reformat CAN 2.0 messages into CAN-FD messages and add authentication information into the gained payload and bandwidth.

➔ Migration to CAN-FD - an opportunity for security

CAN 2.0 limits the payload to 8 bytes, where it is very difficult, although possible, to add limited security information into those CAN messages. CAN-FD offers payload up to 64 bytes, which gives more space for additional information (Fig. 3). Even more, as CAN-FD is able to transmit the data faster, adding authentication information will not puzzle the bus load and a CAN-FD frame with 32 bytes of payload @2Mbit/s can be transmitted over the bus in the same time than a classical CAN 2.0 frame.

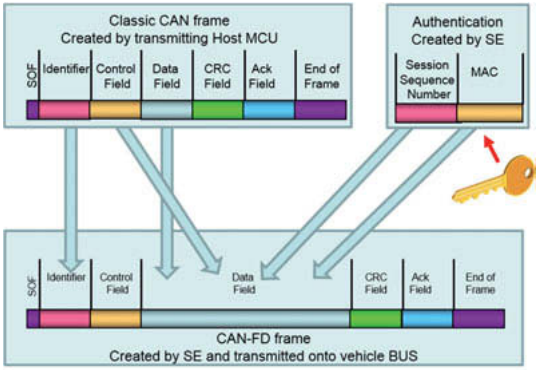


Fig. 3: Migration from CAN 2.0 to CAN FD

Now let's look on an existing CAN 2.0 network. In many cases, not all nodes need to be enhanced by security. The security architecture of the OEM usually defines which nodes and messages are critical regarding security and require authentication & secure boot. Fig. 4 shows an example with 5 nodes, where only nodes A, B and E need to be changed. These 3 nodes can be enhanced by a dedicated secure element handling the CAN-FD framing and authentication/secure boot/secure key storage. The other nodes remain unchanged (apart from the need to be CAN-FD capable). The architecture still allows communication (secured and non-

secured) between any two nodes on this bus. The added secure element can be conFig.d, which communication need to be authenticated. Legacy MCUs continue exchanging messages using CAN 2.0.

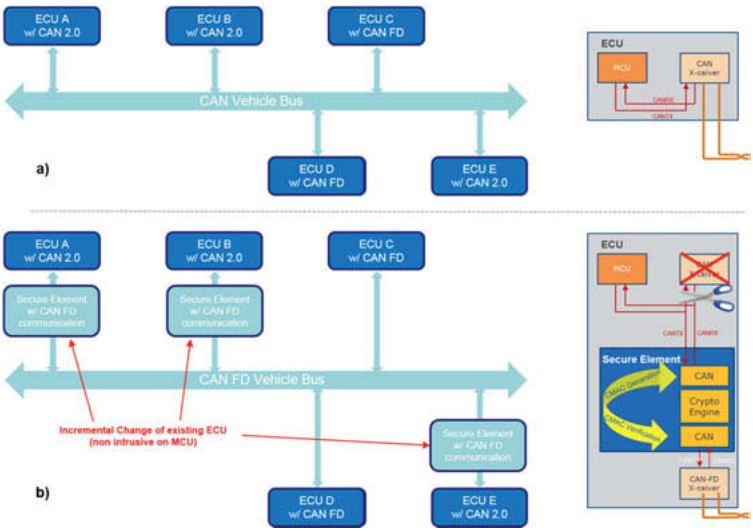


Fig. 4: Incremental enhancement of an existing CAN network

The architecture can be also intermixed with incremental change of an ECU on the one side, and a re-architected ECU with new MCU, new SW and native CAN-FD on the other side – across multiple architectures, and vendors. The user has the freedom to choose the best possible implementation. For interoperability reasons, it is important that any ECU connected to the same CAN bus must support the exact same format for the CAN frame with authentication information. Recently, Automotive Open System Architecture (AUTOSAR) defined a common frame format, which needs to be supported. But some OEMs have already started implementing their own format for authenticated CAN-FD messages, therefore the CAN-FD frame format is highly configurable regarding used cipher for the authentication, and lengths and position of the MAC and sequence number field, to cope with existing solutions. Fig. 5 shows an example frame format, but the various data fields can be conFig.d as described above.

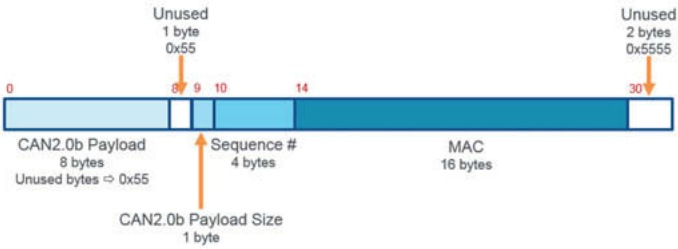


Fig. 5: Configurable CAN-FD frame format

In summary, using the described approach provides a series of advantages:

- **Incremental**: Not all ECUs need to be changed, not all ECUs need to be changed at the same time.
- **Minimally intrusive**: ECUs, which do not require a functional update, can be reused. Besides HW design effort, this also saves significant time and cost for SW retesting and re-qualification. This helps **preserving legacy ECUs**.
- **Interoperable**: Can be combined with any ECU – (i) secure element w/ CAN-FD, (ii) MCU w/ CAN-FD and companion secure element or (iii) MCU with integrated secure element functionality.
- **Compatible**: Can continue with legacy CAN 2.0 communication, where needed.
- **Real message authentication plus secure boot and secure key storage**: Unless other solutions which only use local CAN message filtering, this approach allows a real message authentication combined with secure boot of the ECU and secure key storage. Thereby allowing the maximum possible security implementation for each individual node.
- **Scalable and flexible**: Any ECU can be upgraded with the best possible method.

5. Scalability, Provisioning, Ecosystem, HW Solutions

What we have seen before is an easy way to upgrade an existing system (a legacy system) and enable it for CAN message authentication – without rearchitecting the whole system, or without the need to introduce a more powerful and by that more expensive MCU. But the same concept and principles apply to all in vehicle communication. Authentication and secure boot can be implemented on any communication and for any MCU, using the secure element as a companion device for the math of the algorithm and as secure key storage (Fig. 6).

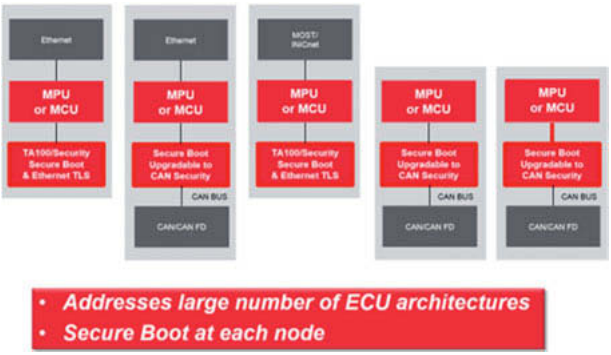


Fig. 6: Scalable security solution for in-vehicle network

When implementing a car-wide security approach, interoperability is a mandatory requirement. The same certificates, or keys and crypto algorithms need to be used in every node. Furthermore, the car maker requires the possibility to define the security architecture for the overall car. He needs a possibility for provisioning the car with his own secrets, when different Tier-1s and different manufacturing sites are involved.

The OEM defines the security architecture, an ability to cryptographically prove the trusted manufactured source of a node via certificate chains, and can allow the Tier-1 and their manufacturing sites to program secrets into the ECU without the Tier-1 knowing the secret information (Fig. 7). A second option is to have the secrets programmed at the semiconductor site inside of a vault – no secret information is exposed to the outside world.

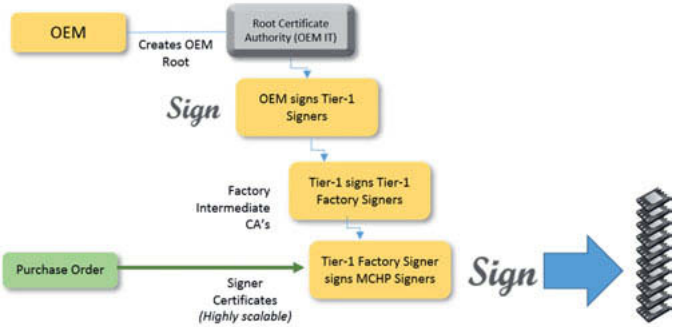


Fig. 7: Root of trust & provisioning

When the ECUs are deployed to the vehicle and the systems are powering up in a provisioning stage, the ECUs need to have a capability to securely exchange keys. Or in the case of asymmetric cryptography they need to be able to support standard based key agreement schemes to create local public/private key pairs, where the private key never leaves the area of the secure key storage.

Microchip Technology's CryptoAutomotive™ Solution: To facilitate this, Microchip Technology offers a companion crypto device family that support the three main pillars mentioned in chapter 3: secret key storage, secure boot & secure FW update, and communication with message authentication and integrity (Fig. 8). The devices provide the required flexibility for customizing secure boot to maintain performance requirements, enable an incremental upgrade of the system and preserve legacy ECUs with its minimally intrusive approach. Provisioning can be supported both, at factory or at OEM, as well as support for required key agreement and key exchange schemes during functional operation.

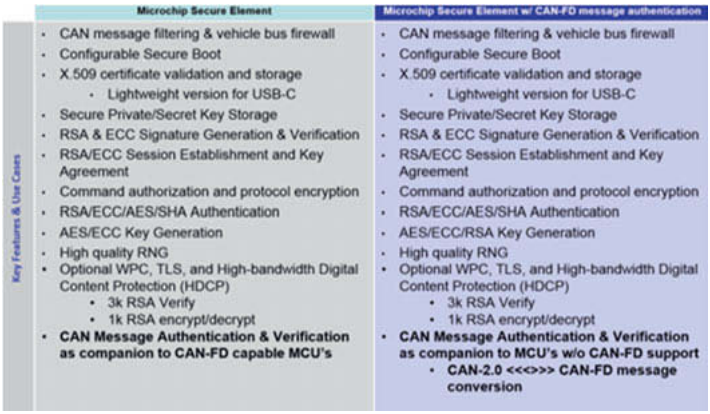


Fig. 8: Microchip secure element with CAN-FD message authentication

6. Summary

The presented approach allows OEMs and Tier-1s to enhance their in-vehicle networking ECUs with authentication, secure boot and FW upgrade, and hardware secure storage capabilities. With an incremental and minimally intrusive approach, legacy ECUs can be preserved

when needed – saving cost and reducing requalification effort. The approach scales across the vehicle's internal communication networks.

Security is added through a secure element adjacent to an existing microcontroller essentially adding an HSM to the microcontroller without redesigning it. With hardware support for the variety of security algorithms and secure key storage, the system is protected against physical and side channel attacks and supports the following aspects:

- Authentic message exchange between nodes with integrity
- Secure key storage protected against physical attacks – secrets must remain secret
- Secure boot and secure firmware upgrade possibilities
- Key management and provisioning infrastructure across different Tier-1 suppliers.

Interoperability is achieved by certificate-based key distribution and provisioning to allow OEMs a multi-tier1 architecture. For small nodes, authentication and secure key storage can be added without re-architecting the ECU.

References

[1] Todd Slack & Mike Jones: Designing for Cost Effective Ethernet Automotive E/E Architecture Against Security Threats, 2017 IEEE-SA Ethernet & IP @ Automotive Technology Day, Munich

[2] Alex Hahn: Enable In-vehicle Network Security with CAN Message Authentication, ConCar Expo 2018, Berlin

Hardware matters: how one chip can impact the security of a connected vehicle

Martin Brunner, Hans Adlkofer,
Infineon Technologies AG, Neubiberg

Kurzfassung

Vernetzte Fahrzeuge beinhalten eine Vielzahl fortschrittlicher Technologien - bedingt durch neue Use Cases für Fahrkomfort, Safety und Mobilitätsdienste als Innovationstreiber, welche wiederum auf der Sicherheit der zugrundeliegenden Funktionen basieren. Durch die Vernetzung mit der Außenwelt und als eingebettetes System sind vernetzte Fahrzeuge dabei sowohl denselben Bedrohungen ausgesetzt wie IT-Systeme als auch Angriffen auf Hardware (HW) Ebene. Dies erfordert den Einsatz von HW-basierter Sicherheit, um sich gegen beide Arten von Angriffen zu schützen. In diesem Vortrag wird die Bedeutung von HW-basierter Sicherheit im Spannungsfeld von automatisiertem Fahren, Vernetzung und den fortschreitenden Anforderungen an die Fahrzeugsicherheit diskutiert. Insbesondere wird aufgezeigt wie ein lokaler Angriff auf ein einzelnes Fahrzeug einen remote Angriff auf viele Fahrzeuge ermöglichen kann. Ausgehend von einer ganzheitlichen Betrachtungsweise werden die Säulen der Cybersicherheit im Fahrzeug aufgezeigt und entsprechende Sicherheitsgrundsätze abgeleitet. Dazu wird der aktuelle Stand der Technik zu HW-Sicherheit im Fahrzeug dargestellt und eine Unterscheidung und Abgrenzung der am Markt verfügbaren Lösungen vorgenommen. Basierend auf konkreten Anwendungsfällen im Umfeld von TCU und Gateway wird gezeigt wie HW-basierte Sicherheit vertrauenswürdige Software in vernetzten Fahrzeugen ermöglicht. HW-Sicherheitszertifizierungen stellen dabei einen Mehrwert dar, indem sie zusammen mit Standards die Integration vereinfachen und somit zu nachhaltigen und zukunftsfähigen Sicherheitsarchitekturen beitragen.

Abstract

Connected vehicles incorporate a wide range of sophisticated technology driven by new use cases enhancing driver comfort, road safety and mobility services, which depend on the security of the underlying functions. Due to the interconnection with the outside world and as an embedded system, a connected vehicle is exposed to both, malicious software activities as faced by traditional IT world systems as well as physical attacks on hardware level. This intro-

duces the need for utilizing hardware-assisted security measures to prevent both kinds of attacks. In this talk, we discuss the role of hardware-assisted security within the relationship between increasing vehicle automation, connectivity and the evolving requirements impacting vehicle security. In particular, we outline how a single, physical attack can lead to large-scale remote attacks. Based on the importance of a holistic approach we depict the pillars of automotive security and derive key security principles and respective mitigation measures. To this end, we sketch the current state of the art in automotive security hardware and compare the solutions currently available in the market. Based on concrete use case examples for TCU and gateway we propose how to leverage hardware assisted security to enable trusted software in connected vehicles on the road towards trusted mobility. We argue that there is added value introduced by security certified hardware and standards simplifying the integration and facilitating sustainable security architectures in current and future vehicles.

1. Vehicle manufacturers on the path toward mobility providers

The automotive industry is currently undergoing the biggest change in its history. Driven by three so-called “megatrends” the traditional business model is upon radical change:

(i) CO2-neutral mobility requires alternate drive systems. This includes the shift from combustion engines powered by fossil fuels towards emission-free driving utilizing alternative drive technologies, in particular electric driving.

(ii) Autonomous driving will change the vehicle design from scratch from both, the technical- and the design-perspective thereby enabling new use cases to the driver and also opening up new business models.

(iii) New mobility concepts will drastically change the role and usage of vehicles, which become shared assets being seemingly integrated into people’s daily life and thus need to be constantly interconnected and “always on”.

In particular OEMs face the edge of a new era as they are on the verge to transform their traditional role from pure car manufacturers to interconnected mobility providers:
The growing symbiosis between the connected vehicle and its surrounding environment offers hereby plenty of opportunities for cross-vehicle improvements. Outstanding examples of these new technologies include V2X communication and advanced driving assistance systems forming the foundation for automated – or even autonomous – cars. These technologies are expected to help preventing a large amount of traffic fatalities thereby significantly contributing to enhance road safety. Incorporating environmental data can help to improve driving strategies and for example reduce fuel consumption. Furthermore, cross-vehicle communication and the

integration of the vehicle with its environment enable new services for OEMs. For example, wireless access to the vehicle via Internet enables remote services such as diagnostics and software over the air updates helping to reduce costly recalls. It also allows to create new business models like on-demand activation of certain features. To leverage these opportunities, OEMs need to establish trusted mobility platforms encompassing not only the vehicle itself but also its corresponding ecosystem and services. This is not limited to manufacturing and selling vehicles, but requires acting as a mobility provider.

2. Advanced security is a prerequisite for trusted mobility

However, this digital transformation on the edge to a new era of mobility has also a dark side: cyberattacks against vehicles and their surrounding ecosystems respectively are on the rise.
Beside substantial changes in business models and value chain the megatrends named above introduce an increasing demand for digitalization and interconnection. This in turn requires an increasing amount of data to be processed and higher bandwidth for the transfer of data. Furthermore, the corresponding automotive architectures need more flexibility. This impacts security, since one the one hand the interconnection introduces new communication paths between automotive systems and arbitrary instances outside of the vehicle, e.g. on the internet, which are potentially unknown or untrusted. Thus, the previously closed systems become part of an open, interconnected system facing the internet thereby being exposed to the same attacks as traditional IT world systems. As experiences from IT and other domains demonstrate, whatever is connected (even though it is only accessible by bridging an air gap) will be attacked sooner or later [1 - 5]. One the other hand flexibility requires software. More and more innovation is realized via software resulting in a very large, still increasing code base. Modern cars are reported to have already reached as much as 100 million lines of code per vehicle which significantly exceeds other complex systems, such as modern operating systems or large internet services [6 – 7]. Beside many other challenges, such as a complex value chain or long life-cycles, this introduces a dramatic increase of complexity. While complexity is considered the enemy of security", complexity in software particularly becomes a security issue since it results in a higher error-proneness thus increasing the number of potential vulnerabilities in the software. In addition, there is a strong asymmetry in the arms race between attack and defense. For example, finding and fixing software bugs may become a challenging, time-consuming task. While developers usually have strict time constraints, an adversary is able to spend plenty of time analyzing the code. Moreover, finding a single (severe) bug may be sufficient. Furthermore, both – attacks and attackers – typically get better over time. Hence, a sustainable protection against cyberattacks resulting from software errors becomes a very

challenging task in this context. In combination with the accessibility of the vehicle from outside and the broad attack surface this poses a significant risk of attacks. As a result, an increasing demand for advanced security can be derived from the megatrends named above.

3. The need for a holistic security approach

Connected and autonomous vehicles (CAV) are based on a multitude of different aspects impacting their design and use, such as a large and complex software base, connectivity (in particular car-to-cloud) or software-over-the-air updates (SOTA). This opens up a broad attack surface, which is difficult to protect facing a constantly changing threat landscape and a state of the art that is evolving quickly. Thus, a holistic security approach becomes paramount to safeguard CAVs in a sustainable manner. To this end, an established best-practice from the IT security domain, the so called “defense-in depth” approach, is required to absorb the risks. This includes to mount multiple, independent layers of defense. As information technology gets integrated into vehicles, cars become prone to attacks known from the IT domain which may also affect the underlying automotive components and their corresponding functions. That is, a successful (cyber-) attack may impact certain critical components, such as brakes or airbag, as well thereby leading into a violation of safety assets. Thus, safety- and security-issues are intrinsically linked and cannot be handled isolated or independent from each other. In addition, there are further facets with overlap to security, such as privacy. As a result, the consideration of the three cornerstones of information security (confidentiality, integrity, and availability – “CIA”) alone is not sufficient in the context of a CAV, since it needs to be perceived as a “cyber physical system” requiring additional protection goals, such as authenticity, robustness, operability and controllability. Instead, a holistic view encompasses four overall automotive security goals [8]:

- Interlock with functional safety
- Protect business cases and intellectual property (IP)
- Protect operations i.e., meet customer’s quality and reliability expectations
- Ensure compliance to privacy and regulatory requirements

These goals are to be applied to the pillars of automotive security as depicted in Figure 1.

1. The protection on ECU-level includes secured data processing and data at rest.
2. Secured vehicle networking forms the foundation for secured communication.
3. This is supported by a corresponding E/E architecture facilitating the central gateway protection and domain separation to isolate untrusted code from trusted software.

4. The protection of the connected car as a whole involves securing the externally accessible interfaces.
5. Sustaining functions, such as a security organization along with corresponding processes maintain all security aspects along the entire life-cycle as a cross-cutting issue.

Following the defense-in-depth approach, these pillars depend on each other. That is, secured ECUs (electronic control units) form the foundation for the in-vehicle network. Secured interconnection, in return, requires trusted devices, which then provide the basis for secured communication. Thus, it is important that all parts along the different pillars are properly interlinked – from external interfaces representing the car as a whole down to security primitives on the chip-level.

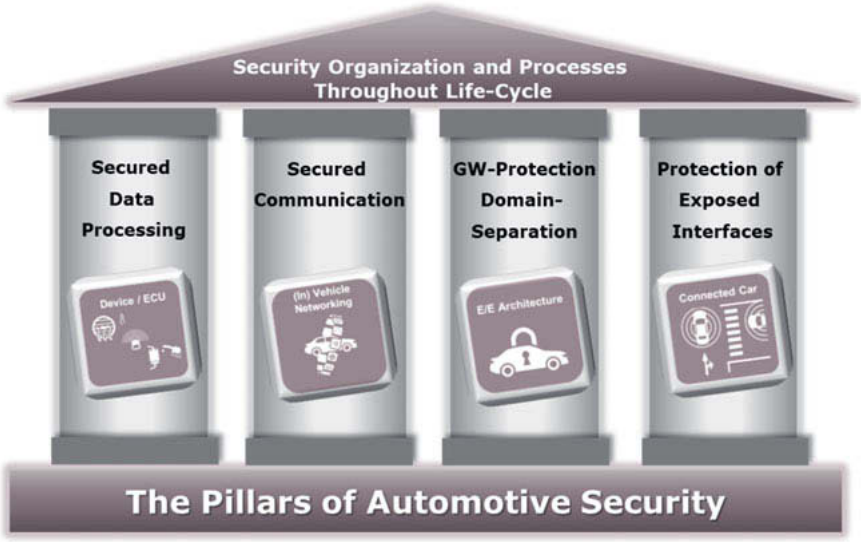


Fig. 1: The pillars of automotive security

4. Hardware-assisted security on the rise

Although there is more and more software, hardware plays an increasingly important role within the relationship between advancing vehicle automation, connectivity and the evolving requirements impacting vehicle security. Given the size and complexity of the software as depicted in Section 2, it becomes infeasible to maintain the security level of the entire code base properly. Moreover, vehicles must be considered to be embedded systems, since there is no restriction

to gain physical access to a car, e.g., via purchase or rental. Thus, it must be assumed that a potential adversary may have full physical access for an unrestricted amount of time. These frame conditions imply, that automotive systems must be resilient to a wide range of attacks that cannot be thwarted by software-only security measures. As a result, protecting software and data with software-based measures (only) is not sufficient anymore. This introduces the need for hardware-assisted security measures. If implemented properly, secured hardware enables trusted software. This becomes particularly important when critical assets in the vehicle must be protected over a long period of time, such as IP or cryptographic keys. The security of vehicle functions often relies on cryptographic keys as a basis. Thus, it is essential to protect the integrity and confidentiality of these keys, because once a key is compromised the underlying security concept may be undermined. Readout and cloning of keys leaves no traces making it impossible to distinguish between genuine and stolen keys later on. Furthermore key handling must be secured through the whole lifecycle including manufacturing and garage updates. Thus, it is essential to implement strong measures for protecting such keys in a vehicle against physical attacks.

4.1. Attacks

Standard semiconductor hardware devices can be attacked on different levels while a variety of different attacks are known [9 – 11]. Beside manipulative attacks on a logical level there are monitoring, semi-invasive and manipulative attacks, which can be either active or passive. They usually require preparation, such as isolation of the chip and opening the package (e.g., via etching).

- Logical attacks target mostly software vulnerabilities and include e.g., attacks on the JTAG port, network- or ECU level or fuzzing.
- Monitoring attacks include for example the analysis of power profiles and probing of I/O signals in order to gain access to sensitive data.
- Semi-invasive attacks are utilized to trigger faults in order to change the program flow, gain access to data or bypass software protection measures. This can be achieved via various means, such as glitches, spikes or laser fault injection.
- Manipulative attacks include probing on the silicon-die, forcing or reverse engineering and refer to access or influence certain circuit paths on the chip.

Hence, the use of so-called “tamper-resistant” chips is essential for protection against physical attacks when storing and processing sensitive data, such as long-term cryptographic keys.

Tamper resistant devices restrict physical access, such that information stored therein is not accessible via external means and any access and interaction has to be done through the Firmware embedded on the chip, which implements appropriate security measures. Tamper-resistant chips can be configured to zeroize their sensitive data (particularly cryptographic keys) once an attack attempt or parameter out of the environmental specification has been detected. Such chips can be leveraged to implement the concept of a so-called hardware-based trust anchor. It provides a tamper-resistant hardware and a protected runtime environment to achieve a high security level. From this “root of trust”, which stores the most valuable and sensitive keys, further keys or security measures can then be derived to realize secured vehicle functions. In particular, a hardware trust anchor can realize secured key storage, related cryptographic operations as well as key management and deployment in insecure environments. With respect to the previously outlined complexity issue, the approach of a hardware trust anchor encapsulates the critical, security-related part and separates it from the rest of the system resulting in a small, maintainable computing base. Thereby it enables the underlying hardware to form the basis for secured vehicle functions.

4.2. Discrete security controller

A common implementation of a tamper-resistant hardware trust anchor are discrete security controllers i.e., it resembles a microcontroller with special security peripherals for security-critical applications, such as payment, access control or passports. Such discrete security controllers contain hardware blocks which provide resistance against hardware attacks. Depending on the type of security controller this hardware may have varying levels of sophistication. In contrast to integrated solutions discrete security controllers are companion chips, which are attached to a host processor (e.g., application processor or microcontroller) via a dedicated interface, such as SPI or I2C. They encompass various protection measures against physical attacks including but not limited to memory encryption, shielding, sensors (e.g., voltage, frequency, temperature, light), code- and data-signatures or error correction codes. For example, common monitoring attacks encompass the observation of certain physical characteristics, such as temperature, time or power consumption. While certain sensitive assets (e.g., data, cryptographic keys, used algorithms) can be derived from standard semiconductor hardware leveraging this information, discrete security controllers implement protection measures against such types of attacks in hardware. Furthermore, security controllers are usually evaluated and certified by a third party according to the international standard Common Criteria (CC) [12]. CC certification provides additional assurance that the solution is secured and that this statement has been extensively tested and certified by a third party. Thereby a certain

protection profile (PP), which has been created by industry members and independent security experts, with a defined coverage is used as a basis. It includes requirements regarding functionality and trust for the respective product type which define the level of detail for the security evaluation. PP are not only available for security controllers but also for software and system solutions. PP are updated every few years to reflect the state of the art. Based on a given PP the corresponding target of evaluation (TOE) and security target (ST) are defined and used for the evaluation. This analysis has to be conducted by a certified lab, while the certification is issued by an approved certification body. Common Criteria distinguishes between seven evaluation assurance levels (EAL). The EAL number refers on the one hand to the level of detail of information exposed to the tester. On the other hand it indicates the testing level and methodology (reaching up to formal verification). Both the report and the list of certified products is publicly accessible enabling transparency and comparability given, certified products. An important aspect is that secured hardware has to be complemented by appropriate software in order to leverage its security capabilities properly. This is achieved via “secure coding”. That is, software which is added to the code if the software running on a given discrete security controller. The amount of secure coding depends on the capabilities provided by the hardware.

5. How a local, physical attack can lead to large-scale, remote attacks

While the security and attack resistance of each individual vehicle remains an important issue, the main concern are large-scale attacks i.e., remote attacks affecting multiple vehicles (in the worst case an entire fleet) at the same time because the effort – effect relation is maximized for an adversary. That is, he gains maximized effect (maximized damage) while investing minimal effort. Large-scale attacks may impact further assets, such as OEM reputation, undermine trust in the infrastructure or technology as a whole and may even cause real-world impact, e.g., costly recalls or a decrease in stock price. While the security of automotive systems has attracted the security community and thus has widely been studied (research dates back several years [13, 14]) such a real-world impact had been first publicly demonstrated in 2015 [15]. To this end, the current focus is on the respective mitigation measures, such as protecting outside interfaces or the security of software updates to prevent attacks on a large scale. However, several attacks which gained public attention in past years shared the fact that they were all primarily software-based and they outlined that an initially local attack may lead to remote attacks [8]. That is, although physical access is needed to perform the initial attack, it may be reproduced remotely or remote access may be established later on. To outline how a local, physical attack can lead to large-scale, remote attacks a simple example is depicted in Fig. 2.

In this scenario the underlying assumption is, that software is stored on standard semiconductor hardware (i.e., without tamper-resistance). First, the adversary gains physical access to the vehicle. For the car as an embedded system, the attack model must consider that a potential adversary may have full physical access, since there are no access restrictions (i.e., a car can be freely bought or rent). After dismounting and detaching the target device, the adversary successfully performs physical attacks and extracts the software from the chip. At this point hardware and software is decoupled i.e., from there on it is a software-only attack. This means, that the adversary can operate without any hardware restrictions and utilize a corresponding set of tools to execute or emulate the runtime environment, analyze the software image, etc. Due to the complexity (thus error-proneness) of the software and the presumably high amount of time he can spend for the attack, he identifies one (or multiple) vulnerabilities which are remotely exploitable. Given that all cars running the same configuration and software version respectively are prone to this vulnerability, the adversary is able to launch remote attacks on a large scale.

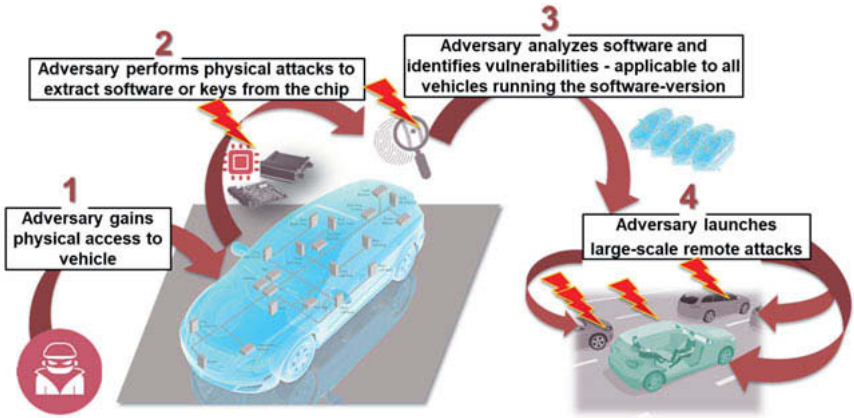


Fig. 2: Simplified example – how a local, physical attack can lead to large-scale, remote attacks

6. Current state of the art in automotive security hardware

‘Hardware security’ is a widespread term, which often refers to embedded devices and microchips that are designed with an emphasis on security in order to physically encapsulate security-relevant functions. Typical use cases include the secured storage of data, the protection of software and cryptographic keys as well as the secured processing inside a certain domain. As depicted above, secured hardware has to be complemented by software, such that today’s

security architectures do not rely on hardware only, but on a synergy between both, hardware and software. While hardware-based security features for automotive security are currently being discussed [16] a common definition for ‘hardware security’ is lacking. Although the same fundamental hardware-based security primitives are used in many cases, there is a variety of terms attributed to ‘hardware security’ encompassing different features sets, such as security controllers / processors / cores / hardware extensions, trusted platform modules, trusted execution environments, integrated secured elements and secured enclaves. One a very high level three basic approaches can be distinguished apart from software-only features.

