Towards Safety Concepts for Automated Vehicles by the Example of the Project UNICARagil

Torben **Stolte**[*]; Robert **Graubohm**[*]; Inga **Jatzkowski**[*]; Markus **Maurer**
TU Braunschweig, Institut für Regelungstechnik, Braunschweig, Germany

Stefan **Ackermann**[*]; Björn **Klamann**[*]; Moritz **Lippert**[*]; Hermann **Winner**
TU Darmstadt, Fachgebiet Fahrzeugtechnik, Darmstadt, Germany

Contact: stolte@ifr.ing.tu-bs.de

*The marked authors are first authors with equal contribution.

Summary

Striving towards deployment of SAE level 4+ vehicles in public traffic, researchers and developers face several challenges due to the targeted operation in an open environment. Due to the absence of a human supervisor, ensuring and validating safety while driving automatically is one of the key challenges. The arising complexity of the technical system must be handled during the entire research and development process. In this contribution, we outline the coherence of different safety-activities in the research project UNICAR*agil*. We derive high-level safety requirements and present the central safety mechanisms applied to automated driving. Moreover, we outline the approaches of the project UNICAR*agil* to address the validation challenge for automated vehicles. In order to demonstrate the overall approach towards a coherent safety argumentation, the connection of high-level safety requirements, safety mechanisms, as well as validation approaches is illustrated by means of a selected example scenario.

1 Motivation

Currently, intensive research targets the operation of automated vehicles at SAE levels 4+ [1]. Still, many challenges must be solved before automated vehicles can be deployed in public traffic. A key challenge is the proof of sufficient safety; however, the exact meaning of the term "sufficient safety" still needs to be determined. It is reasonable to assume that the required safety level is higher than that of human drivers [2]. However, a residual risk always remains, which is inherent to automated vehicles and which is caused by several uncertainties encountered throughout the development and operation, cf. [3].

Due to the complex interaction of a multitude of system components, safety is an emergent property of automated vehicles [4]. The overall safety of the automated operation can only be ensured if developers succeed in managing the complexity arising from the vehicle automation task. Thus, the design of a coherent system-wide safety concept strongly benefits from safety activities that start from the beginning of the research phase of automated vehicles. This is one of the key aspects considered in the project UNICAR*agil*, a joint project carried out by a consortium of eight German universities and six industrial partners.

In the following sections, we will outline the interconnection of different safety activities in the project UNICAR*agil*. After a brief introduction of the project, Section 2 specifies safety aspects considered in UNICAR*agil*, which are the focus of the paper. Section 3 presents an example scenario which is used to illustrate the different safety aspects in the later sections. The process of the initial safety analyses is described in Section 4. The safety requirements derived in these analyses serve as input for subsequent process steps. An approach to derive additional safety-relevant requirements based on road segment characteristics is illustrated in Section 5. Subsequently, Section 6 highlights the interaction of the three central safety mechanisms with each other as well as their connection to the previously derived safety requirements. Lastly, the verification and validation activities concerning the safety mechanisms are outlined in Section 7.

2 Project, system, and process context

In order to further frame the focus of the paper, its context is outlined in this section. After a brief introduction of the project UNICAR*agil*, Subsection 2.1 introduces different safety perspectives taken in the project. In Subsection 2.2, three central safety mechanisms of the project are outlined. Subsection 2.3 relates the paper contents to a development process.

2.1 Project context

The project UNICAR*agil* aims at developing modular architectures for future SAE level 4 automated vehicles. Based on architectural approaches in fields such as mechanics, electrics/electronics, and software, four prototypes are developed which represent different use cases for future automated vehicles. Orthogonal to different architectural viewpoints following Bagschik et al. [4], safety and security aspects are taken into account as one of the key aspects. A comprehensive project description of the project UNICAR*agil* can be found in [5], [6], and [7].

A key challenge of the project is to handle the complexity of the system that arises from different sources. On the one hand, automating a vehicle is a very complex task itself, since vehicle automation causes a much stronger interconnection between the architectural viewpoints. On the other hand, the project goals shall be reached within four years while cooperating in a comparably small, heterogeneous team in an agile environment over different locations throughout Germany.

Although UNICAR*agil* will surely not solve the challenge of building market-ready automated vehicles, one of its main contributions will be in providing approaches to mastering complexity. For instance, one approach to handle the complexity is to shape modules of components that can be considered as independent as possible. Concerning the safety of automated vehicles, we have the opportunity to design safety into the system from the start of research and to consider the interplay of different safety perspectives rather than researching single aspects or components. Moreover, we can report openly about safety approaches as no non-disclosure restrictions apply.

In UNICAR*agil*, we distinguish five different perspectives on safety and security; cf. Klamann et al. [7] following the Waymo Safety Report [9]:

- *Behavioral safety* focuses on the externally observable behavior of vehicles. It defines safe vehicle behavior independent of specific technological solutions. Behavioral safety addresses the potential causes of unsafe behavior beyond functional safety, such as by wrong, missing, or conflicting situation-specific requirements. The perspective of behavioral safety not only comprises but also goes beyond the safety of the intended functionality as defined in ISO/PAS 21448 [10].
- *Functional safety* concentrates on avoiding or mitigating system failures due to faults in the electronic hardware and the software that is running the vehicle following the ISO 26262 standard [11].
- *Crash safety* aims at reducing the impact of collisions on both occupants of the vehicle and other traffic participants potentially involved in a collision.
- *IT security* targets the prevention of intentional manipulation of the vehicle control system, e.g., by attacking the internal or external communication or by manipulating the software.
- *Operational safety* comprises safety aspects that can occur as a result of the interaction between humans and vehicles apart from the actual dynamic driving task (see [1] for definition). For instance, safety issues of automatically opening and closing door systems or during maintenance are parts of operational safety.

The remainder of this paper is based on *behavioral safety* which is technology-agnostic. This safety perspective allows us to derive safety requirements at the vehicle level by purely considering the functionality. In the subsequent development steps, these safety requirements will be allocated to system functionalities as well as partially to hardware and software components. Approaches targeting *functional safety* developed in the project UNICAR*agil* can be found in, e.g., [12], [13], [14], [15], and [16].

2.2 System context

The initial system concept of UNICAR*agil* includes three central safety mechanisms: self-perception, the Safe Halt functionality, and the capability-based route planning. Self-perception is a crucial functionality for safely operating automated vehicles. Driving decisions and trajectory planning cannot rely solely on the perceived environment, but must also take into account the current and future capabilities of the vehicle, see e.g. [17] and [1]. Several internal and external conditions can degrade the nominal performance of the vehicle. Besides physical degradations of, i.a., vehicle actuators or sensors, also environmental conditions such as rain, fog, etc. require an adaption of the vehicle behavior. The self-perception functionality accesses the available diagnostic information of system components and computes quality measures for functionalities necessary for the execution of the dynamic driving task.

The quality measures computed by the self-perception are used twofold in context of this paper, namely for the capability-based route planning as well as for the execution of the dynamic driving task. The capability-based route planning compares the vehicle's (current) capabilities with those required for driving on certain road segments. Thus, routes can be planned that avoid road segments with performance requirements that exceed the vehicle's capabilities. Again, the performance measures stemming from the self-perception are an important input and allow to re-plan routes online if the vehicle's capabilities degrade during run-time such that certain road segments cannot be passed safely.

Moreover, the quality measures computed by the self-perception are an important input for safe dynamic driving task execution. In case of partial degradations, e.g., a reduced sensor range, the vehicle will adapt its performance based on the inputs stemming from the self-perception. Still, the vehicle is able to fulfill the dynamic driving task safely. In case of severe degradations, such as a complete loss of the primary perception system, the vehicle automation system is not able to execute the dynamic driving task. Hence, a mechanism to handle severe degradations is required as fallback. This fallback is the Safe Halt functionality which serves as the final fail-safe procedure in case system parts superimposed on the trajectory tracking fail (therefore, the trajectory tracking functionality is designed in a fail-operational manner in UNICAR*agil*). In such a case, the vehicle will follow a previously computed emergency path leading to a risk-minimized stop. Moreover, collisions with obstacles emerging in the driving corridor are evaded by using a dedicated a backup sensor array.

2.3 Process context

From a development process perspective, this paper addresses different development steps. Following the reference process of ISO 26262, we mainly focus on two development phases which are closely connected via the focus on behavioral safety. The following sections describe the undertaken scenario-based approach using the example illustrated in Section 3. The first focus is on the concept phase, the second focus is on the validation phase which is related to the start and the end of the ISO 26262 reference process, respectively. For the concept phase, we show in the subsequent sections how high-level safety requirements stemming from safety analyses (Section 4) and road segment characteristics (Section 5) are derived and connected to the requirements for the system functionalities of self-perception, Safe Halt and capability-based route planning (Section 6). The undertaken modular validation approach of UNICAR*agil* is outlined in Section 7.

3 Example scenario

To illustrate the safety considerations outlined in the following sections, we use an example scenario as depicted in Fig. 1. We will employ three different variants of this scenario. In all three variants, the ego vehicle starts in the south heading north. The vehicle is driving on a one-way street and is approaching an intersection to the right. It can either continue on the one-way street (variants Ⅰ and Ⅲ) or turn right into a street with one lane in each direction (variant Ⅱ). Coming from the right, another vehicle stops at the stop line and gives way to the ego vehicle. Continuing on the one-way street, the vehicle leaves an open area without buildings and other towering objects. It enters an area with parked cars and multi-story buildings on both sides of the lane. In variant Ⅰ, the vehicle continues on the one-way street. Thereby, it passes a bus which is parked on the right and occupies parts of the driving lane. In variant Ⅲ, the vehicle uses a parking space behind the bus for a safe stop in the operating mode Safe Halt. In contrast, when turning right in variant Ⅱ, the vehicle takes a detour in order to avoid potentially more challenging conditions of variants Ⅰ and Ⅲ.

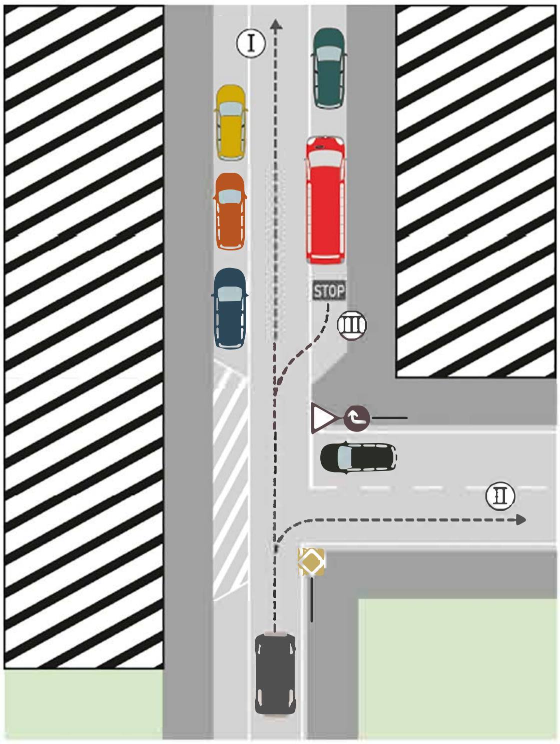

Fig. 1 Abstract bird's eye view illustration of the reference scenario used in this paper. The ego vehicle starts in the south heading north. The three possible variants of the scenario are visualized by the reference paths Ⅰ to Ⅲ (dashed lines). Green and striped areas are open spaces and multi-story buildings, respectively.

4 Fundamental safety analysis

The strategy for systematic hazard identification established in UNICAR*agil* is presented in [18]. Our hazard analysis and risk assessment process is based on externally observable malfunctioning behaviors instead of preliminary and potentially incomplete malfunctions which is the “traditional” approach. In this section, we demonstrate the usage of our hazard identification strategy in the scenario illustrated in Section 3. Based on the identified hazards, we can specify safety goals and indicate some requirements that support the mitigation of emerging risks.

To demonstrate a hazard analysis for the operational scenario depicted in Fig. 1, we assume that the automated vehicle is in its normal operating mode and follows path I without any collision (scenario variant I). In short, the automated ego vehicle follows its lane and keeps an adequate speed throughout the considered time interval. It has right of way at the junction and the other vehicle does not indicate any intention to cross the stop line before the ego vehicle has passed. Therefore, the ego vehicle does not significantly reduce its speed before entering the junction. However, it adapts its lateral position within the lane when passing the parked bus in order to establish an adequate distance.

4.1 Hazard identification

We induce generic malfunctioning behaviors into the described scenario (variant I) and investigate hazards that are a potential outcome of this combination. Generally, any physically possible lateral and longitudinal deviation from a desired behavior is conceivable in a driverless car. Depending on the specified behavior of an individual vehicle in the operational scenario the list of deviations, however, can be reduced to the few that are relevant in the hazard identification process. In the illustrated scenario, for example, the automated vehicle is not expected to show significant acceleration at any point in time. Thus, potentially hazardous scenarios cannot be based on the absence of required acceleration or deceleration. Specifically, possible behavior deviations of the automated vehicle in the scenario are:

- Absence of required lateral position adjustment
- Improper acceleration
- Improper (rapid) deceleration
- Improper course angle changes

At this point, we distinguish between positive acceleration and negative acceleration (deceleration) due to the dissimilar character of the resulting externally visible vehicle behavior and its consequences in the context of a specific scenario. Applying the itemized deviations individually to our scenario at any given time step provides a basis for specifying concrete hazards. At this point, we detail all scenarios that include actual hazards to the health and life of humans or clear violations of traffic law, since these are considered general hazardous behavior.

Analyzing the scenario used in this paper (variant I), we find the following scenarios to include hazards to health and life of vehicle passengers and other traffic participants:

- The vehicle decelerates strongly during lane following.
- The vehicle does not adjust its lateral position when approaching the narrowing caused by the bus that is partially parked in the traffic lane. The inadequate lateral positioning leads to a collision between the vehicle and the parked bus.
- The vehicle changes its course angle towards the roadside, leading to a collision between the vehicle and one or many parked vehicles.
- The vehicle accelerates when driving in the narrow street. The combination of high speed and present uncertainties leads to a collision between the vehicle and one or many parked vehicles.

4.2 Safety goals

The identified scenarios include potential consequences of present hazards (e.g., head and brain injury of vehicle passengers due to the forces of a collision). In the ISO 26262 safety life cycle [11], the examination of these consequences is the next task in the hazard analysis process. Subsequently, risk assessment is performed for the examined scenarios complemented by a specification of safety goals that distinctly mitigate the risk. The assignment of automotive safety integrity levels (ASIL) to safety goals is based on the classification of severity, exposure, and controllability in hazardous scenarios.

Most of the specified hazards due to malfunctioning behaviors in the examined operational scenario are mitigated by a single abstract safety goal:

SG_1 *The vehicle shall move within its lane boundaries during lane following.*

However, the bus that is partially parked in the traffic lane requires us to add another safety goal, to express the required lateral position adjustment:

SG_2 *The vehicle shall pass relevant objects with adequate lateral distance.*

These are just two examples of safety goals. While the presented hazard analysis reveals additional requirements, such as avoiding unnecessary rapid deceleration, we do not discuss additional safety goals for the sake of brevity. Concerning the two specified safety goals, we can conclude that the risks they mitigate in the illustrated scenario likely cause ASIL-C-classifications: We assume a high probability of exposure regarding driving with inner-city speeds while passing parked vehicles in close distance to the lane boundaries (exposure class E4). Additionally, passengers or other traffic participants cannot perform actions to avoid harm when the malfunctioning behavior occurs. This makes the hazardous scenario uncontrollable (controllability class C3). Finally, collisions at presumed speeds of city traffic can result in severe or life-threatening injuries of vehicle occupants (severity class S2).

4.3 Functional safety concept

The formulation of a functional safety concept is the subsequent task in the safety life cycle of ISO 26262. Based on the safety goals and their associated ASIL-classes, the safety concept specifies mitigation strategies and implementation-independent requirements that are allocated to functional components. The interrelations of the different work products and artifacts established in the concept phase are shown in Fig. 2.

4.3.1 Functional component-oriented requirements elicitation

While an introduction of the three work products of the ISO 26262 concept phase potentially indicates a sequential and linear safety development process, in real applications, like UNICAR*agil*, we find the concept phase to manifest features of agile development i.e. artifacts are created iteratively and incrementally. The functional specification and context information of the system under development expressed in an item definition is continuously changed and extended in preliminary design stages. Therefore, the safety requirements also require ongoing revisions, e.g., due to novel architectural assumptions. In previous work, we introduce item refinements (i.e. adaptations of the item definition) that are triggered by safety analyses and potentially have wide consequences for the set of safety requirements [20], [21].

In this paper, we want to provide some examples of functional safety requirements that result from breaking down the two specified safety goals while considering the functional structure of the system. For later reference, consecutive identifiers are used for all specified functional safety requirements, which do not necessarily correspond to the labels internally used in UNICAR*agil*.

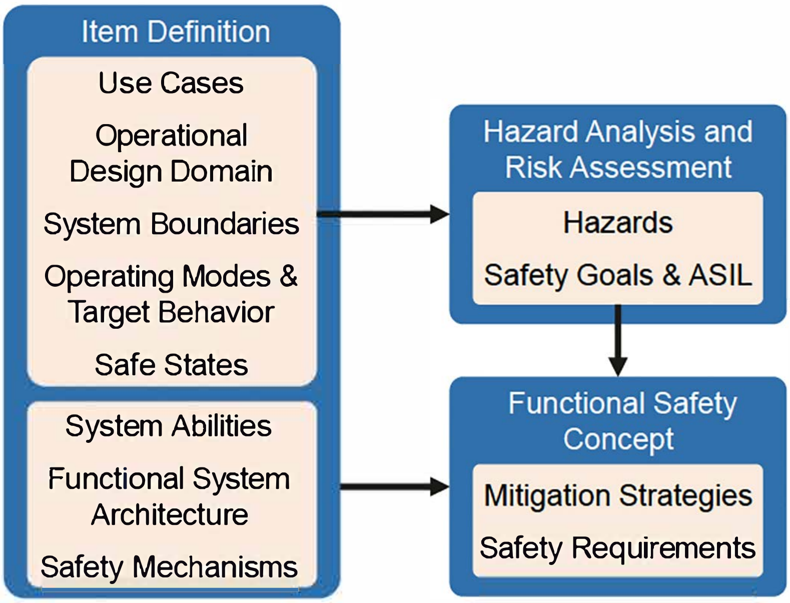

Fig. 2 Functional safety concept design [19].