(i) Security hardware extensions: separate a “secured” from a “normal” domain still sharing the same CPU
(ii) Embedded hardware security modules (HSM): provide a separated, programmable domain which is integrated inside the package of a microcontroller.
(iii) Discrete security controller (c.f. 4.2): are dedicated companion chips, which resemble a microcontroller containing protection measures to provide resistance against physical attacks allowing for tamper resistance and security certification.

In the automotive domain most prominently SHE (Secure Hardware Extension) [17], EVITA HSM (Hardware Security Module) [18] and – more recently – TPM (Trusted Platform Module) [19] have emerged over the past years representing each one of the three.
The feature set for SHE has been specified by HIS (Hersteller Initiative Software), a group of German car vendors, in 2009. The SHE specification defines a set of functions that allows a secured zone to co-exist within an ECU in the vehicle to minimize costs. It was designed for use cases like key storage, secured boot, immobilizing, anti-theft and tuning-protection. Its implementation was done as a fixed state machine inside automotive microcontrollers and encompasses memory for keys, a hardware-accelerator for symmetric cryptography and a true random number generator (TRNG). An example for a microcontroller implementing the SHE specification is the Infineon AUDO MAX SHE and AURIX™ HSM family.
The EVITA research project specified automotive HSM comprising three variants

(i) EVITA Light HSM to protect communication between sensors, actuators and ECUs
(ii) EVITA Medium HSM to protect ECUs and in-vehicle networking
(iii) EVITA Full HSM to protect external communication

A HSM thereby supports one of the three levels of EVITA security classification depending on its performance and feature set while EVITA considered active and passive as well as physical attacks. HSMs are in this context separated, programmable domains inside a microcontroller.

They are separated from the microcontroller via the concept of a bi-directional hardware interface named ‘firewall’. Since this interface is linked to the busses of the host, the HSM can achieve high-performance and actively trigger security functions. Typical use cases involve secured boot, secured onboard communication or intrusion detection. HSMs have their own CPU, RAM, ROM and peripherals and offer security hardware, such as high-performance hardware accelerators for performing symmetric cryptographic operations, internal memory for storing cryptographic keys and critical functions or a TRNG. EVITA light complies thereby with the feature set of the SHE specification, apart from a few exceptions, such as key management. While EVITA Medium runs asymmetric cryptographic operations and hash functions in software, EVITA Full specifies hardware support. EVITA Full support is for example implemented in the 2nd generation of AURIX™ HSM.

The TPM specification is driven by the Trusted Computing Group (TCG), an organization enabling trusted computing with open standards and specifications. A TPM hereby refers to a tamper resistant, discrete security controller that is security certified according to Common Criteria and supplied with firmware compliant with the TCG specifications. The elements of a TPM are defined in an international standard (ISO/IEC 11889). This firmware enables the immediate use of various security features, such as authentication, encryption, signing and verification. An initial key is injected into the device during production within a certified environment and can serve as cryptographic identity of the chip. The TPM can generate, store and administer further cryptographic keys, which are particularly protected against logical and physical attacks, and also be utilized to detect faulty or manipulated software. The firmware can also be updated so that the level of security can be kept up to date throughout the vehicle’s lifecycle (e.g., by enabling crypto-agility). The TPM can be integrated using an open source software stack (TSS stack) for the host processor. Besides a discrete TPM further implementations are possible (e.g., as part of another integrated device or implemented purely in software), enabling a scalable application across multiple ECUs where the use of a discrete security controller may not be feasible. An example for a TPM implementation of an automotive qualified discrete security controller is the OPTIGA™ TPM SLI 9670.

7. Use case examples leveraging hardware-assisted security

For many security-relevant use cases, such as car sharing, V2X communication, EV-charging, black-box, car access, on-demand feature activation or centralized key management, multiple ECUs, which may be distributed across the vehicle, are involved to realize the corresponding function. In order to leverage hardware assisted security enabling trusted software for these functions a security concept comprising different ECUs – which implement different types of

hardware security – is required. Given the scenario in Chapter 5 the ability to sense and react to security incidents in the field becomes paramount. This includes, but is not limited to SOTA where the establishment of a trustworthy connection between the car and the backend is essential. To this end, the cryptographic identity of the vehicle needs to be specifically protected against a variety of threats, including physical attacks. For this the use of a discrete security controller supplied with security certified software as a hardware trust anchor is obvious, since such a system can be attributed a higher trust level than software running on other systems. For performance and storage reasons several operations need to be performed on a central host with the corresponding capabilities. As a result, two entities need to be involved in the sketched example:

(i) A discrete security controller (e.g. OPTIGA™ TPM) residing on a device exposed to outside interfaces, such as a telematics control unit (TCU)

(ii) A central instance (e.g., gateway) with EVITA HSM, such as AURIX™ HSM

In order to realize an implementation for secured backend communication using SSL/TLS and TPM 2.0 utilizing ECC (assuming ECDHE and AES) and the available TSS stack the work split could be realized as follows. The corresponding functions for ECDHE and AES can be used directly by the gateway in order to negotiate the pre-master secret with the server and calculate the session key (for AES) based on the corresponding primitives. For the sake of efficiency the handshake and the SSL/TLS protocol is processed in the gateway to the extent it is necessary. The asymmetric ECDH operation however is processed within the TPM, which holds the long-term asymmetric keys. The subsequent result is used by the gateway to calculate the symmetric shared secret and the symmetric encryption using AES is done in the gateway. Thereby the advantage of both approaches – security of the hardware trust anchor and efficiency of the HSM – can be leveraged.

8. References

[1] Falliere, N., Murchu, L. O., & Chien, E., W32. stuxnet dossier. White paper, Symantec Corp., Security Response, 5(6), 2011.

[2] Li, C., Raghunathan, A., & Jha, N. K., Hijacking an insulin pump: Security attacks and defenses for a diabetes therapy system. In Healthcom, 2011 13th IEEE International Conference on (pp. 150-156). IEEE, June 2011.

[3] Luo, A.: Drones Hijacking - multi-dimensional attack vectors and countermeasures, DEFCON 24, 2016

[4] Felt, A. P., Finifter, M., Chin, E., Hanna, S., & Wagner, D. A survey of mobile malware in the wild. In Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices (pp. 3-14). ACM, October 2011.

[5] Gollmann, D., Gurikov, P., Isakov, A., Krotofil, M., Larsen, J., & Winnicki, A., Cyber-physical systems security: Experimental analysis of a vinyl acetate monomer plant. 1st ACM Workshop on Cyber-Physical System Security (pp. 1-12). ACM, April 2015.

[6] Million Lines of Code: http://www.informationisbeautiful.net/visualizations/million-lines-of-code, accessed Aug. 2019.

[7] Charette, R.N.: This Car Runs on Code, IEEE Spectrum, Feb. 2009, https://spectrum.ieee.org/transportation/systems/this-car-runs-on-code, accessed Aug. 2019.

[8] Corbett, C., Brunner, M., Schmidt, K., Schneider, R. et al., "Leveraging Hardware Security to Secure Connected Vehicles," SAE Technical Paper 2018-01-0012, 2018, https://doi.org/10.4271/2018-01-0012.

[9] P. Laackmann and M. Janke, “Power and Timing Analysis Attacks against Security Controllers,” *Technology Update, Smart Cards*. Infineon Technologies AG.

[10] Grand, J., & Friday, J. Advanced hardware hacking techniques. *DEFCON 12*, 2004

[11] P. Laackmann and M. Janke, “Attack Methodologies on Security Chips,” in *hardware.io*, 2015.

[12] ISO/IEC 15408 Information technology - Security techniques -Evaluation criteria for IT security

[13] Checkoway, S., McCoy, D., Kantor, B., Anderson, D., Shacham, H., Savage, S., Koscher, K., Czeskis, A., Roesner, F., Kohno, T., Comprehensive Experimental Analyses of Automotive Attack Surfaces. In USENIX Security Symposium, August, 2011.

[14] Miller, C., & Valasek, C., A Survey of Remote Automotive Attack Surfaces. http://illmatics.com/remote%20attack%20surfaces.pdf, DEFCON 22, 2014

[15] Miller, C., & Valasek, C., Remote exploitation of an unaltered passenger vehicle. http://illmatics.com/Remote%20Car%20Hacking.pdf, Black Hat USA, 2015

[16] SAE J3101: Requirements for Hardware-Protected Security for Ground Vehicle Applications, WiP, https://www.sae.org/standards/content/j3101/ accessed Aug. 2019.

[17] Escherich, R., Ledendecker, I., Schmal, C., Kuhls, B. et al., SHE –Secure Hardware Extension – Functional Specification, V. 1.1, Hersteller Initiative Software (HIS) AK Security, Oct. 16, 2009.

[18] EVITA Deliverable D3.2: Secure On-board Architecture Specification, Vers. 1.3. 2011.

[19] Arthur, W., Challener, D. and Goldmann, K. “A Practical Guide to TPM 2.0: Using the Trusted Platform Module in the New Age of Security”, Apress, 1st Ed., 2015.

Embedded Intrusion Detection based on AI

A Data-Driven Approach

Dr.-Ing. **Andreas Weichslgartner**, Audi Electronics Venture GmbH, Gaimersheim

Abstract

As modern cars integrate more and more software and become the center of our connected life, the attack surface for malicious attacks also increases. With looming, and already existing, security and privacy regulations, securing the car is a key challenge. To perform the appropriate counter-measurements, first, an attack has to be identified reliably by an intrusion detection system which monitors the communication and computation inside the car. In this paper, we will give insights of our automotive intrusion detection system. We show how Artificial Intelligence (AI) and a data-driven approach can enable an adaptive intrusion detection system which adheres to the future computing and communication landscape of a car. It enables to detect unknown attacks as the observed behavior deviates from normality. Further, we especially focus on the challenges performing AI-based intrusion detection on embedded hardware with constrained resources inside of a modern car.

1. Introduction

In 2014, Charlie Miller and Chris Valasek showed that it is possible to gain remotely access to the infotainment system of a car (a Jeep Cherokee) and from there gain control over the entire vehicle [6]. They were able to manipulate the volume of the radio as well as the brakes and engine. As modern cars evolve more and more to a distributed computing system, they face same security threats as conventional computing systems. This holds especially true with respect to the continuing trend towards increased connectivity and more and more functionality realized in software. A modern car architecture comprises various components with software written in memory unsafe languages (e.g. C/C++) and wireless interfaces such as WiFi, Bluetooth, or NFC.

As the car is a safety-critical system, special precautions need to be considered to ensure the security and, hence, the safety of the automobile. One of such a precaution is an Intrusion Detection System (IDS). As an automotive environment is stricter than enterprise networked systems, it is more meaningful to deploy an IDS. As Miller and Valasek stated in 2018 [7]: *"The*

Ethernet network should be analyzed for anomalies. While, in general, the problem of generic intrusion detection is difficult and often leads to false positives, in this case it works well. That is because this is a network devoid of human users like we are used to in an enterprise environment. All the traffic is periodically generated from machine to machine."

They also emphasize that even a simple IDS could prevent attacks on the CAN bus [7]: *"Like Ethernet, the CAN network traffic should be observed in real time to identify anomalies. All the attacks outlined in the historical section could have been detected (and prevented) with even the most trivial CAN network intrusion detection software."*

2. Intrusion Detection Systems

Since the first introduction of IDS in the 1980s by Dorothy Denning [3], a lot of different IDS approaches were proposed. Overall, the systems can be distinguished in host-based and network-based (see Fig. 1).

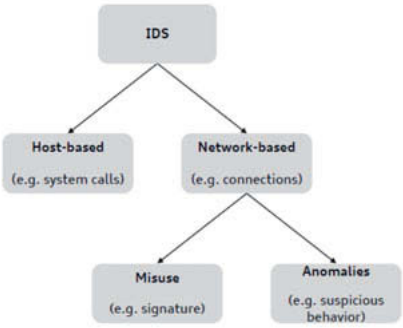


Fig. 1: Classification of IDS based on the monitored system (host/computing or network) and type of detection mechanism (misuse or anomaly detection).

While host-based systems monitor the system behavior, e.g., system calls, system utilization, processes, log files, the network-based IDS monitors the network traffic. Due resource limitations and third party intellectual properties, it may be infeasible to monitor each Electronic Control Unit (ECU) in the car while a network-based IDS (also often referred to as NIDS) can monitor several ECUs when placed in the central gateway. Nevertheless, both IDS version can also be combined, i.e., monitoring the network traffic and the execution behavior of specific ECUs with a high connectivity to the outside such as infotainment systems.

The IDS systems can be further classified into misuse and anomaly detection. Programs like Snort [2] or Bro [1] use various rules to detect known attacks. These rules are generated after an identified attack. Hence, the systems can only detect already known attacks. In contrast, anomaly detection aims to also classify unknown attacks. This is done by defining/learning a normal behavior and categorizing all behavior deviating from the norm as anomaly. Figure 1 summarizes the previously detailed IDS classification. Note here, that not every anomaly must be an attack. The cause of an anomaly can also root in rare events which were not trained or modelled as normal behavior. In addition, malfunctioning devices or new features may result in detected anomalies (see Fig. 2).

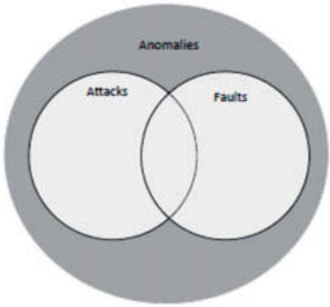


Fig. 2: Venn diagram showing the relationship between anomalies, attacks, and faults.

Further, it can be distinguished between outlier and anomaly. An outlier is a very rare instance of a data point seen in the recorded benign data while an anomaly is a data point which do not belong to the learned distribution (see Fig. 3).

3. IDS State-of-the-Art

Most enterprise network systems rely on firewall-based protection and anti-virus/malware software to protect their networks. The most common IDS which are used, are misuse systems like Bro [1] or Snort [2]. They profit that there are many known attacks for which rules can be provided. As a result these attacks are accurately identified with a low false positive rate.

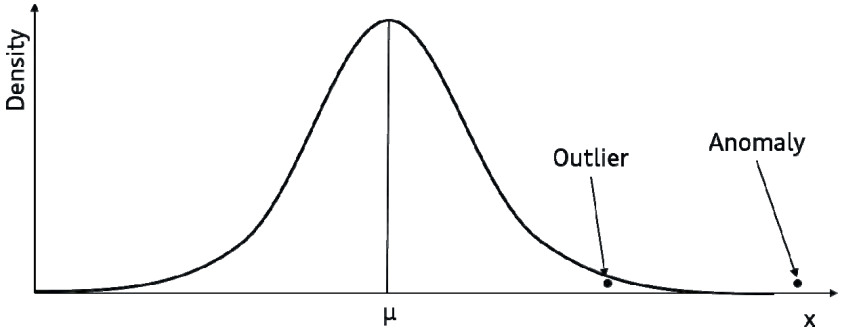


Fig. 3: Diagram visualizing the difference between an outlier and an anomaly: While an outlier has a very rare occurrence, it belongs to the distribution of normal behavior, but an anomaly is typically a data point lying outside the distribution.

Recently, Apache Spot [16] provides machine learning based intrusion detection, monitoring perimeter flows, DNS, and proxy services. While the methodology is interesting the observed targets such as DNS and proxy services do not apply for the automotive context.
In contrast to rich curated rulesets for Bro or Snort for enterprise networks, in the automotive domain, there is not a big corpus of known attacks for which rules can be created.
In academia, anomaly detection for NIDS gained a lot of attention. The range of used algorithm is manifold. However, most of the work suffer from the unavailability of good network data as training model. The KDDCup benchmark [17] gives recorded network traffic with various labelled attacks. But as the data is from 1998 and the attacks evolved since then this benchmark does not reflect the current threat scenario, especially in the automotive context.

4. Using Machine Learning for IDS

While machine learning algorithms, such as the artificial neural network called *perceptron* by Rosenblatt [9], or Support Vector Machines (SVMs) [10] were introduced in the 1960s and 1970s, only recent advancements in hardware (use of Graphics Processing Units (GPUs)) and software tooling, e.g., Tensorflow [15], contributed to the breakthrough and current hype around Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL). In general, AI describes the superset of all systems which reason in some way, where ML is a subset acting as a function approximator fitted with training data. Finally, DL is a specific subfield of ML dealing with artificial neural networks with many layers. See Fig. 4 for an overview of the relationship of the mentioned terms.

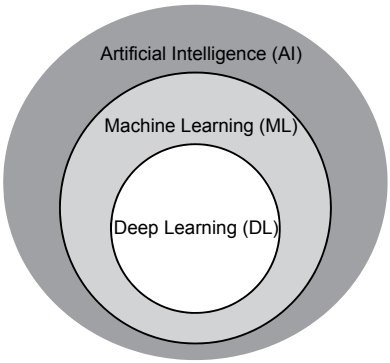


Fig. 4: Diagram showing the relationship between AI, ML, and DL. While AI is a term which includes all systems which reason and make decision, ML refers to algorithms which are trained with data. A subset here is Deep Learning where deep artificial neural networks are used.

In this work, we will focus on ML where a trained function indicates if there is an intrusion in the system or not. If data for known attacks would be available this could be done in a supervised manner, where a feature vector **x** would be labelled accordingly and a function f(**x**) would be trained to perform this decision during run time. However, as these labels are not available in the automotive case and an algorithm trained in a supervised manner can only be trained on known attacks and generalize to similar attacks. This problem is generally tackled with a semi-supervised learning technique known as anomaly detection. Here only one label is available during training (the good case). During run time, feature vectors which adhere to this learned model are classified as normal and feature vectors which do not adhere to it are detected as anomalies and need further inspection to determine if the root cause was an attack, a fault or a normal event unknown during training time (see Fig. 2).

Further, it can be distinguished between point anomalies and contextual anomalies. While point anomalies represent the similarity of feature vectors, the contextual anomalies can only be detected taking the context into account. For example, a feature value might lie in the normal range of values but given the temporal context it might be an anomaly. Therefore, the values before and after the current value have to be taken into account to determine whether an event is normal or not.

For detecting point anomalies, there exists various ML algorithms such as density-based approaches like Local Outlier Factor (LOF) [18] or k-Nearest Neighbours (k-NN) [19] which take

the local neighborhood into account. For example, in Fig. 5 the points (from a feature vector with 2 dimensions) in the grey circle are all close to each other, while the red point (the anomaly) lies far away from the rest of the points. Hence, the density in the grey area is much higher and the distance to the neighbor points is much lower than in the case of the red dot. However, to query the neighborhood points they all have to be saved in memory or a database.

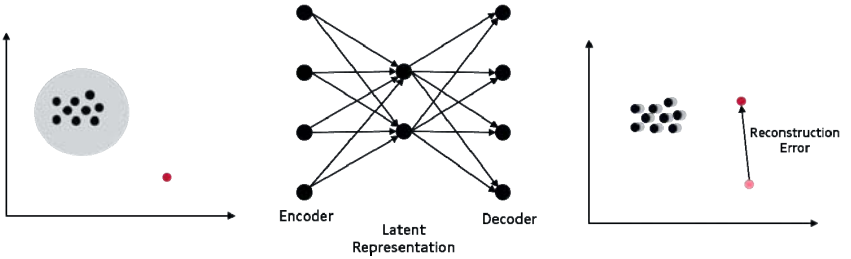


Fig. 5: Point anomaly and detection with an Autoencoder mechanism.

In case of embedded anomaly detection this is infeasible due high memory requirements. Other algorithms are One-Class Support Vector Machines (OCSVMs) [13], Isolation Forest [20] or Autoencoders [21]. As shown in Fig. 5, Autoencoders have the shape of an hour glass and are trained to minimize the reconstruction loss. In other words, an Autoencoder learns a compressed representation of the feature vector and then reconstructs a vector with the same dimensionality as the input vector. This algorithm assumes that there is a redundancy in the features and that the stochastic distribution, from which these vectors stem, can be represented whit less dimensions. If a feature vector is drawn from another distribution, e.g., originating from a malicious attack, the feature vector could only be reconstructed with a high error. Given a threshold for an acceptable reconstruction error, normal behavior can be classified.
Similarly, in the case of contextual anomalies, a regressor can be trained to predict the next time step of a signal from the last *n* preceding time steps. Again, the reconstruction error between prediction and actual value can be used to classify the normality of a value (see Fig. 6 right side). Prominent examples for time series forecasting are Recurrent Neural Networks (RNN) or AutoRegressive Integrated Moving Average (ARIMA).

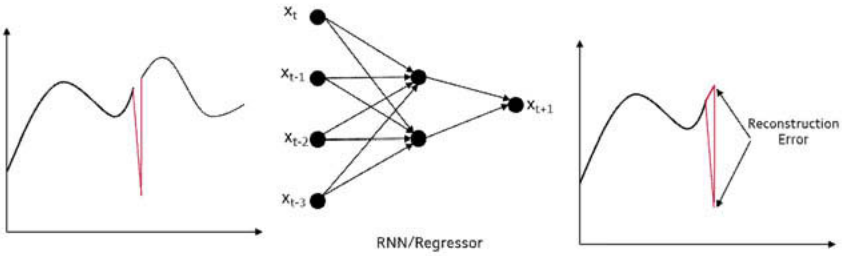


Fig. 6: Contextual anomaly with a regression network predictor.

4.1 Network-based Anomaly Detection

To detect anomalies in automotive network traffic, an anomaly detector has to monitor the network devices. As CAN uses only broadcast messages, this detector can be deployed on any ECU connected on the targeted CAN bus. There it extracts the feature vector from each message on the bus and infer it with its learned model. The features can be extracted from the CAN header, e.g. the ID, from the payload or from metadata such as inter-arrival times. Then, typically this feature vector is checked for point anomalies (see Fig. 7 for an example). Continuous signals from the payload could also be checked for contextual anomalies.

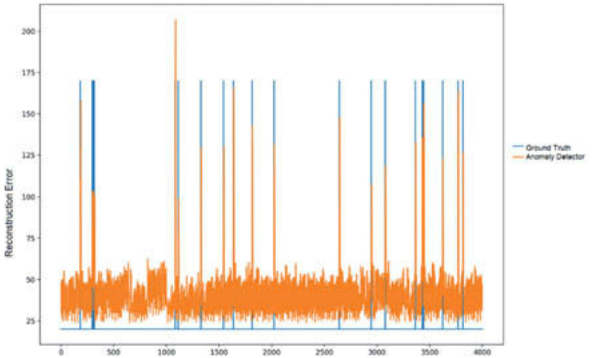


Fig. 7: Reconstruction error for CAN traffic (orange line). While the reconstruction error is normal from 25 to 30, it raises to 100 and more in case of spoofed packets (ground truth indicated by the blue line). Using a threshold of, e.g., 75 could be used to classify abnormal messages.

As Ethernet is a point-to-point connection, the monitoring has to be different in this mode. If the complete traffic should be monitored, then the NIDS must be deployed at central gateway/switch with active port mirroring to observe the whole communication in the specific network. Otherwise a NIDS can be deployed in start/end points of ECUs with connectivity to the outside of the car or on ECUs which are connected to safety-critical actuators.
Further, Ethernet packets consists of various layers of the ISO/OSI model and bigger packet sizes (1516 bytes for Ethernet packets vs. 8 bytes for standard CAN packets). This spans a much bigger feature space than in the CAN case, as each OSI layer can be dissected and flags and information can be extracted. In addition, Deep Packet Inspection (DPI) can be employed to analyze the payload in the highest layer. However, DPI might be infeasible at very high data rates or render useless for encrypted payloads.

4.2 Host-based Anomaly Detection

Besides the network, also the host systems aka the operating systems and processes on ECUs can be monitored by an IDS (see Fig. 1). Here, the contextual anomalies in executed processes, system calls, and log files are of special interest. If an attacker tries to get access a system or a malware is executed, there are usually traces visible in the system. However, as modern operating systems, even in the automotive domain, are getting more complex, it is very hard to define static rules to define the normal state. In contrast, in a data-driven approach, the system can be monitored for some time and this data can be used as a base to learn a model for normality.

4.3 Discussion

In the previous sections we detailed the various algorithms which can be used for anomaly detection. Overall, they all take a feature vector as input and output a single value (normal behavior or anomaly). The algorithms also all have in common, that their model of normality is created from many feature vectors collected during observation. Besides this similarities, the algorithms differ in training time, memory and CPU utilization during run time. In contrast to ML algorithms for image processing which is often executed on powerful GPUs, anomaly detection will probably run on ECUs without these capabilities. As a consequence algorithms such as

deep neural networks might be not feasible for embedded anomaly detection. Also algorithms, such as k-NN are not suitable due the high memory requirements.

5. Conclusions

In this work we presented how AI-based methods can detect anomalies in automotive networks such as CAN or automotive Ethernet, as well in processing units. We first motivated the need of Intrusion Detection Systems (IDS) in the automotive domain and further detailed the possibilities of an AI-based and data-driven approach of anomaly detection to detect unknown attacks inside the vehicle. The advantage of this AI-based approach is that the models of normality are generated by data from observations rather than from handcrafted rules. As modern automotive computing systems are very complex, the need for this kind of data-driven approaches increases more and more. Nevertheless, domain knowledge is still needed to design and craft the features which are fed into these AI-based systems. In addition, they can work hand-in-hand with rule-based systems derived from specifications to secure the car from malicious attacks.

References

[1] The bro network security monitor. https://www.bro.org/.

[2] Snort - network intrusion detection and prevention system. https://www.snort.org/.

[3] Dorothy E Denning. An intrusion-detection model. IEEE Transactions on software engineering, (2):222-232, 1987.

[4] Min Du and Feifei Li. Spell: Streaming parsing of system event logs. In Data Mining (ICDM), 2016 IEEE 16th International Conference on, pages 859-864. IEEE, 2016.

[5] Min Du, Feifei Li, Guineng Zheng, and Vivek Srikumar. Deeplog: Anomaly detection and diagnosis from system logs through deep learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pages 1285{1298. ACM, 2017.

[6] Charlie Miller and Chris Valasek. Remote exploitation of an unaltered passenger vehicle. Black Hat USA, 2015, 2015.

[7] Charlie Miller and Chris Valasek. Securing self-driving cars (one company at a time). Online, 2018. http://www.illmatics.com/securin_ self driving_cars.pdf.

[8] Konrad Rieck. Machine Learning for Application-Layer Intrusion Detection. Dissertation, Technische Universität Berlin, July 2009.

[9] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review 65.6 386, 1958.

[10] Vladimir Vapnik, Alexey Chervonenkis. A note on one class of perceptrons. Automation and Remote Control, 25, 1964.

[11] Chio, Clarence, and David Freeman. Machine Learning and Security: Protecting Systems with Data and Algorithms. O'Reilly Media, Inc., 2018.

[12] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF: identifying density-based local outliers. ACM sigmod record. Vol. 29. No. 2. ACM, 2000.

[13] Bernhard Scholkopf and Alexander J. Smola. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA, 2001.

[14] F. T. Liu, K. M. Ting and Z. Zhou. Isolation Forest. Eighth IEEE International Conference on Data Mining, Pisa, pp. 413-422, 2008.

[15] M. Abadi, et al. Tensorflow: A system for large-scale machine learning. 12th USENIX Symposium on Operating Systems Design and Implementation. 2016.

[16] http://spot.incubator.apache.org/. 2018

[17] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, A detailed analysis of the KDD CUP 99 data set. 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, *ON, 2009*.

[18] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF: identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data (SIGMOD '00). 2000.

[19] Liao, Yihua, and V. Rao Vemuri. Use of k-nearest neighbor classifier for intrusion detection. Computers & security. 2002.

[20] Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. 2008 Eighth IEEE International Conference on Data Mining. IEEE, 2008.

[21] Zhou, Chong, and Randy C. Paffenroth. "Anomaly detection with robust deep autoencoders." Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2017.

Continuous Security Testing for the Automotive Domain[1]

Simon Greiner, Hans Löhr, Paul Duplys,
Robert Bosch GmbH, Renningen

Kurzzusammenfassung

Die Vermeidung von sicherheitsrelevanten Schwachstellen in Software ist ein wichtiges Thema in der Automobilindustrie. Heutzutage erfordert das Auffinden von Softwarefehlern großen manuellen Aufwand und viel Security-Expertise trotz der verfügbaren semi-automatischen Analysetools. Außerdem müssen Security-Testing-Lösungen für den Automobilsektor eine Reihe domänenspezifischer Anforderungen erfüllen; z.B. müssen sie mit heterogener Software und unterschiedlichen Build-Umgebungen umgehen können.

In diesem Bericht geben wir einen Überblick über kontinuierliches Security-Testing und stellen Anforderungen dafür aus dem Automobilbereich vor. Darüber hinaus präsentieren wir eine Plattformarchitektur für kontinuierliches Security-Testing welche die Anforderungen für Automobilsoftware erfüllt. Wir geben Beispiele von Analyse- und Testmethoden die auf der Plattform ausgeführt werden, einschließlich dynamischer Tests und unterschiedlicher Varianten von statischer Analyse. Schließlich berichten wir von experimentellen Ergebnissen mit einem Prototypen der Plattform, den wir bei Bosch Corporate Research entwickelt haben.

Abstract

Avoiding security vulnerabilities in software is a major concern for the automotive industry. As of today, finding bugs in software requires a great amount of manual effort and security expertise, despite the available semi-automated analysis tools. In addition, security analysis solutions for the automotive domain must fulfil a number of domain-specific requirements, e.g., being able to cope with heterogeneous software, varying build environments, etc.

In this report, we give an overview of continuous security testing and present requirements for continuous security testing in the automotive domain. Moreover, we present an architecture of a platform for continuous security testing that accommodates the requirements of automotive software. We give examples of analysis and testing methods that can be run on the platform, including dynamic testing approaches and various flavours of static analysis. Final-

[1] This is a report on previous work. The full paper with more details on our framework has been submitted for publication at escar Europe 2019 [1].

ly, we report on experimental results obtained with a prototype of such a platform built at Bosch Corporate Research.