Initially focusing on the perception system of an automated vehicle, we can identify the need to include elements that are crucial for the fulfillment of the safety goals in its environment model:

FSR_01 *The system shall detect relevant objects in the vehicle's environment.*

FSR_02 *The system shall include relevant objects in the transmitted environment model.*

FSR_03 *The system shall include lane boundaries in the transmitted environment model during normal operation.*

Due to the implementation-independent character of functional safety requirements, the source of the required lane boundary information is not specified at this point such that lane boundaries could be obtained, for example, from map data or detected lane markings. The functional component primarily processing the information of the perception system is the vehicle's guidance system. To fulfill the main safety goals in the examined scenario, the guidance system has to include the specified contents of the environment system:

FSR_04 *The system shall plan paths that realize adequate distance to relevant objects.*

FSR_05 *The system shall plan paths that realize adequate distance to lane boundaries during lane following.*

Additional requirements for the guidance system introduce the mitigation strategy of establishing a Safe Halt functionality as a fallback (cf. Subsection 6.3):

FSR_06 *The system shall permanently plan safe halt paths that lead to a safe stop location in the vehicle's environment.*

FSR_07 *The system shall follow a safe halt path in case of significant degradation.*

In order to detect significant degradations in the UNICAR*agil* vehicles, we aim towards contributions for self-awareness of automated vehicles (cf. Subsection 6.1). This ambition relies on the establishment of advanced monitoring measures in the self-perception system:

FSR_08 *The system shall monitor safety-relevant degradations in all components and communicate the result to other controllers.*

In order to allow the self-perception to perform this task, we additionally have to require all other relevant functional components (localization, environment perception, guidance, and stabilization) to communicate their status information:

FSR_09 *The systems shall indicate their current quality of execution in status messages transmitted to other controllers.*

4.3.2 Requirements elicitation via skill graphs

The safety concept in UNICAR*agil* incorporates ability and skill graphs introduced by Reschka et al. [22] at multiple stages of design. These graphs model skills that a vehicle must possess in order to be able to realize a certain behavior as well as the dependencies between these skills. The required behavior of a vehicle will also be called general capabilities in Section 5. For functional safety concepts, Nolte et al. [23] describe the utilization of skill graphs for a systematic identification of functional requirements.

For variant I of the example scenario in Fig. 1, the vehicle shall follow the lane leading north. In order to realize lane following behavior, the vehicle requires the skill of estimating its position and orientation relative to the lane boundaries. It also requires other perception, action, and planning skills, but for the sake of this example, we will focus this one required skill of lane relative position and orientation estimation. This estimation can be achieved, i.a., by using a highly accurate map of the lane boundaries in combination with a signal from a global navigation satellite system (GNSS). See Fig. 3 for an example of how these skills and their dependencies can be modeled in a graph. This graph segment is part of a larger, more comprehensive graph for a lane following behavior, which encompasses all skills necessary to realize this behavior with the dependencies between them. An example of a complete skill graph for a follow behavior can be found, e.g., in [23].

As stated by Nolte et al. [23], based on skill graphs, functional requirements for each skill can be derived from the safety goals identified during the hazard analysis process. In our example, we can derive safety requirements for the localization system of an automated vehicle. For instance, to stay within lane boundaries and thus fulfill safety goal SG_1, the vehicle has to be aware of its lane-relative pose:

FSR_10 *The systems shall* estimate *the vehicle's position and orientation relative to the current lane with sufficient accuracy*.

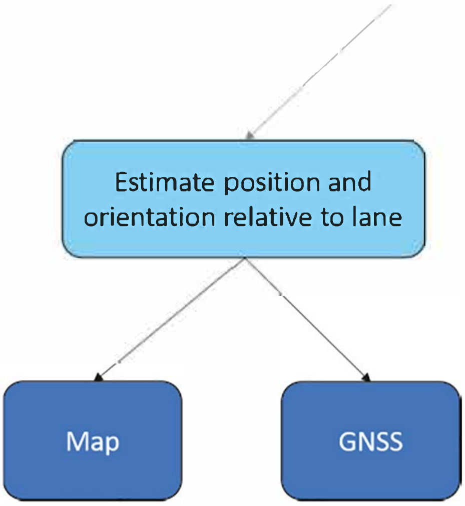

Fig. 3 Example of skills necessary for lane following and dependencies between these skills. Directed edges indicate a "depends on" relationship.

The examples of functional safety requirements deduced and presented in this section provide an insight on how we aim at effectively reducing the risk of a collision with parked vehicles in the illustrated scenario. On the one hand, generic safety goals for safe externally visible safety behavior are translated into performance requirements for individual components. On the other hand, safety mechanisms like the Safe Halt functionality are adopted in the set of safety requirements to establish strategies in order to ensure the fulfillment of all safety goals in cases of failures or problematic degradations of individual components.

5 Road segment categories

Driverless, automated driving on public roads requires a safety approval and a mechanism that identifies drivable routes through a given road network with respect to the safety validation after the approval is done. In UNICAR*agil*, we develop a new road-based approach to support and realize those two applications. Consequently, two main objectives are pursued within this project regarding that approach:

- Pre-approval: Support of the safety validation during the development process of automated, driverless vehicles
- Post-approval: Support of safety during operation

As described in [5], the main idea of the approach is a rearrangement of the parameter space that has to be considered for a safety validation. By sorting and grouping the parameter space into road-segment-based categories, the number of required test scenarios can be reduced. This section explains the road segment categorization and provides examples of how this approach is applied in the project to support safety validation. As an additional benefit of this method, a route-based online safety check can be used as a feature for vehicle operation after the safety approval is done which is outlined in Subsection 6.2.

As stated in Subsection 4.3, an item definition is required during the development process of automated, driverless vehicles. It defines the Operational Design Domain (ODD) [1] which specifies operating conditions regarding the vehicle environment, the externally visible vehicle behavior, and the states of the vehicle. In order to prove safety for an automated driving system, this information is necessary in an early stage of the development process [10]. The vehicle environment provides the motion space for the vehicle representing areas of the road or road elements on which vehicles are allowed to drive. Inside this motion space, there is a permanent interaction between the ego vehicle and the static as well as dynamic environment. This interaction results mainly from actions and reactions of the involved traffic participants or, to some extent, infrastructure objects. Vehicles interact with each other as well as vehicles react to, for example, traffic lights accordingly. In one specific motion space at least one or usually multiple interactions possibly take place. Motion space driving permissions and legally permitted interactions are regulated in national traffic rules (StVO for Germany [24]) and in so-called “common-sense rules”.

Considering the motion space and the corresponding possible vehicle-environment-interaction, requirements with respect to the passability of road segments are derived. These requirements are different from the development process requirements such as the functional safety requirements presented in Subsection 4.3. In contrast to these development process requirements, the requirements in this Subsection aim for a comparison with vehicle capabilities regarding the road-based passability. The actual vehicle capabilities during operation are determined by the self-perception described in Subsection 6.1.

Assuming that there is a certain area of operation for the automated vehicle, a set of requirements would represent the necessary capabilities to be able to drive inside this area. This approach is similar to the usage of an ODD, defining operating conditions under which an automated driving system is designed to function. The effort for a safety validation depends on, i.a., the size and the complexity of the operating area. In order to validate the entire ODD and to prove safety requires a lot of tests as well as driven test kilometers [25]. This results in an enormous test effort, and thus, a safety validation is currently not economically feasible [26].

This new approach subdivides the ODD by sectioning a road network in single road segments. Every road segment requires specific driving capabilities from the automated vehicle. Depending on specific road elements, the resulting motion space with its possible vehicle-environment-interaction leads to a set of requirements for each road segment. Similar road segments in terms of requiring the same capabilities are clustered in categories. These categories represent requirement sets for each group of similar road segments. Consequently, one category represents a subset of the entire defined ODD. By defining reference scenarios for each category, a safety validation can be conducted category-wise. A successful test and validation of one category would automatically include a safety validation for every category-related road segment, even without testing every corresponding segment itself. Thus, parts of the ODD can be approved gradually, resulting in two main benefits. Firstly, the safety approval is possible at an earlier stage of the validation process. Secondly, the overall validation effort is potentially reduced by testing only specific reference scenarios per category.

The example scenario presented in Section 3 contains three streets connected by one intersection. Once an automated, driverless vehicle drives into one of the streets, it has to fulfill the present requirements of the entire section until the next intersection or roundabout. Otherwise, the vehicle is expected to fail to perform the necessary actions leading to a crash in worst case. For that reason, the part of the road network in the example scenario is subdivided into three road segments connected by one node, the intersection. In the following example, road segment RS1 is the street coming from the south, road segment RS2 is the street coming from the east and lastly, road segment RS3 is the street leaving the intersection in northern direction. The intersection itself is neglected in this example.

In order to compare the road-based requirements with capabilities, initially, they have to be derived. Considering the three road segments, there are two different main types of roads. Road segments RS1 and RS3 are one-way single lane roads, whereas road segment RS2 is a two-way road with one lane for each direction. Within road segments RS1 and RS2, parking is prohibited while road segment RS3 provides designated parking strips on each side. These basic road characteristics are represented in Tab. 1 and used for an example application of the categorization approach. Apart from these basic road features, there are many more characteristics influencing the automated execution of the dynamic driving task and, therefore, the requirements for an automated vehicle.

In order to derive requirements for road segments, the dynamic driving task must be considered. An essential maneuver is lane following [27], since a vehicle stays inside a traffic lane most of its driving time. Staying inside a lane directly depends on the width of the lane and the foreseeable narrowing parts. One reason for a narrowing is a parking strip next to the traffic lane due to parking vehicles that protrude in the current driving lane. To be able to follow lanes, the actual driving corridor of a vehicle must be considered. Consequently, one general requirement (GR), which is not road-specific yet, would be:

GR_1 *The vehicle must be able to follow driving corridors with a certain width.*

A requirement is fulfilled by at least one capability. Depending on the dynamic driving task and its complexity, multiple capabilities might be necessary to meet one requirement. There is a difference between the required capabilities and the actual capabilities of the vehicle. A required capability represents the minimum necessary capability level with respect to a suitable scale. In contrast, the actual capability is the current vehicle capability determined by the self-perception described in section 6.1. The current capability lies somewhere within the minimum and the maximum possible capability level depending on the current overall vehicle state.

In order to follow lanes of a certain width, an automated vehicle shall follow an adequate track inside the lane. The vehicle must firstly plan a trajectory resulting in a track and secondly, keep this track with the minimum possible lateral deviation. One general capability (GC) fulfilling general requirement GR_1 would be:

GC_1 *The vehicle keeps the track with a certain maximum lateral deviation.*

This minimum necessary capability is not road-segment-related yet. Nevertheless, a suitable scale has to be found in order to transfer the general capability in a specific minimum capability afterwards. In the best case, the lateral deviation keeping the track is 0. Because of errors and uncertainties, this value most likely can't be reached in practice. In the worst case, the lateral deviation is not limited (theoretically up to infinity). However, these minimum and maximum possible levels represent the scale regarding GC_1 measured in a length scale like meter.

Tab. 1 Road segments characteristics.

Road segment	Road type	Lanes per direction	Parking permission
RS1	One-way	1	no
RS2	Two-way	1	no
RS3	One-way	1	yes

Considering the three road segments stated in Tab. 1, the general requirement GR_1 and the general capability GC_1 can be concretized. The first step is to transfer the general requirement to road-segment-based specific requirements (SpR). In order to do so, information about the width of the traffic lanes is necessary. Adjacent parking strips influence the width of the lane because oversized or adversely parked vehicles might narrow the actual driving lane. Since this restriction is foreseeable, it can be considered in the specific requirement.

The width of the traffic lane inside road segment RS3 is 2.75 m. Since the adjacent parking strips have a width of only 2.00 m, a narrowing caused by parked vehicles is likely. Assuming an overlap with the driving lane of 0.20 m per side results in an effective lane width of 2.35 m while including a safety margin of 0.1 m leads to the required driving corridor width for road segment RS3 of 2.25 m. Therefore, the specific requirement for road segment RS3 would be:

RS3_SpR1 *The vehicle must be able to follow driving corridors with a width of 2.25 m.*

In order to derive the corresponding minimum capability (MC), the vehicle's own width has to be considered. The vehicle in the example scenario has a width of 2.12 m resulting in a potentially lateral free space within the driving lane of 13 cm. Since the planning of the trajectory and thus the track is included in this consideration, any lateral deviation shall not exceed 7.5 cm with respect to the center of the actual available driving space. Hence, one minimum capability fulfilling the specific requirement RS3_SpR1 for road segment 3 would be:

RS3_SpR1_MC1 *The vehicle keeps the track with a maximum lateral deviation of 7.5 cm.*

The process of further specification of the general requirements and capabilities is done in the same way for road segments RS1 and RS2. Since there are no parking strips within these segments, the considered space for overlapping parked vehicles can be skipped. Wrongly parked vehicles on road segment RS1 would block the driving lane completely. Traffic participants wrongly parking on road segment RS2 would only obstruct traffic in an absolute worst case since other traffic participants can use the oncoming traffic lane to pass. Thus, additional space for wrongly parked vehicles is not taken into account. Tab. 2 represents the resulting values for the specific requirements and minimal capabilities for all road segments.

Tab. 2 Values of the specific requirements and minimal capabilities.

Road segment	Lane width	Required lane width SpR1	Maximum lateral deviation SpR1_MC1
RS1	2.50 m	2.40 m	14.0 cm
RS2	2.75 m	2.65 m	26.5 cm
RS3	2.75 m	2.25 m	7.5 cm

Road segment RS2 represents the only street with oncoming traffic in this example. Since parking is prohibited, there is no need for a vehicle driving on that segment to leave the own driving lane. Exceptions are waste disposal vehicles or construction sites that might appear on that road. A constructions site could occupy one driving lane over the entire width. In that case, a vehicle must be able to drive in the opposite traffic yielding to oncoming traffic participants. That same situation applies for a garbage truck that occupies the traffic lane. The field of view for a vehicle that plans to pass those obstacles might be obstructed. This requires a vehicle to crawl into the opposite traffic while permanently observing the oncoming traffic. The requirement changes depending on different speed limits and, thus, various possible ranges of relative velocities between oncoming traffic and the ego vehicle. Therefore, another general requirement for the automated vehicle would be:

GR_2 *The vehicle must be able to follow driving corridors within opposite traffic in certain speed zones.*

This general requirement is only necessary for roads with oncoming traffic. Regarding the present example, the speed limit has to be concretized. Road segment RS2 requires a speed limit of 50 km/h. Thus, the specific requirement for the general requirement GR2 regarding road segment RS2 would be:

RS2_SpR2 *The vehicle must be able to follow driving corridors within opposite traffic in 50 km/h speed zones.*

In contrast to the general requirement GR_1, this requirement only has to be fulfilled for road segment RS2 and only if a construction site or garbage truck is active. Information about the presence of both is tracked via an external information system. Based on that information, road segments change their requirements accordingly. For simplification reasons, the derivation of suitable related capabilities is not presented in this paper.

Overall, this example presents two different requirements. Regarding the lateral deviation following a driving corridor (see general requirement GR_1), the three road segments represent three different required capability levels of the general capability GC_1. The second general requirement GR_2 only applies to road segment RS2 in special cases. If no construction site or garbage truck is identified, the road segment RS2 changes its category to the first one. Considering only those two requirements, two requirement categories could be created. The first one contains the general requirement GR_1 with all road-specific derivatives while the second one contains the same requirements plus the general requirement GR_2 and its derivative. Therefore, road segments RS1 and RS3 are in the first category while road segment RS2 is in the second category. If the first category is tested successfully, the automated vehicle is theoretically able to drive on both road segments. This assumes that the current capability level during operation assumed by the self-perception is high enough for the required lateral deviations. Testing the second category would allow the vehicle to drive on road segment RS2 as well.

This simplified example shows how the categorization approach basically works. In practice, every requirement resulting from the road segments must be derived and suitable capabilities fulfilling them have to be identified. The next step would be the derivation of reference scenarios for the categories. Finally, for a complete safety validation, they have to be tested successfully.

As stated before, this approach can also be used for a safe operation by providing the basis for a capability-based route planning. A short example demonstrating the basic function of this application is described in Subsection 6.2.

6 Safety mechanisms

As described in Subsection 2.2, the self-perception and the Safe Halt functionality will be employed as central elements of the overall safety concept for the prototypes developed in the project UNICAR*agil*. In the following subsections, we demonstrate the connection between these functionalities and the initial safety analysis outlined in Section 4.

6.1 Self-perception

For SAE level 4+ vehicles [1], the system is responsible for monitoring the environment and realizing the entire dynamic driving task. The driver is removed from the control loop and is not responsible and may not even be able to intervene in case of a system failure. Thus, the system must monitor itself and must be able to make decisions that take the system's health and thereby its current capabilities into account.

In an automated vehicle, a variety of aspects must be taken into account in order to obtain a holistic view of the system's overall state of health. This includes the state of the energy supply and its implications for the current mission, the state of the vehicle's hardware and software, the state of the communication network within the vehicle as well as the communication with external systems, and the feasible vehicle behavior based on the vehicle's current perception, action, and planning skills.

For the project UNICAR*agil*, the need for self-monitoring is reflected by safety requirement FSR_08 in Subsection 4.3. As the environment perception system realizes the required monitoring of the environment self-monitoring is realized by the self-perception system [28]. As stated above, self-perception for the system includes a variety of different monitoring aspects. In accordance with the focus of this paper on behavioral safety, we will focus on the aspect of monitoring safe vehicle behavior. Following functional safety requirement FSR_09, the vehicle's (sub-)systems provide information on their own current individual health. Self-representation then draws semantic links between the acquired data to derive an overall model of the system's health [28]. As stated in Subsection 4.3.2, ability and skill graphs are used for the self-representation of the vehicle's behavior [22], [23]. In these graphs, the skills necessary for a certain behavior or general capability and their interdependencies are modeled.

In order to utilize the identified skills for monitoring purposes, the functional safety requirements identified during the development process are attributed to the respective skills. In the example from Section 3, we focused on the skill of estimating a lane relative position and orientation for the lane following behavior in variant I of the example scenario. The information about the current estimation accuracy of position and orientation to the lane is provided by the respective (sub-) system(s) (functional safety requirement FSR_09). This link between the capability architecture (e.g., realized as ability and skill graphs) and the software architecture has been described by Bagschik et al. [4]. The accuracy estimation is then evaluated by the self-representation with respect to the identified functional safety requirement FSR_10. When the estimation accuracy does not fulfill the requirement, we consider the accuracy as insufficient and the skill of lane relative position and orientation estimation as degraded. One approach for such an accuracy assessment for the localization of an automated vehicle was presented by Reid et al. [31]. Such a degradation in localization accuracy may occur, e.g., due to high buildings or vegetation obstructing a GNSS signal or causing multipath effects. This would likely be the case for the lane leading north in the example scenario. This quality assessment of the skill of lane relative pose estimation is combined with the quality assessment of all other skills within the skill graph and aggregated into a performance assessment of the lane following behavior at the root of the graph.