1 Introduction

Security is a key property for connected vehicles because they are inherently exposed to remote attacks. With up to a hundred Electronic Control Units (ECUs) connected to the in-vehicle network, a modern car essentially runs on software. From the security perspective, software is typically a very critical component for two reasons. First, real-world software is usually very complex, and complexity ultimately leads to security issues. Second, software – in the form of, e.g., a communication stack or an API – is the first (and, oftentimes, the only) thing a remote attacker interacts with. As a result, avoiding security vulnerabilities in software is a major concern for the automotive industry.

As of today, finding bugs in software requires a great amount of manual effort and security expertise, despite the available semi-automated analysis tools. In addition, security testing solutions for the automotive domain must fulfil a number of domain-specific requirements in order to be practically useful, e.g., being extensible and able to cope with heterogeneous software, varying build environments, low-level languages, non-standard libraries, etc.

For cost and efficiency reasons, it is essential to detect vulnerabilities as early as possible in the development cycle. Therefore, methods and tools to analyse and test software with regard to potential security issues during (not only after) the development phase are urgently needed. Continuous integration (CI) frameworks are used increasingly for automotive software, for instance, to execute unit tests on a regular basis. We stipulate that security analysis should also be integrated in CI workflows, to enable early detection of software vulnerabilities. Existing tools for continuous security analysis either focus on relatively homogeneous enterprise software – and hence, they cannot easily be applied to embedded software – or they provide specific (limited) analysis methods without the flexibility to include novel approaches. To accommodate the requirements of automotive software, we need an extensible framework that can cope with heterogeneous software and development environments. Moreover, it must be easy to extend the framework with new software analysis methods. We envision a large variety of such analysis methods such as dynamic testing approaches, static analysis, and combinations of several techniques.

As a solution, we propose CrATE, our framework for continuous security analysis and testing. CrATE provides a scalable, extensible, and flexible platform to integrate various software analysis and testing methods in a CI workflow. Although our focus is on software security

and robustness, other automated quality assurance methods could also be included in CrATE.

Outline. In the remainder of this report, we present requirements for continuous security testing that are specific to the automotive domain. Moreover, we present the architecture of a platform for continuous security analysis that accommodates the requirements of automotive software. We give examples of analysis and testing methods that can be integrated in our platform, including dynamic testing approaches, various flavours of static analysis, and combinations of several techniques. Finally, we show selected experimental results obtained with a prototype of such a platform built at Robert Bosch Corporate Research.

2 Automotive Requirements

In contrast to enterprise software and web applications, automotive software possesses some distinct characteristics, leading to specific requirements for our software security analysis framework.

Automotive software is (deeply) embedded, i.e., very hardware-dependent, using low-level programing languages (e.g., C, C++), and highly dependent on complex build- and execution environments. Hence, a software analysis framework must support low-level languages and heterogeneous environments.

Safety relevance, in particular, where human life is concerned, leads to the need for highly reliable and robust software. This implies that measures to ensure safety and security are extremely relevant. Thus, a software analysis framework should be suitable for the integration of a variety of tools.

Safety engineering norms, standards, and best practices require safety and quality assurance measures, which lead to high costs for software changes and updates. The cost of such changes increases, when they occur in later stages of the development or product life cycle. Therefore, vulnerabilities must be detected in the earliest possible development phase. This implies that the platform must be easy to include in continuous integration work flows. Hence, a high degree of automation and a low false positive rate of the analysis tools are essential.

Moreover, “batch processing” software (i.e., programs that get input, process the data, and then terminate) is rare in the embedded domain. In contrast, stateful, long-running programs (processing messages or sensor data, executing control loops, etc.) are common. Thus, we need analysis methods for stateful programs.

Typically, there are separate software development and security teams, which means that the tasks of security experts and software developers must be separated, with a clearly de-

fined interface between them. It must be possible to apply complex analysis methods without too much burden on the software development team.

A rapidly changing tool landscape and the use of non-standard libraries and APIs (compared to enterprise software) require a flexible platform, where new tools and methodologies can be integrated easily. In addition, source code is available at Tier 1 and software suppliers, i.e., methods are not limited to black-box approaches.

The automotive industry tends increasingly towards the use of virtualization and Software-in-the-Loop (SiL) as the basis for software testing, and employs container technology (e.g., Docker). If development and testing teams are already using such techniques for other purposes (e.g., unit and component testing), they should also be used for a security analysis and testing framework.

Current and anticipated trends. Moreover, the importance of robust and secure software is increasing, because of changes in the automotive industry. Most notable are increased connectivity and higher software complexity in recent automotive software. But we also expect changes in industry regulations and standards that will pose increased requirements to robustness and security testing, e.g., by demanding penetration tests before deploying connected ECUs. In addition, the limited update capabilities of on-board software during the lifetime of a vehicle and the fixed software state at deployment time lead to the necessity of thorough testing and quality assurance before the start of production. Moreover, the support timeline for software is increasing (security monitoring and maintenance after deployment are relatively new requirements in automotive). These trends emphasize the need for a comprehensive software security analysis solution.

3 The CrATE Platform

We introduce the CrATE (Continuous security Analysis and Testing) platform, a prototypical implementation of a platform which meets the requirements discussed in the previous section. Here, we limit the consideration of the framework to the CrATE tool, an extensible, highly tailorable tool for security analyses in a continuous integration development process.

The architecture of the CrATE tool defines an entry point for the analysis of software, a workflow for the specific tasks that are common for the analysis and tests of embedded projects, and a reporting interface, which provides the results to the project members, e.g., developers, architects, and security specialists. Methods integrated into CrATE are typically tailored to specific projects and bundled into so-called Bottles.

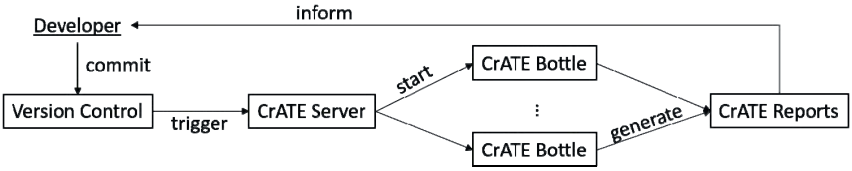


Fig. 1: Components of the CrATE Platform

3.1 CrATE Continuous Integration Architecture

The CrATE tool consists of the components CrATE server, a set of CrATE Bottles, and CrATE reports, as shown in Fig. 1. The main entry point for a continuous integration of CrATE into a concrete project's process is some Version Control Sytem (e.g. git). After some trigger, which is typically the commit of a new software version or a time trigger, the CrATE server copies the content of the version control system to its local storage. The repository contains the source code of the project's software, as well as some configuration files for CrATE and potentially additional information which is required by a particular analysis or test method. Depending on the configuration, the CrATE server starts one or more analyses of the software in parallel. After the analyses have terminated, the CrATE reports component transforms the output from the analyses into a human-readable form or other representations which can be integrated with additional tools.

The CrATE server is based on the commonly used Jenkins server[2], and the workflow is implemented as a so-called Jenkins pipeline.

The project-specific tool bundles (Bottles) perform the actual analysis of the software. Bottles run in virtualized environments. Different analysis methods and embedded software projects typically have specific dependencies on their runtime environments. The virtualization allows us to easily encapsulate these potentially contradicting dependencies per Bottle. Details on Bottles are described in the next section.

After each analysis has terminated, the format and level of detail of the results are specific to the analysis method used. The CrATE reports component transforms these results into a consistent form, as required by the project.

3.2 CrATE Bottle Workflow

CrATE Bottles implement analysis-specific and project-specific bundles. They are the central architectural element to tackle the requirements for embedded software development pro-

[2] https://jenkins.io/

cesses of typically heterogeneous build environments, cross-domain tooling, easy integration of new methods into project development processes, and virtualization and Software-in-the-loop solutions. The workflow of a Bottle consists of the steps pre-processing, analysis, and reporting, as illustrated in Fig. 2.

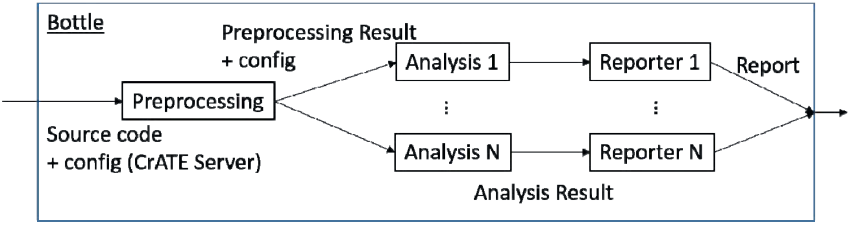


Fig. 2: Workflow of a CrATE Bottle

The pre-processing step takes as input the content of the Version Control System of the project and generates general information about the analysis target. The concrete form of this information depends on the analysis method implemented in the second step. For dynamic analysis, the pre-processing step typically implements a compilation of the target, for the application of formal methods, the pre-processing can generate a formal representation of the source code in some suitable format.

In the second step, the actual analysis of the software is performed. The concrete analysis can be software fuzzing, analysis of program dependency graphs for finding vulnerabilities, formal verification of path conditions, and others. (Some examples are discussed in more detail in the next section.)

Often, the pre-processing step can generate information which can be used by different analyses using the same method. For example, a single compilation of a software can yield several executables, while dynamic analysis, e.g. software fuzzing, can be performed on each of these executables. Therefore, several analyses can be started after one pre-processing step in parallel.

In the third step, the method-specific results from the analysis are interpreted and prepared for reporting findings to developers and other interested people involved in software development. The results from different analysis methods are typically very different. While software fuzzing, for example, provides concrete inputs for the software for each finding, static analysis provides correctness proofs for formulas extracted from the software. Therefore, after each analysis step, the CrATE reports component provides a more consistent view of

the results of each analysis step. These results can be used to generate human-readable reports or as inputs to bug tracker and other tools used during software development. The concrete format depends on the needs of a particular project.

In order to meet the widely different requirements and dependencies of each project and analysis methods, we implemented the different steps of Bottles as Docker containers[3], a light-weight virtualisation solution. Each container encapsulates an installation of the Linux or Windows operating system and thus allows to run several analyses with different, potentially contradicting, dependencies on one server.

4 Analysis Methods in CrATE

CrATE can support a large variety of software analysis methods. Here, we exemplify this flexibility by discussing the integration of three different methods: a simple static checker for source code, a more involved static analysis based on code property graphs, and software fuzzing as a dynamic method.

4.1 Static Code Checking

Relatively simple static source code checkers are used in the automotive domain to ensure that the source code adheres to best practices, coding rules, and guidelines, such as the automotive-specific MISRA standard[4]. The purpose of these rules is to avoid programming errors, in particular, in safety-critical software. However, there are also coding rules that focus on security issues, for instance, the rules defined in MISRA C:2012 Amendment 1, or the SEI CERT Secure Coding Standard[5]. Static analysis tools are used to check adherence to the subset of those rules that can be checked statically on a purely syntactic basis. Examples for such code checkers are the open-source tools cobra[6] and flawfinder[7], but there are also commercial alternatives, for example Checkmarx CxSAST[8].

A CrATE Bottle for such static checkers is relatively straight forward: pre-processing is usually not even necessary, so we only need a Docker container for the actual analysis, and one to report the results. The analysis step can be implemented by simply installing the respec-

[3] https://www.docker.com/

[4] https://misra.org.uk/

[5] https://www.sei.cmu.edu/

[6] https://spinroot.com/cobra/

[7] https://dwheeler.com/flawfinder/

[8] https://www.checkmarx.com/

tive tool in a Docker container, adding a script that starts the tool with appropriate configuration, and providing configuration information for CrATE. The reporting component can be as simple as a script that collects the findings from the output of the tool in a human-readable report document, or it can involve some further post-processing like filtering, sorting, etc.

4.2 Static Analysis Based on Code Property Graphs

Code property graphs [2] provide the basis for more powerful static source code analysis methods than syntactic checking. The basic idea is to integrate graphs which are well-known in compiler construction – like abstract syntax trees, control flow graphs, and program dependency graphs – into one code property graph (CPG) and add data flow semantics. The CPG can then be stored in a graph database, and a graph query language can be used to search for patterns, for instance, to find taint-style vulnerabilities, where attacker-controlled input is used in some critical operations. Since CPGs provide us with more sophisticated information about the program than purely syntactic properties, they offer possibilities for more powerful code analysis, in particular, they enable the search for patterns that indicate potential vulnerabilities where knowledge about dataflow is required.

CPG-based vulnerability discovery can be integrated into a CrATE Bottle by constructing the CPG in a pre-processing step, and querying the CPG in (potentially several) analysis steps. The pre-processing step contains a fuzzy parser (based on so-called island grammars) that takes source code or even incomplete source code fragments, constructs the CPG, and stores the result in a graph database. For the analysis step, scripts with a collection of graph queries can be executed to search the CPG for indications of potential vulnerabilities. Different collections of CPG queries could form different analysis steps which can be run in parallel. A final reporting step could build a report from the findings. The main challenge, both from a practical and a scientific point of view, is to formulate good search queries that can find vulnerabilities in real-world source code automatically, without reporting too many false positives.

4.3 Software Fuzzing

Software fuzzing is a dynamic testing method, where a program is executed many times with seemingly random input in order to find robustness problems, for instance crashes. Coverage-guided fuzzers try to maximize code coverage during the test executions by instrumenting the source code during compilation (to obtain coverage information later on) and then varying the fuzzing input during test runs according to some heuristic based on previous inputs and coverage information. To fuzz software components such as the implementation of

an API, a test harness has to be written that takes inputs from the fuzzer and injects it into the software component in such a way that the fuzzing input can be processed. Examples for state-of-the-art coverage-guided fuzzers are american fuzzy lop (afl)[9] and libFuzzer[10].

We integrated software fuzzing into CrATE Bottles in the following way: The pre-processing step is a docker container with a complete build environment for the embedded software. In this build environment, the original compiler has been replaced by a fuzzer-specific compiler (e.g., afl-gcc) that instruments the code as required for coverage-guided fuzzing. Moreover, a project-specific test harness had to be included to connect the fuzzer to the fuzz target (i.e., the software under test). To facilitate the implementation of the test harness, we provide an abstract wrapper (an abstract class), where only some project-specific methods have to be implemented by the development team. The result of the pre-processing step are binaries that can be executed by the fuzzer in the subsequent analysis steps. Several analysis steps can be executed in parallel, e.g., with different configurations of the fuzzer, or with different fuzz targets generated by the same pre-processing step. For our first proofs-of-concept, we kept the reporting very simple, i.e., we just collect the output of the fuzzer (e.g., the crashes found by fuzzing).

In the following section, we report on some of our results from fuzzing embedded software.

5 Fuzzing Embedded Software with CrATE: Preliminary Results

We used the CrATE platform in combination with the dynamic analysis method software fuzzing to analyse the programs of three software development projects. We applied it to an inter-process communication middleware for automotive systems, an open source logging library used in embedded automotive software, and a network communication library developed for automotive ECUs. We discovered security-relevant findings, which could be resolved and did not have an effect on software in the field. We ask the reader for understanding, that we cannot provide all details on the findings and results due to publication rules in our organization.

[9] http://lcamtuf.coredump.cx/afl/

[10] https://llvm.org/docs/LibFuzzer.html

5.1 Inter-Process Communication Middleware

The communication middleware is a stateful program, i.e., the behaviour of the program for given inputs heavily depends on the current state of the program reached by previously processing other inputs. Reaching different, interesting initial states for software fuzzing during setup of the program requires a system expert and a developer. Further, this particular program exclusively requires resources which are offered by the operating system the middleware runs on.

The concept of our CrATE platform proved very useful when analysing this program. For one, in order to analyse the program in several initial states, we built several executables during pre-processing, each initializing the program in a different state. During the analysis step, we could analyse these executables in parallel. Further, the dockerization of the analysis phase allowed us to encapsulate the resource requirements for which the program, and therefore each executable, required exclusive access.

5.2 Open Source Logging Library

We also applied our platform on an open source library for logging information during runtime of an embedded system. Since we did not want to place configuration files into the official project repository, we mirrored the official repository into our internal repository. As a result, we were able to easily switch between different versions of the software for analysis.

5.3 Network Communication Library

The most recently analysed project is the implementation of a network communication protocol, which is used in ECUs. The program is stateless, but relies heavily on structured input which fits the format defined by the protocol. We used a wrapper which starts the program and takes care of transforming the input from the fuzzer into a format as defined by the protocol. This wrapper is stored in the project's repository and thus easily modifiable by developers of the project team.

6 Related Work

Security testing in the automotive domain. A recent article on secure automotive software by Pike et al. [3] refers to the Motor Industry Software Reliability Association's "Development Guidelines for Vehicle Based Software" that recommends approaches for improving the software quality including software security. In the area of automated testing, the guidelines recommend using fuzzing (which is also mandated by SAE J3061) and property-based testing.

Static code analysis – a popular testing method used in the automotive industry – refers to any source code analysis performed without actually executing the program. A widely-used static code analyzer in the automotive domain is Astree[11]. Astree, which is tailored towards safety-critical embedded code, is based on abstract interpretation and analyzes programs written in C. While the company AbsInt offers a Jenkins plug-in to integrate Astree into a project-specific Continuous Integration/Continuous Delivery (CI/CD) pipeline, we are not aware of any commercial solution offering Astree as a service.

Continuous security testing. Recently, many companies across the IT industry started adding tools for static and dynamic security analyses to the CI/CD pipeline [4]. This trend, which has a strong link to DevOps, is referred to as continuous security testing, continuous security validation, security as code, and sometimes as DevSecOps, a larger field concerned with the integration of security processes and practices into DevOps environments [4].

Continuous security testing aims for a wide coverage of vulnerabilities by employing multiple security analysis methods [5]. According to [5], most suitable methods for continuous security testing are those that perform tests whose results do not require further investigation, i.e., whose results are either of a pass/fail nature or can be meaningfully interpreted as such. In addition to individual, stand-alone tools, there are also tool suites and frameworks specifically designed to run in the CI/CD context. Popular representatives of such frameworks are the open source framework Gauntlt[12] and F-Secure's Mittn[13].

Essentially, the main technical theme in continuous security testing is the integration of the aforementioned analysis tools into the CI/CD pipeline on a per-project basis. As a result, continuous security testing is not really easy to use for developers and the integration of new analyses for the security analysis of automotive software poses a problem, except for organizations with their own DevSecOps team.

Security testing as a service. The paradigm of automated security testing as a service is a natural evolution of continuous security testing. Google just recently open-sourced ClusterFuzz[14], a distributed fuzzing infrastructure. It provides end-to-end fuzzing automation including the execution of the fuzzer itself, triaging of the crashes, minimizing the testcases, identifying the code changes that introduced the bug, verification of fixes, as well as bug filing and

[11] http://www.astree.ens.fr/

[12] http://gauntlt.org/

[13] https://github.com/F-Secure/mittn/

[14] https://github.com/google/clusterfuzz/

closing for issue trackers. Feature-wise, ClusterFuzz depends heavily on the Google Cloud Platform services like Google Cloud Storage.

BlackBerry Jarvis[15] is a commercially offered service for binary static application security testing (SAST) geared towards automotive and embedded applications. The service is hosted on Amazon's AWS and supports static analysis of binaries.

7 Conclusion

In this report, we discussed requirements for security analysis of embedded automotive software and presented CrATE, our solution for continuous security analysis and testing. CrATE provides a scalable and extensible platform for security analysis, robustness testing, and other quality assurance techniques. We explained how different analysis methods can be integrated in the same framework and reported on preliminary results from our initial experiments with fuzzing embedded software. We found that acceptance and demand for continuous security testing in development projects is very promising.

In addition to the technical platform, we also developed a process for introducing software fuzzing into development projects based on CrATE, which cleanly separates the duties of system experts, developers, and security experts. For more details, on both the technical platform and the CrATE fuzzing integration process, see [1].

Based on our experience, we plan to add further testing and analysis techniques to our framework and investigate how well the Docker containers can be re-used in other contexts. Finally, we will use our framework to systematically compare the performance of different security analysis techniques on real-world automotive embedded software.

[15] https://www.blackberry.com/us/en/static-application-security-testing/blackberry-jarvis/

References

[1] Greiner, S., Löhr, H., Duplys, P.: Scalable Security Analysis for Automotive Software. Embedded Security in Cars – escar Europe 2019, submitted

[2] Yamaguchi, F., Golde, N., Arp, D., Rieck, K.: Modeling and Discovering Vulnerabilities with Code Property Graphs. IEEE Symposium on Security and Privacy, 2014, pp. 590–604

[3] Pike, L., Sharp, J., Tullsen, M., Hickey, P.C., Bielman, J.: Secure automotive software: The next steps. IEEE Software 34(3), 2017, pp. 49–55

[4] Bird, J.: DevOpsSec. O'Reilly Media, Inc., 2016

[5] Kuusela, J.: Security Testing in Continuous Integration Processes. Master thesis, Aalto University, 2017

AUTOSAR Adaptive Platform

A standardized SW platform for intelligent vehicles with functional safety and data integrity

Dr. **Günter Reichart**, **Michael Niklas**,
AUTOSAR partnership, Aschheim near Munich

Abstract

The AUTOSAR development partnership has grown to more than 270 partner companies since the first AUTOSAR Classic Platform specification was released more than twelve years ago. This classic platform is widely adopted in the automotive industry and can be found in many modern vehicles. But it was soon recognized that this approach will be not sufficient to meet the coming challenges of future vehicles. Highly automated driving with high demands for functional safety as well as of security will lead to massive changes of the E/E-architecture in vehicles both for hardware as well as for software. Communication between vehicles or between vehicles and infrastructure demand for very high levels of data integrity and communication security. AUTOSAR develops since 2015 an intelligent and flexible infrastructure SW which is named AUTOSAR Adaptive Platform. The cornerstones of this adaptive platform have been consolidated in spring 2019 and provide developers a sound basis for project realization. Further features will come with the next releases in a 6-12 month schedule. The platform runs on high end computing hardware and supports parallel processing on many core systems and GPUs. Since AUTOSAR has its roots in the automotive industry the highest priority of safety and security features is as self-evident as the compatibility to systems based on the AUTOSAR Classic Platform. An appropriate software framework for safe and secure for next generation vehicles is now available.

1. Introduction

Cars continue to turn into real cyber physical systems – just connecting to the internet and exchanging data with smartphones is state of the art. Future cars will be connected to almost everything: Smart homes, roadside infrastructure and even vehicles around them – they become a part of the internet of things.

Highly automated driving IoT and cloud services New processor technologies

Fig 1: Use-cases driving the further development of the AUTOSAR Standard

Another trend beside the increasing connectivity is the vision of autonomous driving. Further enhancements of today's advanced driver assistance systems pave the way towards highly automated driving and autonomous parking, see Figure1.

The realization of these new features also adds new requirements on the hardware and software infrastructure, hosting these functionalities. Besides the existing requirements such as functional safety and security the software architecture has to support e.g.

- high-end manycore processors,
- hypervisor technology
- updates and upgrades over-the-air with the ability of self-configuration,
- communication with backend-systems or
- dynamic deployment/allocation of applications, as well as
- dependability to realize autonomous vehicles.

An evaluation by the AUTOSAR consortium showed that these new requirements cannot be realized by today's software architectures where almost all vehicle internal communication is done via deeply embedded controllers with dedicated functionality to meet OEM requirements like start up times or functional safety. With the increasing complexity there has to be a move from the traditional signal based communication paradigm to a service oriented concept allowing a much better structuring of cross domain tasks. In general a software infrastructure is required that is much more flexible as today's but can meet high safety and security demands. It has to be highly available and capable to adapt itself to specific application requirements at a given point in time. It has to support dynamic allocation of functions and self configuration. Consequently today's architectures will be complemented by a new platform that comes along with operating systems designed for high-performance computing which needs to be enhanced by dependability features like e.g. safety, availability, security or real-time.

Nevertheless the well-known characteristics of deeply embedded system will remain. The combination of these trends results in a revolution of today's E/E architectures.

2. Key aspects of the new E/E architecture

The following three key aspects characterize tomorrows E/E architectures:

A Integration of heterogeneous software platforms

Networking architectures of today's cars can be clustered into different domains for infotainment and connectivity, chassis, powertrain, etc. While infotainment ECUs are typically using Linux or commercial general purpose operating systems, the AUTOSAR Classic Platform is the standard for deeply embedded ECUs. With the new use cases and the increasing demand also from deeply embedded applications for computing power a third type of ECUs will arise with different characteristics that has to be interconnected with existing E/E architectures. Especially the handling of across domain functionality requires very powerful central computing units with effective measures to meet the high safety and security requests. For these ECUs a new SW architecture compatible with classic AUTOSAR platform is an urgent need.

B Service oriented and signal based communication

The traditional automotive communication is still based on the idea of ECUs providing signals to other ECUs as broadcast. This paradigm fits very well for control data of limited size, which has to be communicated cyclically.

Advanced applications like highly automated driving with higher payload demands e.g. to exchange dynamic length lists of objects detected by a set of sensors and Ethernet as a communication system require more sophisticated protocols. The concept of service oriented communication is based on applications that provide a service on the communication system and other applications that subscribe to this service. The data will only be sent to the subscribers. The combination of service oriented communication with the existing signal based paradigm is the second key aspect of future E/E architectures and a demanding challenge from a methodological point of view.

AUTOSAR as the global standardization consortium for automotive software architectures adopt to these trends and provides a consistent standard for these aspects.

C New standard defined

For deeply embedded systems realizing typical power train and chassis functionalities the AUTOSAR Classic Platform will remain the first choice. Such applications are characterized

by high-demands on safety, real-time and determinism while running on low cost hardware. Meanwhile AUTOSAR provides for these applications a well-proven and mature software platform including a widely used methodology, which supports all of today collaboration models.

To support dynamic deployment of customer applications, to support dependable applications and to provide an environment for applications that require high-end computing power AUTOSAR is currently standardizing a second software platform which is called AUTOSAR Adaptive Platform. Based on existing standards, the idea is to benefit as much as possible from developments in other areas (e.g. consumer electronics, automation industry), while still considering automotive-specific requirements such as functional safety.

3. The AUTOSAR Adaptive Platform

Figure 2 depicts the overall functional architecture of this new platform. With Release 17-03 in the year 2017 one of the main goals for the AUTOSAR Adaptive Platform was to provide application developers a stable programming interface the so called AUTOSAR Runtime for Adaptive Applications (ARA).

The interface consists of a standardized interface for accessing operating system functionalities and a communication middleware, which allows data exchange with local and remote applications as well as to the Adaptive AUTOSAR Services. The second goal is to support the basic functionality needed to smoothly integrate the platform into existing E/E architectures based on Ethernet.

To achieve this goal the following functional clusters are specified: The core of the AUTOSAR Adaptive Platform is the operating system based on POSIX. The OS can be used from the application via a subset of POSIX according to IEEE1003.13 [1], namely PSE51 and provides the application developer commonly used functionality such as signals, timers, semaphores, signals and thread handling. If an application restricts itself to the use of this subset it shall be portable to any AUTOSAR Adaptive Platform. From the functional point of view the operating system defines some of the essential differences of the AUTOSAR Adaptive Platform compared to the AUTOSAR Classic Platform. Applications are not bound any more to a very strict and static scheduling and memory management but are free (within well-defined boundary conditions) to create and destroy threads and to allocate memory depending on their current need.

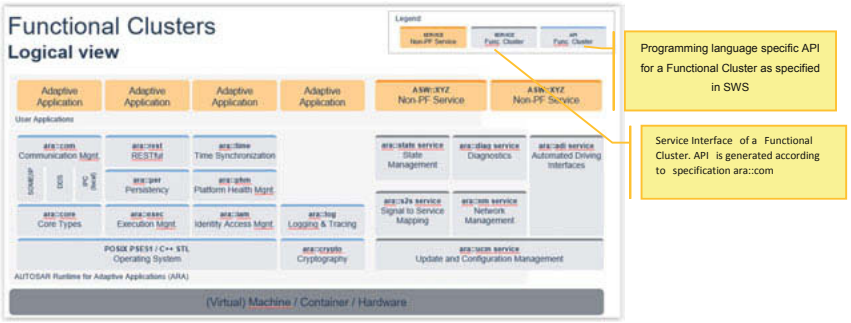


Fig 2: Functional Architecture of the AUTOSAR Adaptive Platform

The functional cluster Execution Management is responsible for startup and shutdown of the ECU and the applications by maintaining the application states. Together with the Platform Health Management it has to take care that the necessary resources for the applications are available to trigger degradation or similar strategies so that the ECU can react in a critical situation according to a well-defined strategy.

The communication middleware realizes and abstracts the service oriented communication between local applications as well as to applications residing on other ECUs. This includes the interaction with the AUTOSAR Adaptive Services.

The support of Ethernet based communication systems is achieved, based on SOME/IP as bus protocol, which also enables applications on the AUTOSAR Adaptive Platform to establish communication relationships with ECUs running the AUTOSAR Classic Platform.

The architecture contains also Log and Trace features for debugging of an AUTOSAR Adaptive ECU as well as persistency and the adaptation of the most important diagnostics features to a POSIX based architecture.

The goal for the Release 17-03, to enable the first series projects to start developing ECUs and applications based on the AUTOSAR adaptive standard was widely achieved. For the following releases of the Adaptive Platform a roadmap of major additional features has been defined and the releases are published in 6-12 month schedule. This includes e.g. more support for fail-operational systems, enhanced safety and security features and the integration of Car2X communication protocols. Due to the main use-cases of the Adaptive Platform special focus will be on the support for dependability.

4. SUPPORT FOR THE INTEGRATION OF HETEROGENEOUS PLATFORMS

The methodology defined by AUTOSAR is the only standardized solution available today, which allows describing the software architecture together with the network topology in a unified machine-readable format. Up to now this concept considered only the ECUs based on the AUTOSAR Classic Platform and the principle of signal-based communication. For the future the most important challenge is to support a seamless integration process of different platforms. With the introduction of Ethernet and a service-oriented communication paradigm based on SOME/IP in the Release 4.1.1 of the Classic Platform, AUTOSAR already made a big step to support this use case. In terms of SOME/IP, a service can be seen as the functional representation of an ECU on the bus independent of the underlying software platform. Therefore, a formal description of a service has to be independent of a specific platform, be it the AUTOSAR Classic Platform, the AUTOSAR Adaptive Platform or a non-AUTOSAR platform. The key point is that each platform realizes the service in the same way w.r.t. to its behaviour and the representation in the on-the-wire format. This means that AUTOSAR has to provide answers for the following aspects:

- Description of the relevant communication relationships on software level
- Provisioning of standardized communication protocols.