This information about the system's current skills can subsequently be used by the vehicle's behavior generation to adapt the vehicle's behavior to its current skills. In the example scenario, a heavy degradation of the skill to estimate position and orientation relative to the lane may lead to the vehicle not being able to safely follow the north-leading lane anymore due to the constriction of the lane caused by the parked bus partially protruding into the lane. This may cause the vehicle's behavior generation to select a different behavior other than lane following and may lead to the vehicle stopping in the lane.

In case of a severe system degradation due to unforeseen events, the vehicle may not be able to uphold nominal automated system operation. Based on the information on such a degradation by the self-perception system, a safe halt of the vehicle will be triggered which will transfer the vehicle into a risk-minimized state. This safety mechanism is described in Section 6.3.

As demonstrated in the example of lane following, a skill-based self-perception provides information about the system's health. It is, however, only a very small part of an overall monitoring framework. For the monitoring of the vehicle's current skills, the necessary skills with their interdependencies must be modeled for all possible vehicle behavior, e.g., abstracted in the form of driving maneuvers [27]. Necessary requirements for the skills within a certain vehicle behavior must be derived and be connected to appropriate metrics that can be used to assess whether a requirement is currently fulfilled. These individual metrics of the skills have to be aggregated throughout the skill graph into a performance assessment of the respective vehicle behavior. Such a skill-based model can provide information about the system's health on a behavioral level but also on the level of the individual skills and can thus provide system health information on different levels of abstraction.

As part of a holistic monitoring framework, such as a representation of the vehicle’s current skills and its integration into automated vehicles provides information about the system's health and can contribute to vehicle safety if integrated into vehicle behavior generation. Additionally, such a representation can be used to plan and re-plan routes by evaluating the current skills of the vehicle against the capability requirements of the route segments, cf. Section 5.

6.2 Capability-based route planning

The knowledge about road-based required capabilities and current vehicle capabilities can be applied to realize a capability-based route planning as additional safety mechanism. One part of this approach is the road segment categorization described in Section 5 serving as the basis for a safety comparison. Every road segment contains a set of requirements and correspondingly a set of required vehicle capabilities. With the knowledge about the current vehicle capabilities represented by the self-perception described in Section 6.1, a route can be planned with respect to the current road-segment-availability.

In general, only road segments related to the approved road categories after a successful safety validation are considered within the route planning. This ensures that the vehicle doesn’t operate outside its approved ODD where it’s likely to fail to perform the dynamic driving task and provoke a crash in the worst case. The remaining road segments represent the potentially available road network to plan a driving mission. In order to plan a route that can be driven by the automated vehicle, the current capabilities must be compared with the required capabilities resulting from the road segment requirements. If the vehicle doesn’t have any degradation, no restriction applies to the potentially available road network. In case of degradations, meaning the loss of capability levels, the available road network is reduced. Losing important capabilities can even result in a discontinuous network that is not sufficient for planning the desired driving mission.

Depending on the use case of the mission, a route always contains one starting point, one or more intermediate points and one endpoint. The route planner has to take all that points into account while considering only the approved and currently available road segments. In general, the route is planned as efficiently as possible. Thus, the planning algorithm basically considers the two following prioritized optimization criteria:

- First priority: Available capability-based route
- Second priority: Shortest / fastest / most economical route

After initially planning the route, the vehicle can start the driving mission. As soon as any capability level changes, the route is checked with respect to the current availability. After a change of any capability level, three options depending on the change are generally possible:

- The downgrade of capability levels results in less available road segments: Calculation of a new route considering detours.
- The upgrade of capability levels results in more available road segments: Calculation of a new route considering fewer detours.
- The down-/upgrade of capability levels does not change the available road segments: Current route remains the same.

In the example scenario illustrated in Section 3, the automated vehicle generally has the option to cross the intersection following path Ⅰ or turn right driving the path Ⅱ. The following part gives an example, how the capability-based route planner works, considering the example explained in Section 5 of keeping the track with a maximum lateral deviation.

The initial route is planned with a current maximum lateral deviation of 7.0 cm, which fulfills the first category containing road segments RS1 and RS3. The second category is not applicable yet, since no construction site or garbage truck is identified on road segment RS2. Therefore, road segment RS2 is changed to the first category. Based on the assumption that path Ⅰ is more efficient than path Ⅱ, the planned driving mission follows path Ⅰ. Before the vehicle reaches the intersection, a degradation of the capability level would be identified by the self-perception and reported to the route planner. The new capability level only provides a maximum lateral deviation of 12.0 cm due to a change of surrounding buildings visible on the left-hand side of the vehicle (Fig. 1). This change of the capability level results in an unavailability of road segment RS2, requiring a maximum lateral deviation of 7.5 cm. Consequently, path Ⅱ is no longer available for the driving mission. As a consequence, the route planner changes the route according to the first priority optimization criterion of an available capability-based route. Now, the new route follows path Ⅱ requiring only 26.5 cm of maximum lateral deviation regarding track-keeping.

As a safety mechanism, the capability-based route planner is applied before start of and during a driving mission. However, the capability-based route planer is only able to plan or change routes that are successive of the currently driven road segment. Degradations that result in a non-fulfillment of the currently required capabilities can’t be handled. Such degradations are managed by the vehicle automation using either the self-perception functionality or, in case of severe system degradations, the Safe Halt as a fallback system described in the following section.

6.3 Safe Halt

In order to overcome the missing human fallback as required for SAE level 4+ vehicles in [1], the automated fallback system Safe Halt [30], an essential safety mechanism resulting from the safety activities in UNICAR*agil* project is introduced in this section. This fallback system aims to enable the vehicle to be transferred into a risk-minimized state at all times. The functional requirements for the Safe Halt functionality are derived from the functional safety requirement FSR_07 in Subsection 4.3. The fallback system is activated if the abilities of the primary driving functions do not suffice to fulfill the dynamic driving task [1]. The evaluation of the current abilities is based on the self-perception presented in Subsection 6.1.

The emergency maneuver performed during the Safe Halt procedure relies on a pre-calculated emergency path, cf. [32]. The destination of the emergency path is a risk-minimal stopping location, which is located outside the moving traffic if possible. The emergency path is planned simultaneously to the desired reference trajectory (see functional safety requirement FSR_06 in Subsection 4.3) by the trajectory and transmitted to the fallback system.

6.3.1 Functional requirements for Safe Halt

Based on the functional safety requirement FSR_07 and the emergency maneuver performed during the Safe Halt procedure, some of the functional requirements for the fallback level are derived.

The emergency path is planned on the basis of the current situation of the vehicle environment. However, since the functional safety requirement FSR_04 is also valid during the emergency maneuver and the vehicle is operating in a dynamic environment, the occupancy of the emergency path shall be monitored for a collision avoiding emergency maneuver.

SH_FR1 *The automatic fallback system shall monitor the occupancy of the emergency path.*

SH_FR2 *A perception system for monitoring the occupancy of the emergency path shall be available even in cases of severe degradation of the primary perception system.*

SH_FR3 *The perception system shall be suitable for all use cases of Safe Halt and all dynamic driving states.*

Based on the emergency path and the occupancy information from the emergency path monitoring, the fallback system generates collision-free trajectories for trajectory control.

SH_FR4 *The automatic fallback system for the dynamic driving task (DDT) shall generate collision-free emergency trajectories based on the course of the emergency path and the occupancy information from the emergency path monitoring.*

During the performance of the emergency maneuver, the fallback system is also responsible for enhancing the conspicuity of the vehicle (e.g., lighting).

SH_FR5 *The automatic fallback system for the DDT shall enhance the conspicuity of the vehicle as soon as the emergency maneuver is performed.*

Severe degradation of the primary driving functions of the automated vehicle can lead to uncontrolled vehicle behavior and thus to critical driving situations. In order to avoid these situations and to reduce the risk for the passengers and the vehicle environment, the fallback system is activated with a maximum latency.

SH_FR6 *The automated fallback system for the DDT takeover process shall have a maximum latency.*

For the successful performance of the emergency maneuver, the abilities of the fallback system must be sufficient. To verify this, the abilities of the fallback system shall be monitored even when the fallback system is not active.

SH_FR7 *The automatic fallback system for the DDT shall always monitor its abilities.*

The fallback system Safe Halt is an essential part of the safety architecture of the automated vehicle, as it aims at enabling a functionally degraded vehicle to be transferred into a risk-minimized state in any situation. To ensure that this function operates properly, the fallback system has to be fail-safe:

SH_FR8 *The automated fallback system for the DDT shall be fail-safe.*

6.3.2 Functional description of Safe Halt in UNICAR*agil*

Based on the functional requirements derived in Subsection 6.3.1 for a fallback system and additional requirements resulting from the modular software [32] and hardware architecture of the vehicle in UNICAR*agil*, this section provides the functional description of the fallback system Safe Halt.

In order to provide the availability of the fallback system (see functional requirement SH_FR8), the relevant hardware and communication architecture of the vehicle in UNICAR*agil* is fail-safe. Independent from the primary perception system, a secondary perception system is used by the fallback system so that the monitoring of the emergency path occupancy is possible even in case of severe failures of the primary perception system (functional requirement SH_FR2).

The vehicles in UNICAR*agil* can perform a 360° direction of movement. However, the maximum velocities are demonstrated only in the longitudinal direction of the vehicle: up to 20 m/s forwards and up to 10 m/s backward. The fallback system uses embedded hardware and, thus, does not provide the computing power as known in high-performance computers with integrated graphics cards. Hence, the raw sensor data is processed by the sensor hardware. Due to the required sensor range in the main driving directions, radar sensors are installed in the front and rear of the vehicle. The two radar sensors are supplemented by a 360° sensor system, which monitors the 360° motion direction. Driving with large sideslip angles is only permitted at low vehicle velocities, thus a comparatively small sensor range is required. Due to the required sensor range, ultrasonic sensors and cameras with fisheye lenses are used as a secondary 360° perception system (functional requirement SH_FR3).