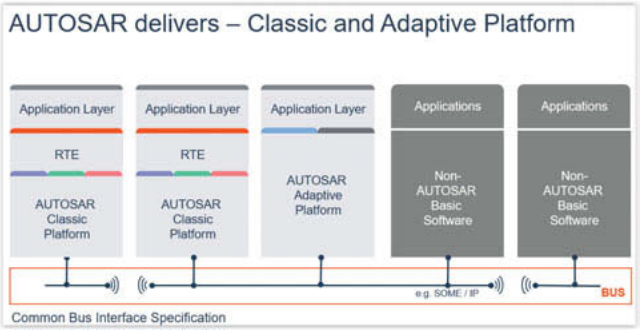


Fig 3: Integration of different platforms

The main difference between platforms supported by AUTOSAR and other platforms is then the support of a methodology, which enables tool vendors to provide automated configuration of ECUs running a software stack, which complies with the standard.

5. AUTOSAR STANDARDS AND TIMELINE

Previously, AUTOSAR released all of its specifications as one bundle, however, it became increasingly difficult for users to track relevant changes to the standard, especially with the introduction of the Adaptive Platform. To keep the standard manageable, useable and to enable flexibility AUTOSAR decided to split up its results into several standards for different feature sets. An AUTOSAR standard is now defined as a consistent set of AUTOSAR deliverables, which are released at the same time. AUTOSAR deliverables can, but are not limited to be of the following kinds: textual explanations, textual specifications, test specifi-cation, source code, other formal or semi-formal textual formats (e.g. ARXML, UML models, XML Schema)

The AUTOSAR Classic Platform release R4.4.0 includes, among others, an extended serialization for Data Structures in SOME/IP with tag, length and value encoding (TLV). It simplifies the communication and improves the compatibility between the AUTOSAR Classic and Adaptive Platform. Moreover, a Transport Layer Security (TLS) allows the use of well-established services in non-vehicle domains.

One of the major achievements featured in the AUTOSAR Adaptive Platform Release R18-10 is the support of the Internet Protocol Security (IPSec) to secure data communication allowing a more secure communication between Adaptive platforms and their surrounding platforms. Better support for Diagnostics over Internet Protocol (DoIP) is in accordance with ISO 13400-x. In addition, release R18-10 includes the specification of the network binding to support the Data Distribution Service (DDS) protocol.

Most important, release R18-10 comprises the harmonization of Classic Platform and Adaptive Platform enabling the Network Management protocol to utilize network management across platforms. Moreover, a unique time base shared among all entities of both platforms has been introduced through the first step of the Time Sync harmonization.

The last consolidation release was R19-03. This synchronized release of the three AUTOSAR Standards shall enable the application of a new joint methodology to apply AUTOSAR as a whole properly in E/E architectures. So, the work group organization was harmonized in 2019 to achieve backward compatibility of the Classic and Adaptive standards by joint work groups. The approach of structuring AUTOSAR's deliverables into standards also opens up the new possibility to deal with future use cases and to establish the new AUTOSAR Adaptive Platform beside AUTOSAR Acceptance Test and the AUTOSAR Classic Platform. Common parts of these standards such as bus protocols and common aspects of the methodology are released as a separate standard called AUTOSAR Foundation.

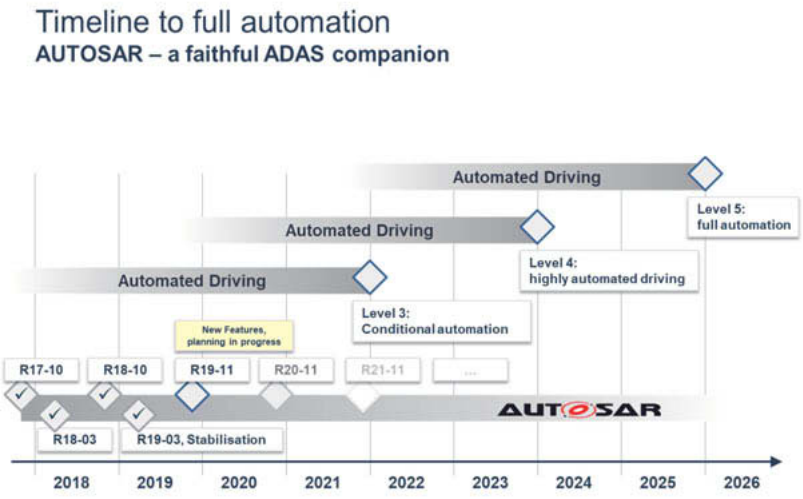


Fig 4: AUTOSAR Release Schedule

With the introduction of the so-called foundation standard the interoperability between the two AUTOSAR platforms is specifically addressed. Foundation contains common requirements and technical specifications (for example protocols) shared between the AUTOSAR platforms. The revision of the Foundation and the release FO R1.5.1 came along with the Adaptive Platform release R19-03. The release FO R1.51. supports the changes that are implemented in the release R19-03.

6. CONCLUSION

For over more than 15 years AUTOSAR has demonstrated to be the best established and well-suited organization to coordinate and drive the standardization for software infrastructures and platforms meeting the requirements of automotive electronics.

Upcoming demands and new functionalities will pose further challenging requirements on future E/E architectures. With the current set of available specifications for the classical and

the adaptive platform developers have a comprehensive set of documents to realize advanced software architecture solutions for the next generation vehicles.

7. ACKNOWLEDGMENT

This work was supported by the members of the AUTOSAR Steering Committee and the AUTOSAR Project Leader Team.

9. REFERENCES

[1] http://www.autosar.org

[2] IEEE1003.13; IEEE Standard for Information Technology - Standardized Application Environment Profile (AEP) - POSIX® Realtime and Embedded Application Support

[3] Slansky, Lorenz Daimler AG, Scharnhorst, Thomas, AUTOSAR Development Partnership, AUTOSAR für intelligente Fahrzeuge. HANSER automotive Special 2017.

Service-Oriented HPC Communication Standard for Vehicle Lifecycle Management

Dr. **Ansgar Schleicher**,
DSA Daten- und Systemtechnik GmbH, Aachen

1 Abstract

Vehicle electronics topologies are changing rapidly. New functions, like Car2X, ADAS and autonomous driving require application level software of considerable complexity to be executed within the vehicle. High Performance Controllers/Computers (HPCs) with considerable computing resources are found in most of today's vehicle platform designs. Vehicle topologies with multiple HPCs employ Automotive Ethernet for high-bandwidth connectivity between the HPCs.

Hence, the next generations of vehicle electronics architectures are going to be hybrids of regular Electronic Control Units (ECU), networked by CAN, LIN or FlexRay and HPCs networked by Automotive Ethernet. These new architectures result in new challenges regarding the management of the vehicle lifecycle, which includes the diagnosis, (re-)coding and flash re-programming of these hybrid vehicle electronics.

Today's gold standard to handle this life-cycle is composed of the OBD-connector (ISO 15031-3) to which an external test device is physically connected, which then communicates with the ECUs in the vehicle via the Unified Diagnostic Services (UDS – ISO 14229-1) protocol on a CAN(-FD) (ISO 11898-1), or in few modern vehicles via Ethernet. This protocol is focused on diagnosing faults in the electronics of the vehicle and allows for (low data volume) updates to the configuration and software of the ECU. A diagnosis of the software within an ECU is not possible. European and US legal regulation rely on this gold standard or very similar technology for emissions regulation (EUR 5/6, CARB), independent aftermarket access (EUR 5/6, Right-to-Repair, EPA) and Periodical Technical Inspection (2014/45/EU).

This technology is in great contrast to (a) the technical possibilities an HPC-equipped connected vehicle (3G/4G) is offering, (b) the diagnostic needs of an HPC hosting multiple virtual machines running multi-threaded application level software, (c) the widened spectrum of use cases in the digital lifecycle of a vehicle, which includes onboard- and remote use cases in addition to today's proximity use cases, which require a vehicle to be driven to a dealership. Examples include remote diagnosis, Firmware-over-the-air (FOTA) updates and predictive maintenance.

This contrast leads to individual implementations in the industry, which again lead to proprietary solutions that are costly and may risk market and legal acceptance. A group of automotive companies including three major OEMs has decided to initiate a standardization committee within the ASAM to standardize the future communication access to connected vehicles with hybrid electronics in the above sense and publish it as an ISO standard.
The new standard is going to cover the following aspects and requirements:

- Use of existing technology for service-oriented architectures (e.g. REST, JSON),
- Coverage of all usage scenarios: onboard, proximity and remote,
- Integration of existing authentication and authorization technology,
- Integration of legacy protocols ensuring one consistent access to hybrid electronics,
- Coverage of new lifecycle needs for HPCs regarding software update and diagnosis of multi-threaded complex software modules.

2 Introduction into future vehicle electronics topologies

Vehicle electronics engineering has focussed on the development of new control functions for an isolated mechatronic system as well as customer-facing functions in the comfort and infotainment domain for the past 20-30 years. Today's developments evolve around a vehicle that can perceive its environment through image, radar and potentially even lidar data and that is able to communicate with vehicles nearby (Car2Car communication), with the infrastructure (Car2Infrastructure communication) and internet-based services via mobile broadband networks. This enables a variety of new function areas, like

- ADAS
- Autonomous Driving
- OTA-services, like Firmware-Over-The-Air, Coding-Over-The-Air, Remote Repair
- Predictive and Preventive Maintenance

These function areas cannot be developed on today's electronics platforms, composed mainly of Electronic Control Units (ECUs) and the CAN (or other low-bandwidth bus systems like LIN or FlexRay). New technology has been introduced to enable the implementation of above function areas, namely High-Performance Controllers (HPCs) and Automotive Ethernet (AE). This new technology is not a mere evolution of the previous generation of employed technology: Where ECUs with their single-core designs, limited memory and networking capabilities are clearly focussed on cost-effective, reliable, embedded computing of control functions, HPCs enable the development of full software functions within the vehicle. Multiple cores of even different design (e.g. RISC, DSP), Gigabytes of memory and Gigabit networking capabilities both wired and wireless, bring advanced computing power to a vehicle and allow the execution

of modern operating systems like Linux, QNX, Android or AUTOSAR Adaptive. Even multiple and different virtual machines on top of a hypervisor can be designed within a single HPC.

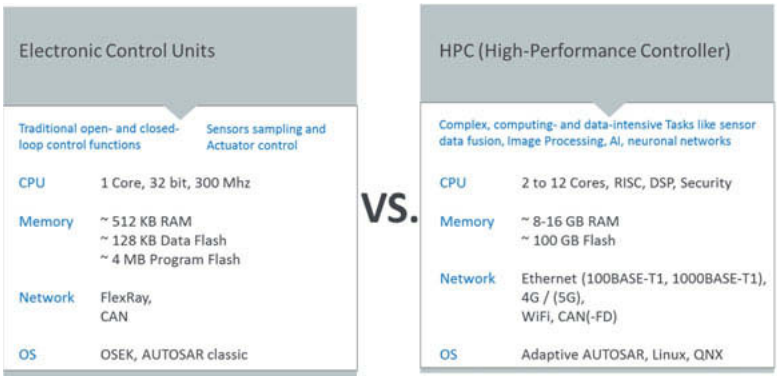


Fig. 1: Comparison of ECU and HPC

At the same time the onboard networking capabilities develop rapidly. Where the CAN bus with its maximum bandwidth of 1Mbit/s and 8 bytes of payload data per frame was designed to exchange signal (sampled sensor) data between multiple ECUs, Automotive Ethernet is designed to adapt well-established LAN-technology for cost-effective usage within the vehicle and enhance the bandwidth by a factor of 100 or even 1000.

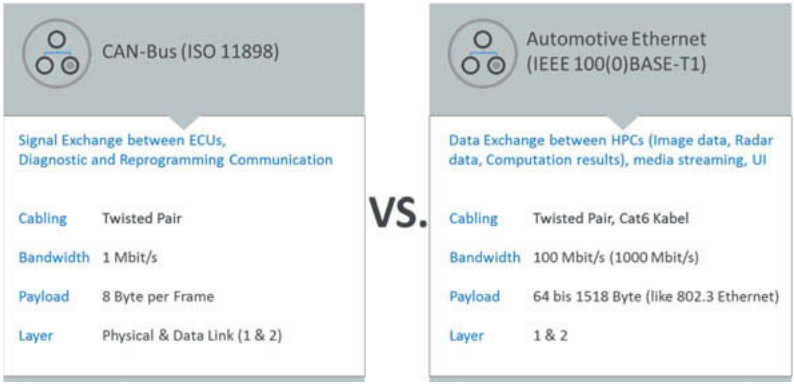


Fig. 2: Comparison of CAN and AE

Resulting next-generation vehicle topologies will feature one or multiple HPCs networked via AE and connected to the vehicle environment with multiple varying technologies like WiFi, 4G/5G, Car2X etc. ECUs will remain to be part of the vehicle topology. Interconnected via CAN, CAN-FD or FlexRay they will still execute the majority of the control functions. Modern vehicle architectures segregate different domains of the vehicle like Powertrain, Chassis, Comfort. This segregation will even be enhanced by HPCs taking over the role of a domain controller for one or multiple domains, replacing today's common central gateways by AE-networked HPCs.

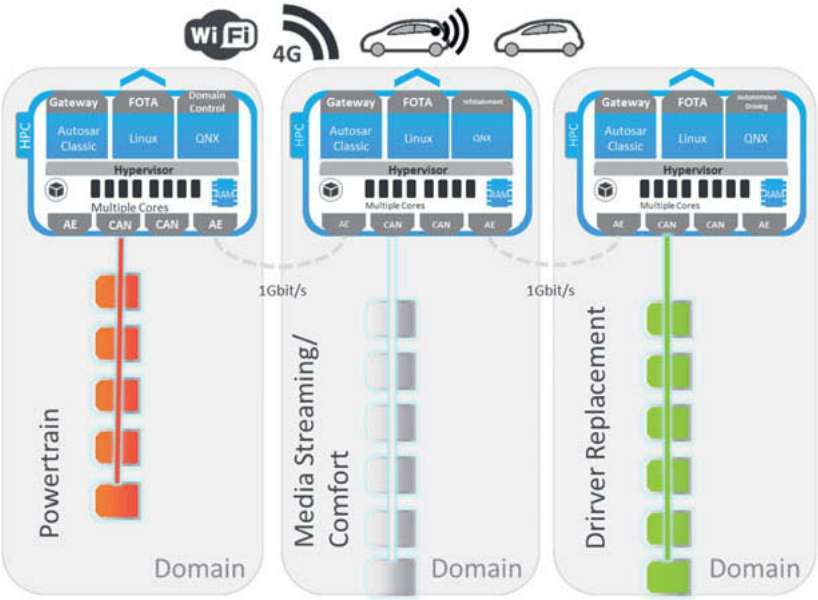


Fig. 3: Future hybrid, connected E/E architectures

This combination of new HPC- and AE-based technology with well-established ECU-/CAN-based technology results in a new hybrid electronics design (ref. Fig. 3) that inflicts new challenges during the lifecycle of the vehicle. One of these challenges is addressed within this paper: The challenge of performing high-standard diagnosis, firmware/software update and configuration of such vehicles in the engineering, production and service phase of the lifecycle.

3 Limitations of today's offboard communication technology

A strong and wide-spread technology has been developed to provide offboard communication capabilities for (HPC-less) vehicles [5]. Many aspects of this technology have been standardized and many of these standards are referenced by national law. It is not the aim of this paper to give a full-fledged overview of the different derivatives of this technology. Rather, the mainstream technology stack is described to establish a sound motivation for a more advanced technology stack for vehicles featuring a hybrid electronics design.

Physical access to the vehicle is enabled through the OBD-Connector, which has been standardized as ISO 15031-3 [8]. The standard specifies the mechanical details of this connector and the pin assignment of some of the connector's 16 pins, while other pins can be assigned at the OEM's discretion. Current generations of vehicles feature a CAN on the standardized pins 6 and 14, which is used as the standard means to communicate with the ECUs in the vehicle via a central gateway. The most employed communication protocol for offboard communication on this CAN is UDS (Unified Diagnostic Services, ISO 14229-3 [3]).

Most proprietary and standardized or even legally required solutions to access the vehicle for emissions check, periodical technical inspection, diagnosis, repair and update are somehow based on this or very similar technology.

To communicate with the ECUs within the vehicle via this access, the same protocol (UDS) is implemented on the employed offboard device (e.g. a diagnostic tester, a Vehicle Communication Interface - VCI). Please note that, while UDS standardizes the protocol services structure, it does not standardize the data content transported with each service. To a limited extent this data has been standardized for specialized use cases (like emissions check), but most data transported via UDS or similar offboard protocols is OEM-proprietary. To simplify communication with ECUs, standardized software stacks and APIs have been defined that enhance the level of abstraction for the development of vehicle communication applications.

One standardized API is the so-called PassThru-Interface (SAE J2534-1/2 [7]), which was originally established to enable the re-programming of emissions-related ECUs within the independent (US) aftermarket but has by now been employed for many offboard communication use cases, including diagnosis.

A more elaborate and feature-rich software stack has been designed and standardized by ASAM and ISO. At the heart of this standard is an XML-based data format that specifies the complete diagnostic, configuration and re-programming capabilities of a vehicle (platform) in a human and machine-readable format, called ODX (Open Diagnostic data eXchange,

standardized as ASAM MCD-2D V2.2.0 alias ISO 22901-1 [4],[6]). This data is interpreted by a software module that again provides a standardized API (called MVCI Server) towards offboard applications that allows for more efficient application development as it abstracts from all protocol, topology and encoding details (ASAM MCD-3D V3.0.0 alias ISO 22900-3 [2]). To provide a sound basis for this MVCI Server, an additional API has been standardized to abstract from any type of hardware that is used to connect to the vehicle's OBD-2 connector and provide a common interface across all possibly supported offboard communication protocols, the D-PDU API (ISO 22900-2). For a schematic overview of this architecture, please refer to Fig. 4.

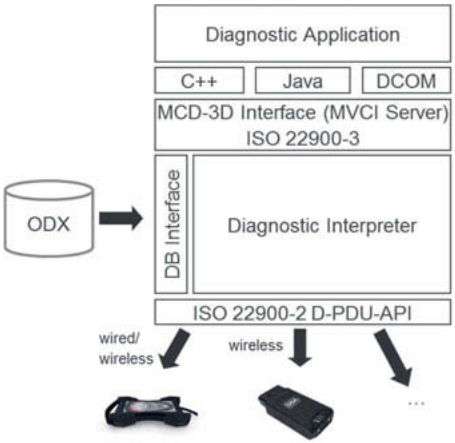


Fig. 4: ASAM MCD software stack

This offboard communication technology has been designed for three major use cases:

1. ECU/vehicle diagnosis, i.e. finding errors in the vehicle electronics and supporting their repair
2. ECU re-programming, i.e. updating the firmware of an ECU
3. ECU configuration, i.e. providing parameter sets to correctly conFig. the ECU for a specific market, for the correct vehicle variant or similar

At the time these APIs and stacks were designed (between 2002 and 2009), ECUs were usually single-core devices, running statically scheduled control functions, sampling sensors and controlling actuators. The individual control function software in these designs is assumed to be correct. Hence, all vehicle diagnosis today focuses on the correct function of the electronics, not the software. ECU self-diagnosis analyzes circuits connected directly to the ECU and stores an event, or DTC (Diagnostic Trouble Code) in the ECU's memory, which can later be read by an offboard device through UDS. An offboard device itself can request sampled sensor values, statically saved values (like IDs) and can trigger actuators externally to verify correct behavior of the electronics. An access to the individual control function (or log or trace files of the executed software is not part of the UDS protocol and thus not part of today's diagnostic capabilities.

For hybrid electronics designs, these offboard capabilities are insufficient. HPCs execute multi-process, multi-threaded, dynamically scheduled software potentially within multiple virtual machines on shared resources. The runtime behavior of an HPC is thus a lot more like a modern virtualized server environment than like a regular ECU. A multi-HPC design has more similarities with a networked server rack than with a cluster of ECUs connected via CAN. The type of failures within these designs will have a lot more in common with complex software systems than with today's vehicles and may include deadlocks, race conditions, access violations, memory leaks etc. By consequence, capabilities to diagnose the software of an HPC will become necessary to support the vehicle's lifecycle. Additionally, the software binary required to update an HPC's software will be much larger in size than the binary of an ECU. This data volume needs to be handled within the lifecycle.

4 Usage Scenarios for future Vehicle Communication Access

The previous two sections have introduced modern HPC-based vehicle electronics designs and the current available and legally supported technology to communicate with the vehicle electronics through an offboard device.

On the one hand, we introduce new vehicles as powerful computing devices with full wireless internet connectivity and advanced digital features. On the other hand, the available access to the vehicle is based on a connector design and protocol technology based in the 1980s that has evolved but not been replaced in past decades. Without a change in technology, usage scenarios for vehicle offboard communication will remain to be proximity use cases, where a worker or technician physically connects an external device to the OBD-2-connector of the vehicle and performs his or her task with the vehicle. However, many new usage scenarios have been developed in recent years that cannot be implemented with this communication paradigm. Besides today's well-established proximity use cases, we introduce onboard and remote use cases that enable a much more advanced vehicle lifecycle management for future vehicle generations.

Onboard use cases are highly relevant to monitor a vehicle's behaviour while the vehicle is in use. The reasons for this type of monitoring can be manifold: (a) a new vehicle model introduced in the market that shall be closely monitored to identify and mitigate any remaining problems with vehicle quality, (b) a dedicated monitor to detect the root cause of a sporadic problem in the vehicle, (c) long-term data collection and analysis of vehicle fleets to derive valuable failure-rate data of components to increase quality in future vehicle platforms, (d) specific monitors for most critical or expensive parts to exchange as a basis for preventive maintenance use cases (with special relevance for commercial vehicles).

The advantage of onboard monitors is that internet connectivity does not need to be available continuously and data can be recorded and pre-processed on-board, saving bandwidth and data costs.

Remote use cases are most likely the most complex and manifold cases for future vehicle lifecycle management. Remote use cases also address vehicle usage scenarios that will gain importance like Ride Hailing, Fleet Operation, Car Sharing etc. In all these scenarios an efficient lifecycle management is business critical. Thus, remote use cases include: (a) remote repair case preparation, where dealership performs remote analysis of vehicle, before it enters the garage, to prepare parts, staff etc. (b) remote trouble shooting, where service technician can remotely assess vehicle health status and recommend next steps to driver or even repair problems through FOTA (c) continuous vehicle fleet monitoring to establish permanent health status, identify service needs and deactivate vehicles in fleet, if necessary (d) remote activation or deactivation of functions in the vehicle, potentially selling new features to driver remotely.

5 New Standard Interface SOVD for future Vehicle Communication

Because many automotive players already identified these new use cases, it is time to provide a new, state-of-the-art access mechanism to the vehicle electronics. This avoids multiple proprietary and incompatible implementations.

The new standard's aim is to define and standardize interfaces and services to cover all above use cases in one flexible architecture, instead of defining interfaces and services separately for each of them. It is also important to define an architecture that allows for the integration of classic diagnostics for ECUs that will be integrated with the HPCs to a hybrid electronics design. The architecture needs to leave enough room for OEM design decisions and needs to cover vehicle platforms that may only have one HPC to those that have multiple.

The general idea of an architecture is, that every HPC can provide a diagnostic service that fully covers the diagnostic capabilities of the domain this HPC controls. This service is based on technology that can be used by

- processes executed onboard the vehicle (onboard use case)
- processes executed on an external test device that is connected to the vehicle via a local network technology (proximity use case)
- processes executed remotely connected to the vehicle via a mobile broadband network or through any internet-based access (remote use case)

In addition, the architecture is intended to leave room for designs where only one HPC is the central diagnostic access for the complete vehicle (even if multiple HPCs are present) and for designs where multiple or even all HPCs within the vehicle provide a separate diagnostic service for their respective domain.

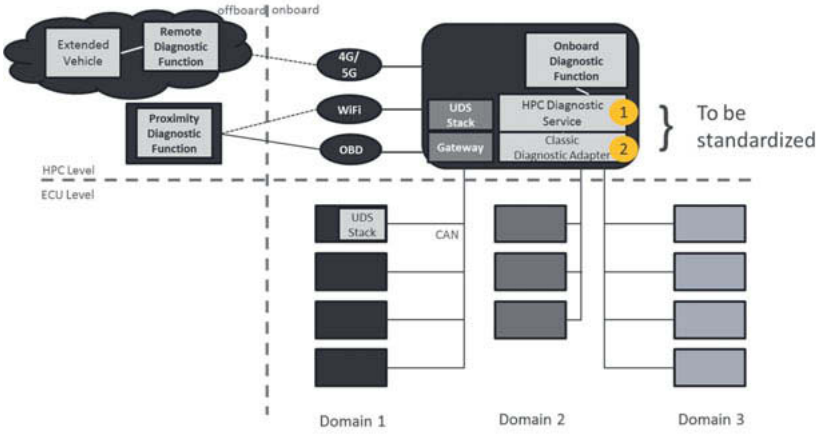


Fig. 5: Two core interfaces

Fig. 5 shows the general idea of the concept within a single-HPC vehicle network. Today's common ECU diagnostics via protocols like UDS are implemented by a Classic Diagnostic Adapter that connects to the common ECUs to integrate their diagnostic capabilities into the overall architecture and make them accessible for the HPC Diagnostic Service, while simultaneously permitting smooth migration into HPC-based architectures. Of course, the HPC could also take over the full functionality of a gateway and thus provide classic diagnostics via OBD-connector for the full vehicle or a common ECU-based gateway could be connected to the OBD connector to achieve the same result. This OBD-based access to the vehicle will have to be maintained for legislative reasons for years to come. Additionally, the HPC itself should also supply standard UDS diagnostic capabilities (ECU-side), so that it can be diagnosed with classic diagnostic means for legislative reasons.

On top of this, the HPC offers an HPC Diagnostic Service, which is implemented with a service-oriented paradigm. This service can uniformly be accessed by Onboard Diagnostic Modules (e.g. performing a monitoring function for a critical component in the vehicle), by a diagnostic tester with Proximity Diagnostic Modules that is connected via WiFi or Ethernet to the vehicle

and is used by a service technician to analyze the vehicle or even by Remote Diagnostic Modules that connect to the vehicle via mobile broadband network. To allow for this wide variety of modules to access the same service, it will be crucial to design authorization and authentication mechanisms that fine-tune the level of access a particular module will have within the HPC Diagnostic Service. For example, it could be prohibited for a Remote Diagnostic Module to perform an actuator test. This authorization mechanism should also take the current vehicle status into account. For example, it could be prohibited for a Remote Diagnostic Module to change the coding of the HPC, while the vehicle is driving. The same use case could be permitted, if the vehicle is not moving.

The new standard shall include API level definitions for the HPC Diagnostic Service (see item 1 in Fig. 5) and the Classic Diagnostic Adapter (see item 2 in Fig. 5).

The design of the HPC Diagnostic Service needs to ensure that communication is possible via non-secure networks with varying latency and varying available bandwidth.

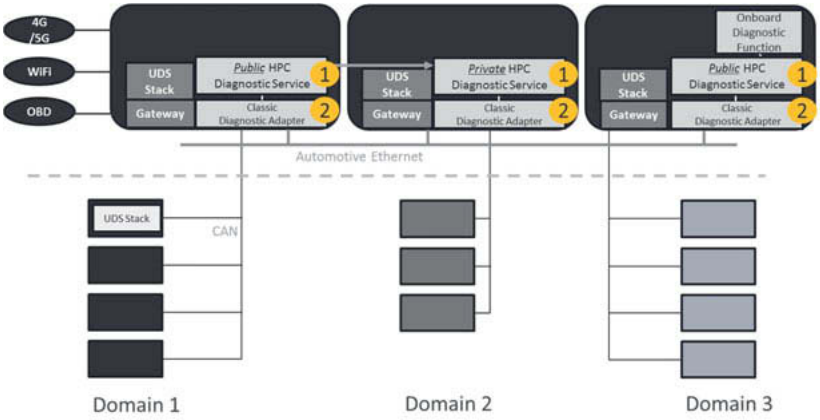


Fig. 6: Multi-HPC-scenario

In a multi-HPC scenario, the architecture may vary depending on the design goals of the OEM. Fig. 6 shows a scenario, where multiple HPCs provide a Diagnostic Service for their domain. Some of these services may be defined as *private* and some as *public*. This terminology is taken from common IT language, where *private* means non-visible and non-accessible for entities outside a defined context. In this case: non-visible and non-accessible for any application apart from other HPC Diagnostic Services within the same vehicle. *Public* means visible and

accessible for entities outside the defined context, which includes applications outside of the vehicle. Please note that *public* does not imply the access by any given entity. Authentication and Authorization procedures apply to fine-tune access levels to the HPC Diagnostic Service. Given sufficient authorization, *public* HPC Diagnostic Services can be accessed by Proximity or Remote Diagnostic Modules. Even Onboard Modules can access services of multiple HPCs to perform their onboard monitoring task. For classic diagnostics to ECUs a regular ECU-based gateway should be employed or one of the HPCs takes over the role of the gateway to provide diagnostics compliant with current legislation.

Depending on OEM's design rules this concept allows for single-point-of-contact HPCs, where one HPC implements the only public Diagnostic Service, while all other HPCs in the vehicle network implement private Diagnostic Services. At the other extreme, a design could permit a public HPC Diagnostic Service within every HPC. And any combination is also feasible. Fig. 6 shows such a mixed approach, where two HPCs publish their own public Diagnostic Service, while one HPC provides a private one.

6 Technical Aspects of SOVD standard

The SOVD standard must be a sufficient means to communicate with the vehicle for onboard, proximity and remote use cases and for hybrid electronics designs. It is also important to ensure a migration path from current technology to SOVD exists. As explained in Section 2, future generations of vehicle electronics platforms will be based on current diagnostic technology to a great extent and add HPC technology as domain controllers. Correspondingly, the current technology for offboard communication must be integrated with the SOVD approach. The following subsections are an excerpt of design guidelines to consider during the development of SOVD.

6.1 Communication Paradigms

Today's standardized diagnostic tester stack based on D-PDU-API and ASAM MCD-3D is a stateful software stack. The core reason is, that classic ECUs maintain state within a diagnostic session through session handling services and security access and this state needs to be reflected within the tester software stack ECU-individually. The current diagnostic tester stack also covers and abstracts from older protocols (like KW2000-on-K-Line) that are stateful in themselves, which means the protocol status must be reflected within the diagnostic stack. The HPC Diagnostic Service is supposed to be standardized as a stateless service. Therefore, it has been agreed that the stateful communication to classic ECUs needs to be completely handled underneath the Classic Diagnostic Adapter Interface (cf. Fig. 7).