Fig. 4 illustrates the interfaces and the system architecture of the fallback system Safe Halt. The object information from the secondary perception systems are fused by the fallback system. With the knowledge of the course of the emergency path and the intended speed profile, future vehicle poses can be calculated. With the vehicle dimensions (length, width, and height, supplemented by a safety margin), the vehicle pose is described in a spatial-temporal dimension. Based on the object states of the secondary perception systems, this dimension is monitored for collisions (functional requirement SH_FR1) [34]. If a collision is predicted, the velocity is adjusted to avoid or at least mitigate the collision (functional requirement SH_FR4).

The trajectory generation of the fallback system generates an emergency trajectory for the trajectory controller based on the course of the emergency path, the included speed profile, and the collision mitigating velocity input of the emergency path monitoring. The specifications of the emergency trajectory and the reference trajectory for automated driving are identical. Therefore, it is not necessary to develop a separate trajectory controller for the fallback system.

Even if the fallback system is not activated, the secondary perception system monitors the occupancy of the emergency path and generates a collision-avoiding trajectory. The modular software architecture of UNICAR*agil* allows switching (*switch* in Fig. 4) between the reference trajectory and the emergency trajectory. Hence, the software and hardware of the fallback system are active even without severe degradations of the primary driving functions. When the fallback system is requested, the input of the trajectory controller is altered by switching to the emergency trajectory.

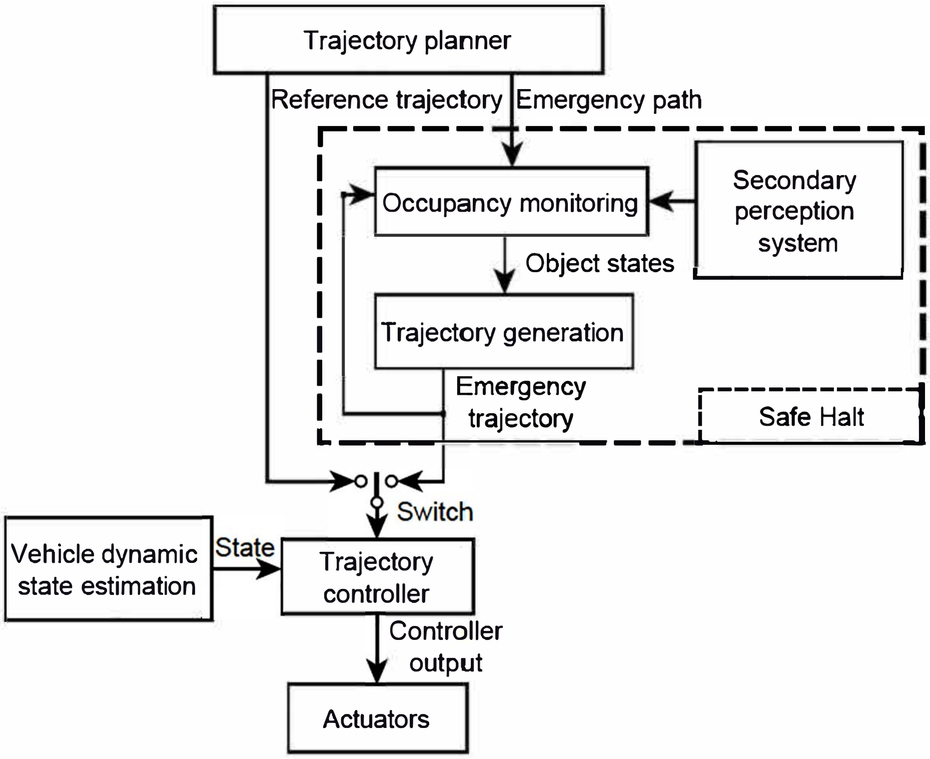

Fig. 4 System architecture of the fallback system Safe Halt (dashed: system boundary of the Safe Halt module).

With this architecture, it is possible to switch to the fallback system with minimal latency (functional requirement SH_FR6) and to monitor the abilities of the fallback system at all times (functional requirement SH_FR7). Simultaneously, the fallback system enhances the conspicuity of the vehicle by activating the lighting (functional requirement SH_FR5).

6.3.3 Interfaces of Safe Halt

The Safe Halt fallback system is integrated into the overall vehicle architecture by means of three interfaces, two primary (depicted in Fig. 4) and one secondary:

- The *emergency path* includes discrete reference poses (position and orientation) that the vehicle will track during the emergency maneuver. The emergency path ends at a safe stopping location. Each discrete reference pose is assigned a velocity in the direction of the path. In addition, further information can be attached to the discrete reference poses, e.g., at which locations the direction indicators are to be activated.
- The *emergency trajectory* generated by the fallback system meets the specifications of the reference trajectory for automated vehicle operation.
- The conspicuity during the emergency maneuver can be enhanced by providing an interface to the corresponding vehicle functions (e.g., direction indicators and hazard warning lights).

6.3.4 Safe Halt in the example scenario

This section describes the performance of an emergency maneuver based on the example scenario presented in Section 3.

In the scenario, the vehicle approaches the intersection from the south. The end of the path Ⅲ is selected as the safe stopping place and, thus, the destination of the currently valid emergency path. The vehicle is in automated operation. The independent perception systems and the services of the fallback system are active. A collision-free emergency trajectory is generated based on the emergency path, the included speed profile, and the detected collision objects.

Based on the information of self-perception, serious degradation of the primary driving functions is revealed. In this case, the driving mode switches to the fallback system Safe Halt. From now on, the emergency trajectory is fed to the vehicle's trajectory controller. The independent perception system monitors the occupancy of the emergency path and adjusts the reference velocity accordingly. At the beginning of the emergency maneuver, the conspicuity of the vehicle is enhanced, for example, by switching on the hazard warning lights. Subsequently, the vehicle tracks the emergency trajectory until it reaches its end.

7 Modular safety validation

In ISO 26262 [11], the safety validation of the analyzed item is described in parts 4-8 and is recommended to be done on a vehicle level (in the following called system level for the sake of generalization). This provides the opportunity to test the item in its intended environment. Therefore, interactions between the item and its environment, i.e. other components in the vehicle and the natural environment or other traffic participants, are tested during the safety validation. These interactions are set conditions due to the existence of the operating components or other surroundings in the test environment. Thus, these interactions do not need to be analyzed in more detail than necessary for the development of the function of the system.

The consideration of all possible interactions in the system tests is beneficial for the safety validation but requires having the whole system available for testing. Furthermore, minor changes of the item or its surrounding components or the environment may influence the function of the item. Therefore, a new safety validation for the whole system becomes necessary leading to an economically unviable process. Additionally, tests on system level are in many cases less controllable and less observable than on modular level, not only due to the higher complexity of an integrated system but also of the interfaces between components that may not be available on system level.

In UNICAR*agil*, as described in [5], a modular safety approval is seen as an opportunity to overcome these disadvantages of safety validation on system level. ISO 26262 [11] recommends modular testing in part 5 (chapter 10) and part 6 (chapters 9, 10, and 11) in the form of unit and integration testing. The described methods address the verification of different software and hardware components. For a modular validation, one difference to these verification processes is that all the influences on system level already need to be propagated to a modular level.

Influences can be found by analyzing the module regarding its interfaces. Interfaces are separable into desired and imposed interfaces. Desired interfaces can be found in the specification of a module, while the further specification of this interface for validation could be necessary. Imposed interfaces are usually known as well, but some influencing factors transported by these interfaces may not be found without analyzing the module and its environment more precisely. A document for the definition of the module, similar to an item definition, supports this process. In “traditional” processes integration and system testing, which we want to avoid, uncovers these factors.

When all influences on modular level are identified, modules need to be validated consistently to the defined safety concept on system level. As described in part 4 of ISO 26262 [11], safety goals are the input for the safety validation. In Section 4.3, we further developed functional safety requirements. While the system architecture has also been considered in this process, the safety requirements are already assignable to modules of the system. Therefore, it might be a matter of course to test the modules against these functional safety requirements which are also listed as possible input for the safety validation in part 4 of ISO 26262 [11]. Nevertheless, we recommend using safety goals for the validation process in the first place. In ISO 26262, the safety goals are the main input for the safety validation on system level. Using them in modular validation ensures direct traceability from safety at system level down to modules.

For this purpose, Klamann et al. [7] describe a method to breakdown safety goals to modular level. The method uses Fault Tree Analysis (FTA) and System Theoretic Process Analysis (STPA) [35] as favored methods to derive potential causal factors on modular level from safety goals on system level. Afterwards, safety requirements on module level are defined to prevent these causal factors. The methods used in Section 4.3 may have already generated these safety requirements. Even though, the fault oriented processes may find additional ones due to further developments of the system and its modules or due to the different view of a test engineer. Furthermore, the results (e.g., potential causal factors) are essential for finding challenging test cases. For the case that these methods have already been used in the concept phase, the results can be reused in the safety validation but should be adapted to the chosen modular architecture.

Using the method from [7] derives, as an example, the following causal factor (CF) from safety goals SG_1 and SG_2:

CF_1 *Deviation from localized ego pose out of required range.*

CF_2 *Deviation of estimated error of ego pose out of allowed range.*

These causal factors are initially allocated to the vehicle dynamic state estimation module. Further analyses of the module revealing more specific causal factors based on causal factors CF_1 and CF_2 show that other modules may be the source of failure as well:

CF_1_1 *Accuracy of angular rate out of allowed range.*

CF_1_2 *Accuracy of GNSS position out of allowed range.*

CF_1_3 *Accuracy of wheel speed out of allowed range.*

CF_1_4 *Misguided assumptions in the internal filter model.*

CF_2_1 *Unidentified GNSS multipath interference.*

CF_2_2 *Unidentified GNSS non-line of sight signal.*

CF_2_3 *Misguided assumptions in the error estimation model.*

The third causal factor is in the response of the dynamic module (in UNICAR*agil* this is the wheel, brake, engine, and steering system [5], [14]) what shows that modules need to be analyzed in-depth to find further essential causal factors and to reach sufficient traceability between safety goals and causal factors.

Causal factor CF_2 is particularly safety relevant since an estimated error out of a specified range may cause the ego vehicle to plan paths causing collisions. The following derived causal factors inherit this attribute and are therefore analyzed further in this paper. From the analysis of the vehicle dynamic state estimation, it is known that buildings cause GNSS signal blockage, multipath interference, or non-line of sight signals. Thus, the accuracy may exceed the required range while this may not be identified. The presented scenario in Section 3 includes this challenge, containing several tall buildings on both sides of path I . Further challenges may be derived from the intended use cases for a module. Therefore, the module definition requires detailed information about the intended use cases and the environment of the module.

While the scenario provides a specific use case for the whole vehicle, which can also be part of a road category as described in Section 5, it can be reduced to the identified influencing factors for the vehicle dynamic state estimation in modular testing. Exemplary functional test cases (FTC) which can be derived from these influencing factors triggering causal factors CF_2_1 and CF_2_2 are:

FTC_1 *The vehicle dynamic state estimation is receiving faulty GNSS signals due to multipath interference while moving with different velocities and accelerations.*

FTC_2 *The vehicle dynamic state estimation is receiving faulty GNSS signals due to non-line of sight while moving with different velocities and accelerations.*

Functional test cases are defined conveniently to functional scenarios [35] and describe test cases only semantically, while logical test cases further describe the range of values in it. Concrete test cases define the exact value for each parameter.

Test cases for misguided assumptions (e.g., causal factor CF_2_3) can be derived by finding challenging values or combinations of values for the error estimation model. A sensitivity analysis is a proper method for this [37], but requires a working prototype of the module. Furthermore, it is beneficial when this prototype is as final as possible to achieve a high validity of the sensitivity analysis. Therefore, in the first place, expert knowledge about the general function of the model or even deeper knowledge about the implemented error estimation model should be used to determine challenging test cases.

Provided knowledge about the vehicle dynamic state estimation module reveals that turns with low accelerations and the special maneuvers of the UNICAR*agil* vehicle due to its four-wheel steering system may be challenging for the estimation of the pose. Therefore, the following exemplary functional test cases that trigger causal factor CF_2_3 are added:

FTC_3 *The device of the vehicle dynamic state estimation module is slowly rotating without (high) lateral/longitudinal acceleration.*

FTC_4 *The device of the vehicle dynamic state estimation module is moving slowly and with low accelerations in lateral direction.*

The test cases reveal that providing valid test data may be challenging for the case that the module providing the data in the vehicle is not part of the testbed. Therefore, analyzing and specifying the interfaces are essential for generating the necessary test data.

In the next steps, the functional test cases need to be translated into logical and concrete test cases. Thus, it needs to be specified, e.g., what kind of signal blockage and multipath effects may occur or how the stimuli for these faulty signals look like for functional test cases FTC_1 and FTC_2. In the last two exemplary functional test cases (FTC_3 and FTC_4) it needs to be specified, e.g., which velocities and acceleration are assumed as low to challenge the module. These steps require specific knowledge about the exact signals or ranges of the module and its environment and are thus not addressed in this paper.

The method of Klamann et al. [7] further develops the necessary pass-/fail-criteria for the evaluation of the outcome of these test cases. The identified causal factor of the dynamic module is not addressed in these tests but has to be part of the modular validation of the dynamic module. For the vehicle dynamic state estimation, the pass-/fail-criteria is the allowed deviation from the actual pose and the allowed deviation of the estimated error from the real error. The criteria are also derivable by the road segment requirements provided in Section 5 and can, therefore, be used for a road segment-specific safety approval of a module.

This section provides a traceable process using the behavioral safety concept generated in the previous sections as input. Test cases are generated on a modular level by analyzing possible deviations from the intended behavior caused by modules. This process supports the concept of the modular safety approval for automated vehicles. Applying it to the UNICAR*agil* project will evaluate the described approach and reveal required adaptions.

8 Conclusion and outlook

In the previous sections, we outlined different approaches of the research project UNICAR*agil* towards safety of SAE level 4 automated driving functionalities. These safety approaches target ensuring and validating safety during an agile, rapidly progressing development process as well as during run-time.

The process activities and run-time safety mechanisms are strongly interconnected which contributes to the complexity of the vehicle automation task. Hence, managing the interconnection is key for ensuring and validating safety. In this paper, we have demonstrated the interconnection of the different safety approaches from a requirements perspective using an example scenario. The process perspective focuses on the concept phase as well as on validation activities before the release of the automated driving functionality. In this context, requirements are derived based on initial safety analyses and investigations of the operational design domain. The validation approach undertaken in the project UNICAR*agil* leverages the strict modular design to reduce the usual validation complexity. Moreover, we have shown how three key safety mechanisms (self-perception, capability-based route planning, and the Safe Halt functionality) interconnect and how these mechanisms embed into the development activities.

Overall, the project UNICARagil allows us to research the safety of automated vehicles without restrictions of evolutionary system designs. The next steps in the project UNICAR*agil* from a safety perspective can be summarized twofold. First, we will further integrate the safety activities outlined in this paper with those regarding the other safety perspectives (cf. Subsection 2.1), such as *functional safety* or *operational safety*. Second, we will implement and demonstrate the safety mechanisms outlined in this paper (cf. Section 6) in the vehicle prototypes developed in the project.

9 Acknowledgment

The authors would like to thank Marcus Nolte and Grischa Gottschalg for the valuable discussion of the paper's contents as well as Sonja Luther and Ibrahim Aftab Khan for the intensive language review. This research is accomplished within the project "UNICAR*agil*" (FKZ 16EMO0285, FKZ 16EMO0286). We acknowledge the financial support for the project by the Federal Ministry of Education and Research of Germany (BMBF).

10 References

[1] SAE, 2018.
Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, Society of Automotive Engineers, Standard J3016.

[2] Junietz, Philipp, Steininger, Udo, and Winner, Hermann, 2019.
Macroscopic Safety Requirements for Highly Automated Driving.
In: Transportation Research Record: Journal of the Transportation Research Board. Vol. 2673, No. 3.
DOI: 10.1177/0361198119827910.

[3] Maurer, Markus, 2018.
Hochautomatisiertes und vollautomatisiertes Fahren.
In: 56. Deutscher Verkehrsgerichtstag 2018, Deutscher Verkehrsgerichtstag - Deutsche Akademie für Verkehrswissenschaft, Editor.
Goslar, Germany: Luchterhand Verlag.
ISBN: 3-472-09576-8

[4] Bagschik, Gerrit, Nolte, Marcus, Ernst, Susanne, and Maurer, Markus, 2018.
A System's Perspective Towards an Architecture Framework for Safe Automated Vehicles.
In: 2018 IEEE International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, pages 2438–2445.
DOI: 10.1109/ITSC.2018.8569398

[5] Woopen, Timo, Lampe, Bastian, Böddeker, Torben, Eckstein, Lutz, Kampmann, Alexandru, Alrifaee, Bassam, Kowalewski, Stefan, Moormann, Dieter, Stolte, Torben, Jatzkowski, Inga, Maurer, Markus, Möstl, Mischa, Ernst, Rolf, Ackermann, Stefan, Amersbach, Christian, Leinen, Stefan, Winner, Hermann, Püllen, Dominik, Katzenbeisser, Stefan, Becker, Matthias, Stiller, Christoph, Furmans, Kai, Bengler, Klaus, Diermeyer, Frank, Lienkamp, Markus, Keilhoff, Dan, Reuss, Hans-Christian, Buchholz, Michael, Dietmayer, Klaus, Lategahn, Henning, Siepenkötter, Norbert, Elbs, Martin, von Hinüber, Edgar, Dupuis, Marius, and Hecker, Christian, 2018.
UNICAR*agil* – Disruptive Modular Architectures for Agile, Automated Vehicle Concepts.
In: Aachener Kolloquium 2018, Aachen, Germany.
DOI: 10.18154/RWTH-2018-229909.

[6] Woopen, Timo, van Kempen, Raphael, and Eckstein, Lutz, 2020.
UNICAR*agil* - Where we are and where we are going.
Accepted for: Aachener Kolloquium 2020, Aachen, Germany.