For certain use cases it may also be important to allow for stateful functions within the HPC Diagnostic Service. However, such functions shall be subject to higher authorization credentials and thus reserved e.g. for OEM engineering use. In these cases, also subscription services (i.e. continuous supply of data by the vehicle to registered users) could be permitted, while in all main use cases, communication shall be based on a request/response communication paradigm.

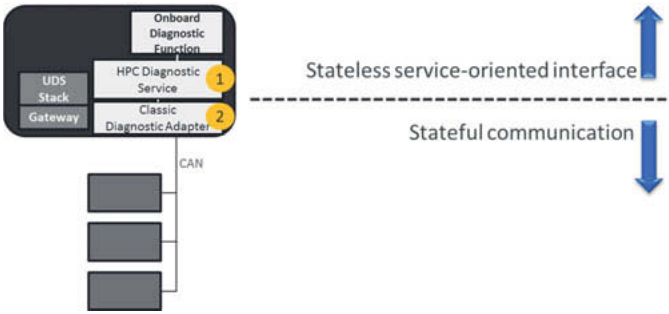


Fig. 7: Stateful vs. stateless communication

The data transferred within the data structures of the Classic Diagnostic Adapter as well as the HPC Diagnostic Service shall be on a physical level and carry clear semantics. That means, for example, that a list of sensor values can be requested from the HPC Diagnostic Service, which will be delivered with key/name, value, precision and unit. These values will not require any further post-processing on the client side of the service.

The service design for both interfaces shall take an efficient remote access into account. Thus, communication should not be too frequent, nor should individual messages be too small. For example, an external device could request a list of sensor values from multiple HPCs and/or ECUs with one single request/response communication roundtrip.

The HPC Diagnostic Service shall also allow to communicate with varying context, with context e.g. ranging from vehicle, to domain, to function, to individual HPC or ECU. The service could allow to request data in mixed contexts within one request (e.g. Battery Voltage from ECU A and ECU B, door lock status from FUNCTION Central Door Locking, Vehicle Speed from VEHICLE) and receive one response with all result data. Naturally, every HPC can only serve the information within its scope of control (ref. Fig. 8).

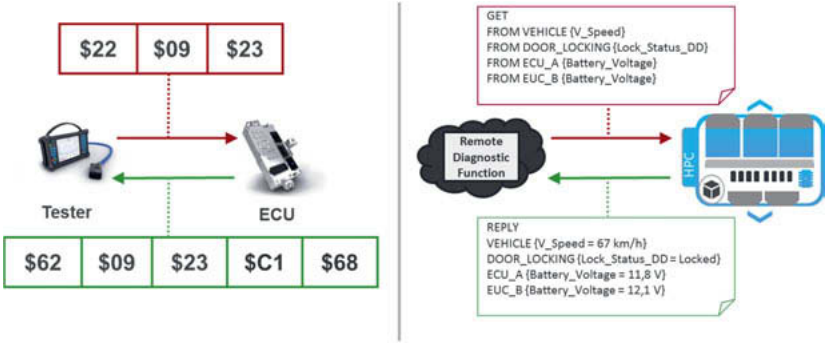


Fig. 8: UDS hexadecimal vs. SOVD service-oriented communication

6.2 Standardize Interfaces not Implementations

The SOVD standard shall only define the two mentioned interfaces, not their respective implementations. This leaves room for multiple different implementations depending on the set of use cases to be implemented, the capabilities of the underlying hardware and network etc.

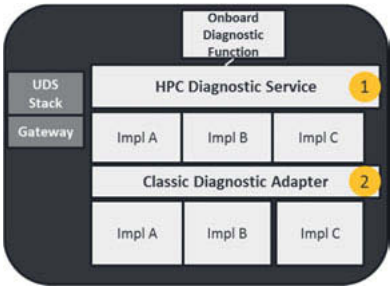


Fig. 9: Interface Standardization

Fig. 9 shows how different implementations of the two interfaces can exist and each combination needs to be possible. To implement the full standard, an HPC Diagnostic Service implementation based on Impl B can for example be based on a Classic Diagnostic Adapter implementation Impl C.

6.3 Compatibility with other (ASAM) standards

The two interfaces shall be completely independent of other standards in the diagnostic or onboard software domain, e.g. it shall be independent of VCI interface standards like SAE J2534, ISO 22900-2 (D-PDU-API) or RP1210. It shall also be independent of current diagnostic interface standards like ISO 22900-3 (MVCI Server alias MCD-3D). It shall also abstract from all existing diagnostic protocols like ISO 14229-3 (UDS). At the same time, however, the implementation of the two services shall be possible based on these standards and nothing in their standardization shall prohibit the mapping onto these existing standards.

On the other hand, the implementation of the two services shall also be reasonably possible without the basis of these standards.
Please refer to Fig. 10 for an illustration of three example implementations of the Classic Diagnostic Adapter, one based on the ASAM MCD-3D interface with an underlying D-PDU-API, one directly based on an SAE J2534 interface and one based on a completely proprietary implementation.

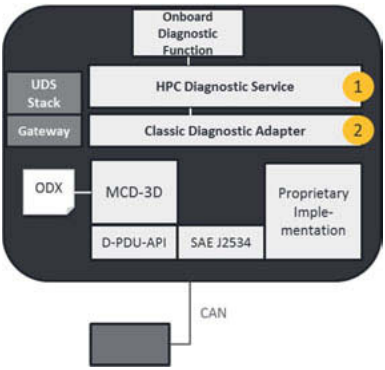


Fig. 10: Application of existing standards

6.4 No invention of base technology

The complete standard shall be based on existing technology. It shall not define a new remote service technology or protocol or data format description for data to be transferred via the standardized services. It is a core task of the technical workgroup to select the best-in-class technology that fulfills the use cases as specified.

6.5 Diagnosing Software

HPCs are powerful computing devices that are regularly equipped with Hypervisors and execute one to multiple guest operating systems running in separate virtual machines. Software implemented for HPCs is usually based on operating systems like Linux, QNX, AUTOSAR Adaptive or Android. It is commonly multi-process, multi-threaded software that is dynamically scheduled and shares resources with other software within the same guest virtual machine. It can therefore be anticipated that HPC software will be faced with all potential problems of server or desktop-based software, like Deadlocks, Race Conditions, Load and scalability issues, memory leaks etc. Therefore, the HPC Diagnostic Service requires capabilities to diagnose the behavior of software functions. This shall include functions to access data for analysis of the following:

- Threading, Deadlocks, Race Conditions
- Watchdog, Watchdog Status, Watchdog Activity (also past activity)
- Performance, Load, Memory Footprint, Network Load, Network Latency
- Log Files, Trace Files, Post-mortem Files (crash files)

It is not assumed that these functions, or the data supplied by these functions is analyzed by a worker along the production line or a service technician in the dealership. Assumption is, that

this data will be analyzed by centralized service, which includes experts on HPC software and is supported by efficient tools.

7 Conclusion

It is time to standardize a new technology to access connected vehicles with hybrid electronics designs. Current technology is focussed on electronics diagnosis and assumes perfect software within the vehicle. It also solely addresses proximity use cases. Within this paper we have shown an approach to a new standard that allows for the integration of current technology for offboard vehicle access, while at the same time modern communication technology is employed to enable onboard and remote use cases that will become highly relevant for future vehicle generations. Standardizing this vehicle access is important to ensure common and controlled access technology across all vehicles.

The standardization of SOVD has already been initiated and receives wide support within the industry.

8 References

[1] ISO 22900-2:2017 Road vehicles -- Modular vehicle communication interface (MVCI) -- Part 2: Diagnostic protocol data unit (D-PDU API)

[2] ISO 22900-3:2012 Road vehicles -- Modular vehicle communication interface (MVCI) -- Part 3: Diagnostic server application programming interface (D-Server API)

[3] ISO 14229-3:2012 Road vehicles -- Unified diagnostic services (UDS) -- Part 3: Unified diagnostic services on CAN implementation (UDSonCAN)

[4] ISO 22901-1: Road vehicles -- Open diagnostic data exchange (ODX) -- Part 1: Data model specification

[5] Zimmermann, W.; Schmidgall; R: Bussysteme in der Fahrzeugtechnik Protokolle, Standards und Softwarearchitektur. Wiesbaden: Springer Fachmedien 2014

[6] Schleicher, A.: Impacts and Benefits of ODX in the Diagnostic Tool Chain, SAE World Congress 2007, also in: Vehicle Diagnostics, 2007-SP-2137, SAE 2007 Transactions Journal of Passenger Cars: Electronic and Electrical Systems V116-7

[7] SAE J2534-1: Recommended Practice for Pass-Thru Vehicle Programming, 2002

[8] ISO 15031-3:2016 Road vehicles -- Communication between vehicle and external equipment for emissions-related diagnostics -- Part 3: Diagnostic connector and related electrical circuits: Specification and use

[9] ISO 13400-2:2012: Road vehicles -- Communication between vehicle and external equipment for emissions-related diagnostics -- Part 3: Diagnostic connector and related electrical circuits: Specification and use

How to Improve Automotive Testing in an Agile Development Process

A Review of Popular Testing Methods and Overview of Advanced Automated User Interface Testing

David Robinson, Altia Europe GmbH, Nuremberg

Abstract

With the emergence of fully digital instrument clusters, head-up displays, radios, passenger displays and full integrated cockpits, increasingly complex production graphics software and production code is on the way. Both software and code will be arriving with rapidly expanding features and depth.

Since development of these displays passes from artists to engineering down to embedded hardware experts, there are countless ways for errors to be introduced. Agile development is a sure method to help development teams confirm delivery of the highest quality software for these automotive applications. The fast iterations that result from an agile development practice can quickly derail automotive test teams. Even in a non-agile development project, about half of the total development effort is consumed by graphical user interface testing.

How can automotive OEMs and Tier 1s ensure that the user interfaces running on their production hardware is a perfect match to their developer's intent design? With so many screens to review and compare, it is physically impossible to manually run through all those screens and features, using only the human eye to spot minute differences and achieve comprehensive testing. Such testing will be plagued by human error, false positives and severe schedule overruns.

In this paper, Altia will review popular methods for automotive user interface testing to help OEMs and Tier 1s to meet their demands. Manual systems, camera-based test systems and more technically advanced approaches will be considered. Each solution offers tremendous potential as well as problematic pitfalls. Additionally, Altia will present a state-of-the-art method that compares baseline user interface designs with the same user interface running on real-

world embedded hardware. This method provides a pixel-by-pixel comparison of the model in a fraction of the time of other test methods.

Designers, Engineers and the Need for Testing

In graphical user interface (GUI) design, two very different groups must collaborate. Designers create. Graphical layouts, complex animations and user interaction flows are their specialties. Engineers must understand data, APIs, state machines and requirements specifications – as well as the limitations of the embedded systems that will ultimately run the application. Both groups have specific and important contributions for embedded GUI development for production devices. As technology improves, GUI complexity is also increasing at a fast rate. Designers and Engineers must team up to manage processing power, availability of touch screens and user expectations.

Within a GUI development project – from design through to implementation – Designers and Engineers need to work together to confirm that the Designer's intent matches the end GUI on hardware. With such disparate skill sets and areas of focus, it can be challenging for these two teams to bridge the gap between their disciplines. Short agile development cycles help keep these teams communicating and on the same page. Testing of their rapidly released user interfaces becomes an extremely important step for teams to take to uncover the errors that arise during development sprints as part of an embedded GUI implementation. For example, the color and style of the product must be checked so that they remain consistent with the specification. Dynamically driven content must be validated for accuracy. User interaction should be reviewed to confirm it remains consistent with the flow diagram. GUI animations must be tested to ensure they match the Designer's documented vision. Font types and sizes should be analysed so that they are implemented properly throughout the application, including fonts for multiple languages. Additionally, adding new, more complex technologies like 3D to a GUI drives the requirement for more rigorous testing.

What Kinds of Errors Will Testing Uncover?

When conducting GUI testing, there are several areas where issues can be found. Graphics rendering errors can happen between software and hardware pipelines. Font engines can act differently between Adobe® Photoshop® in Windows and whatever font rendering technique is used on the embedded hardware. If not ported properly, installed fonts can be dissimilar and features of those fonts – hinting, kerning and shaping (particularly for complex languages like Arabic and Thai) – may not display properly. Differences in system configurations between

development and final hardware may also occur as a result of issues with software configuration management, like an operating system or compiler configuration. The GUI model might also attempt to leverage features that the embedded hardware does not fully support, like the number or type of layers or the color format or configuration.

The ability to find the source of these failures is critical and can be time consuming. GUI development teams implement a variety of testing methods for their GUIs.

Manual Testing

Manual Testing is a very traditional way of testing user interfaces. Imagine a real person sitting at a screen trying to detect differences manually. This is the simplest and most common test solution; it includes no automation. Testing is done by a human interacting with a touchscreen and documenting their visual comparison of a Designer's intent GUI and the same GUI on hardware.

The manual testing process is a natural evolution from people testing code to people testing user interfaces. Manual testing can be a good test of human factors, since humans are inherently involved in the system. Testing the human factors side of the usability of a GUI that is difficult to characterize with software.

There are downfalls to manual test systems. Tests are typically simple and often constrained to the designed use case. Successful manual test programs require a large number of human resources for comprehensive results, which is a costly endeavor. This test method is prone to human error. Manual testing is very subjective – and it is often impossible for humans to detect all error cases. Testers can get tired and their senses can become numb to repetitive test cases. Results from one tester to the next can be inconsistent, as well. Color blindness, for example, will impact how a tester perceives the colors on the GUI during the test. Test interaction can also be inaccurate due to variations in human touch based on variables like fingertip size, pressure and conductivity. The benefit of human factors testing with this method is also hard to capture since manual tests typically are not set up to record such data.

An obvious way to improve upon manual testing is to replace the human factor with a camera interface and software.

Camera-Based Test Systems

In a camera-based test system, a camera is used to monitor the GUI while software is used to compare images. Inputs for camera-based tests can include robotic touch or a software API. For this type of test method, consistent lighting and screen tuning – contrast, brightness, color – is critical.

There are real benefits for this type of test, particularly over a manual test method. The capture aspect of this type of test can be adapted to test many types of GUIs. If test conditions are controlled, this type of test is moderately repeatable. Camera-based testing eliminates the issues of tester fatigue and manual variation. These types of tests can also be automated and quite easily reproduced.

This test method comes with a significant list of challenges, however. Test environment variations – lighting, screen calibration and camera calibration -- must be tightly controlled or the test will deliver poor or inconsistent results. Cameras will often have difficulty resolving colors from a display since it has fine resolution. A camera may see individual RGB pixels (red, green, blue) instead of the resulting blended color that a human eye or software will see. With this method, color matching is extremely difficult and complex. Because of the camera resolution, text is also difficult to acquire and test with this method.

Camera-based testing will often yield false positives which results in wasted time, plus additional investigation and test cycles. Compared to other test methods, this type of test system is rather rudimentary as it is typically only able to provide simple pass or fail results -- such as the absence or presence of an entire graphics object. Camera-based tests are costly in terms of setup time. Testers must take significant time to teach the machine, especially when it comes to robotic inputs like where to press buttons on the GUI. Those robot-based interactions can also block the camera capture during navigation. When robotic inputs require that action happens between the screen and the camera, that robotic interaction interferes with the camera imaging – resulting in missed animations.

Frame Buffer Capture Method

A more sophisticated solution for testing GUIs involves replacing the camera with software. This test method consists of interrogating and copying video memory that is sent to the display. It requires custom code to extract the data for analysis – both a custom frame buffer capture API and a communication path to store and extra data from the embedded system into another

device for testing and comparison. Input for this time of testing can occur manually or via robot or API.

The frame buffer method captures a fine level of detail – what should actually be going to the screen. Additionally, algorithms for data comparison can be tuned for better results.

One downfall to this test method is that it is complicated to run. In order to test this way, the graphics application is paused so that the frame buffer can be copied. Without pausing, the graphics application might try to write to the same frame buffer testers are attempting to capture. This may lead to graphical artifacts like tearing. Therefore, testers must sync the pausing of the graphics application, capture the frame buffer and then let it run. Tuning the test to pause the graphics application demands a custom build or software hooks for the test to be successful, which means that that the GUI testing is being performed on different software. This variation to the software could cause false results or bad testing.

An additional downfall to the frame buffer capture method is that the test does not detect what is actually displayed on the screen. What is in the frame buffer must go through the display controller. Low level driver errors, configuration issues, bit-flipping or color data truncation issues are possible, so what a tester gets in memory is different from what is shown on the GUI. This type of testing method is typically a very data heavy process to transport from the micro to the host PC. It is very difficult to run on low-end processors. Not only does this interfere with the execution of the application itself, but this testing also requires additional hardware resources outside of the application. Furthermore, on a low-end processer, there is not enough RAM to store a copy of the buffer to be transmitted externally and stored.

Altia Advanced Automated Testing

As an evolution of these previous methods, Altia has responded to the market and converged on a robust solution path by developing a state-of-the-art setup, process and system for advanced GUI testing.

The Advanced Automated Testing method offered by Altia compares reference GUI designs with the identical GUI running on production embedded hardware. A custom hardware interface captures the raw data stream and intercepts it as it goes to the display and reconstructs the frame buffer dynamically. It then captures all the data and goes through an algorithm to

convert that data stream back into the frame buffer that is going to the display. Custom software performs a pixel-by-pixel analysis of the model based upon the reconstructed frame buffer.

Because this test method is fully automated, it can be completed in a fraction of the time of many other test methods. Such fast test cycles enable GUI development teams to test often and test with a great deal of variation. This test method also has no impact on the software running on the device -- only the data inputs to make the GUI run on the screen are required. Thus, this is the only GUI test method that guarantees the data sent to the display is used for testing. The Altia automated test method allows comparison of reference images from the designer's intent to the identical GUI running on real hardware. It is an extremely fast way to achieve a pixel-perfect analysis of the current graphical state of the system. It also has zero impact on the software; the final graphics application can be tested with absolutely no software hooks.

Altia Automated Testing Details

The Altia test method begins with a new or existing hardware capture interface – this can be an LVDS, HDMI or parallel RGB interface. Those tests are run, and the data stream is captured. The test captures real data and recreates the image frame buffer from the display. Then an automatic image comparison tool is run.

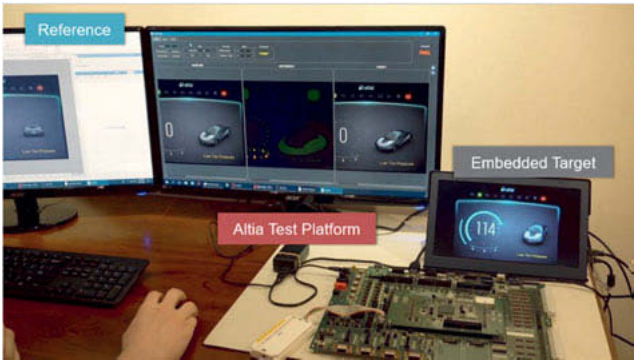


Fig. 3: An example of an Altia Automated Test setup. A custom hardware-software capture rig (off screen) compares the reference design GUI (left) to the GUI on the embedded target (right). Differences found during test are displayed on the screen in the center of this setup.

The test will capture synchronized screenshots of both the designer's intent reference GUI and the system-under-test. Graphical comparisons and analysis of each screenshot will be performed before finalized test results are delivered. Software test parameters are completely tuneable and can be customized on a case-by-case basis. The system-under-test is not impacted in any way during the testing. This ensures that test conditions accurately reflect conditions in the field without artificially inducing any additional test factors. Altia Automated Testing utilizes an entire suite of sophisticated reports and automated GUI analysis tools used to investigate and root-cause any failures.

The Altia Automated Test process begins with a GUI development model. The GUI development team works with Altia to identify and create a test plan, which can evolve over time. From that test plan, testers identify and extract reference frames from the designer's intent development model. Embedded code for the production-intent GUI is run on hardware while a custom hardware and software capture rig runs in parallel to extract the GUI-on-hardware results. Those results are collated and compared; visual difference heat maps are provided to show variations between the designer's intention and production intent GUIs.

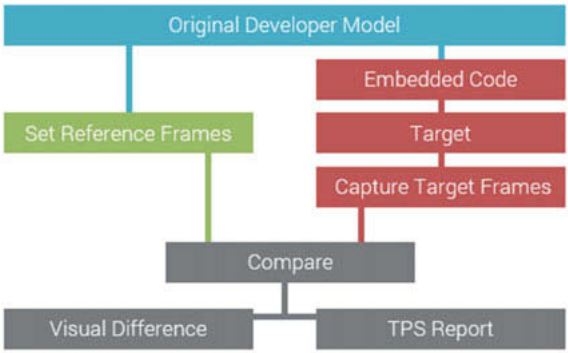


Fig. 2: The Altia Automated Test begins identifying and extracting reference frames from a designer's intent model. In parallel, embedded code is run on the target hardware and a custom hardware and software capture rig extracts the results. The final report includes a visual difference map with all errors found during test.

The Altia test method will catch not only obvious differences – like missing glyphs – between the reference GUI and production intent GUI, but also very subtle differences in the shadows of a graphic or an improperly drawn font. The heat maps provided demonstrate the severity of a variance. Green represents a minor difference; yellow represents a more significant difference; red represents a severe difference or shift.

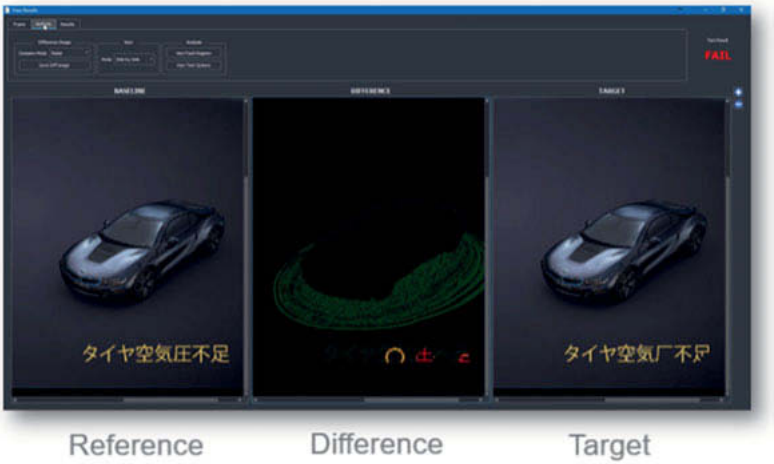


Fig. 3: In an Altia Automated Test, the reference design is compared with the design on target. Differences are shown in a heat map format – green for minor differences, yellow for more significant differences and red for severe differences.

Considerations for Altia Automated Test

With the Altia Automated Test method, initial setup time can be longer than a manual test, for example. This time can be made up during testing since the process is fully automated. This testing technology catches visual discrepancies that would be missed by other methods. For example, this test will find font rendering issues due to variations in embedded font engines and font shapers, as well as gradients and shadows due to compression and rendering algorithms. Test tolerances can be tuned to accommodate such differences.

Conclusion

GUI testing allows Designers and Engineers to effectively evaluate their user interface during agile development and perfect their GUIs before products go to production. For a process that creates many minimally viable products, automated testing is the only way for these teams and this process to stay on track. Automated testing enables accurate and efficient testing early and often – which provides for the best possible GUI for production. Altia's Automated Test process provides a proven example for how agile teams can achieve clear, actionable and accurate GUI testing results.

800V Fast Charging is Reality

From the Vision in 2015 to Reality in 2019

Dipl.-Ing. **Otmar Bitsche**, Dr. Ing. h.c. F. Porsche AG, Weissach

Abstract

During ELIVE 2015 we presented the Mission-E concept study, our vision of a fully electrical vehicle. The target was to build a vehicle with an electrical range of 500 km (NEDC) and a fast charging capability of 400 km in 20 min (NEDC). The 800 V powertrain was identified as the key technology for a high performance sports car with fast charging capability. In this paper I would like to speak about how the vision became a reality.

1. Why the 800V Technology is necessary

In order to be ready for a fully electrical vehicle, we had to ensure that the customer expectations could be achieved.

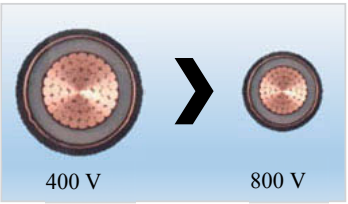


Fig. 1: Cable Cross Section

One of the major needs for the customer is the ability to drive long distances without limiting charging times. From the vehicle perspective the short charging time requirement leads to a higher charging power demand. In order to support a charging power above 150 kW and convenient usage of the charging connection at the same time, it was necessary to increase the charging voltage.

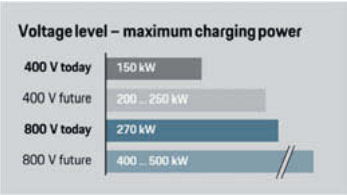


Fig. 2: Charging Power

The need to build a high performance sports car with repeatable acceleration led to a power requirement of more than 600 hp (440kW). To solve the trade-off between cable cross-section, heat-loss and weight of the high voltage system, it was also necessary to increase the voltage level.

2. How we realized the 800V powertrain

High-Voltage - System

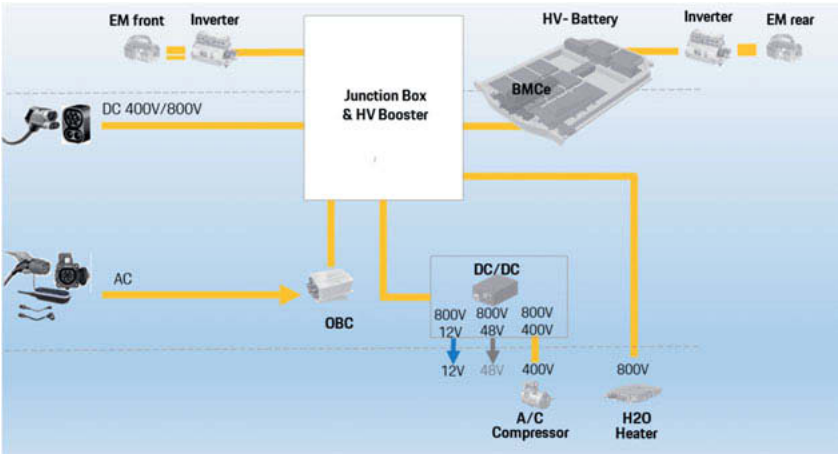


Fig. 3: HV- System

In order to support the higher voltage level, the power electronics for the high voltage components had to be modified, for example the DC/AC converter for the electrical drive engines and the electronics of the H_2O heater.

The 12V and the 48V System is powered by a combined DC/DC converter from the 800V battery. The 48 V voltage level is mainly used for the active chassis system.

In order to ensure compatibility with existing charging infrastructure, it was necessary to develop the HV-Booster. It enables the Taycan to charge the battery from a lower voltage level by increasing the voltage level up to 800V.

Electrical machines and DC/AC converter

Fig. 4: Electrical machine with DC/AC

The propulsion is delivered by two permanent magnet synchronous motors (PMSM) with integrated power electronics. An integrated transmission is used to combine a high torque during acceleration with a high terminal velocity.
The optimized water cooling enables the PMSM to deliver a reproducible maximum performance. The test confirmed 26 accelerations (0-200 km/h) with no change in performance.

High-Voltage – Battery

The driving performance on a test track also depends on a low center of gravity. This ensures a higher apex speed (lateral acceleration) and better handling.

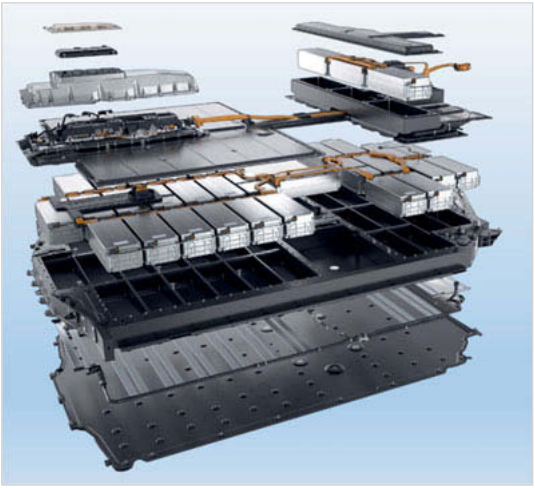


Fig. 5: HV - Battery System

Therefore, the battery was designed as an underfloor system.
The battery system consists of 33 modules with 12 cells each.
Electrically the 396 cells are connected in a 2P/198S pattern.
The weight of the battery system is around 630 kg.
In order to ensure a comfortable seat position in the second row, the battery modules had to be rearranged to build the so called “Fussgarage” to ensure more legroom.
The battery cells and the cooling system have been optimized to support high performance driving and fast charging.

3. Charging

From the customer perspective charging should be as easy as fueling.

However there are many different hardware standards for the charging connection as well as software protocols for the communication between the charger and the vehicle.

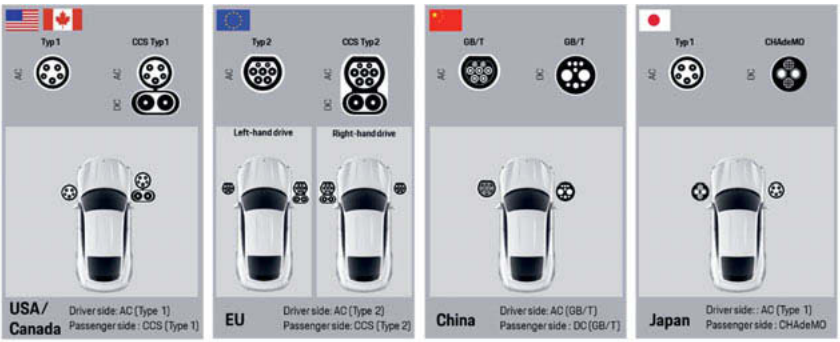


Fig. 6: Charging Interfaces

Charging- Infrastructure

The key enabler for long distance mobility is the charging station coverage along the highways. To ensure the area-wide infrastructure IONITY and the “electrify America project” started to rollout fast charging stations in Europe and the U.S.

IONITY

The European Highway-Charging Network enables long-distance travelling.

By 2020 there will be 400 fast-charging stations installed along the European transport corridors (6 charging poles per station). Every charging pole will support the fast-charging technology with 800V.

Electrify America:

Fig. 7: Charging Infrastructure (Electrify America)

In order to ensure a good charger-coverage the highway stations will be installed every 110 km on average and no more than 190 km apart.

In the first cycle a network of ~4.700+ non-proprietary electric chargers will be installed.

The Highway stations will be equipped with chargers capable of delivering maximum power levels from 150 kW to 350 kW.

Porsche Charging Service

Fig. 8: The Porsche Charging Service offers access to more than 200,000 charging points

Fig. 9: Intelligent Charging Functions – Operating via App

The Porsche Charging Service offers: a single access to charging facilities from different providers, billing, station availability, charging power and further information shown either in the navigation system or via our customer app.

4. What is the result?

The name “Taycan”:can be roughly translated as “lively young horse”, referencing the imagery at the heart of the Porsche crest, which has featured a leaping steed since 1952.

Fig. 10: The Taycan

The Taycan demonstrates that it is possible to build a fully electric sports car with no limitations. The constant torque and absence of engine noise lead to a new driver experience that cannot be described only by the performance characteristics.

Table 1: Taycan characteristics ()

Top Speed	260 km/h
Acceleration 0 to 100km/h	2,8 s
Power	up to 460 kW
Torque (Launch Control)	1050 Nm
Range (WLTP)	381-450

Perspective

By 2025 Porsche will have invested more than 6 billion euros into E-Mobility.

By this time more than 50% off all new Porsches produced will be electrically powered.

Addressing the challenges in designing fail-operational architectures for autonomous driving platforms

Tailoring fail-operational systems based on production experience in the aerospace industry for the automotive use cases

Dr. **Stefan Poledna**, TTTech Auto AG, Vienna, Austria

Abstract

The automotive industry has set out to master automated driving use cases, which require a fail-operational performance. Based on concepts which have previously been developed for industries like aerospace and space, this trend is now moving into the automotive industry. With one pronounced difference – the stricter constraints on the system costs as they are aimed at mass market production. This calls for a redefinition of the key practices and an implementation with the aim on cost reduction.

The automotive industry is evolving at high speed

Driverless cars are the biggest innovation and evolutionary leap in the automotive industry since the invention of the car. As an underlying paradigm shift within the E/E vehicle architecture stands the transition from dedicated hardware appliances used to execute a specific function, towards a software-defined, highly modular, converged, service-oriented platform. This centralized architecture, built on generic high-performance computing hardware, can deliver highly automated driving use cases and enables software function re-use across different car models, variants and lines. The suggested architecture helps OEMs to overcome complexity and safety challenges of driverless software development and maximize the value of investments, while making their business and their technology agile. The number of required automated driving software functions is skyrocketing, as well as the complexity created by this paradigm shift, which affects the required underlying hardware, the operating systems and the required communication. At the same time, this brings freedom to the OEM or Tier 1 companies to select the best-in-class solution elements coming from third parties and to then integrate them on one converged platform with high transparency, quite contrary to the approach of sourcing a complete closed-box system from one supplier. All these elements need to deliver

upon the high automotive-grade requirements and perform safety-critical and real-time functions in a fail-operational manner to safely perform the task of autonomous driving and to gain public acceptance.

Challenges brought on by the industry evolution

Time-to-market

Introduction of software-defined functions is unlocking faster innovation cycles, but it also creates tremendous levels of complexity resulting from integration, validation and testing of the software-defined functions. Closed-loop and manually integrated systems cannot be easily deployed across multiple SoCs. Even small changes require re-testing and re-validation of the entire environment, reducing agility and making the prospect of introducing new levels of automation extremely complex, time consuming and prohibitively expensive.

Safety

Taking a manual approach to building and orchestrating driverless software potentially limits a company's ability to meet essential safety requirements. This is because there is often no way to ensure that all essential safety-related tasks performed by diverse applications that run on different SoCs are scheduled appropriately, based on their mission-criticality and on a global basis. This can negatively impact the ability of the system to react in real-time to ensure vehicle and driver safety.

Fail-operational requirements

Starting from level 3 automated driving, solutions require a fail-operational design, which means that the E/E system must continue operation under all circumstances - even in the presence of a fault. In contrast to traditional fail-silent solutions in the automotive industry, where the driver took over in case of a fault, it is now necessary to build solutions based on fail-operational concepts previously met in the aerospace and space industry. This calls for a redefinition of key practices and an implementation without having the full redundancy of all systems, while still delivering an efficient and safe solution at reduced costs.

High complexity of the end-to-end solution

Increased complexity of the software functions used in vehicles is redefining the solution architecture design, as well as how the automotive industry collaborates. New companies such

as application providers enter the industry as disruptors, increasing the amount of parties involved in such programs. High levels of complexity are brought about by numerous applications with different ASIL levels and a highly heterogenous environment.

Scalability & versatility

Closed loop and manually integrated systems limit OEMs' ability to re-use software for different vehicle lines and their variants, spanning from basic models to luxury cars, potentially requiring them to 're-invent the wheel' for delivering a sound solution for different vehicles. This effectively multiplies the investments needed to deliver specific levels of automation targets, while also reducing agility.

Real-time processing

Networks for real-time systems have stringent end-to-end latency requirements. They have to ensure that extremely complex schedules work correctly in practice and that all required resources are available, so they can be executed in an automated way with constant availability. Real-time processing capability is one of the most prevailing challenges, since it requires many different functions from different layers to perform in a real-time manner. The compliance with automotive standards alone is not enough to ensure this key capability is covered across a global, centralized system.

Guarantee of service

It is challenging to ensure guarantee of service for such highly granular, complex and heterogenous systems in order to deliver a reliable and safe automated driving experience. For use cases such as sensor fusion layer or emergency braking, it is critical that data from specific applications arrives at the right target locations at the right time and with always predictable end-to-end latency, no matter of the system load.

General automated driving system architecture

In order to enable automated driving, the vehicle must sense its environment, process the inputs and actuate the behavior of the vehicle in a real-time manner. A set of diverse sensory inputs from sensors such as radars, cameras, LIDARs are processed to identify elementary pieces of information about the vehicle surroundings, while using high definition maps as a base and a cloud connection to further expand the perception about the surrounding. This information then travels through the sensor pre-processing and the perception layer and is further fed into a central fusion layer to compile a complete and consistent representation of

the surrounding of the car and the trajectories of all objects in real-time. The clusters of applications from path planning, controlling and actuation of the vehicle functions are building upon these inputs. They devise the driving strategy and implement the navigation and control algorithms to drive the vehicle by controlling steering, braking and the drive train. Besides the automated driving software functions, a wide array of applications from powertrain and chassis domain control to infotainment and beyond will also be hosted on the same platform. Some of the applications' software functions require the highest safety measures, while for other functions with lower safety requirements the execution on performance hosts with higher computing capabilities is appropriate. The illustration below depicts the reference automated driving system architecture with the high complexity of all solution elements that are required to deliver upon highly automated driving use cases.

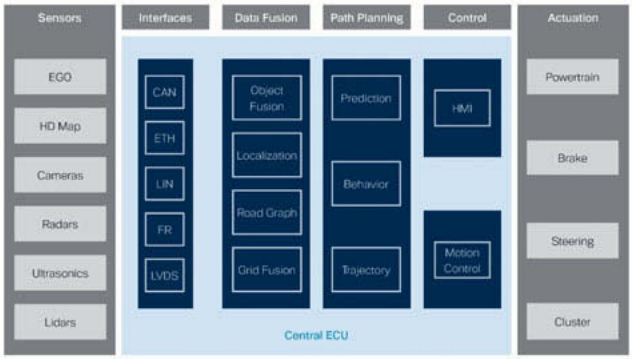


Fig. 1: A reference automated driving system architecture of a driving function with data processing from left (sensor inputs) to right (actuation outputs)

End-to-end safety approach

The safety topic is an often underrated, yet highly critical aspect of automated driving. It is absolutely necessary in order to gain public acceptance. More than 20 years of profound experience and know-how gained from the aerospace and space industry have allowed us to make it one of our core competencies. With the transfer of this safety know-how into the automotive development, we have brought neutral safety experts on board for the automated driving projects and have developed solutions for the automotive use cases. Now multiple solutions tackling the industry's most prevailing challenges are available, like the ASIL D safety software platform MotionWise, which can evolve all the way to the suggested In-Car Compute

Platform (ICCP) with the next generation electronics, supported by the ongoing development of the Safety Co-Pilot application. The combination of MotionWise, ICCP and the Safety Co-Pilot offer an end-to-end and future-proof safety approach.

Suggested fail-operational architecture design

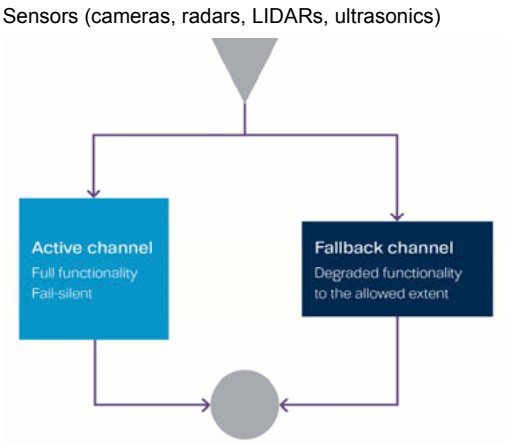


Fig. 2: Fail-operational architecture consisting of two independent channels. Here, an active channel usually executes the function and a fallback channel takes over in case of a severe failure of the active channel.

In the automated driving domain, the inability to execute a function needs to be considered as a dangerous failure. This is specifically applicable to the software-defined components without any mechanical fallback, which often calls for a complete redesign of the fail-silent solution architectures, in order to meet the fail-operational requirements and ensure reliable and safe behavior. In case of the failure of the function, the vehicle at least needs to be brought into a state of acceptable risk, which will not typically be reached by a blindly performed braking maneuver, which would lead to a vehicle standing still on the current lane. This cannot be considered safe in all driving situations, for example if the car is stopped on the far-left lane of a highway or on a railroad crossing. This suggests that at least a degraded version of the driving function has to be executed, in order to bring the vehicle to the next safe place, such

as a parking space or at least the emergency lane. The function can be degraded in the areas of comfort, speed of execution of that function, as well as the cooperativeness with other road users. However, safety-critical features must be available at all times. Fig. 2 shows a simple representation of a fail-operational architecture. Out of the many possible architecture designs, this scheme of a fully performant primary active channel with full functionality and a smaller, simpler fallback channel with degraded functionality seems to be most suited for the cost-sensitive automotive industry, since it is intended for the mass market. The definition of how much the function is allowed to degrade in case of different faults will be made based on respective use cases of the observed automated driving tasks. For example, if a driver can take over at any time, the system might degrade to a level where only safety-critical features need to be guaranteed. A shuttle service, on the other hand, might need a fallback that is capable of bringing the passenger to the desired destination and bringing the vehicle to the repair shop afterwards.

Special care must be taken when decomposing the ASIL of the two sub-systems - the active and the fallback channel. This is because the classical fail-silent system usually does not come with high availability requirements and capabilities of the overall system, since the underlying QM sub-systems protected by ASIL safety mechanisms only check for the correctness of the service, but do not ensure the highest possible availability of the services. When a symmetrical reduction of the ASIL assigned to the two redundant channels is desired, then the active channel must also inherit availability requirements. This calls for a mandatory re-evaluation of existing fail-silent systems, which might also require an in-depth redesign of core features. For example, an end-to-end communication protection is often used to achieve a high ASIL for data correctness on QM communication channels. This often used safety mechanism is not sufficient to also guarantee the availability of the communication with the required ASIL quality. As we move towards highly automated driving use cases, these new availability requirements reach into the core of already existing solutions of operating systems, microcontrollers, communication busses, etc. A more detailed proposed fail-operational architecture is depicted in Fig 3.

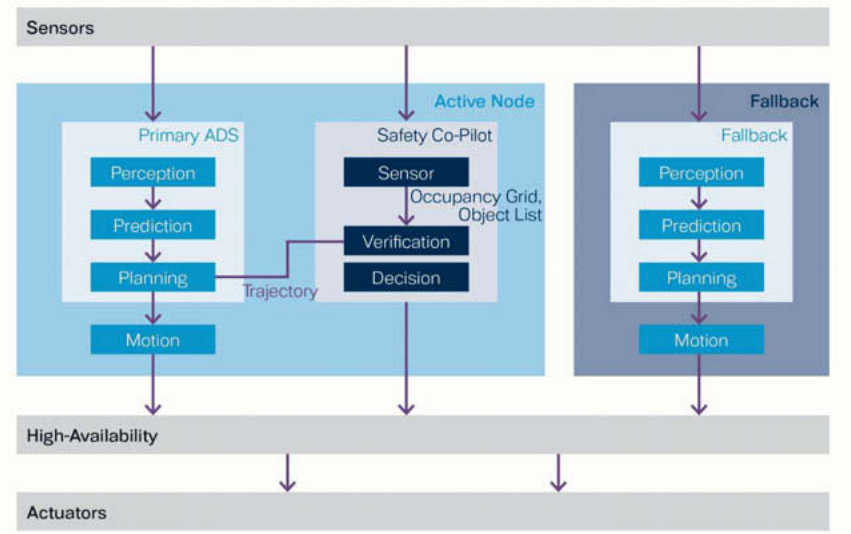


Fig. 3: Suggested fail-operational architecture in more detail. Correct behavior of the active channel is achieved by checking the output of the primary automated driving system (ADS), using an independent checker called Safety Co-Pilot (SCP).

A complex primary automated driving system (ADS) performs the automated driving task safely and cooperatively with other road users. An independent checker function called Safety Co-Pilot (SCP) checks the safety of the primary channel's output. Whenever it is unsafe, the primary channel will be shut down and the fallback channel's trajectory will be followed. The fallback channel permanently produces a conservative trajectory that minimizes risk and, depending on the use case of the automated driving function, brings the vehicle to a safe situation, such as the closest parking space. In any case, a suitable platform is needed to host the driving functions of different safety criticality, degraded or not.

Addressing the challenges with the safety software platform MotionWise

MotionWise is a series-proven, fully integrated, automotive-grade safety software platform suitable for automated driving use cases level 2 to 5, advanced driver assistance systems, chassis and electric or hybrid drivetrain control systems, infotainment and beyond. It acts as the safety

mastermind within a vehicle architecture, handling the high complexity of electrical and electronic systems (E/E systems).
With its platform-centric approach, MotionWise helps organizations move away from slow, costly and complex ‘closed loop’ systems with inefficient ‘self-integrated’ practices. As a result, this speeds up the time-to-market for new automated driving functionality. It is designed to ensure the seamless shift towards a highly modular, service-oriented architecture. Using the MotionWise platform, OEMs can integrate, test, validate and schedule any number of components and applications, helping them to reduce development, testing and validation complexity and to ensure that essential safety and mission-criticality requirements can be met in both single and multi-SoC environments.

Key capabilities of MotionWise:

Real-time orchestration
The MotionWise time-aware platform architecture uses innovative global scheduling technologies and algorithms, together with design tools to support smooth application and communications network traffic scheduling in order to fulfil end-to-end real-time guarantees. Built on deterministic technologies, MotionWise delivers always available services with guaranteed latency across a highly heterogenous environment – regardless of the system load. Real-time processing with deterministic scheduling uses sophisticated computation chains, allowing safe execution and safe behavior of the complete system in an automated way.

Open integration platform
MotionWise provides a future-ready software architecture that abstracts hardware and operating systems. This provides the modularity, portability and standardized interfaces required to deploy, test, validate and centrally manage all of the software components needed for any level of automation, resulting in versatility and great flexibility of the solution which is fit-for-purpose.
Our open architecture with unified and open standardized APIs, which are AUTOSAR classic & adaptive compliant, also provides a unified management environment for heterogeneous solution elements, including multi-core SoCs with different underlaying operating systems.
The platform also ensures predictability with regards to resource consumption, runtime, data flow latencies and application task sequences, allowing OEMs to guarantee the highest levels of performance for their environments in a real-time manner. Globally available automotive, development and support services achieve greater service agility.

Safety by design

Built on fail-operational principles, MotionWise delivers constant availability and the highest levels of performance for mission-critical functions in line with ASIL D safety requirements according to the ISO26262 standard. By ensuring freedom from interference for all safety-critical functions with encapsulating applications from their peers, a safe environment is created, where applications with different safety and real-time requirements coexist and interact.

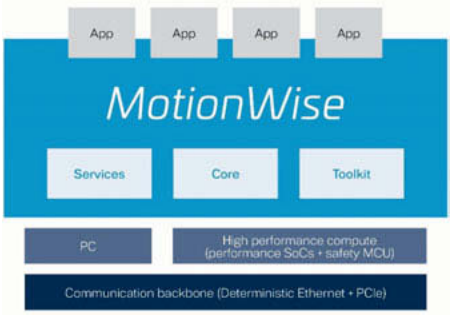


Fig. 4: MotionWise consists of core features, optional services and tools.

In-Car Compute Platform

Current computing and network architectures in vehicles are built around several functional domains, e. g. for powertrain, chassis, infotainment, energy and autonomous driving and ADAS. Although such a domain-based architecture presents the first step towards consolidation and the reduction of the number of ECUs, it still has disadvantages. Guaranteed safety of the autonomous vehicle requires close interaction and streamlining of all domains and it is not possible to reduce safety features to one domain only. Having multiple domains (and correspondingly domain ECUs) implies multiple development, testing and maintenance costs. Besides, it is difficult to adapt or add new cross-domain applications expected by the customer. The *In-Car Compute Platform (ICCP)* aims to consolidate a maximum of automotive functions from different domains inside a single highly integrated, high-performance ECU and thus drive the transition from current, hardware-oriented vehicles to the software-defined vehicles of the future. In particular, the requirements of these functions vary concerning ISO 26262's ASIL, as well as their real-time response requirements. The orchestration also contains the *Smart-I/O* computation layer that is primarily responsible for the interaction with sensors and actuators

(as shown in Fig. 6). For fail-operational properties, the architecture will need to contain redundant ICCP ECUs connected over dual redundant, high-speed backbone links (as seen in Fig. 6). The connection to (critical) smart-I/O should also be redundant. Fail-operational architectures are realized through suitable redundancy and the interconnection of the replicas' redundant elements by means of a deterministic network. ICCP itself defines redundancy on multiple levels, for example that the failover to a second (or third) ICCP ECU will only be triggered under extremely unlikely failure conditions.

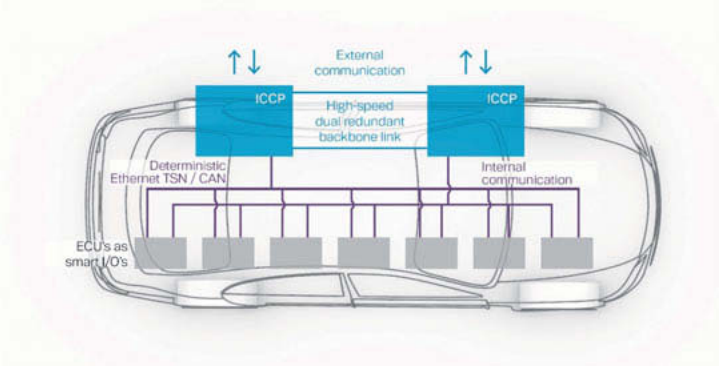


Fig. 6: ICCP architecture

Fig. 7 shows a high-level ICCP architecture that is able to host applications from different functional domains (including cross-domain applications). The architecture includes shared services, e.g. for safe and secure integrated diagnosis, monitoring and logging, as well as for shared management capabilities, e.g. for safe and secure re-configuration and balancing.

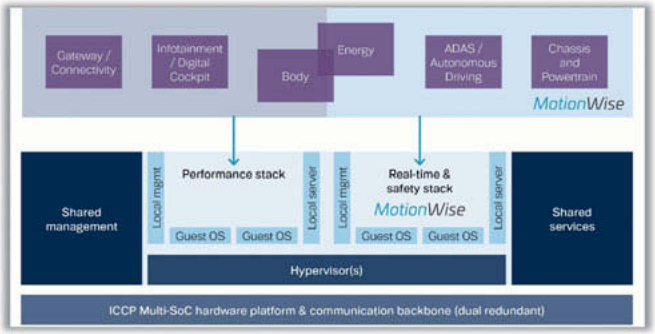


Fig. 7: High-level ICCP software architecture.

Benefits of the MotionWise safety software platform

Fast, simple integration of best-in-class applications

MotionWise abstracts hardware and operating systems, creating a converged, open, AUTOSAR-compliant architecture with a unified management environment out of heterogeneous elements. This provides the modularity, portability and standardized interfaces required to deploy, test, validate and centrally manage software components.

Real-time orchestration of applications

The time-aware platform architecture uses innovative global scheduling technologies and design tools to support smooth application and communications network traffic scheduling in order to fulfil end-to-end real-time guarantees. Built on deterministic technologies, MotionWise delivers always available services with guaranteed latency – regardless of the system load.

Guaranteed safety compliance and performance

MotionWise delivers safety by design on a global basis by using unique technology for scheduling and overcoming resources constraints. This ensures constant availability and the highest levels of performance for mission-critical functions in line with ASIL D safety requirements according to the ISO 26262 standard. Freedom from interference is ensured by encapsulating every application from its peers, thus creating a safe environment, where applications with different safety and real-time requirements coexist and interact.

Fail-operational performance

Safe operation for highly automated driving use cases is continued even after failure of a component. Constant availability requirements are fulfilled by the way we design the solution to ensure redundancy for safety-critical parts. This is achieved with sophisticated software platform architecture and with the HW architecture that is best sourced from various vendors.

Minimized development and deployment risks

With a platform-centric approach for highly automated driving use cases, MotionWise minimizes development and deployment risks for OEMs, while offering the full utilization of the existing resources and the capability to resolve scheduling bottlenecks. This allows for significant increases in automation and improvements in resource utilization.

Real-time capabilities

Networks for real-time systems have stringent end-to-end latency requirements. One cost-efficient way to meet them is time-aware scheduling and communication, using computation chains to guarantee end-to-end latency. With a time-aware platform architecture using advanced global time synchronization and deterministic scheduling, we can ensure end-to-end real-time guarantees.

High flexibility and versatility achieved with scalable solution

With MotionWise, OEMs and Tier 1 suppliers are introducing one solution which can be used in multiple vehicle lines and variants and in different models as required, with reduced integration, testing or validation effort or cost. It can tackle various use cases, from chassis control domain, to automated driving, powertrain and beyond.

Accelerate time-to-market with our series-proven solution

MotionWise is the first series-proven safety software platform enabling level 3 - 5 automated driving on the market. Adopting a platform-centric approach to building driverless software capabilities allows the reuse of the software in the future and across multiple SoCs, vehicles and models, achieving faster time-to-market at reduced costs.

Increased competitive edge providing a seamless roadmap to future full automation

MotionWise is laying the foundations for a fast, cost-effective introduction of highly automated driving features as a natural continuation of current work and investments. It enables the re-

use of software investments to add emerging functionality in the future in a cost-effective and agile way.

Boost returns on investments with a platform-centric approach

Building software-defined systems for automated driving requires vast investments. With a platform-centric approach, MotionWise speeds up software deployment for automated driving use cases and harnesses investments to keep the path open for future automation developments, allowing software re-use in multiple vehicle lines, variants and models.

Reduce costs and achieve cost predictability

With a platform-centric implementation, no large and unpredictable up-front investments are required to develop driverless solutions in-house. OEMs can instead harness a smart licensing model to add new features to their platform and pay only for the software they are using, which increases cost predictability.

Conclusion

As we move towards highly automated driving use cases, requiring fail-operational behavior of the system, we suggest building solutions based on a proven, software-defined, highly modular and highly integrated platform. This centralized architecture, built on generic high-performance computing hardware combined with a well-designed amount of redundancy, can deliver highly automated driving use cases and enables software function re-use across different SoCs and car models, while still delivering an efficient and safe solution at reduced costs. In order to accelerate time-to-market and reduce development costs, we have designed the MotionWise safety software platform. This helps organizations move away from a slow, costly, complex and iterative integration process to a platform approach that speeds up time to market for new functionalities, guarantees safety, and allows software investments to be reused for highly automated driving projects, resulting in a seamless roadmap to full automation in the future. Paired with our Safety Co-Pilot, MotionWise delivers an end-to-end safety approach, capable to evolve all the way to a software-defined In-Car Compute Platform (ICCP) in the future.

Boost Safety & Styling for vehicle lighting – Individualization and new Functionalities

Dr. **Michael Kleinkes**, Dr. **Wolfgang Pohlmann**, Dr. **Carsten Wilks**,
HELLA GmbH & Co. KGaA, Lippstadt

1. Abstract

High Definition Solid State Lightsources (HDSSL) will boost new functionalities such as light on demand, new animations and light distribution safety features, unreached in the past.
Micro-LED-Displays enable new features on LED-Displays in front and rear lighting by new technology approaches, which are highly attractive to customers and boost safety & styling significantly.

2. Introduction

High Definition Solid State Lightsources (HDSSL) are supposed to become *the* light sources for future HD headlamp lighting modules. Based on the funded research project “µ-AFS” [1][2] this technology is about to take a significant step forward. Smaller pixel sizes, larger light emitting surfaces and, thus, orders of magnitude higher number of pixels, higher brightness, with application specific aspect ratio and higher functional ASIC integration will allow new use cases with high light quality and performance. Typical applications are the full adaptive headlamp, visualize specific areas of safety or interest for the driver inside the car.
Automotive exterior displays are the next evolution step for sophisticated signal functions. This evolution has started with the implementation of the 1D-animated turn indicator [3], followed by 1D animations all over the car body, and will be continued with 2D displays for a large variety of use cases. Exterior displays will be placed at front, rear and wherever they are useful. They will not only add new features to the car stylist’s tool box. Following the automobile megatrends individualization and digitalization displays will enable new digital product concepts (e.g. functionality-on-demand), visualize information to the outside of the car, and communicate with pedestrians and other human road users.

3. HDSSL Light Sources for Front Lighting

HDSSL Light Sources in general are based on a super-integration of one single silicon application specific integrated circuit (ASIC) supplying several thousand LED pixels individually. For white light emission these LEDs are coated with “phosphor” layer(s) converting a fraction of

the blue light, which is generated by the InGaN LEDs semiconductor, into overall white light. New solutions to establish Full-field-of-view applications are requested to provide total light emitting surface (LES) size of 30mm² - 45mm² with an aspect ratio of approx. 4:1 and pixel sizes in the range of 35µm - 55µm. This results in overall pixel numbers far above 10000 per light source.

(a)

(b)

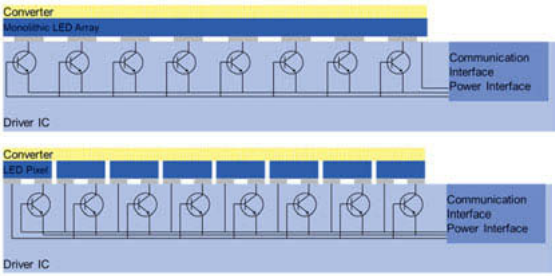


Fig. 1: Schematic cross section of different HDSSL light sources with (a) monolithic and (b) separated LED pixels carried by a Driver IC (ASIC)

Especially for the LED pixel layer different designs are generally feasible. On the one hand the LED layer could be one single monolithic chip with several pixel areas defined by current barriers inside the epitaxial structure. On the other hand, the light can be generated by separate individual LED pixels with a certain distance. This approach is widely known as “µLED technology”. These options are displayed in figure 1. In addition, intermediate solutions for these two technologies are conceivable, while pixels are only separated to a certain depth or the separated space between the pixels is filled with specific material.
With respect to technical and commercial aspects the described options potentially have reciprocal ad- and disadvantages listed in table 1.

Table 1: Comparison of different LED pixel layer designs for HDSSL light sources

	Monolithic Pixelated LED Array	**Single Pixel LED Array**
Potential Advantages	- Efficiency - Pixel position accuracy	- Contrast - ASIC/LED-Interface reliability
Potential Disadvantages	- Contrast - ASIC/LED-Interface reliability	- Efficiency - Pixel position accuracy

With smaller pixel size the external quantum efficiency (EQE) of the single pixel LEDs is reduced compared to monolithic LED, because parasitic effects at the single pixel edges increase [2]. In addition, the total thermal path or interface area of a monolithic approach can be larger with one common electrode for all pixels whereas single pixels would need two electrodes in any case. Intrinsically, the position accuracy of the monolithic pixels is higher compared to the single pixel solution. In contrast, the monolithic solution may face challenges in ASIC/LED interface reliability with increasing size due to tensile stresses especially for high aspect ratios in combination with an imbalance of the coefficient of thermal expansion (CTE) between silicon and InGaN.

The super-integration of LED and ASIC allows significant improvement of functionality; the HDSSL can be seen as an advanced white emitting display. Firstly, the light flux emitted by each individual pixel can be adjusted by pulse width modulation (PWM) and/or current control of individual pixels or partial areas with several pixels, respectively.

With higher pixel density and total number of one HDSSL the PWM generation is supposed to be integrated into the silicon device to reduce the necessary connections and data rates. Hence, it is highly important to implement appropriate parallel and serial digital data interfaces of the component for both, video and control data transmission according to the various system architectures and configurations of different car manufacturers.

For current control and communication interfaces the ASIC is supposed to combine analog and digital circuits in so-called mixed signal ASIC. Additional features such as pixel error detection, thermal sensor integration, supply voltage control, adjustable PWM frequency and others can be integrated into the ASIC by significantly enhancing its usability inside the head lamp.

Important evaluation criteria of HDSSL light sources are the maximum achievable luminance, the contrast between the pixels, the homogeneity of the luminance and the colour as well as the overall electro-optical efficiency.

4. Display technologies

The display technologies currently under discussion for automotive applications are

- LCD, a TFT display controlling the polarization of transmitted light via a liquid crystal,
- AMOLED, an active matrix of some micrometer small OLED pixels,
- LED matrix systems, arrays of top- or side-emitting LEDs with LED size down to 0.3 mm, providing monochrome, white or even multicolour (RGB) light with high luminance,

- Micro-LEDs with size less than 0.3 mm, either produced by monolithic fabrication remaining on their original waver or transferred to a receive substrate, e.g. via laser lift-off or another mass-transfer process [4].

These technologies strongly differ in their performance, which is mainly described by parameters like resolution/pixel pitch, luminance and colour. In the end the use case of the intended application will always define which technology is suitable.

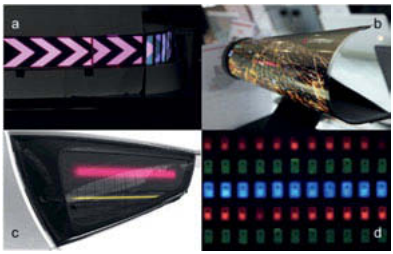


Fig. 2: Display technologies for automotive signal applications: a) LCD, b) AMOLED [6] c) LED matrix, d) Micro-LED [9]

5. Functional Use Cases - Front

The use cases of HDSSL can be split up in use cases which support safety and those which have a major styling impact.

C1: The welcome / farewell-use case visualizes the resolution and flexibility of HDSSL-Headlamp systems to the user in an animated fashion. This is highly impressive not only for the lighting experts but also for other road users. Its major benefit is a styling purpose.

C2: The glare-free high beam reaches by HDSSL an unprecedented resolution by an additive light source LED Module, which flexibly adapts the illuminance on traffic signs and minimizes also the shadow area around oncoming and followed traffic participants. The night driving safety is significantly increased by this HD function.

C3: The optical lane assist supports the driver in its estimation of the width of the vehicle and warns other road users from leaving their driving track, too. This minimizes adjustment movements of the steering wheel, lowers the driver's stress level and adds as optical feature to increased traffic safety.

C4: The animation / dynamization is a mixture of safety and styling, as the switching from low beam to high beam is done in a sparkling appearance. This enhances the awareness of better light at night drives and, thus, adds even more to driving safety appeal.

C5: The 3D light distribution [7] is a bending light with a more specific adaptation to the road surface / adaptation to the curvature / to road crossings and bends or branching. It adapts to the needed driving direction of the vehicle.
C6: Safety contributing light distributions enable e.g. specifically illuminated bicycle variable pathways incl. safety distances to neighbouring traffic. This light function requires a higher resolution sensor system for the surrounding monitoring, e.g. HELLA Nano Radar.
F1: Future HD light distributions allow for symbol projection onto the road surface, if legal preconditions would be changed to do so.

Fig. 3: HDSSL Front Lighting Functions

6. ALiSiA for HD lighting function development

The development of advanced lighting functions for HD systems is a quite complex process, as many components as sensors, ECUs and lighting modules have to be chosen or developed and merged to one functional system. This may lead to lengthy discussions on concepts and components, followed by complex alignments on the required interfaces. Once a mock-up vehicle has been build-up, lighting functions can be tested and applied with numerous night drives. However, here, scenes are not reproduceable and the lighting function behaviour can only be evaluated subjectively by the participants. Moreover, the contributions of singular components remain unclear, as only the resulting lighting behaviour can be judged.
HELLA created the ALiSiA (Advanced Lighting Simulation Architecture) to face these challenges. ALiSiA is a PC tool that allows the real-time simulation of dynamic lighting function behaviour on a standard PC. The basic concept is shown in figure 4.
ALiSiA provides a user interface, where input scenarios for each lighting function can be chosen. Input scenarios consist of video and CAN data having been recorded synchronously during real test drives. Moreover, relevant parameters for the chosen lighting function can be adjusted. During simulation, the CAN data is used to provide the lighting function algorithms with the required input data and the behaviour of the driver electronics and the headlamp are simulated on a functional base.

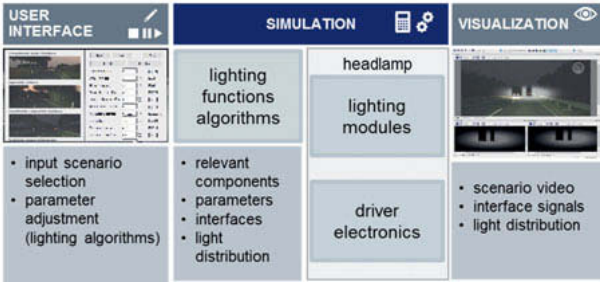


Fig. 4: ALiSiA concept

The calculated output light distribution is visualized together with the input video data, as shown in figure 5 for a glare-free high beam use case. It is also possible to visualize relevant interface data like e.g. detected objects positions. This allows to judge the contributions of each component and hence creates the required transparency. By using predefined scenarios, the lighting function behaviour can be parameterized on an objective base, as sequences can be replayed and the impact of parameter changes becomes clear.

Fig. 5: ALiSiA simulation of a glare-free high beam use case for a HD front lighting system

The advantages of using ALiSiA within the development of HD lighting functions are obviously the ones that always apply for the usage of simulations: Frontloading by early system simulation, effective analyses, enhanced transparency and development speed up for all phases.

The last aspect is especially relevant for the application phase, where ALiSiA can be used to pre-adjust the required lighting function behaviour and hence to save many night drives.

7. Functional Use Cases - Rear

Animated Signal Light

The use of animations for conventional signal-light functions tail light, stop light, and turn indicator pursues more than only one objective. On the one hand it at-tracts the attention of other road users and thus, enhance the contribution to safety. On the other hand, animation can always be used as additional styling feature.

While this application can be implemented with the use of just the two colours red and yellow, the main requirement is that the light intensity must fulfil legal requirements. Although the required luminance values strongly depend on the size of the light-emitting area, as a rough rule of thumbs 10^3 cd/m² can be considered as lower luminance limit for tail light, as well as 10^4 cd/m² for turn indicator and stop light. For example, in theory a 2mm-pitch array of top-emitting LEDs can provide up to 10^5 cd/m² in yellow or super red. Even after considering limiting factors like thermal management and current carrying capacity, such a display can provide more than 10^4 cd/m², enough light to fulfil the requirements of all conventional signal-lighting functions. On the other hand, a high-resolution AMOLED display providing a maximum luminance of less than 10^3 cd/m² can hardly support a tail light. In general, a correlation between resolution and achievable light output can be found, as shown in figure 2: While displays providing high resolution like LCD and AMOLED suffer from low luminance values, the minimum pixel pitch of LED arrays is limited by LED size and assembly technologies.

Fortunately for the given use case high resolution is not required, since the minimum viewing distance of other road users is larger than 5 meters. A study with test persons was performed at HELLA in collaboration with the L-LAB. Rear lamp displays were evaluated with different technologies and various resolution values. As one result of this study most of the test persons rated the LED matrix systems with pixel pitch of 1.6 and 2.5 mm as good regarding resolution and desirability. Only for a LED display with larger pitch this rating was less positive [8].

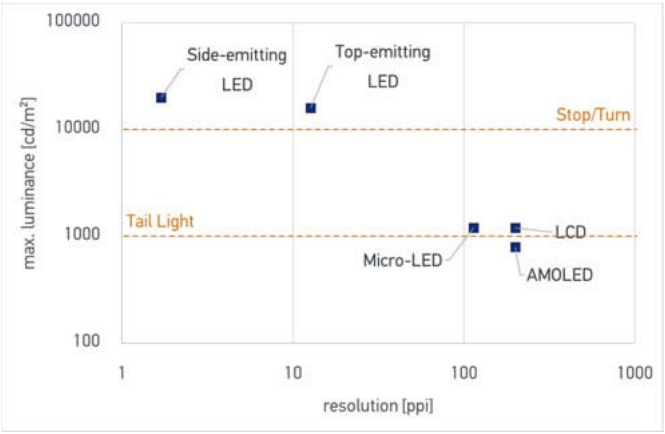


Fig. 6: Display technologies with typical luminance and resolution values.

Personalized Welcome Animation / Advertising Video

Signal lighting functions are on the way to transform into new styling possibilities and safety features by HD technology features.

In near future car makers will offer digital products to their customers like personal animations/videos showed anywhere on the car body. Car sharing/rental companies can use this feature to present their brand on all cars of their vehicle fleet.

Regarding the requirements for the display it must be considered that the end-customer pays an additional fee for this digital product. So, a high-quality animation is expected, with multiple colours, RGB if feasible. Furthermore, the short viewing distance requires high resolution, allows low luminance on the other hand.

The wiping direction indicator was one of the first starting points of animated signal functions. This 1D signal functions will be extended to car body lighting functions, going all around the whole vehicle in a 1D or even partly 2D manner. The illuminated front grill or company logos start the journey from 1D to 2D styling features at the front. The highly sophisticated animations at the rear lights, based on 2D signature lighting go far beyond the pure minimal needed standard light functions. Their animation will track further attention to them, minimizing the reaction time of following traffic. Styling and individualization of the vehicle is enhanced, based on several homologated scenarios, by different signature animations. This adds up to the attractiveness of the vehicle and keeps it updated during the operation of the vehicle by over the air updates on demand or functions on demand.

Fig. 7: LED-Display Functions – a first set of potential use cases

D1: The Display shows a personalized welcome or farewell scenario, in which the signal could vary in size, location, intensity or velocity hovering over the display for styling purposes.
D2: Rental cars or e-Mobility Services provides are highly interested in doing advertisement and special offers directly on the parted vehicle.
D3: Logos, Symbols and Smileys could be used for branding or vehicle to pedestrian / human road users communication purposes e.g. thanks for giving right of way. This supports safety, rises mood and increases traffic flow in real life traffic jam conditions.
D4: Animations with moving shapes, geometry, structure, orientation, distance or changing shapes, dynamically pulsing signals like heart beat rhythm combined with stripe shaped tail, brake or turn light signals are illustrated here for styling purposes.
D5: The interaction with the vehicle for functions on the demand, SOS-Help-Signals, connectivity provider to the cloud or other W-LAN, 5G communication means and the energy charging status of electric or plug-in hybrid vehicle are shown on the display. The economical driving style status could be shown to other road users to animate others to lower their energy consumption. Those functions support safety, connectivity and ecology.
D6: Special weather conditions, like heavy rain, snow, icy roads, heavy flooding or strong stormy winds could be visualized based on cloud-based weather data and/or boosted by local vehicle sensor information to warn other road users as safety support.

GAIN for animation design

There is a need for improvement of animation design procedures with regards to animation design, simulation and realization. In order to cover this, HELLA has developed the approach “GAIN” (Graphical Animation INterpreter). It mainly consists of two parts: A software application for animation design and an embedded software solution (GAIN embedded interpreter). These two parts fill the gaps within the animation design process and provide the required consistency.

GAIN animation design app

The GAIN animation design app can be used from the first concept phase up to the final design of animations for series purposes. It is a powerful tool, which can be applied for the alignment on simple animations e.g. for wiping direction indicators, just as well as for complex animations across various lighting modules all around the car. Due to the limited extent of this article, only some key facts are presented subsequently.

The idea of the GAIN animation design procedure is to describe the required animation within a virtual coordinate system based on so called “animation objects” and their transitions from one state to another. The animation design process is supported graphically by the usage of timelines. Hence, it is as intuitively to apply as commonly used video cutting tools. Animations having been specified in this manner can be simulated and the 3D visualization allows the evaluation of the animation from different viewpoints, considering also the surrounding parts (figure 8, bottom right).

Despite the 3D visualization being neither physically correct nor photorealistic, as only the properties of 3D surfaces are changed over time, the GAIN app allows realtime visualization of animations from every perspective, which already helps significantly for the alignment on animation patterns and durations.

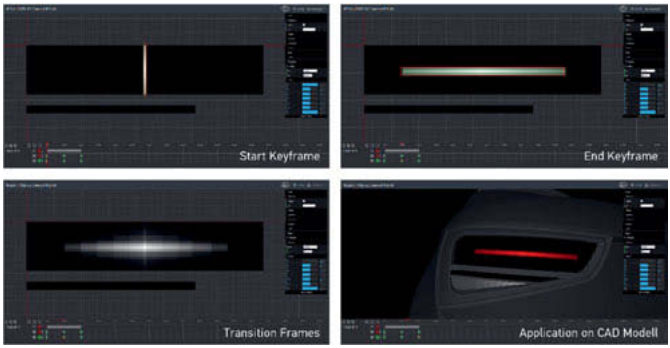


Fig. 8: Animation design within the GAIN app

The GAIN animation design app allows the export of animations in different formats, e.g. tabular with one value for each LED channel for each time instance. Customized formats are available, too. Moreover, there is a special export format with low memory needs for the usage with the GAIN embedded interpreter, which is described below.

GAIN embedded interpreter

The GAIN embedded software solution is optimized for efficient realtime computation of light animations. Here, the idea is to use a special, object-based format, which is generated automatically by the GAIN app, to describe the animations and to compute the required control values during operation. Especially for large systems with many individually controllable light sources and complex animations this is advantageous regarding the required memory space. Even if it puts a computational strain on the overall system, the realtime computation of animations is the most efficient way balancing the costs between computational power, flexibility and memory footprint for the control of large systems like the here presented pixel rear lamp concept.

The GAIN app provides the possibility to export animations directly and ready to be used for target systems with the GAIN embedded interpreter software module.

Improved animation design process

The GAIN animation design app can be used during the whole development process. It is easy and intuitively to use – users can focus on how animations shall look like and not how they can be realized. The outcomes can be used in every development stage, whereas the animation design remains centrally located within one tool The GAIN embedded interpreter completes the HELLA tool chain for a continuous animation design process.

8. System aspects relevant to the implementation

Thanks to the high number of pixels, both technologies HDSSL headlamps as well LED-Displays open the door to new functions and thus increase the added value for drivers. When realising HD lighting systems, dealing with the high number of pixels in realtime is the main challenge from an electronics viewpoint. The transmission and processing of these large data volumes significantly influence the allocation of the individual functions and function groups to the particular system components. A distinction is made between a central and a distributed E/E architecture depending on where the focus is placed for pixel data processing.

Furthermore, some specific constraints must be taken into account. For example, for thermal reasons and due to the lack of space, the LED power driver usually must be mounted outside of the lamp housing. Due to the high-frequency data transmission to the HD actuator, the control electronics must be integrated directly in the light module. For transmitting the video signals from the vehicle electrical system to the light module, an interference-proof interface must be selected.

Central system architecture

In this architecture scenario, the pixel light matrix is calculated in a central vehicle control unit. For example, a high-performance processor is used to generate a grayscale pixel matrix with the resolution of the HDSSL light source or LED display. These grayscale values calculated in the vehicle electrical system control unit are transmitted to the control electronics in the light module with as little alteration as possible, figure 9. Transfer rates in the Mbit/s range are achieved because of the 8-bit grayscale gradation, for example, in addition to the display refresh rate of 60 Hz. This large amount of data requires sufficient computing performance to be present in the vehicle in order to achieve the requirement for real-time system capability.

The task of calculating and transmitting pixel matrices represents particularly for headlamps a considerable challenge for the system component chain. The HDSSL module and other light modules, such as the module for the area in front of the vehicle, must also be synchronously controlled to generate a complete light distribution. For example, the vehicle electrical system control unit takes over the task of the light master for the in-sync control of all LED driver modules in the headlamp. In this case, system protection with regard to functional safety and defence against cyber-attacks is challenging and cost-intensive.

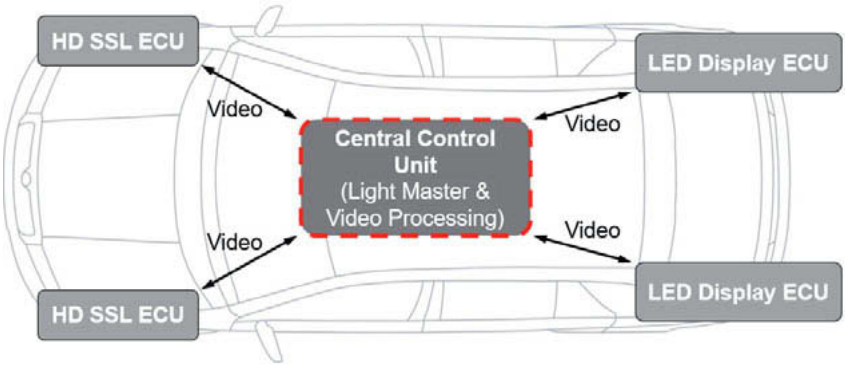


Fig. 9: Exemplary realization of a central E/E system architecture – block diagram

The transmission of pixel data from the vehicle electrical system to the light module in the headlamp or in the rear lamp is equally complex. Conventional data interfaces have long since reached their limits with respect to bandwidths and real-time capabilities. As a potential way, solutions from infotainment systems can be mentioned here. Similar to the control of high-

resolution flat screens in the centre console or in the instrument cluster, video interfaces suitable for automotive applications are being discussed for transferring the pixel data to the headlamp.

Distributed system architecture

In this realisation scenario, the pixel light distribution is created in an ECU in the headlamp or in the rear lamp. Only sensor data or light function indexes are transmitted from a master control unit in the vehicle to the ECU in the lamp. In this case, the transmission of high volumes of data between the vehicle and the lamp is eliminated. Standard interfaces that are currently widely implemented, such as CAN bus, can therefore be used. As explained in figure 10, the architectures of an HDSSL headlamp system and a conventional LED headlamp are no longer different.

With respect to data processing, the computing focus shifts from the vehicle electrical system control unit to the ECUs placed locally in lamps. These task of the calculation of the grayscale pixel-matrices requires the use of high-performance, multi-core processors that are intended to guarantee real-time graphics processing of the digital light distributions.

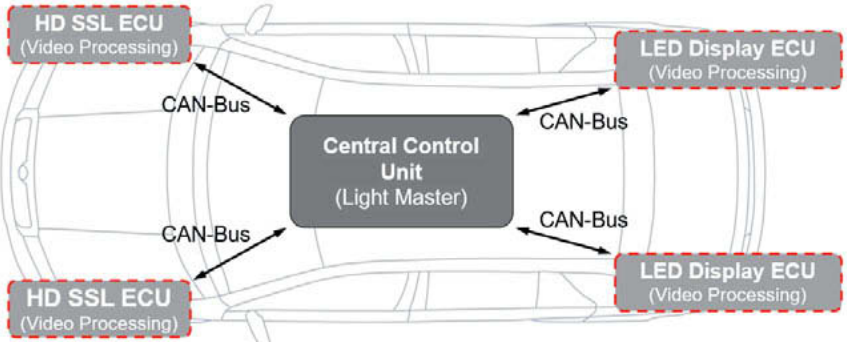


Fig. 10: Exemplary realization of a distributed E/E system architecture – block diagram

E/E architecture of the future

The current trend in automobile lighting development is toward the use of pixel-based applications. When realising system solutions ranging from one hundred to multiple thousands of pixel

segments, controlling the transmission and processing large volumes of data represent significant challenges from an electronics viewpoint. A distinction is made between a central and a distributed E/E architecture depending on where the focus is placed for pixel data processing. The consequences resulting from these two contrasting architecture scenarios are summarised in [10].

The selection of the suitable E/E architecture for an application depends on many factors. In addition to technical aspects like actuator technology and functionalities, the existing vehicle architecture, system capacities and change effort and costs also play a significant role. Automotive lighting is at the brink of the first series applications for HD headlamps and LED display systems. So it is too early to talk about clearly identifiable trends. It is to be expected, however, that both architectures will be used in parallel for a while. Distributed architectures are state of the art and will be used first, especially for display applications. Here, static textures or dynamic animations can be locally stored in the rear light and displayed with a graphic microcontroller. Later, central system architectures with the use of domain computers will be used more frequently due to the adaptation of vehicle electrical systems. Thanks to their flexibility regarding software updates, new functionalities such as Pay-Per-Use, Software-on-Demand or product upgrades after SOP can be realised.

9. Summary and Outlook

High Definition Solid State Lighting (HD-SSL) and LED-Displays offer new possibilities for digital lighting at front and rear lighting, which benefit on safety, on styling, on animation and on communication for the vehicle owner but also for other road users. Further developments in the area lighting are very fast but a holistic approach including customer needs, the development of the E/E architecture and also the legislation has to be taken into account to develop successful future products.

References

[1] S. Grötsch, A. Pfeuffer, T. Liebetrau, H. Oppermann, M. Brink, R. Fiederling, I. Möllers, J. Moisel: Integrated High Resolution LED Light Sources in an AFS/ADB Headlamp. 11th International Symposium on Automotive Lighting (ISAL), 2015

[2] I. Möllers, J. Moisel, R. Fiederling, S. Grötsch: „An efficient and high-resolution ADB headlamp based on micro-integrated LED arrays“ in VDI-Optische Technologien in der Fahrzeugtechnik, Karlsruhe, 2016

[3] M. Hamm: Safety Improvement by New Matrix and Turn Indicator Func-tionality, ISAL 2013

[4] K. Ding, V. Avrutin, N. Izyumskaya, Ü. Özgür, H. Morkoç: Micro-LEDs, a Manufacturability Perspective, Appl. Sci. 2019, 9, 1206

[5] S. Konoplev, K. Bulashevich, S. Karpov: From large-size to micro-LEDs: scaling trends revealed by modeling, Phys. Status Solidi A (2017), DOI: 10.1002/pssa.201700508

[6] LG Display, press release, Jan 04, 2016

[7] C.Wilks, B.Kubitza: HD-Headlamp Technologies and Development Process: From simulation to demonstration under real traffic, ISAL 2017

[8] B. Willeke, C. Hohmann, D. Mundt, A. Köhler, A. Thoma: HD Technologies: New Functions and Possibilities for Signal Lighting, ISAL 2017

[9] L. Hou: LEDinside, May 02, 2017

[10] J. Roslak, C. Wilks: High-resolution LCD Headlamps - Challenges for Electronics Architectures, ATZ elektronik worldwide, 06/2017, p. 46-51

CAN FD Light

A novel communication bus supporting digitalization and customization of automotive lighting for the broad market

Fred Rennig, **Jochen Barthel**, **Marianna Sanza**, **Donato Tagliavia**,
STMicroelectronics Application GmbH, Aschheim-Dornach near Munich

Abstract

With the pervasion of LEDs and nowadays with Organic LEDs (OLEDs) automotive lighting has developed into an eye catching design element that gives the car an unique style and offers perceivable value to end customers. Light functions require more and more dynamic and individual control not only to attract end customers by enabling customized and animated light patterns, but also to increase safety by adapting instantaneously beam shapes and light/signal patterns to the actual driving situation, including hazardous warning. Such systems require a broad evaluation of car sensors and infrastructure (car2car car to x) information and at the same time a fast, robust and reliable communication interface from the central gateway to the light control unit hosting the light source (LED, OLED) as well as the LED/OLED Driver IC.

The CAN FD Light communication bus is derived from the well-known CAN FD standard bus, it can control luminosity and run all relevant diagnostic tasks from up to 4096 individual light spots with a high bus bandwidth. It includes all relevant CAN FD protections and safety measures to make the bus system easily ASIL B compliant. The CAN FD Light communication bus can be used with existing hardware in microcontrollers on the market. On the LED/OLED driver side it is integrated without the need for additional components like crystals and external driver circuits. Its tailored protocol structure ensures an efficient data management and minimizes the system overhead in terms of hardware and software. CAN FD Light integration offers modern and safe premium automotive lighting solutions with a lean implementation.

This paper describes the innovation of automotive lighting with CAN FD Light bus and how it helps to achieve the automotive market requirements with a lean and robust solution.

Introduction

Automotive Lighting underwent a long and impressive evolution. In the old days, life was much simpler: there were very few switches in the car and if you got lucky the directional light switched on at the exact spot you wanted it to. If a bulb didn`t work, the horns of nearby drivers announced you were about to be blindsided. Nowadays, defective light sources are almost history. With the introduction of new light sources, replacing classical bulb solutions, lighting became not only more reliable, but vehicle developers and lighting designers are finding completely new and varied creative freedom to emphasize lighting as a brand and quality-recognition feature. LEDs with their extended lifetime have become a standard and new light sources such as OLEDs (Organic LEDs), Lasers and high resolution beams composed of many individual controllable light spots and pixels are gaining importance.

Fig. 1: Modern Rear Light

The electronic architecture of a classical bulb driven exterior lighting application was quite simple (Fig. 2). A central body control module, typically located close to the dashboard in the car managed the exterior lighting control among various other tasks. In this module, a semiconductor switch with built-in protection and diagnostic capability (Smart High Side Driver) provides power to the load by connecting the vehicle power supply with the light. For each light function a separate high-side switch is used. The light function can be dimmed by controlling the switch in Pulse-Width-Modulation (PWM) mode. A malfunction, such as a worn open light filament can be detected by supervising the current flowing from the high-side switch to the load. The switch output is connected to the lamp through the wire harness, typically 0.75mm² running across the car, one cable for each light.

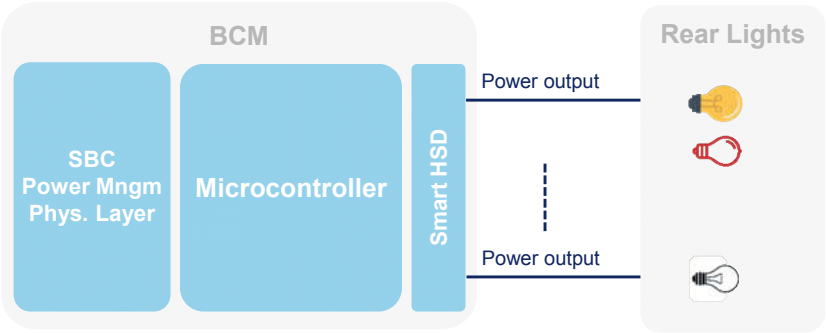


Fig. 2: Electronic architecture of a classical exterior lighting system

With the introduction of LEDs and nowadays organic LEDs (OLEDs) first the light control topology had to be changed into a constant current regulation concept. The situation becomes the more complex the more individual light spots and pixels have to be controlled.

Modern light functions require light patterns adjusted to environmental-, customized- and traffic conditions. It is evident, that for light functions composed out of tens, hundreds or even more pixels, each pixel with its individual current and dimming setting, it is neither economically nor technically practical to draw one cable to each pixel through the car. Therefore a decentralized solution is required, physically separating the (O)LED drivers and sources from a central entity managing the lighting control and supervision.

The data exchange between the central entity and the (O)LED drivers is handled by an automotive compliant communication bus. It transports information about how to link data with the individual light source, the (O)LED current settings, the PWM dimming ratios, environmental conditions of the lighting system and various diagnostic data. Current setting and PWM dimming ratios in dynamic animated light scenarios are refreshed in time steps of 10 – 50 ms, with a depth of minimum 8 bit per light spot to ensure a smooth and homogenous evolution of the light pattern without visible steps. To ensure a short fault reaction time diagnostic information per light spot has to be refreshed in time steps of 50ms – 200ms. Therefore a high speed bus with a data rate > 250kB/s becomes easily mandatory if hundred or more individual light spots have to be controlled.

Moreover a communication bus in automotive has to comply with requirements and automotive specific challenges that must be respected in safety critical light functions like braking lights, turn indicators, rear light and front light systems. Therefore the communication bus

controlling these functions must assure their correct operation according to automotive safety integrity levels (ASIL) as defined in ISO 26262 even under electromagnetic influence and electrostatic discharge. In turn they must not cause any distortions of other systems in the car.

Fig. 3 shows the basic block diagram of such a system.

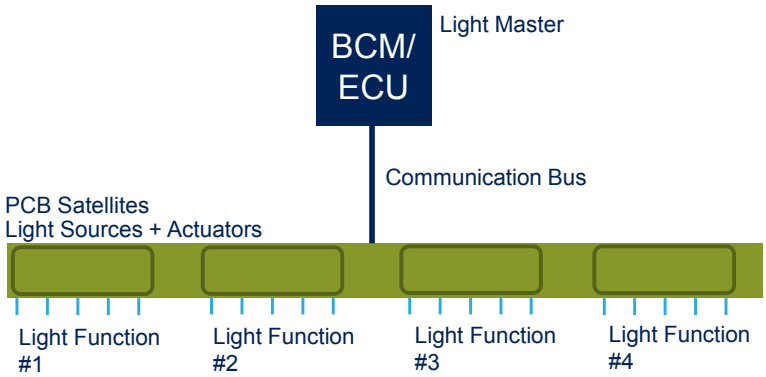


Fig. 3: Electronic architecture of a modern (O)LED exterior lighting system

CAN FD Light as a communication bus

Over the past decades the automotive industry has gained extensive experience with the control area network (CAN) and has over many years developed implementation methods and testing capabilities to ensure the robust and safe operation of this network. These experiences have greatly influenced the newest CAN FD communication bus standards. These standards and with them the conformance tests of standard compliance, automotive system compliance, such as electromagnetic compliance, electrostatic discharge and other distortions, and interoperability tests assure a very high level of safety and quality.

Additionally the industry has developed soft- und hardware solutions to leverage these special features of the CAN network which makes it one of or perhaps the most successful bus network in the car.

Due to these characteristics and the already existing infrastructure for implementation and evaluation at the car manufacturers it is appealing to use this communication bus also for lighting applications.

While in the past the physical dimension of an automotive lamp was well restricted to the edge of the car, nowadays in modern cars we can find lamp segments spanning the entire rear with light sources distributed all-over. To reduce cabling effort within the lamp, light sources and actuators are positioned close together on the same printed circuit board. The light actuator / (O)LED Driver however has to communicate with the vehicle infrastructure to receive configuration and control data on the one hand, and to send status and diagnostic data on the other hand. To reduce the complexity of the electronics in the lamp, this is typically managed by one central ECU communicating with the lighting actuators through a communication bus. Single ended communication interfaces like SPI are tailored for short distances with sender and receiver typically located on the same printed circuit board. For spanning larger distances and for communication between different devices a differential bus with twisted pair wiring like the CAN FD bus offers high electromagnetic immunity, low electromagnetic emissions and robustness against electrostatic discharge; an important argument considering the safety implications of exterior light functions.

With the introduction of CAN FD for data rates beyond 1 Mb/s and 64 byte user data per frame enough bandwidth for upcoming lighting applications is available. Our studies have shown that 1 Mb/s offers enough bandwidth even for complex lighting scenarios and animations.

Unfortunately the universality and broad range of use cases of the CAN bus require features that go along with expenses that would impact the overall system cost and are not needed in lighting applications. While a CAN bus is designed for exchanging data between any connected nodes at any time it is necessary to implement a scheme to manage bus access based on access priorities. This scheme is called arbitration and is a way to negotiate between bus participants the access right to the bus. This negotiation has to work between independently acting bus participants on different positions on the bus network. Therefore the arbitration has to cope with inaccuracies in the individual sending data rate and different delays between the actors. To make this work reliably under the previously described automotive conditions a very tight specification of the data rate accuracy has been set up. Some car makers require 0.1% of data rate accuracy. This specification calls for a precise frequency generation which is usually achieved with either a crystal or accurate ceramic resonator, which both are for automotive grade costly.

In a lighting system as shown in Fig. 4 we do not have this requirement of obtaining bus access at any time with the need of access negotiation. Instead we have a controller defining the light functions and actuators setting the various light segments to a desired brightness. The controller also determines when to request diagnosis information and when to set con-

figuration data. This system architecture allows to implement a master – slave communication scheme with the controller as master and the actuators as slaves removing the need for arbitration on the bus line. The communication data rate is then determined by the master and the slaves synchronize to this frequency, which removes the need for accurate frequency determining components like crystals in the actuators which reduces the system cost significantly.

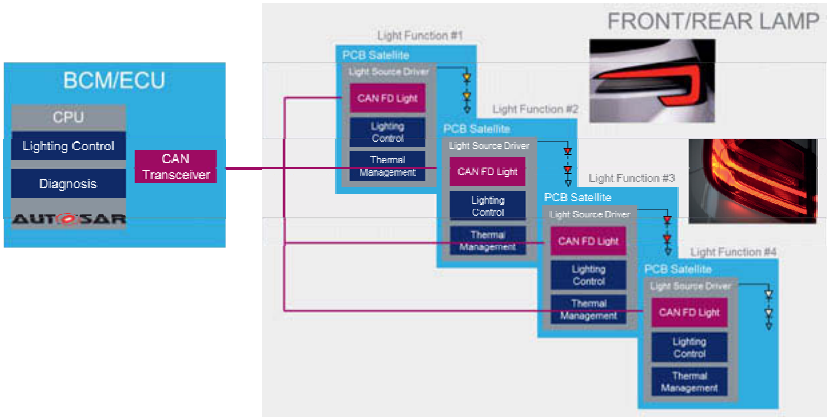


Fig. 4: Light control system

This master slave structure allows to keep the bus characteristic of the network. This permits to send messages to several slaves or even all of them within one CAN message increasing the utilization of the network.

A CAN FD based frame format has been chosen since it is designed for 64 bytes of data within one frame which is received by all actuators at the same time. Since arbitration is not needed anymore the message identifier can solely serve as an identifier and does not contain message priority information anymore. The message integrity is preserved by the cyclic redundancy check (CRC) as defined in the ISO 11898 protocol standard. Furthermore focusing on basic protocol functions allows a reduced cost implementation within the LED drivers while at the same time preserving the support of existing CAN FD hardware protocol controllers in existing microcontrollers.

Innovation by CAN FD Light: use of existing infrastructure while reducing system complexity

With the upcoming of complex animated light functions the need for a bus network to control these came up. Light patterns with more than 600 individual light sources with refresh rates of less than 50 ms have to be generated while at the same time diagnosis data has to be requested to ensure fault reaction times of less than 200 ms. This network has not only to provide the sufficient bandwidth but has also to ensure the integrity of the transmitted data. The implementation at the light driver side has to be low cost with no external components in small pin count packages since many driver circuits are used for the large number of lights. Since the light functions can be generated in low cost microcontrollers with limited computation power support by already existing hardware modules is beneficial. This is especially needed with regard to complex data integrity validation functions like the cyclic redundancy check with more than 16 bit and the encoding to ensure a sufficient edge density during data transmission for synchronization. A bus network where all participating devices are connected to the same wire is advantageous not only from the cost point of view with the need of only one transceiver per device but also for wiring flexibility and the low latency in comparison with structures like e.g. daisy chaining. For the envisioned application a data rate of 1 Mb/s is required. This data rate must be achieved without the need for frequency generating external components like crystals due to their high cost in the automotive environment.

This has been realized by using a master slave structure with a master that provides an accurate data rate to which the slaves synchronize. In the light system the master is the controller that generates the light pattern and sends the individual element brightness to the driver circuits. The data can be the on-/off time relation (PWM) or LED current settings. The master slave structure removes the need for arbitration because the slaves answer only upon master request which is addressed to only one slave that answers. Therefore the bus access is entirely controlled by the master.

On a bus network messages have to be clearly addressed to reach their receivers and to mark their sender. Additionally their content type has to be signalled in the message. This ensures that no misinterpretations of destinations and message purpose occur. Together with the cyclic-redundancy check these measures provide a high level of data integrity and safety.

So in conclusion a fitting network must be low cost with a minimum number of external components, it must use a bus architecture for maximum flexibility to cover a broad range of application needs, an addressing scheme labelling the destination, source and content type of the message, sufficient data payload size to support the high number of individual lighting

sources with low latency and minimum overhead, a high data rate to allow animations and diagnosis data exchange, safety provisions in case the message exchange fails and ways to detect corrupted data so the system can react within the given failure reaction time for functional safety.

A frame providing these features has an addressing part that can be extended to enclose message content, control bits to indicate the size of the data payload and a check sequence to make data bit failures detectable. The data frame must also include coding to ensure a sufficient high edge density. With a master slave architecture arbitration is not needed, also direct error signalling on the bus is not required since a master slave architecture can work with message timeouts like watchdogs given that erroneous messages can be detected. Ideally it can be examined with existing developing and debugging hard- and software, implemented with already built-in hardware modules to reduce the computation load of the microcontroller core and with standard MCAL software drivers in the AUTOSAR operating system.

The well-known CAN FD protocol provides all these features with additional overhead that is not needed for lighting applications. For example the data can be transferred at the same data rate as the control field, so no bit rate switching is needed. Arbitration is dropped, 11 bit addressing is sufficient and various control bits are not used. Error frame transmission and handling are also not needed, even though they can be treated if they occur. Since the CAN FD network is standardized and widely used conformance and interoperability tests are in place to ensure its operation in the harsh automotive environment. Hardware protocol controllers are available taking care of the message integrity checks and the insertion of stuff bits to ensure the needed edge-density for synchronization.

With a data payload of 64 bytes per frame 64 individual light sources with eight bit resolution can be set with one data frame. At a data rate of 1 Mb/s the CAN FD frame has a transmission time of less than 700µs. Therefore for programming 640 light sources 10 frames are needed with a total time of less than 7 ms. At a refresh rate of 20 ms this leaves 13 ms for diagnosis messages. Under the assumption that 16 bytes diagnosis data are transferred and received a diagnosis request with an answer takes about 400µs which makes it possible to get diagnosis data of 32 devices within these 13 ms. With a failure response time of maximum 200 ms enough time is left to validate the diagnosis data and to react properly in case a failure is detected.

The used data frame looks very close to the CAN FD frame shown in Fig. 5.

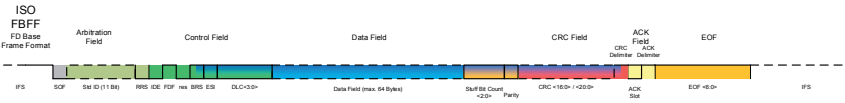


Fig. 5: ISO CAN FD Frame Format

The protocol, for data exchange uses a master – slave scheme. Therefore the satellites, which are the slaves, send only upon request from the master device. A master / slave scheme does not require a collision resolving method since during normal operation collisions are avoided. A collision is treated as error.

The master sends data to the slaves in defined intervals, which is received by all of them. This data stream can be used by the slaves as network heartbeat or watchdog. If it is not received within a defined time slot the slaves can enter their fail-safe (or limp-home) mode.

Data (e.g. diagnosis data) from the slaves is requested by the master using dedicated command frames. Only one addressed slave answers this request within a given time frame. This answer can be used by the master to detect the availability of the slaves.

The protocol uses only CAN FD format frames without bit rate switching and with standard identifier. Frames that are not supported by the implementation or are erroneous are ignored. Therefore all bits intended for changing these operation modes can be kept at their fixed values.

Fig. 6 shows the block diagram of the master as referred in Fig. 4. The protocol master uses a standard CAN FD protocol controller that may be already implemented in the automotive microcontroller in hardware or software. An additional CAN FD protocol extension software controls this protocol controller to make it act as CAN FD Light master and implements the master side of the CAN FD Light protocol

A standard CAN FD transceiver is used as the physical interface of the CAN FD network.

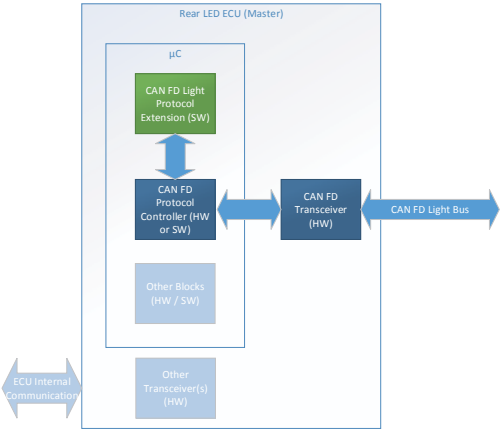


Fig. 6: Master Block Diagram

The CAN FD Light slave is located in the LED satellites as shown in Fig. 4 and there inside the light actuator device (Fig. 7)

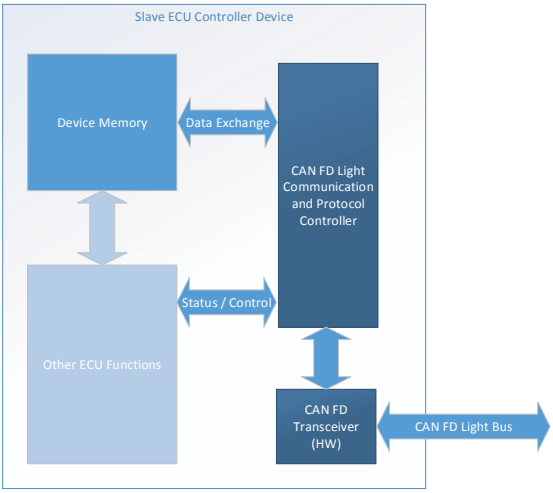


Fig. 7: Slave in light actuator device

The implementation is done in hardware to avoid embedded processors for running software. It consists of three parts as shown in Fig. 8:

- An embedded standard CAN FD transceiver according to ISO 11898-2
- A CAN FD Light protocol controller with a synchronizing oscillator for generating and sampling the data bits
- A communication protocol controller that sends and receives data to and from the protocol controller and controls the communication

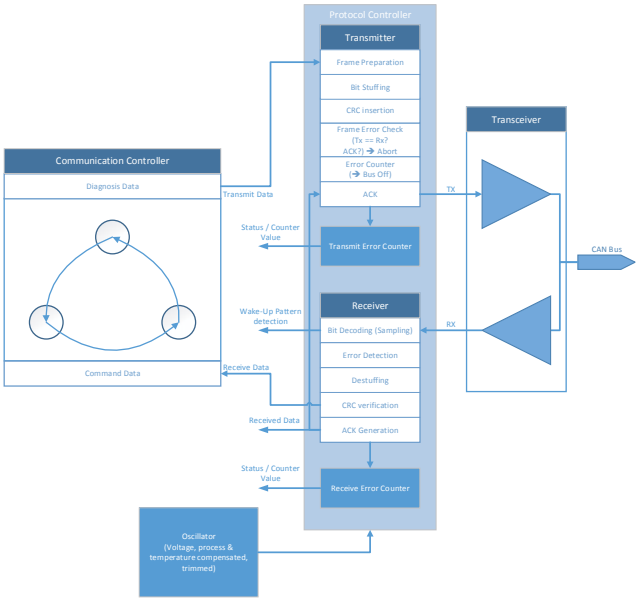


Fig. 8: CAN FD Light Communication and Protocol Controller

A bus network gives the advantage that a data frame sent by the master reaches all bus members at the same time. Therefore it is possible to use broadcast frames to set many light sources with a small addressing and integrity check overhead. The slaves then pick the data intended for them.

Data like diagnosis data is requested by data frames that are addressed to a dedicated device with a unique address. All other slaves ignore these messages. With these unicast messages registers in the addressed slaves can be programmed and data from dedicated

memory contents can be requested. Upon the reception of the frame the slave answers with its data addressed to the master. All other network participants ignore this communication. Apart from the destination and source device identifier memory addresses and content is transmitted to allow read and write operation to specific memory locations.

Application examples: How CAN FD Light can simplify the transition from traditional to dynamic LED / OLED lighting

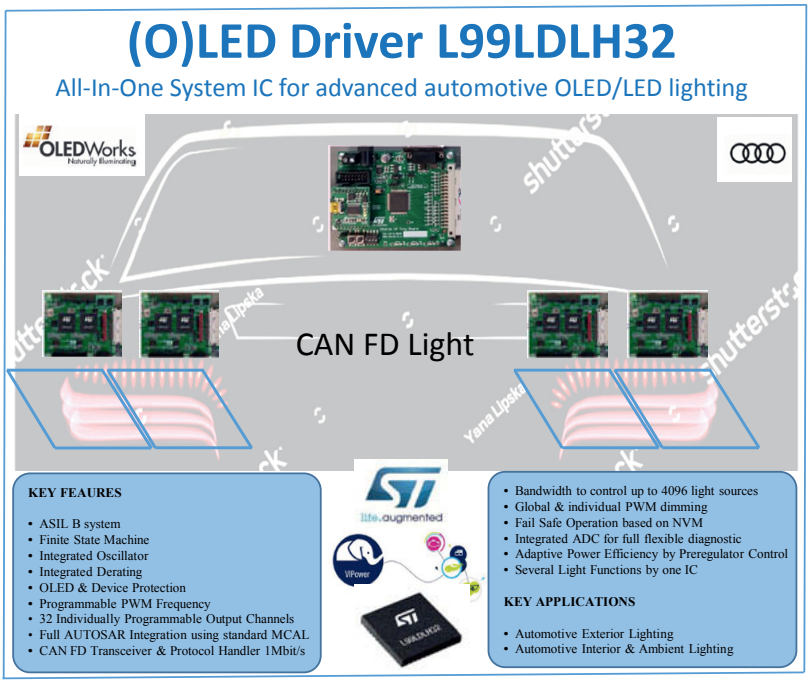


Fig. 9: CAN FD Light Demonstrator

The CAN FD Light offers the proper infrastructure to realize modern, customized, dynamic, high bandwidth automotive lighting functions with a lean, compact and robust solution. First LED / OLED Drivers integrating the CAN FD Light interface are available as prototypes. These drivers offer an “All-in-one” solution, integrating communication interface, power management, programmable logic and state machine, digital computing power, non-volatile memory, control and multichannel output driver stages, protection and complete diagnostic

coverage in one IC. All functions can be real time conFig.d, controlled and supervised through the CAN FD light interface enabling a system topology as shown in Fig. 4.

Conclusion

CAN FD Light is communication interface for controlling light functions with a high data rate for new complex light systems and animations with the flexibility needed for modern illumination architectures anywhere in the car.

Due to its specific features like master – slave communication and the ability to synchronize on the master data rate which enables a full integration into the light drivers without expensive additional components it allows a very cost efficient system implementation in the car.

Additional functions like broadcast messages to all bus participants and unicast diagnosis messages to access specific registers in dedicated devices make it a safe and reliable communication bus with a high fault detection capability with a low failure reaction time.

Since it is compliant to CAN FD and uses the same physical layer transceiver it provides the same interoperability and robustness. The communication can be used with the AUTOSAR drivers used for CAN FD and can be evaluated using the same already available tools.

These features make CAN FD Light a suitable bus communication system for the demanding requirements in modern car light systems.

References

[1] CAN with Flexible Data-Rate Specification Version 1.0 (released April 17th, 2012 - BOSCH)

[2] INTERNATIONAL STANDARD ISO 11898-1 (second Edition , 2015-12-15)

[3] Romeo Letor, Donato Tagliavia, Prof. Angelo Raciti, «STM is enabling a new era in automotive,» in AEIT 2018 Seminar, Catania, 2018.

[4] Dr. Arthur Mutter, Florian Hartwich - Robert Bosch GmbH, «Advantages of CAN FD Error detection mechanisms compared to Classical CAN,» in iCC 2015, 2015.

[5] «CIA - Cyclic redundancy check (CRC) in CAN frames,» [Online]. Available: https://www.can-cia.org/can-knowledge/can/crc/.

[6] «CIA - Cyclic redundancy check (CRC) in CAN frames,» [Online]. Available: https://www.can-cia.org/can-knowledge/can/can-fd/.

Digital Light – Function & Design on Demand utilized for Car2X Communication

Dr. **Michael Kruppa**, Dr. **W. Thomas**, AUDI AG, Ingolstadt

Abstract

Utilizing Car2Car communication modern mobility concepts will improve in typical standards like safety and comfort. While driving this evolution of connected machines, it is also very important to keep a high level of Car2X communication to non-machine interfaces like human beings. Communication with pedestrians or human drivers is needed to establish a healthy acceptance of autonomous driving vehicles. For this communication pathway light plays a very important role. Modern vehicles have to implement an exterior communication content easy to understand and easy to recognize during daylight and nighttime. Therefore, new light sources and optimized car architectures are required to provide this Car2Human communication while maintaining legal requirements. Audi's approach to digital light is the first step into a revolution of exterior lighting. Combining the benefit of digital light sources and displays with highly connected cars will completely change the way light is used in exterior lighting.

Car2X Communication

In the United States a self-driving car has successfully completed a test drive in total darkness. The test car, a Ford Fusion Hybrid, used laser-based radar as its main road-sensing system and finished a winding track in the Arizona desert. The so-called Lidar system relies on high-resolution 3D maps. It complements the cameras installed on the outside of cars, which help navigating during daytime. Thanks to Lidar technology, self-driving cars are not reliant on visibility conditions. So why do we need automotive lighting in the future at all?

But before discussing this provoking question there is at first to clarify the different levels of automated driving. Only in the highest level we really have autonomous driving cars, but in the near future we will still have all cars equipped with modern front lighting systems.

SAE (standard J3016) and VDA published a very detailed description about automation levels where they distinguish between assisted, piloted and autonomous driving. All existing systems at the moment are assisting systems. Even if it is only working in traffic jam situations. Level 5 with fully autonomous driving cars will be a technology for the next decade. Nevertheless we already have to discuss about future needs and how the in between steps for technologies and regulations should or could look like.

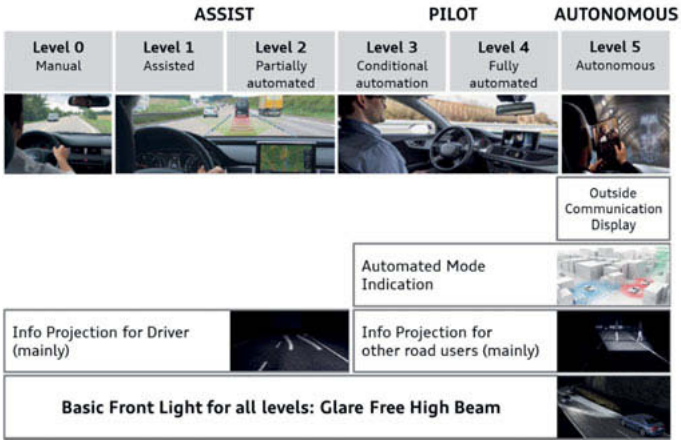


Fig. 1: Overview of different levels of automation according to SAE/VDA

Figure 1 shows that we will still need our well known automotive lighting for the piloted level with conditional automation or fully automated driving. Nevertheless even with autonomous driving cars we will need our existing lighting for all situations where the driver has the wish to pilot the car by himself.

Nevertheless in all cases it makes sense to indicate the driving mode. Other road users need an indication to at least have the chance to realize that the car is controlled by computers and not from the obviously inattentive driver. This helps the other road users for inspiring confidence in this new technology.

A non-discussible need for communication between pedestrians, bicyclists or manual driving cars is in every open situation where all participants have to decide about e. g. the priority at crossings, pedestrian walks or other cases. Classical communication between drivers and road users outside the vehicle, such as making eye contact, nodding one's head or giving hand signs may will no longer be possible. Looking for contact with the driver will cause confusion in case of a distracted driver who is not concentrated on the road traffic.

The car-to-car communication between autonomous driving cars will not be the specific problem in the future. There are already standardization discussions for the signals and information via automotive WLAN. The HMI between pedestrians etc. and automated vehicles is not standardized and even the communication way is completely open. Possible HMI could be sound or speech, projections on the road or on the windshield, LED-displays on the roof, in the front grill, the hood, the doors or the windshield or LED light strips around the vehicle or similar to a

CHMSL in the front or in the rear. Last but not least even portable or wearable devices like smartphones, activity tracker, smart watches, or smart glasses could be addressed by the car. For an overall evaluation of the different HMI possibilities the pros and cons have to be discussed not only for every OEM, but also for every market or homologation area. Visual contact alone is a small risk for a successful communication, sound could help. Visual contact depends on weather conditions, environmental brightness, viewing angle and glare. The communication could be 2-path-exchange, bidirectional or even for several persons accessible. The complexity of the information should be low, the meaning should be unambiguous and clear. Distance, infrastructure and signal noise are also points to take into account.

Projection and Display Light for Car2X Communication

Fig. 2: Possible lighting signals in the windshield and in the front, but what is the meaning of the signals? Is it the right color? Is it the right wording?

Considering the importance of visual communication light should play a central role to secure and reassure other road users. The meaning and the appearance of the signals and information should be standardized for all markets. In a world of globalization it is even a must. Writing is to depending on the font, Chinese or Russian types are not understandable for the most European people and vice versa.

There is a lot of research necessary to define the right signals in any situation to increase safety. Afterwards standardization and even marketing for the new information and safety signals is necessary. People will have to learn this new signals. The more intuitive they are the better.

The technology for new signals may vary in the future. Lamps could be the most conventional technology for indicating any light signal. LEDs would help to reduce package, power consumption and combined with more intelligent electronic the signals could even get dynamic or adaptive to any environment conditions. High definition projection systems similar to video

projectors with DMD, MEMS or LCD technology are new upcoming technical possibilities to project even more complex and adaptive signals on the road.

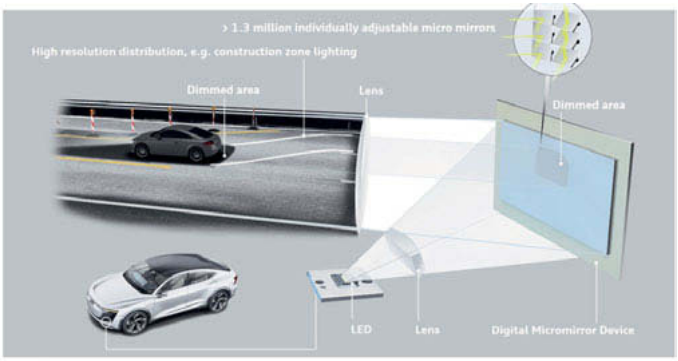


Fig. 3: DMD for front light projection

The projection area can extend the communication area in a much bigger radius around the car than a signal just on the car.

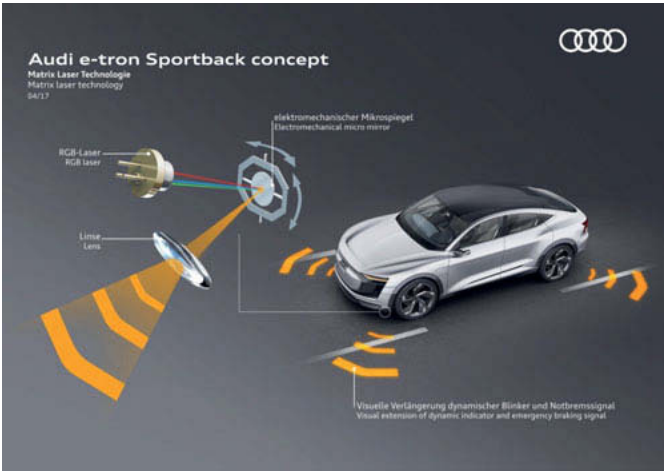


Fig. 4: Visual projection with Matrix Laser Technology (RGB laser and MEMS)

The advantage of a MEMS projection system with RGB laser is the possibility to project colored signals according to the lighting regulations in white, amber or red. Also the intensity can be varied easily.

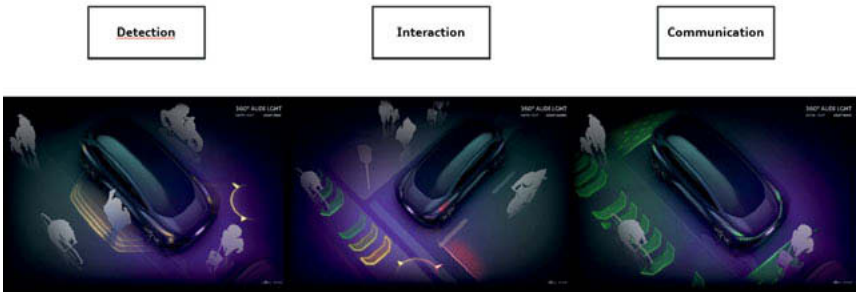


Fig. 5: Vision of future Car2X communication with projected light

In the front the technology for AFS and ADB becomes higher resolutions by increasing numbers of LEDs that can be addressed singular. For the rear we can expect higher resolution by new technologies too. LCD or DMD has been shown in several pre-development studies, but also OLED-displays are under evaluation and will be addressed in the chapter underneath. These technologies will lead to more complex communication opportunities directly on or close to the outer shape of a car. Years ago Audi showed the vision “The Swarm” with possible communication to the environment by moving light points, that could change color, intensity, speed and direction to show conventional signals like braking, direction indicating and just tail light with all the time moving and changing light dots. “The Swarm” was even able to show the following driver where the car is going to drive. All signals were intuitively to understand, although the display format was not according existing lighting regulations.

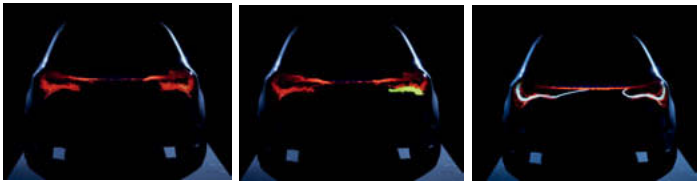


Fig. 6: Audi “The Swarm” display for communication on the rear with dynamic moving light dots instead of static reflectors and lenses

Most existing light source can be enriched with above mentioned functionalities. Regarding rear combination lighting display lighting is of higher importance than projection lighting. Therefore a recently introduced lighting technology is best in class performing for the needed approach of Car2X communication – OLED.

DIGITAL OLED = Most flexible Display Technology for Personalization and Car2X communication in Taillighting

The feature of high contrast at small distances between segments within one OLED tile can be beneficially used for new applications. Today, only a low number of segments (N<5) is used, like in the AUDI A8 where segmentation is used for novel animation effects. New innovative approaches by Audi and OLED suppliers like OLEDworks and LG-Display are modifying the complete OLED into a highly segmented display. While utilizing existing processes and materials to produce this segmented light source is fulfilling all known reliability requirements for automotive exterior lighting.

Fig. 7: OLED Evolution from Audi TTRS toward Digital OLED @ Audi

More than 50 segments per OLED tile are opening up more or less an infinite number of combinations of different segment states. Having those installed in taillight applications it is possible to change the shape and appearance of the taillight signature by simple digital information. Hence, the OLED light source enables the personalisation of the tail light design. Therefor it is not needed to change the hardware of a rear combination lamp; plenty of designs can be covered due to specific OLED and OLED-segment design, as illustrated in Figure 8.

Fig. 8: Schematic depiction for software based variation of OLED tile illumination

In contrast to display technologies, each segment can be designed individually and no regular (grid-) pattern is needed. Hence, significantly different segment shapes and sizes within one OLED panel are possible. Thus, precise patterns can be directly created by the segment shape instead of using a multitude of tiny OLED segments requiring complex driver electronics and suffering from edge aliasing effects. In addition, high luminance values of ≥ 2.000cd/m² at deep

red colour coordinates (λ_{dom} ≥627nm) are already possible, outperforming display based approaches in this application, by far.

By following this approach, it is also possible to develop one OLED module that can be used in many different cars while still being able to provide an individual design of every taillight signature. This is opening up a huge potential to initiate a dramatic cost down process for OLED applications also in A- and B-segmented cars.

Next to design driven modifications of the taillight signature, it is now also possible to take this highly segmented taillight to display further information in the rear of a car, e.g. for following traffic or pedestrians. While sticking to the possibilities that are provided by current regulation, digital information within the car or provided by the swarm of connected cars and infrastructure can help to improve the safety and gain trust in the field of autonomous driving cars. Considering alternative technical solutions for this rear car communication – LED displays – are unable to compete with the package, homogeneity and contrast ration and flexible segment configuration of a multi segment OLED panel that we name *DIGITAL OLED*. The functional use of this feature has just started and will leap-frog all other standard approaches utilizing LEDs.

After having installed the mentioned high number of segments the real revolution is accomplished by combining the multi-segmentation with the possibility to use flexible substrate technologies. Today, all mass production OLED tail lamps are utilizing planar glass substrates limiting the degrees of freedom in integrating the 2D OLED panels into a 3D curved lamp design. To further utilize the complete package of Audi tail lamps, OLEDs need to follow the wrap around of the car. Hence, the application area of OLEDs in exterior lighting can be greatly increased and the entire lamp can be covered with OLED panels.

For this, flexible substrate technologies are needed and have to be adapted for the usage in automotive lighting applications. The flexibility of the substrates can be used as means to bend the OLED in the lamp production, creating a 3D OLED module that fits the curvature of the car.

Fig. 9: Principle illustration, flexible OLED following curvature of car shape and lamp curvature of AUDI TT

In combination with its viewing angle independent color point stability and homogeneity, light weighted thin flexible OLEDs will be a unique light source for tail lamps which can never be mimicked by any other light source. The most flexible display for exterior lighting is born: *FLEXIBLE-DIGITAL OLED.*

First successful developments of OLEDWorks in combination with willow-glas from Corning are opening this path into a new display lighting future.

Summary and Outlook

Automated Driving and Digitization will revolutionize our automotive lighting business. New technologies will extend the communication with other traffic participants. This will lead to increasing traffic safety. On the other hand the light distribution will get completely personalized for safety but also for entertainment and commercial possibilities. Enriching the user experience will lead to more attractive lighting products and will lead to higher take rates and faster market penetration. This again will bring the modern lighting technology easier into market and can then increase traffic safety for all participants. All these ideas could be an opportunity for the automotive industry, if development or regulation takes too long then IT-companies and service companies will take over this market and the chance to finance increasing car safety with new business plans will be gone.

References

[1] Huhn W.: The Influence of Future Mega Trends on Car Lighting, Joanneum Research Zukunftskonferenz Graz, 2017

[2] Huhn W.: Autonomous Driving, Front and Rear Trends, GTB Lighting Forum, 2017

[3] Burkert A.: Interview, ATZ 02/119 (pages 22-26), 2017

[4] Pape T.: Light Years Ahead, Encounter Technology 1/2016 (page 54-59), 2016

[5] Rabenau P.: Flexible OLEDs in Rear Lighting, ATZ 02/119 (pages 16-21), 2017

[6] Thomas W., Rabenau P., Lendle R. : OLED Technology, Electronic Automotive 12/2015 (pages 42-47), 2015

[7] Hamm M.: Today‘s and Future Technologies meet Rules, Regulations and Reality, DVN Workshop Rochester, 2017

[8] Omerbegovic S., Kammann T., Funk C., Neumann C.: Artificial intelligence for future light-based assistance systems, 11th ISAL, 2015

[9] Berlitz, S.: Automated Driving and Vehicle Lighting, OICA GEE, 2016

[10] Kruppa, M., and Thomas, W.: “An OLED Taillight Revolution – From Point Light Sources to Area Light Sources” Vision 2018 – Proceedings (2019)

See you soon!

Save the date

ELIO Marketplace 2020

October 13-14, 2020, Baden-Baden

www.eliv-marketplace.de

ELIO 2021

October 20-21, 2021, Bonn

www.eliv-congress.com