[7] Buchholz, Michael, Gies, Fabian, Danzer, Andreas, Henning, Matti, Hermann, Charlotte, Herzog, Manuel, Horn, Markus, Schön, Markus, Rexin, Nils, Dietmayer, Klaus, Fernandez, Carlos, Janosovits, Johannes, Kamran, Danial, Kinzig, Christian, Lauer, Martin, Molinos, Eduardo, Stiller, Christoph, Wang, Lingguang, Ackermann, Stefan, Homolla, Tobias, Winner, Hermann, Gottschalg, Grischa, Leinen, Stefan, Becker, Matthias, Feiler, Johannes, Hoffmann, Simon, Diermeyer, Frank, Lampe, Bastian, Beemelmanns, Till, Kempen, Raphael van, Woopen, Timo, Eckstein, Lutz, Voget, Nicolai, Moormann, Dieter, Jatzkowski, Inga, Stolte, Torben, Maurer, Markus, Graf, Jürgen, v.Hinüber, Edgar, and Siepenkötter, Norbert, 2020.
Automation of the UNICAR*agil* Vehicles.
Accepted for: Aachener Kolloquium 2020, Aachen, Germany.

[8] Klamann, Björn, Lippert, Moritz, Amersbach, Christian, and Winner, Hermann, 2019.
Defining Pass-/Fail-Criteria for Particular Tests of Automated Driving Functions.
In: 2019 IEEE International Conference on Intelligent Transportation Systems (ITSC), Auckland, New Zealand, pages 169–174.
DOI: 10.1109/ITSC.2019.8917483

[9] Waymo, 2017.
Waymo Safety Report - On the Road to Fully Self-Driving.
Available: https://storage.googleapis.com/sdc-prod/v1/safety-report/waymo-safety-report-2017-10.pdf.

[10] ISO, 2019.
ISO/PAS 21448: Road vehicles – Safety of the intended functionality, International Organization for Standardization, International Standard ISO/PAS 21448.

[11] ISO, 2018.
ISO 26262: Road vehicles - Functional Safety, International Organization for Standardization, Geneva, Switzerland, International Standard ISO 26262:2018.

[12] Goth, Marcus, Keilhoff, Dan, and Reuss, Hans-Christian, 2020.
Fault Tolerant Electric Energy Supply System Design for Automated Electric Shuttle Bus.
In: 20. Internationales Stuttgarter Symposium, Stuttgart, Germany.

[13] Niedballa, Dennis and Reuss, Hans-Christian, 2020.
Concepts of Functional Safety in E/E-Architectures of Highly Automated and Autonomous Vehicles.
In: 20. Internationales Stuttgarter Symposium, Stuttgart, Germany.

[14] Martens, Timm, Pouansi, Brice, Li, Minglu, Henkel, Niclas, Wielgos, Sebastian, Schlupek, Martin, and Eckstein, Lutz, 2020.
UNICAR*agil* Dynamics Module.
Accepted for: Aachener Kolloquium 2020, Aachen, Germany.

[15] Ernst, Rolf, Ahrendts, Leonie, and Gemlau, Kai-Björn, 2018.
System Level LET: Mastering Cause-Effect Chains in Distributed Systems.
In: IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, pages 4084–4089.
DOI: 10.1109/IECON.2018.8591550.

[16] Gemlau, Kai-Bjorn, Peeck, Jonas, Sperling, Nora, Hertha, Phil, and Ernst, Rolf, 2019.
A new design for data-centric Ethernet communication with tight synchronization requirements for automated vehicles.
In: IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, pages 4489–4494.
DOI: 10.1109/IECON.2019.8927070.

[17] Maurer, Markus, 2000.
Flexible Automatisierung von Straßenfahrzeugen mit Rechnersehen.
Dissertation. Universität der Bundeswehr München, Munich, Germany.

[18] Graubohm, Robert, Stolte, Torben, Bagschik, Gerrit, and Maurer, Markus, 2020.
Towards Efficient Hazard Identification in the Concept Phase of Driverless Vehicle Development.
Accepted for: 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA.

[19] Graubohm, Robert, Stolte, Torben, Bagschik, Gerrit, Steimle, Markus, and Maurer, Markus, 2019.
Functional Safety Concept Generation within the Process of Preliminary Design of Automated Driving Functions at the Example of an Unmanned Protective Vehicle.
In: Proceedings of the Design Society: International Conference on Engineering Design, Delft, The Netherlands, pages 2863–2872.
DOI: 10.1017/dsi.2019.293.

[20] Stolte, Torben, Bagschik, Gerrit, Reschka, Andreas, and Maurer, Markus, 2017.
Hazard analysis and risk assessment for an automated unmanned protective vehicle.
In: 2017 IEEE Intelligent Vehicles Symposium (IV), Redondo Beach, CA, USA, pages 1848–1855.
DOI: 10.1109/IVS.2017.7995974.

[21] Graubohm, Robert, Stolte, Torben, Bagschik, Gerrit, Reschka, Andreas, and Maurer, Markus, 2017.
Systematic Design of Automated Driving Functions Considering Functional Safety Aspects.
In: 8. Tagung Fahrerassistenz, Munich, Germany.

[22] Reschka, Andreas, Bagschik, Gerrit, Ulbrich, Simon, Nolte, Marcus, and Maurer, Markus, 2015.
Ability and skill graphs for system modeling, online monitoring, and decision support for vehicle guidance systems.
In: 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, South Korea, pages 933–939.
DOI: 10.1109/IVS.2015.7225804.

[23] Nolte, Marcus, Bagschik, Gerrit, Jatzkowski, Inga, Stolte, Torben, Reschka, Andreas, and Maurer, Markus, 2017.
Towards a skill- and ability-based development process for self-aware automated road vehicles.
In: 2017 IEEE International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan.
DOI: 10.1109/ITSC.2017.8317814.

[24] Straßenverkehrs-Ordnung vom 6. März 2013 (BGBl. I S. 367), die zuletzt durch Artikel 1 der Verordnung vom 20.April 2020 (BGBl. I S. 814) geändert worden ist, 2020.

[25] Wachenfeld, Walther and Winner, Hermann, 2016.
The Release of Autonomous Vehicles.
In Autonomous Driving: Technical, Legal and Social Aspects, Maurer, Markus, Gerdes, J. Christian, Lenz, Barbara, and Winner, Hermann, Editors.
Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, pages 425–449.
ISBN: 978-3-662-48847-8

[26] Amersbach, Christian and Winner, Hermann, 2019.
Defining Required and Feasible Test Coverage for Scenario-Based Validation of Highly Automated Vehicles.
In: 2019 IEEE International Conference on Intelligent Transportation Systems (ITSC), Auckland, New Zealand, pages 425–430.
DOI: 10.1109/ITSC.2019.8917534.

[27] Reschka, Andreas, 2017.
Fertigkeiten- und Fähigkeitengraphen als Grundlage des sicheren Betriebs von automatisierten Fahrzeugen im öffentlichen Straßenverkehr in städtischer Umgebung.
Dissertation. Technische Universität Braunschweig, Braunschweig, Germany.

[28] Nolte, Marcus, Jatzkowski, Inga, Ernst, Susanne, and Maurer, Markus, 2020.
Supporting Safe Decision Making Through Holistic System-Level Representations & Monitoring – A Summary and Taxonomy of Self-Representation Concepts for Automated Vehicles.
To be submitted to: IEEE Transactions on Intelligent Vehicles.

[29] Ulbrich, Simon, Reschka, Andreas, Rieken, Jens, Ernst, Susanne, Bagschik, Gerrit, Dierkes, Frank, Nolte, Marcus, and Maurer, Markus, 2017.
Towards a Functional System Architecture for Automated Vehicles.
In: arXiv:1703.08557 [cs]. Available: http://arxiv.org/abs/1703.08557.

[30] Ackermann, Stefan and Winner, Hermann, 2020.
Systemarchitektur und Fahrmanöver zum sicheren Anhalten modularer automatisierter Fahrzeuge.
In: Workshop Fahrerassistenz und automatisiertes Fahren, Walting, Germany.

[31] Reid, Tyler G. R., Houts, Sarah E., Cammarata, Robert, Mills, Graham, Agarwal, Siddharth, Vora, Ankit, and Pandey, Gaurav, 2019.
Localization Requirements for Autonomous Vehicles.
In: SAE International Journal of Connected and Automated Vehicles. Vol. 2, No. 3, pages 173–190.
DOI: 10.4271/12-02-03-0012.

[32] Mokhtarian, Armin, Kampmann, Alexandru, Alrifaee, Bassam, and Kowalewski, Stefan, 2020.
The Dynamic Service-oriented Software Architecture for the UNICAR*agil* Project.
Accepted for: Aachener Kolloquium 2020, Aachen, Germany.

[33] Wang, Lingguang, Wu, Zhenkang, Li, Jiakang, and Stiller, Christoph, 2020.
Real-Time Safe Stop Trajectory Planning via Multidimensional Hybrid A*-Algorithm.
Accepted for: 2020 IEEE International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece.

[34] Ackermann, Stefan, Winner, Hermann, and Buchholz, Michael, 2019.
Modul und Verfahren zur Absicherung von Solltrajektorien für automatisiertes Fahren.
German patent application, official file number: 10 2019 125 401.9

[35] Leveson, Nancy G. und Thomas, John P., 2018.
STPA Handbook.
Cambridge, MA, USA.
Available:
http://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf.

[36] Bagschik, Gerrit, Menzel, Till, Reschka, Andreas, and Maurer, Markus, 2017.
Szenarien für Entwicklung, Absicherung und Test von automatisierten Fahrzeugen.
In: 11. Workshop Fahrerassistenz und automatisiertes Fahren, Walting, Germany.

[37] Nolte, Marcus, Schubert, Richard, Reisch, Cordula, and Maurer, Markus, 2020.
Sensitivity Analysis for Vehicle Dynamics Models – An Approach to Model Quality Assessment for Automated Vehicles.
Accepted for: 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA.