D4.6

Report on the validated hybrid and real-world testing and validation techniques

Project short name
SUNRISE

Project full name
Safety assUraNce fRamework for connected, automated mobIlity SystEms

Horizon Research and Innovation Actions | Project No. 101069573
Call HORIZON-CL5-2021-D6-01

Funded by the European Union

ccam-sunrise-project.eu/

Dissemination level	Public (PU) - fully open
Work package	WP4: CCAM V&V framework
Deliverable number	D4.6: Report on the validated hybrid and real-world testing and validation techniques
Deliverable responsible	Philipp Legran, ika
Status - Version	Final - V1.0
Submission date	27/02/2025
Keywords	CCAM, virtual, hybrid, real-world, physical, testing validation, assessment, V&V, limitations, simulation, ViL, DiL, test bench, scaled model, providing grounds, public roads

Authors

Role	Name
Main author	Philipp Legran (ika)
Contributing authors	Bernhard Hillbrand (VIF) Bryan Bourauel (BASt) Amélie Reher (BASt) Eloy Collado (IDI) Thaddaeus Menzel (IDI DE) Thomas Amler (IDI DE) Martin Skoglund (RISE) Mateen Malik (RISE) Anders Thorsén (RISE) Mohammed Shabbir ALI (VED) Emre Kaynar (AVL) Anastasia Bolovinou (ICCS) Shabbir Ali (VED) Xizhe Zhang [Jason] (UoW)

Quality Control

	Name	Organisation	Date
Peer review 1	Bernhard Hillbrand	VIF	24/01/2025
Peer review 2	Stefan de Vries	IDI	06/02/2025
Peer review 3	John-Fredrik Grönvall	Chalmers	10/02/2025

Version history

Version	Date	Author	Summary of changes
0.1	06/11/2024	Philipp Legran	First draft of document structure
0.2	23/01/2025	All authors	Initial addition of deliverable contents
0.3	23/01/2025	Philipp Legran	Editing for internal review
0.4	21/02/2025	All authors	Edit based on internal feedback
0.5	21/02/2025	Philipp Legran	Formatting and general editing
0.6	21/02/2025	Philipp Legran	Merging updates from section 2.2 and 2.3
0.7	21/02/2025	Philipp Legran	Updating section 5.3
0.8	21/02/2025	Philipp Legran	Small adjustments
0.9	26/02/2025	All authors	Final Updates based on second review
1.0	27/02/2025	Philipp Legran	Final version for submission

TABLE OF CONTENTS

EXECUTIVE SUMMARY 11

1 INTRODUCTION 12

1.1 Project introduction 12

1.2 Purpose of deliverable 14

1.3 Intended audience 15

1.4 Deliverable structure and relation to other parts of project 16

2 HYBRID AND REAL-WORLD TESTING - ADDRESSING LIMITATIONS OF THE HARMONISED V&V SIMULATION FRAMEWORK 18

2.1 Introduction to Hybrid and Real-World Testing 18

2.2 Summary of the Harmonised V&V Simulation Framework for Virtual Testing 19

2.3 Limitations of Virtual Testing 20

2.3.1 General Limitations 20

2.3.2 Limitations for the Subsystem "Subject Vehicle – Sensors" 20

2.3.3 Limitations for the Subsystem "Subject Vehicle – AD function" 22

2.3.4 Limitations for the Subsystem "Subject Vehicle – Vehicle Dynamics" 22

2.3.5 Limitations for the Subsystem "Environment" 23

2.4 Classification of Hybrid and Real-world Testing Techniques 23

3 PHYSICAL TESTING ON PROVING GROUNDS AND PUBLIC ROADS FOR CCAM VALIDATION 26

3.1 Black Box Testing on Proving Grounds and Public Roads 26

3.1.1 Physical and Virtual Components on Subsystem Level 29

3.1.2 Advantages in Comparison to Virtual Testing 29

3.1.3 Limitations in Comparison to Virtual Testing 30

3.1.4 Additional Validation Needs 30

3.1.5 Exemplary Implementation and Setup – UC 2.1 30

3.2 Scaled Model Testing 33
3.2.1 Physical and Virtual Components on Subsystem Level 33
3.2.2 Advantages in Comparison to Virtual Testing 34
3.2.3 Limitations in Comparison to Virtual Testing 34
3.2.4 Additional Validation Needs 36
3.2.5 Exemplary Implementation and Setup – UC 4.1 37
4 VEHICLE-IN-THE-LOOP TESTING FOR CCAM VALIDATION 41
4.1 Vehicle-In-the-Loop on Proving Grounds 41
4.1.1 Physical and Virtual Components on Subsystem Level 42
4.1.2 Advantages in Comparison to Virtual Testing 42
4.1.3 Limitations in Comparison to Virtual Testing 43
4.1.4 Additional Validation Needs 43
4.1.5 Exemplary Implementation and Setup 43
4.1.6 Exemplary Implementation and Setup – UC 1.3 44
4.2 Vehicle-In-the-Loop on Testbenches 48
4.2.1 Physical and Virtual Components on Subsystem Level 49
4.2.2 Advantages in Comparison to Virtual Testing 51
4.2.3 Limitations in Comparison to Virtual Testing 51
4.2.4 Additional Validation Needs 51
4.2.5 Exemplary Implementation and Setup 52
5 DRIVER IN THE LOOP TESTING FOR CCAM VALIDATION 57
5.1 Introduction to Driver-in-the-Loop Testing 57
5.2 Driver-In-The-Loop on Proving Grounds 57
5.2.1 Physical and Virtual Components on Subsystem Level 59
5.2.2 Advantages in Comparison to Virtual Testing 59
5.2.3 Limitations in Comparison to Virtual Testing 60
5.2.4 Additional Validation Needs 60

5.2.5 Exemplary Implementation and Setup 60

5.3 Driver-In-The-Loop in Simulators 61

5.3.1 Physical and Virtual Components on Subsystem Level 62

5.3.2 Advantages in Comparison to Virtual Testing 63

5.3.3 Limitations in Comparison to Virtual Testing 63

5.3.4 Additional Validation Needs 63

5.3.5 Exemplary Implementation and Setup 63

6 SUMMARY OF ADVANTAGES AND DISADVANTAGES OF HYBRID AND REAL-WORLD TEST METHODS 65

7 CONCLUSIONS 68

8 REFERENCES 70

LIST OF FIGURES

Figure 1: Safety Assurance Framework stakeholders 13
Figure 2: Workplan of the SUNRISE Project 14
Figure 3: Architecture of the harmonised V&V Simulation Framework 20
Figure 4: Black box testing schematic 26
Figure 5: Black box testing process 27
Figure 6: Cut-In Scenario from Use Case 2.1 31
Figure 7: Set-up for proving ground testing with crashable soft car target and ego vehicle 31
Figure 8: Set-up for public road testing with ego and target vehicle 32
Figure 9: Representation of an actual full-size truck and its scaled-down version that is used in UC4.1 testing. 35
Figure 10: Preliminary set-up of for testing with the small-scale model truck used in UC4.1 testing. ..38
Figure 11: Schematic sensor setup for backing function under test. 39
Figure 12: IDIADA VIL approach 44
Figure 13: Real ego of CAV of VED 45
Figure 14: Visualisation screen inside the ego CAV showing received CPM information on the map. 46
Figure 15: SUMO GUI of the XIL simulator. 46
Figure 16: Components of a VIL Testbench Setup 49
Figure 17: Radar Stimulator configuration with modular absorber system with variable airgap 50
Figure 18: System Architecture for VIL on Testbench for Adaptive Cruise Assist 52
Figure 19: Real setup of the VIL Testbench 54
Figure 20: Comparison of the ground truth distance and the measured distance 55
Figure 21: Comparison of the ground truth lane and line position and the measured position 56
Figure 22: Driver in the Loop on Proving Grounds 58
Figure 23: Structure of the Real-Dynamic Driving Simulator 60
Figure 24: ADS simulation reference architecture 62
Figure 25: The 3xD immersive driving simulator in WMG, University of Warwick 64
Figure 26: WMG's 3xD used for gantry study 64
Figure 27: WMG's 3xD used for the SAVVY project 64

LIST OF TABLES

Table 1: Categorizing Hybrid and Real-World Testing Methods by Subsystem Implementation 24
Table 2: Advantages and Disadvantages of Hybrid and Real-World Testing 65

ABBREVIATIONS AND ACRONYMS

Abbreviation	Meaning
2D	Two dimensions or two dimensional
3D	Three dimensions or three dimensional
5G	Fifth Generation of mobile networks
ACA	Adaptive Cruise Assist
ACC	Adaptive Cruise Control
AD	Automated Driving
ADAS	Advanced Driver Assistance System(s)
ADS	Automated Driving System
AEB	Automatic or Autonomous Emergency Braking
API	Application Programming Interface
AR	Augmented Reality
C-ITS	Cooperative Intelligent Transport Systems
CAM	Cooperative Awareness Message
CAV	Connected Automated Vehicle
CCAM	Cooperative, Connected and Automated Mobility
CNN	Convolutional Neural Network
CPM	Cooperative Perception Message
DIL	Driver-In-the-Loop
ECU	Electronic Control Unit
ETSI	European Telecommunications Standards Institute
FMU	Functional Mockup Unit
FPGA	Field Programmable Array
GNSS	Global Navigation Satellite System
GPS	Global Positioning System

GPU	Graphics Processing Unit
GUI	Graphical User Interface
HIL	Hardware-In-the-Loop
HMD	Head-Mounted Display
HMI	Human-Machine Interface
HW	Hardware
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
ISMR	In-Service Monitoring and Reporting
ISO	International Organization for Standardization
ITS	Intelligent Transport Systems
LCA	Lane Change Assist
LiDAR	Light Detection And Ranging
LKA	Lane Keeping Assist
MIL	Model-In-the-Loop
MQTT	Message Queuing Telemetry Transport
NVH	Noise, Vibration, and Harshness
OBU	OnBoard Unit
ODD	Operational Design Domain
OEM	Original Equipment Manufacturer
ORAD	On-Road Automated Driving
OSI	Open Simulation Interface
PG	Proving Ground
Radar	Radio Detection And Ranging
ROS	Robot Operating System
RSU	RoadSide Unit
SAE	Society of Automotive Engineers

SAF	Safety Assurance Framework
SCDB	SCenario DataBase
SIL	Software-In-the-Loop
SUNRISE	Safety assUraNce fRamework for connected, automated mobIlity SystEms
SUT	System Under Test
SW	Software
TTC	Time To Collision
UC	Use Case
UI	User Interface
UN	United Nations
UNECE	United Nations Economic Commission for Europe
V&V	Verification and Validation
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything
VIL	Vehicle-In-the-Loop
VTD	Virtual Test Drive
VUT	Vehicle-under-Test
VR	Virtual Reality
WIFI	Wireless Fidelity
XIL	Everything-In-the-Loop

EXECUTIVE SUMMARY

Safety assurance of Cooperative, Connected, and Automated Mobility (CCAM) systems is a crucial factor for their successful adoption in society, yet it remains a significant challenge. It is generally acknowledged that for higher levels of automation, the validation of these systems by conventional hand-crafted test methods is not feasible. Furthermore, certification initiatives worldwide are looking for harmonized safety assurance approaches to enable safe large-scale deployments of CCAM systems.

The **SUNRISE** project develops and demonstrates a **CCAM Safety Assurance Framework** (SAF). The overall objective of the SUNRISE project is to accelerate the large-scale and safe deployment of CCAM systems by focusing on three core elements: a Method, a Toolchain and a Data Framework. The **Method** is established to support the SAF safety argumentation, and includes procedures for scenario selection, sub-space creation, dynamic allocation to test instances and a variety of metrics and rating procedures. The **Toolchain** contains a set of tools for safety assessment of CCAM systems, including approaches for virtual, hybrid and physical testing. The **Data Framework** provides online access, connection and harmonization of external Scenario Databases (SCDBs), allowing its users to perform query-based extraction of safety relevant scenarios, allocation of selected scenarios to a variety of test environments, and reception of the test results.

This deliverable focuses on **hybrid and real-world test methods** that are suited for the execution of scenario-based test cases. Scenario test case execution plays a central role in the SUNRISE Safety Assurance Framework (SAF). The execution environment is given a concrete scenario and a system under test (SUT) and determines the outcome of the test case. While this process can be fully virtualized, this deliverable explores options to execute the test case partially or completely in the real world.

Hybrid and real-world test methods for scenario execution offer a valuable alternative to virtual simulation for the execution of test cases in the SAF. This deliverable is relevant to both internal and external stakeholders who need to make an informed decision when choosing an execution environment. The content supports this decision process by providing an overview of the variety of hybrid and real-world test methods and explaining the **advantages and disadvantages** of different compositions of hybrid and real test method components.

The main content is a review of established and novel hybrid and real-world methods. In this deliverable, the context of real-world testing is limited to scenario-based testing. This deliverable plays a key role as a **knowledge base** for the selection of an execution platform that is suitable for the test case at hand and provides guidance for the implementation of hybrid and real-world execution platform. This deliverable sets a knowledge foundation and creates awareness for the inclusion of hybrid and real-world test methods into the scenario-based SAF of the SUNRISE Project.

There remains a need for further standardization and harmonization of individual hybrid or real-world test methods to promote modularity, comparability, and interchangeability.

1 INTRODUCTION

1.1 Project introduction

Safety assurance of Connected, Cooperative, and Automated Mobility (CCAM) systems is a crucial factor for their successful adoption in society, yet it remains a significant challenge. CCAM systems need to demonstrate reliability in all driving scenarios, requiring robust safety argumentation. It is acknowledged that for higher levels of automation, the validation of these systems by means of real test-drives would be infeasible. In consequence, a carefully designed mixture of physical and virtual testing has emerged as a promising approach, with the virtual part bearing more significant weight for cost efficiency reasons.

Worldwide, several initiatives have started to develop test and assessment methods for Automated Driving (AD) functions. These initiatives already transitioned from conventional validation to a scenario-based approach and combine different test instances (physical and virtual testing) to avoid the million-mile issue.

The initiatives mentioned above, provide new approaches to CCAM validation, and many expert groups formed by different stakeholders, are already working on CCAM systems' testing and quality assurance. Nevertheless, the lack of a common European validation framework and homogeneity regarding validation procedures to ensure safety of these complex systems, hampers the safe and large-scale deployment of CCAM solutions. In this landscape, the role of standards is paramount in establishing common ground and providing technical guidance. However, standardising the entire pipeline of CCAM validation and assurance is in its infancy, as many of the standards are under development or have been very recently published and still need time to be synchronised and established as common practice.

Scenario Databases (SCDBs) are another issue tackled by several initiatives and projects, that generally tends to silo solutions. A clear concrete approach should be used (at least at European level), dealing with scenarios of any possible variations, including the creation, editing, parameterisation, storing, exporting, importing, etc. in a universally agreed manner.

Furthermore, validation methods and testing procedures still lack appropriate safety assessment criteria to build a robust safety case. These must be set and be valid for the whole parameter space of scenarios. Another level of complexity is added, due to regional differences in traffic rules, signs, actors and situations.

Evolving from the achievements obtained in HEADSTART and taking other project initiatives as a baseline, it becomes necessary to move to the next level in the development and demonstration of a commonly accepted **Safety Assurance Framework** (**SAF**) for the safety validation of CCAM systems, including a broad portfolio of Use Cases (UCs) and comprehensive test and validation tools. This will be done in **SUNRISE**, which stands for **S**afety ass**U**ra**N**ce f**R**amework for connected, automated mob**I**lity **S**yst**E**ms.

The SAF is the main product of the SUNRISE project. As the following figure indicates, it takes a central role, fulfilling the needs of different automotive stakeholders that all have their own interests in using it.

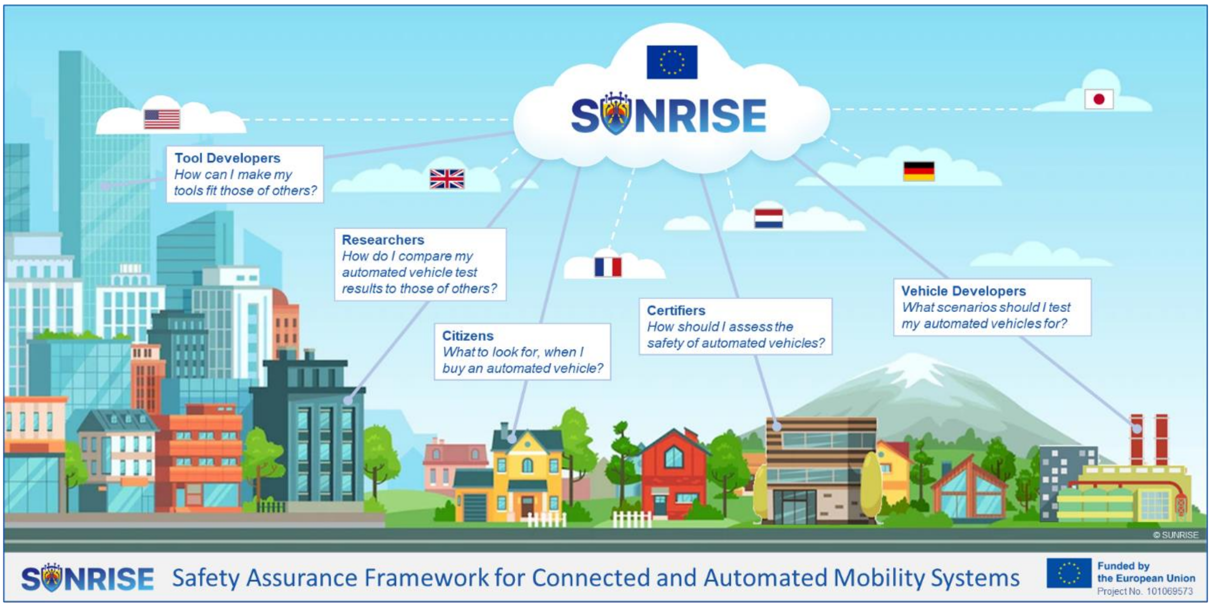

Figure 1: Safety Assurance Framework stakeholders

The **overall objective** of the SUNRISE project is to accelerate the safe deployment of innovative CCAM technologies and systems for passengers and goods by creating demonstrable and positive impact towards safety, specifically the EU's long-term goal of moving close to zero fatalities and serious injuries by 2050 (Vision Zero), and the resilience of (road) transport systems. The project aims to achieve this objective by providing a SAF consisting of three main components: a Method, a Toolchain and a Data Framework. The **Method** is established to support the SAF safety argumentation, and includes procedures for scenario selection, sub-space creation, dynamic allocation to test instances and a variety of metrics and rating procedures. The **Toolchain** contains a set of tools for safety assessment of CCAM systems, including approaches for virtual, hybrid and physical testing. The **Data Framework** provides online access, connection and harmonization of external Scenario Databases (SCDBs), allowing its users to perform query-based extraction of safety relevant scenarios, allocation of selected scenarios to a variety of test environments, and generation of the test results. The SAF will be put to the test by a series of **Use Cases demonstrations**, designed to identify and solve possible errors, gaps and improvements to the underlying methods, tools and data.

Following a common approach will be crucial for present and future activities regarding the testing and validation of CCAM systems, allowing to obtain results in a standardised way, to improve analysis and comparability, hence maximising the societal impact of the introduction of CCAM systems.

The following figure shows the general workplan of the SUNRISE project.

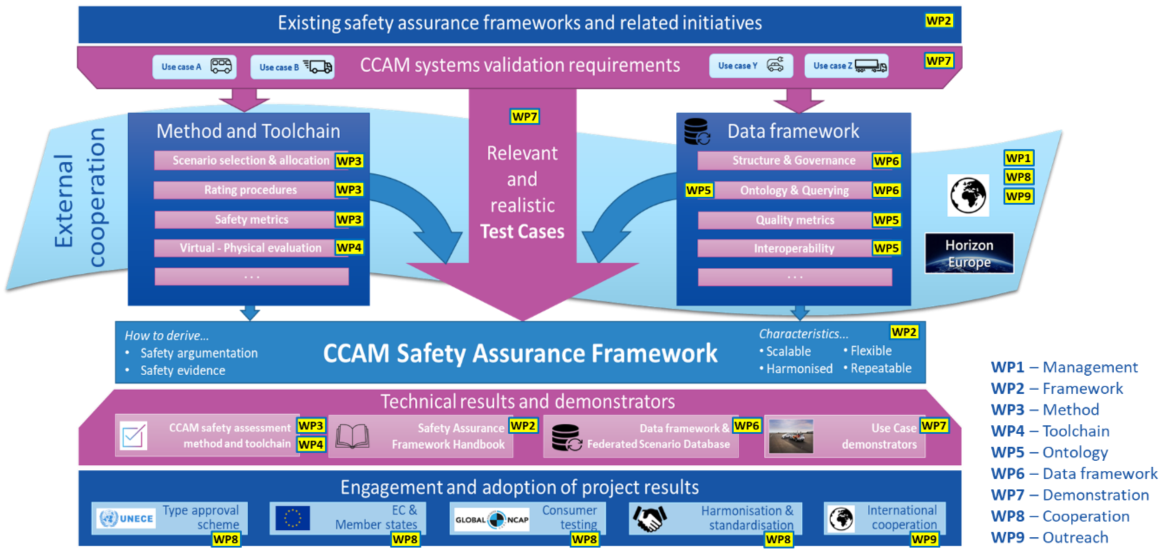

Figure 2: Workplan of the SUNRISE Project

1.2 Purpose of deliverable

This deliverable serves several important purposes within the SUNRISE Safety Assurance Framework, consolidating knowledge of testing methods and enhancing shared expertise among the project partners. It provides both an overview of hybrid and real-world test methods, specifically examining how these methods can be applied to scenario-based testing in SAF.

The primary goal of this deliverable is to consolidate the expertise from a diverse range of project partners and to disseminate this knowledge both within and outside of the project. By providing a detailed overview of hybrid and real-world test methods for execution in the SAF, the deliverable serves as a knowledge baseline that brings these methods into context and highlights their similarities and differences.

In line with SAF's overarching goals, this deliverable supports the allocation process for hybrid and physical execution platforms. It provides the foundation needed to understand and implement hybrid and physical testing methods within the SAF, furthering the project's objectives of running scenario-based tests in validated test environments.

The deliverable aims to significantly enrich the project's knowledge base, especially for those stakeholders who may not be fully aware of the capabilities and potential benefits of hybrid and real-world testing methods. The anticipated outcome includes improved decision-making and broader adoption of these methods, enabling stakeholders to leverage their unique assets and capabilities in the context of SAF applications.

The scope of this deliverable includes:

- An overview of the general Limitations of the Harmonized V&V Simulation Framework
- Scenario-based real-world testing on test tracks and public roads

- Hybrid methods that combine real-world and virtual testing on both test tracks and test benches
- Human-in-the-loop testing on test tracks and in driving simulators
- Additional Validation Needs of hybrid and real-world testing in comparison to virtual testing

The deliverable does not cover unstructured open-world testing or fully virtualized environments outside of these contexts. The validation of simulation components of the Harmonised V&V Simulation Framework is not covered in this task or deliverable. The validation of simulation components is instead covered in detail in deliverable D4.5.

This deliverable identifies general limitations of the SUNRISE Harmonised V&V Simulation Framework for virtual validation, highlighting where hybrid and real-world testing methods can enhance coverage of the Operational Design Domain (ODD). Each method is presented with an analysis of its advantages and disadvantages. Furthermore, the validation needs in comparison to pure virtual testing are discussed. Additionally, specific implementations and examples of testing methods are shown to provide a thorough and nuanced view of the options available. These examples also include implementation and validation approaches for core features based on requirements from Task 7.1 [1].

The goal of WP7 is to demonstrate various implementations of SAF modules and methods across different use cases. However, the hybrid and real-world methods explored do not aspire to deliver fully validated results and were still in development during Task 4.6. As a result, this deliverable does not include method validation results but instead outlines pathways for individual hybrid and real-world methods toward validation.

This deliverable is designed to be a practical resource for project partners and external stakeholders. Within Task 3.4 [2],it contributes to the allocation of test runs to appropriate test environments. In Task 7.3 , the deliverable is essential for the use-case-based demonstration of the SAF, supporting partners on choosing and implementing testing methods best suited to their specific use cases. Overall, this deliverable strengthens the SAF by facilitating informed, strategic decisions about the execution platforms and enabling partners and stakeholders to meet validation requirements.

1.3 Intended audience

This document is aimed at both project partners and project externals who already include or are considering hybrid or real-world methods in the CCAM validation process. The target audience includes stakeholders developing hybrid or real test methods as well as application-oriented stakeholders who should be aware of the differences, advantages and disadvantages of fully virtual, hybrid or real-world test methods.

This deliverable provides project partners and external stakeholders with a common understanding of hybrid and real-world testing methodologies, and how these methods enhance coverage of the ODD by compensating for limitations in virtual simulation. The testing method descriptions, comparisons, and practical examples in this deliverable enable stakeholders to select the most suitable testing methods for their specific use cases. The knowledge gained from this deliverable empowers stakeholders to leverage their unique assets and capabilities more effectively within the context of SAF applications. Understanding the unique validation needs of hybrid and real-world testing methods ensures that their testing processes produce reliable results.

1.4 Deliverable structure and relation to other parts of project

This deliverable is divided into four main chapters, each focusing on specific types of testing methodologies for connected, cooperative, and automated mobility systems.

Chapter 2: Hybrid and Real-World Testing – Addressing Limitations of the Harmonised V&V Simulation Framework introduces the basic concepts of hybrid and real-world testing. It highlights the motivation for exploring different testing methodologies by addressing the limitations of purely virtual testing. The chapter sets the stage for the discussion that follows by grouping the methods according to their core approaches. Finally, a classification approach for comparing test methods is discussed.

Chapter 3: Physical Testing on Proving Grounds and Real Roads looks at methods that evaluate whole systems under predominantly physical conditions. This includes testing on public roads and proving grounds. Secondly, reduced scale test environments are discussed, which provide a test environment of complete systems in reduced dimensions to gain insight into the behaviour of CCAM systems.

Chapter 4: CCAM Validation in the Loop sets the focus to methods for testing CCAM systems in hybrid environments. The Vehicle-in-the-Loop test methods are divided into two main categories: Vehicle-in-the-Loop testing on proving grounds and Vehicle-in-the-Loop testing on test benches. Both approaches allow detailed analysis of system performance while maintaining a high degree of flexibility and control of over the test conditions.

Chapter 5: Driver-In-the-Loop Testing for CCAM Validation examines methods that involve the human driver or passenger in the testing process. These approaches are essential for understanding how CCAM systems interact with and accommodate human users and provide a valuable tool for benchmarking CCAM systems against the human driver as a reference.

Chapter 6: Summary of Advantages and Disadvantages of Hybrid and Real-world Test Methods presents the key findings from Chapters 3,4 and 5 in a tabular format. The table in this chapter provides a clear overview of the advantages and disadvantages of various testing methods.

This deliverable integrates with several other project activities and outputs in the SUNRISE project. It builds on the definitions of the harmonized V&V Simulation Framework established in D4.4 [3]. This simulation framework is the basis for the investigation of limitations of virtual simulation and the basis for the definition of hybrid and real-world test methods.

The contents from this deliverable are closely linked to the contents of D4.5 [4]. Hybrid and real-world methods described in this deliverable can be employed to generate data for toolchain validation methods defined in D4.5. Conversely, applying the validation guidelines from D4.5 helps to identify scenario gaps that can be addressed using the hybrid or real-world testing methods described in this deliverable.

This deliverable supports project activities, particularly in WP7, by establishing a common understanding and consistent terminology for hybrid and real world test methods. In task T7.3, several of the hybrid and real-world testing methods described in this deliverable are employed as part of proof-of-concept SAF demonstrations. Task T7.3 relies directly on this deliverable as a knowledge base for hybrid and real-world testing methods and specific validation considerations, that are described for each testing methods.

2 HYBRID AND REAL-WORLD TESTING - ADDRESSING LIMITATIONS OF THE HARMONISED V&V SIMULATION FRAMEWORK

2.1 Introduction to Hybrid and Real-World Testing

The SUNRISE Safety Assurance Framework relies on scenario-based testing, where test cases are executed virtually, physically or through a hybrid of both methods. Testing methods must strike a balance between efficiency, accuracy and reproducibility.

Virtual testing takes place in a simulation environment and offers advantages over real time testing in terms of cost, effort, speed and reproducibility. The behaviour of relevant systems is replicated by parameterized models. However, for virtual testing to provide reliable results, the simulation tool chain must be rigorously validated for the specific use case.

In the scope of this document, **real-world testing equals physical testing**, such as full CCAM system testing on proving grounds or public roads. A common approach to real-world testing involves collecting data from a fleet of vehicles operating in public traffic. However, in the SUNRISE context real-world testing is adapted to the scenario-based SAF, by staging scenarios to assess system safety under controlled conditions. Within this deliverable we refer to this kind of scenario-based testing as public road testing. Since arranging a scenario on public roads and obtaining the necessary permits can be resource-intensive, it often makes more sense to replicate these scenarios on closed proving grounds. Proving ground testing can be used to expose the system under test (SUT) to real-world environmental factors and traffic participants. While virtual testing is often preferred due to its efficiency and repeatability, there are situations where physical testing provides greater value. These include cases that are difficult to accurately model and validate in simulation, such as wet road surfaces or other complex conditions. Physical testing is also beneficial when assessing a black box CCAM system that only exists in the real physical world.

In practice, the most efficient testing approaches can be a combination of physical and virtual methods, known as **hybrid testing**. This approach allows the benefits of both methods to be used to effectively address specific test requirements. Hybrid testing integrates physical elements, such as a car driving on a test track, with virtual components, like simulated traffic interacting with the system under test.

To understand where physical and hybrid methods are particularly valuable, the following chapter introduces the SUNRISE Harmonised V&V Simulation Framework for virtual testing and explores the potential limitations of virtual methods. Subsequently different kind of hybrid and real-world test setups are defined.

2.2 Summary of the Harmonised V&V Simulation Framework for Virtual Testing

A goal of SUNRISE is to define a harmonised V&V simulation framework for virtual validation of CCAM systems, as a core element of the SUNRISE SAF that should be applied by a wide variety of users. Therefore, it must be based on (open) standards and it should be possible to use non-commercial tools. In the SUNRISE deliverable D4.4 [3] partner simulation frameworks were analysed as well as the information of some relevant projects.

To this purpose, the interfaces and data formats of the individual subsystems were compared with each other and differences and similarities were analysed. The aim of the task was not to provide as complete a list of data formats as possible but to propose a harmonised V&V simulation framework. The analysis showed that data formats from the ASAM family are used almost universally. The OSI standard was found to be the most commonly used interface. Both were described in D4.4 [3]. With this information, an approach for a harmonised V&V simulation framework was presented.

The approach consists of a so-called base layer consisting of 4 interconnected subsystems:

- Subject Vehicle – Vehicle Dynamics
- Subject Vehicle - Sensors
- Subject Vehicle – AD function
- Environment

In this approach, the base layer is the core element that can be harmonised, because these four subsystems are essential for all simulations. That is the reason why it is possible to use standardised interfaces between these subsystems. The framework can be extended by the user in 4 dedicated dimensions:

- The Target Operational Design Domain
- The Vehicle Sensor set-up
- The Software architecture
- The Hardware architecture

More information on this approach can be found in D4.4 chapter 4.5 [3].

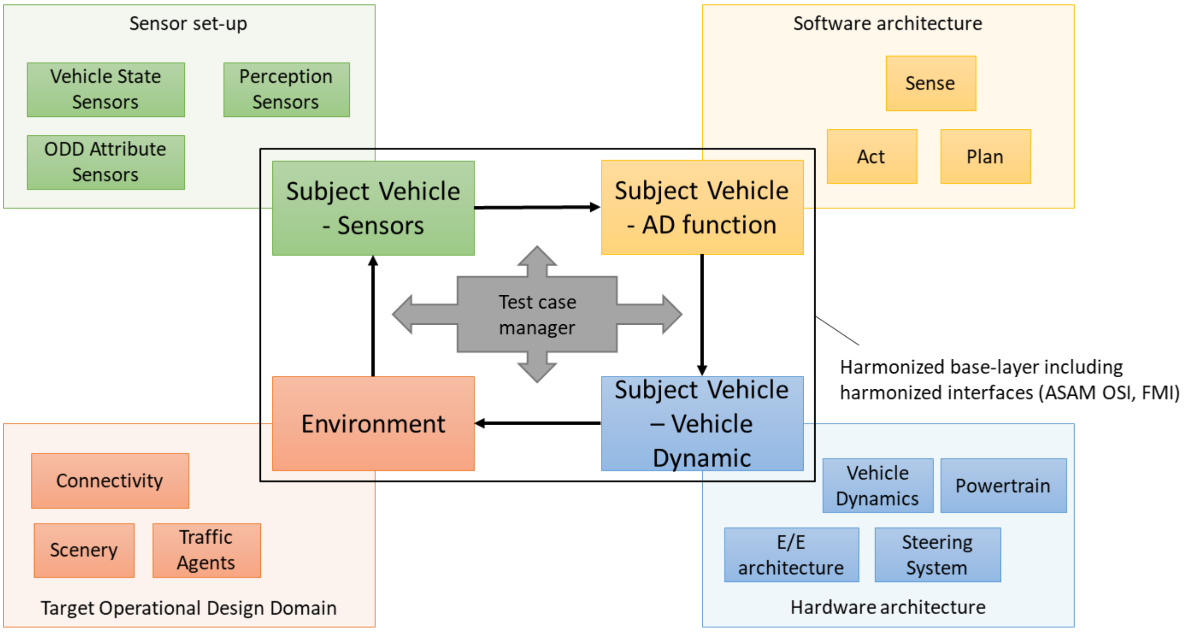

Figure 3: Architecture of the harmonised V&V Simulation Framework

2.3 Limitations of Virtual Testing

2.3.1 General Limitations

Virtual testing has a prominent role in safety-critical systems' testing due to no safety concerns involved and efficiency. With computer simulations, high quality training data can be generated at a fraction of a cost of traditional manual annotation while scenario and ODD (Operational Design Domain) variations can automatically be generated. In addition, new co-simulation environments allow testing external system aspects such as connectivity. Nevertheless, testing AD systems in simulated environments introduces three main shortcomings:

a. latencies in data exchange can be introduced by co-simulation models resulting in **a simulated SUT with different timing behaviour** than the SUT in real world,
b. the semantic gap between the real world and the simulation (with respect to aspects such as AD functions models' fidelity and driving environment fidelity), also known as **sim2real gap**, which is hard to quantify.
c. data exchange protocols for connected AVs may differ in real-world deployment scenarios (due to selected HW constrains) and in simulation, resulting in simulation results not easily transferable to real-life.

In the following sections, we further explain how these shortcomings can be present in different subsystems of a simulation pipeline and we add more practical limitations where applicable.

2.3.2 Limitations for the Subsystem "Subject Vehicle – Sensors"

In driver assistance functions, the perception layer, which is based on processing of the sensor data, has a crucial role. In order to simulate different sensor setups, different types of sensors used in automotive industry have to be modelled like Radar, LiDAR, Camera, Ultrasonic, GPS, IMU. The main limitation of this subsystem for all types of sensors is the discrepancy between behaviour of the sensor in simulation and behaviour of the sensor in real world due to how the sim2real gap referred above, affects these sensor models:

The next generation of sensor simulation products coming to the marketplace is capable of producing highly realistic visualization using technologies such as 3-D remodelling, physics-based rendering, ray tracing and dynamic lighting. The detail of different terrains, environmental lighting (i.e. haze, shadows), lens flare, glare effects, dynamic materials (i.e. rain, snow, fog), and much more can be achieved with 3-D photo-realistic quality, further advancing sensor-realistic simulation. While more realistic virtual sensors are now offered by commercial simulation tool providers, the sim2real gap remains and the degree of its effect depends on the accuracy of the sensor model. The sim2real gap also depends on the type of sensor being simulated since some sensors like cameras are more well studied than others in simulation environments.

A good sensor model is able to react to the lighting and reflectivity properties of the rendered environment (realistic textures, shadows etc.) in a way similar to what a real sensor device would react to a real condition on the road. However, this is extremely hard to be achieved when sim2real gap exists, i.e. both the sensor model and the environment model cannot represent all properties of the physical world. One possible way to bridge this gap and have credible simulations would be to have in the future an established partnership between simulation suppliers on one hand and Tier1s on the other hand in order to generate sensor surrogate models that will not explicitly represent how the physics of the sensor work but will be equivalent with how the real sensor works in terms of object detections. Another method for a simulation to be closer to its reality counterpart, is to use XiL (everything in the loop). But also, in the XiL environment in order to close the sim2real gap, one of the requirements is to have the sensor's noise model.

Vehicle-to-everything (V2X) can be considered as an external off-board vehicle 'sensor' and it is considered in SUNRISE UCs. To simulate V2X situations, multiple collaborative vehicles and roadside units (RSU) within effective collaboration distance are simulated. This is done through a co-simulation approach where driving simulation is paired with a network simulator usually through some dedicated interfaces or APIs. Existing open-source network simulation tools support compliance with C-ITS ETSI standards and offer different network parameters that can be tuned. The sim2real gap is also present here since emulating network aspects/ especially in dense urban environments is not always possible and hence network models used in simulation can over-simplify environment effects on V2X data exchange.

Other practical limitations of this subsystem include:

- **Sensor models based on physics** reproduce the measurement principles of the sensor (i.e. camera uptakes, radar wave propagation) and are used to simulate phenomena such as haze, glare effects or precipitation. They can generate raw data streams, 3-D point clouds, or target lists. Because these models address physical effects, their complexity level is much higher. Calculation typically takes place on a graphics processing unit (GPU) and hence **running the simulation requires a GPU equipped machine**. These models are typically validated in a SIL or HIL test setup.

Sophistication of the models may differ and the practice is that those models are **not shared** across suppliers or research community, hence **comparison among different setups and same sensor is difficult**.

- The creation of **sensor noise models for various ODD conditions** requires new datasets from sensors operating in those conditions (incl. rain, fog etc..). This is partially covered by European research projects like RoadView but these are generally missing or **not public**.
- High resolution sensor data **logging from multiple sensors in simulation** requires dedicated software and hardware.

2.3.3 Limitations for the Subsystem “Subject Vehicle – AD function"

In this part of the simulation and for pre-deployment AD functions, the sim2real gap is not so significant since the AD stack to be simulated, that usually involves perception/prediction, planning and control sub-modules, can be very close or even identical to the in-vehicle AD stack implemented e.g. in Robot Operating System (ROS). However, if the AD function in the real vehicle is intended to run on an embedded environment (deployment), this embedded AD stack should be interfaced with the simulation environment something which is not usually supported by the current simulation tools. This is a limitation.

Regarding the close-to-real time performance and timing limitations more generally (as described in the introductory part), this subsystem can be considered as OK. With new HW in computer graphic chips, near-to real time performance of the AD function in simulation can be achieved. Moreover, simulation SW timing issues can be overpassed since AD function can be tested on pre-recorded data during a simulation “Data replay” mode.

New limitations apply specifically for this subsystem for interactive AD functions, where control of the vehicle is shared between the AD system and the driver, driver-in-the-loop testing is also needed. In those system-user interaction environments many limitations still exist mainly related with the ability of the driving simulators to create an immersive environment for the user in order for the experiments to be engaging and the results to be representative of what could happen in a real driving cabin of a real vehicle.

2.3.4 Limitations for the Subsystem “Subject Vehicle – Vehicle Dynamics"

Commercial tools for vehicle dynamics modelling and simulation have been long used by the automotive industry to support their testing pipeline and especially before introducing new ADAS functions. Requirements for realistic vehicle dynamics simulation are even stricter for AD systems under test. The limitations here relate again to sim2real gap and the ability of engineers in reproducing in models the real physics of the vehicle parts and their relation to the wind and the terrain effects (estimate traction to wet surfaces, estimate the longitudinal and the lateral stiffness of a tire etc.). To overcome those limitations, the idea and need for vehicle-in-the loop simulation in test tracks has emerged as possible complement of virtual testing which in turn brought new testing requirements.

2.3.5 Limitations for the Subsystem "Environment"

New photorealistic driving environment simulations are now available by commercial (e.g. rFPro) and open source (e.g. CARLA using Unreal Engine 5+) tools allowing for realistic representation of reflectance and texture properties of 3D surfaces and allowing for an immersive simulation environment for human-in-the-loop testing. While this leads to a variety of tools being offered, a main limitation of this subsystem is that the degree of fidelity supported by various simulation tools may differ a lot and is not always comparable, leading to results that are not easily comparable and thus making the role of the CCAM regulators very important and quite hard for instance, each tool should be validated for its fidelity w.r.t:

- Weather Effects: Simulations should account for weather conditions such as rain, snow, fog, or extreme temperatures. These environmental factors can significantly impact sensor performance (e.g., camera, lidar, radar) and vehicle dynamics.
- Day/Night and Lighting Variations: Changes in lighting conditions due to the time of day, shadows, or artificial lighting should also be modelled to ensure the AD system performs well under various visibility conditions.
- Road Surface Modelling: The simulation should include the impact of road conditions such as wet, icy, or unpaved roads on vehicle handling and sensor performance.

Simplification of reality based on the above factors can for instance involve idealized road geometries like straight roads, simple intersections and neglecting minor road features or buildings representations that differs from real-world data requiring AD function re-training for simulation and quantification of sim2real gap which is not trivial. This results in a major limitation i.e. not being able to virtually test realistic ODD conditions also limiting the required granularity in the synthetic ODD representation.

Finally, as already discussed in the introduction, the more complex the environment model (high fidelity models), the more computation power is required to run and hence if real-time rendering during simulation execution is required this can be difficult to achieve.

2.4 Classification of Hybrid and Real-world Testing Techniques

This chapter will classify and categorize various test methods that play a role in the safety assurance process. In order to provide an overview and to achieve a common understanding of testing nomenclature, it is shown which aspects of the listed test methods can be implemented virtually and physically.

The possible test methods cover a broad spectrum and are implemented to varying degrees in the real world and in the virtual world. Table 1 illustrates which test methods contain various combinations of real and virtually implemented subsystems that were presented in chapter

2.2. These subsystems are extended by the human driver, who was omitted from the virtual validation process and can interact with or evaluate the system under test.

When a vehicle drives in real traffic and scenarios are staged, each component is implemented in real life. This method, known as “Real Road Testing”, come closest to real natural driving situations. If situations are played out on the test track, known as “Proving Ground Testing”, some conditions such as the road surface and road layout are idealized, but all components are still implemented in the real world.

Table 1: Categorizing Hybrid and Real-World Testing Methods by Subsystem Implementation

Testing Method	Subject Vehicle **Sensors**	Subject Vehicle **AD Function**	Subject Vehicle **Vehicle Dynamics**	**Environment**	**Human Driver**
Proving Ground and Public Road Testing	Real-world	Real-world	Real-world	Real-world	Real-world
Scaled Model Testing	Real-world	Real-world	Real-world	Real-world	Not implemented
Vehicle in the Loop – Proving Ground	Real-world or Virtual	Real-world	Real-world	Real-world or Virtual	Not implemented
Vehicle in the Loop – Testbench	Real-world or Virtual	Real-world	Virtual	Real-world or Virtual	Not implemented
Driver in the Loop – Proving Ground	Real-world or Virtual	Real-world	Real-world	Real-world or Virtual	Real-world
Driver in the Loop – Driving Simulator	Virtual	Real-world	Virtual	Virtual	Real-world

Real-world   Real-world or Virtual   Virtual   Not implemented

An interesting approach that also makes it possible to implement every component in real physical form, but in a scaled-down version, is called “Scaled Model Testing”. In this way, costs can be reduced, but it is no longer possible to use a real driver to test ADAS systems or obtain human feedback.

The “Vehicle-in-the-Loop - Proving Ground” method is derived from the Proving Ground tests mentioned above. In the vehicle-in-the-loop approach, parts of the test situation are implemented virtually and the real vehicle is integrated into the simulation loop. For example, other road users as well as route information and perceptions can be simulated virtually and fed into the perception of the CCAM software.

If the vehicle-in-the-loop method is transferred to a test bench, the entire environment, i.e. both traffic and static objects, can be simulated. The information can be fed digitally to the vehicle under test via data interfaces or physical via simulators such as radar, Global Navigation Satellite System (GNSS) or motion stimulators. While the environment exists virtual, the environmental interaction is replicated in the physical world.

In addition to the test methods that can be used as parts of the SAF to execute test cases, human-in-the-loop methods also play an important role in validation. A distinction is made between two different types of test methods. One is the classic driving simulator test, which simulates the environment and vehicle behaviour, but offers an interface to a real driver through a vehicle mockup. Alternatively, there is the “driver-in-the-loop test on the test track”, which makes it possible to combine the flexibility of driving simulation with real driving dynamics.

The various advantages and disadvantages resulting from the different combinations of real and virtual components are explained in more detail in the following chapters.

3 PHYSICAL TESTING ON PROVING GROUNDS AND PUBLIC ROADS FOR CCAM VALIDATION

This chapter focuses on two main physical testing approaches: Black Box Testing with production vehicles (chapter 3.1) and Scaled Model Testing (chapter 3.2). It should be noted that the testing approaches discussed could also be applied to prototype testing (white box) by OEMs, as they provide a foundation for evaluating the performance and functionality of CCAM systems at various stages of development. In addition, white box testing allows the access to internal parameters of the system to be accessed for a more detailed analysis and optimization.

3.1 Black Box Testing on Proving Grounds and Public Roads

Physically testing of Cooperative, Connected and Automated Mobility (CCAM) systems is a critical step in ensuring their safety, reliability and performance under real-world conditions. To guarantee the safety of these tests critical scenarios will be commonly realized on a proving ground. In addition, tests can be also carried out in realistic infrastructure on real roads. A black box testing approach is often employed to assess the system's functionality without requiring internal knowledge of its architecture or implementation. This method focuses solely on the inputs, outputs, and observable behaviour of the system.

The physical validation of a CCAM system with testing in a black box approach is generally used when a production vehicle or a prototype in its final development stage is to be evaluated. Especially testing in the field of consumer protection organizations, type approval authorities or market surveillance of a CCAM system requires the use of testing without the support of the Original Equipment Manufacturer (OEM). This means that the technical service or the test lab which executes the tests has no further insights into the system's technology.

In this approach the CCAM system also called "Vehicle under Test" (VUT) is seen as a black box, where the input is the scenario which is staged and the output is the observed performance of the system reacting to the staged scenario (see Figure 4).

Figure 4: Black box testing schematic

To validate the performance of the system, the VUT is equipped with cameras, additional sensors and measurement systems, mainly an inertial measurement unit combined with D-GNSS (Differential-Global Navigation Satellite Systems) to measure the vehicle dynamics parameters, like acceleration, speed and position. The measurement devices have to be selected to measure the relevant data channels to allow the calculation of the performance metrics (e.g. Time-To-Collision or impact speed) to evaluate the system's performance. Depending on the test scenario to be assessed, additional test tools such as representative dummies, targets and their propulsion system can be included and synchronized with the VUT to stage a safety critical scenario.

Test objectives on CCAM systems are likely to include the requirement that it's safety performance should significantly exceed that of human drivers. For a holistic view of the similarity between a system and a human driver, the individual judgement of a human driver can be used as an additional performance criterion. The objective should be to determine the relevant metrics and their threshold values at which a human driver judges the system performance to be inadequate, i.e. less safe than that of a human driver, and as a result, interrupts or stops the operation of the system. In order to identify causal factors in the system performance in such situations, eye-tracking can be used to make assumptions about the human driver´s decision making process.

A proving ground as a test site provides a safe environment in which safety-critical test cases can be performed that could cause a possible impact with a dummy or target. On the other hand, proving grounds are limited to the performance of the ADs system due to the linkage of the functionality of the ADs function to specific road types (e.g. 'Geofencing' and/or as defined in the ODD). To test the full functionality of the ADs system, tests on public roads are required in addition.

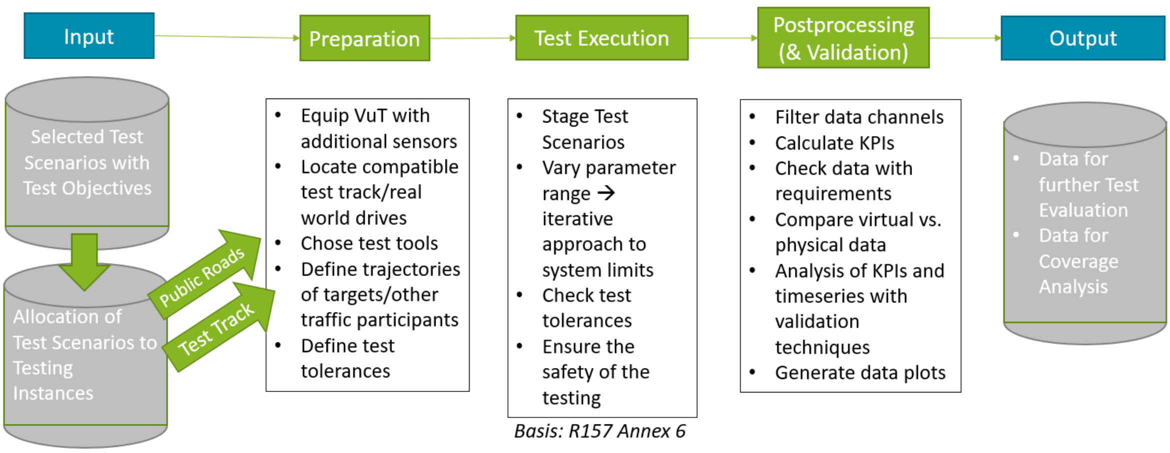

Figure 5: Black box testing process

The process including the preparation, test execution and postprocessing of the developed black box testing approach can be seen in Figure 5. This process is directly linked to the Safety Assurance Framework and is located in the "Execution"-block. It describes the steps involved in applying black box testing to a specific set of scenarios that were allocated to the "test track" and/or "public roads" test instances. The output is data that is further processed in the "Analyse"-block of the SAF (e.g. for test evaluation and coverage).

This approach also ensures that an important part of a precise diagnostic of CCAM systems is fulfilled, that is called content validity. Content validity is given if a test addresses the measurement of the intended construct in every of its aspects completely [5]. To define the construct, the expected feature behaviour in the selected test scenario needs to be fully operationalized by test objectives and adequate metrics. Test scenarios have been allocated to test instances due to methods described in WP3. In the preparation phase a test vehicle will be equipped with additional sensors and a measurement and data acquisition system and the planning of tests will be done. For that a suitable proving ground or route on public roads has to be chosen. The trajectories of the other objects in a scenario have to be defined and programmed in the test tools (dummies, targets and their propulsion systems). During test execution, the scenarios will be staged and synchronized to the VUT. After test execution, the recorded data channels will be post processed and metrics calculated that allow an assessment of the system.

To reach a higher validity of the diagnostic of the CCAM system, besides content validity, the convergent validity of testing should be addressed. Convergent validity is given, if the results of two different measurements of the same construct match [5]. In the case of public road testing of CCAM systems, the opportunity to simultaneously assess the objective safety performance of the system together with a subjective judgement of a CCAM system user arises. The objective could therefore be to assess if a user judges, purely subjective, the system performance to be less safe than that of a human driver and as a result interrupts or stops the operation of the system. In order to identify causal factors in the system performance in such situations, eye-tracking can be used to measure the user’s visual attention allocation in those situations. This makes it possible to determine whether the user based his decision to interrupt the system operating on its poor performance and what kind of situational factors contributed to this decision. Including a user in public road testing makes it therefore possible to assess the convergent validity of the testing, by matching the system performance metrics to the user behaviour and their acceptance of the CCAM system performance. When the judgement of the CCAM user should be included in the testing, it is important to create a sufficiently large number of samples, to avoid a biased system performance judgement by outliers. In general, the measurement of the performance metrics should be highly reliable to ensure a precise diagnosis of the CCAM system. This means that the demonstration of their stability is inevitable and also therefore a sufficient large number of test repetitions for each scenario is needed [6].

A well designed and developed test for CCAM systems that meets the highest possible quality requirements makes it possible to predict the system behaviour, also outside the test scenarios. Public road testing is a good approach for achieving this goal, when conducted as described and based on a representative sample. Apart from that and mentioned already in Chapter 2 public road testing has the advantage that it exposes a CCAM system to real-world environmental factors and traffic conditions. Only there CCAM systems safety advantages or disadvantages are clearly visible, because random interaction effects with the environment and surrounding traffic can take place.

3.1.1 Physical and Virtual Components on Subsystem Level

Environment

On a proving ground the environment is physical, but the conditions are standardized built up with ideal infrastructure (no natural environment: no obstructions, optimized asphalt). Possibilities exist to use a test track with equipment that simulates additional environmental conditions, like a rain or fog simulator [7] [8].

For public road testing the environment is also physical and allows the full functionality of the ADs function of the CCAM system. Safety critical aspects of the surrounding traffic must be taken into account for the test execution. It is recommended to choose an iterative approach to the system limits and begin the testing with safety uncritically scenarios.

Sensors

Due to the use of a real prototype or series production vehicle the sensors are physical.

AD function

Due to the use of a real prototype or series production vehicle the AD function is physical.

Vehicle Dynamics

Due to the use of a real prototype or series production vehicle the vehicle dynamics are physical.

3.1.2 Advantages in Comparison to Virtual Testing

In general, black box testing enables testing a wide range of scenarios without modifying the system and ensures safety-critical functionalities work as intended regardless of the underlying implementation. Testing a CCAM-system in a black box approach means that you validate a production vehicle which has an implemented ADS function and all systems perform as if they are under normalized conditions on public roads. The perception will be fully tested and the ODD is the real one as for the series vehicle. Another advantage of the physical black box testing is that no virtual, dedicated vehicle model with the dynamic performance parameter of the VUT and its sensor system is needed. This vehicle model would be only available from the vehicle manufacturer. Black box testing is mostly independent of support by the manufacturer of the vehicle. Finally, black-box testing is essential for validation, ensuring that the system meets specified requirements and functions reliably on public roads and proving ground environments, providing a more thorough and realistic assessment than virtual testing alone.

As mentioned in the introduction of Chapter 3 black box testing can also be applied to prototypes, but there is a higher likelihood of obtaining additional insights into the system (also called white box). Since prototypes often come with more knowledge from the developers about the internal workings, they may unintentionally influence the testing process, leading to a less pure black box approach. This deeper understanding could result in the collection of more internal metrics, which might affect the objectivity of the test.

3.1.3 Limitations in Comparison to Virtual Testing

While effective, black-box testing has its limitations: It may not uncover deep, architecture-specific flaws, and it requires a well-designed set of test cases to ensure comprehensive coverage. Additionally, the physical testing of CCAM systems often demands significant resources, including specialized test tracks and equipment. It is also possible to test a black-box as a virtual testing system (Functional Mockup Unit - FMU) but that is an exception. One limitation is the availability of resources: test track rent is expensive; the test instance is limited in the usage of a small number of targets at the same time. Virtual testing can run many tests in a short amount of time with quickly changing the environmental conditions for each individual test. This fact can help in developing a software feature. In physical black-box testing you are very limited in test capacities and can only execute a limited number of tests per day.

The performance of an ADs system is limited on proving grounds due to possible “Geofencing” combined with the ODD of the production vehicle. Geofencing in the context of CCAM systems refers to the creation of virtual geographic boundaries using GPS or other location-based technologies. It is used to define specific areas where a vehicle is allowed to operate or restricted from entering. Geofencing is commonly applied to enhance safety, compliance, and operational efficiency for vehicles.

If you are testing on public roads, you are more limited due to the surrounding traffic, limitations of infrastructure and safe test scenarios to not endanger any other. Edge cases at the limits of the CCAM cannot be performed on public roads. For some test you may need a permission of the state or owner of the chosen route.

3.1.4 Additional Validation Needs

When testing on a test track in a black box approach depending on the function there won’t be the functionality of the ADs system due to the linkage of the ODD to specific road types and other infrastructural limitations. Systems with Society of Automotive Engineers (SAE) Level 3 or higher and SAE Level 2 systems with ‘hands off’ functionality are nowadays limited in their performance on the test track due to possible “Geofencing” (see limitation chapter). Verification test runs on public roads are needed to validate the performance seen on the test track. A two-step method, as described in EU R157 [9], can be used to validate CCAM systems. First, testing occurs in a controlled environment, such as a test track, to ensure safety. In the second step, the system is tested on public roads to confirm its reliability in real-world conditions.

3.1.5 Exemplary Implementation and Setup – UC 2.1

An exemplary implementation of the validation method described in chapter 3.1 will be demonstrated with a link to Use Case 2.1 from the demonstration work package 7 [10]. This contains several ALKS scenarios (e.g. Cut-Ins, Cut-Outs) aligned with the UN Regulation R 157 [9]. The regulation describes a two-step procedure in which safety critical tests will first be executed on a test track, and after if these tests are passed, tests will be executed on public roads.

To demonstrate the methodology an exemplary state-of-the-art vehicle with an ADS function was rented and equipped with additional sensors. As the scenario setup consists of two vehicles (see Figure 6) a second vehicle will be also set up for testing.

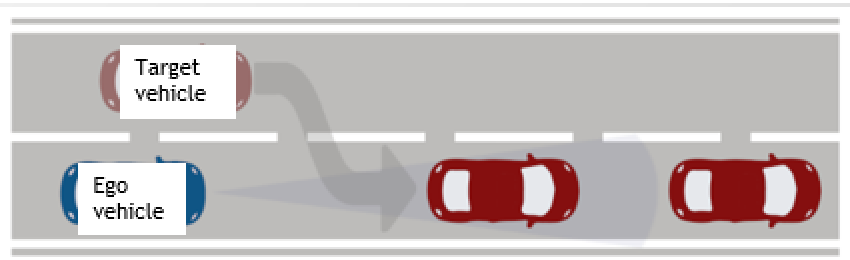

Figure 6: Cut-In Scenario from Use Case 2.1

For proving ground testing the second vehicle was represented by a soft car target propelled by an ADAS platform (see Figure 7).

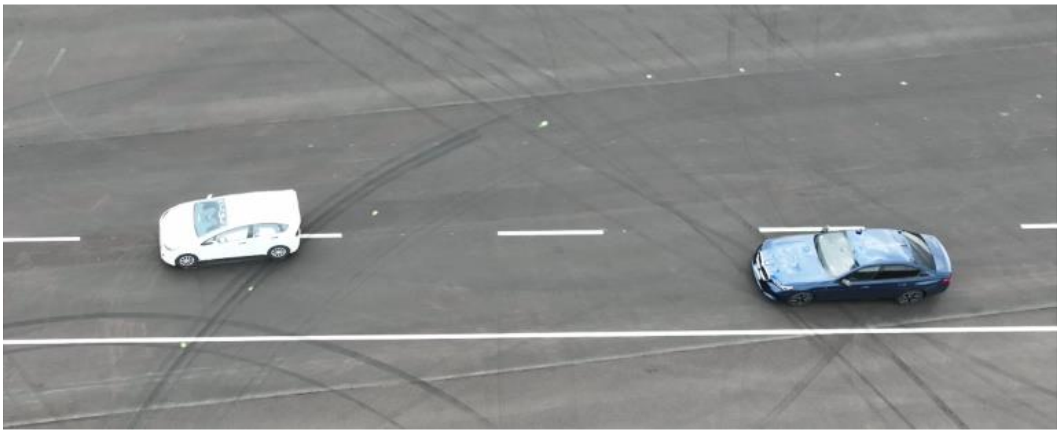

Figure 7: Set-up for proving ground testing with crashable soft car target and ego vehicle

For real-world-testing the target vehicle was represented by another series vehicle which was driven by a test driver (s. Figure 8). There were no driving robots used because of the forbidden usage of these vehicles on real roads. To ensure the data connection between the two vehicles a Wireless Fidelity (WIFI) network has been set up with which the relative positions between these vehicles have been measured.

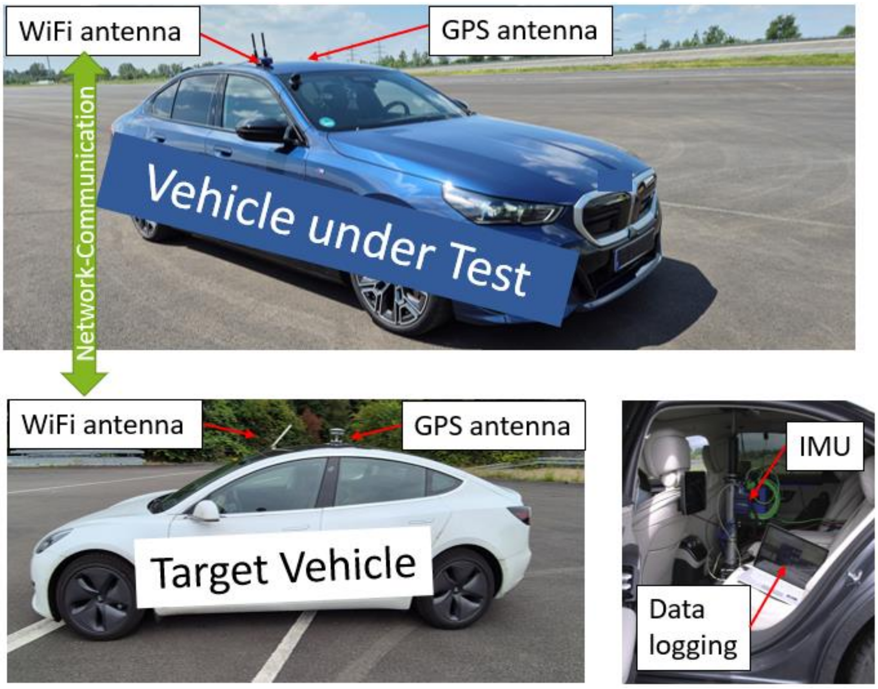

Figure 8: Set-up for public road testing with ego and target vehicle

Validation Requirements

The high-level validation requirements from D7.1 [1] for an Automated Driving System (ADS) in this specific Use Case 2.1 encompass several essential areas, including functional safety, scenario generation, test frameworks, and user perspective validation. These requirements are critical to ensuring that the ADS operates safely, efficiently, and in compliance with regulations. To validate these requirements, physical CCAM testing in a black-box approach, where the internal workings of the system are not considered, can be employed effectively.

For functional safety, physical tests can validate the system's ability to keep the vehicle in lane, detect obstacles in real-time, and adapt to speed limits or road conditions by measuring the vehicle's external behaviour, such as steering, braking, and speed adjustment. Obstacle detection and lane-keeping can be tested on proving grounds, where the system's responses are monitored and compared to expected outcomes. In terms of scenario description, physical testing ensures the vehicle complies with regulations like UN-R 157 [9] and handles various driving tasks such as lane-keeping and curve navigation. Testing in public road conditions, such as maintaining speed limits and executing lane changes, helps verify these requirements. The test framework requires the vehicle to have a perception system and map data integration. These can be validated through public road testing, where the vehicle's ability to detect obstacles and track lane markings in various conditions is observed. Simulation models can be validated by comparing public road and proving ground test outcomes with simulated predictions, ensuring the system's performance aligns with the expected results. For user perspective validation, a configurable KPI dashboard enables easy evaluation of results. During physical testing, the dashboard can monitor key performance indicators like time-to-collision (TTC), acceleration, and vehicle speed, ensuring the system's performance meets the required standards.

In summary, physical CCAM testing in a black-box approach allows validation of the ADS requirements by observing its behaviour on proving grounds and public roads, ensuring it meets safety, regulatory, and performance standards.

3.2 Scaled Model Testing

A CCAM system involves many interconnected subsystems and components, including sensors, actuators, decision and control algorithms, and a communication system. Full-scale testing [11] of these systems, which involve the intended system and the real-world test environment, poses significant challenges. These challenges are related to the high cost of testing, safety risks to the testers and the surrounding environment, and limited accessibility of the real-world test environment [11, 12]. Virtual testing does not pose same challenges, but may fail to accurately capture the complexities and dynamics of real-world environments to its full extent, such as sensor interactions, physical constraints, and communication delays. Further, physical tests are necessary to validate the correctness of the virtual test.

To overcome above mentioned challenges with full-scale testing, testing with scaled models have been proposed as an intermediate solution where the scaled down version of the actual system is used for testing, verification and validation of the virtual test results [11]. For test runs suitable for scaled models, the physical tests can be performed to much lower cost and to much lower safety risk compared with testing with full-scale vehicles. Consequently, the small-scale test platforms bridge the gap between purely simulation-based testing and real-world test setups, offering a practical and flexible approach for CCAM system evaluation.

3.2.1 Physical and Virtual Components on Subsystem Level

In this section, we briefly discuss the key components used in scaled model testing and the fact that some choices would need to be made to facilitate this type of testing. Note that, scaled model testing does not require any virtual elements; therefore, the environment and sensors are fully physical.

Environment

However, since the system under test is reduced in dimension, some elements of the environment would also need to be set up to take this reduction into account. For example, the images captured by a camera installed on a scaled vehicle are most likely taken from a different angle than if the images were taken by a full-size vehicle or in a simulation environment. Therefore, the environment might also need to be adapted, for example through reduced scaled modelling of the infrastructure. On the other hand, there are other aspects of the environment such as lighting, weather conditions and even perhaps the communication that may not need to be scaled and could be kept in the same way as done in the virtual environment.

Sensors

To keep the testing costs and complexities to a reasonable level, the sensors used might be the ones used in a real production vehicle or those that could represent these sensors in a realistic manner. It is also worth noting that, due to the reduced dimension of the test vehicle, it might not be physically possible to mount certain sensors that are used in a real production vehicle on a scaled vehicle.

AD function

The AD function is typically a representation of the one used in a real production vehicle. It might need to be adapted to take into account the design choices made when facilitating scaled model testing. Using the libraries and tools available by operating systems such as ROS (robot operating system), one might be able to close the gap between the implementation of AD function used in the virtual, scaled, and real-world vehicle testing.

Vehicle Dynamics

The same as for the AD function, the vehicle dynamics of scaled vehicles are generally a representation of the ones used in a real production vehicle. However, the representativeness of the vehicle dynamics models is tightly connected to the fidelity of the design choices made some of which might have been made due to cost-related reasons, while others might have to do with the complexity of having a fully representative vehicle dynamics model.

3.2.2 Advantages in Comparison to Virtual Testing

Small-scale test platforms provide a controlled and cost-effective environment for real-world testing, enabling researchers to conduct safe, repeatable experiments without the logistical challenges of full-scale testing [12]. Small-scale testing offers an effective means to validate results from virtual test environments.

In addition to their affordability and convenience, these platforms enable detailed evaluation of both high-level system behaviour, such as decision-making and coordination, and low-level interactions, such as sensor-actuator performance and vehicle dynamics in the real-world environment. By replicating key aspects of real-world conditions on a smaller scale, these platforms deliver valuable insights into system performance while minimizing risks and resource usage.

Small-scale testing, while highly valuable for bridging the gap between simulation-based testing and full-scale tests, also has certain limitations that we will describe in detail in the next section.

3.2.3 Limitations in Comparison to Virtual Testing

In this section, we describe the limitations of the small-scale testing compared to the virtual testing also known as simulation-based testing.

Physical Constraints

As opposed to simulation-based testing, small-scale test platforms are comprised of scaled down versions of the actual system and real-word test environment. Figure 9 shows an example of a scaled-down version of the actual full-size truck that RISE used for testing and verification of a braking function related to UC4.1. Due to a real-world nature of small-scale testing, the tests are bound by certain physical limitations such as test space, hardware scalability as shown in the figure below and its precision of scaled-down components and systems. These factors restrict the scope of the small-scale testing.

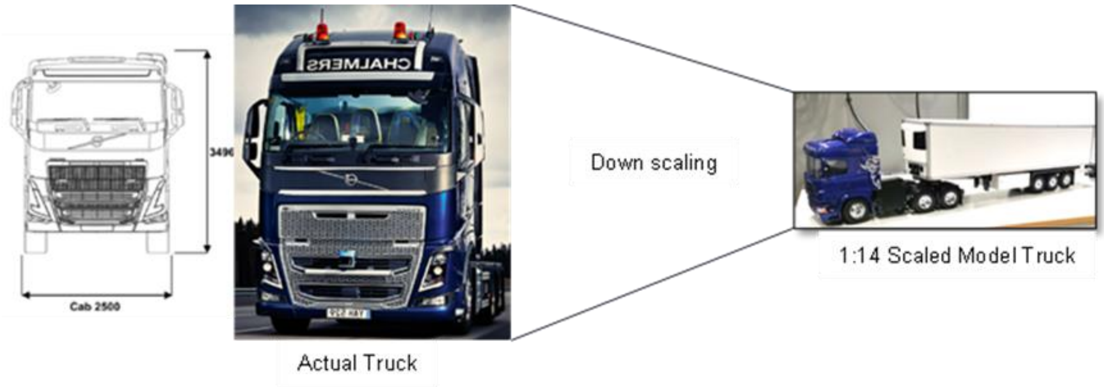

Figure 9: Representation of an actual full-size truck and its scaled-down version that is used in UC4.1 testing.

Cost and Test Setup Effort

Building and maintaining a small-scale test platform requires investment in hardware (e.g., scaled-down vehicles, sensors, actuators) and infrastructure (e.g., indoor environments). This can be more resource-intensive than running tests in purely simulation-based environments [13].

Limited Scale

Even though small-scale test platforms can replicate interactions with multiple agents, they cannot match the scale of virtual testing, where tens of thousands of vehicles or nodes can be modelled effortlessly in a simulation environment [13].

Simplified Environments

Small-scale test environments are simplified representations of real-world scenarios. Certain factors, such as weather conditions, road textures, or external interferences such as surrounding traffic, may be difficult to recreate physically at a small scale but can be modelled virtually with high fidelity.

System, Sensor and Actuator Downscaling Challenges

While small-scale testing can replicate system, sensor and actuators and their interactions, the scaling may introduce inaccuracies. For example, the dynamics of scaled-down systems such as a vehicle, its sensors or actuators may not fully capture the behaviour of their full-scale counterparts, leading to discrepancies in the test outcomes. Integrating the actual sized sensors into the small-scaled system imposes an extra challenge to the small-scale testing.

Flexibility and Reproducibility

Virtual testing allows for easy modification of scenarios, parameters, and environments, making it more flexible. Small-scale testing requires physical reconfiguration, which can be time-consuming and less adaptable to rapidly changing test requirements such as weather and varied traffic scenarios.

Data Collection

In simulation-based testing platforms, it is possible to generate detailed and comprehensive logs for every aspect of the simulated target system such as vehicle dynamics, communication protocols, sensors and their interactions, and decision-making algorithms. This includes logging parameters and capturing data with high granularity and frequency that is often not feasible in small-scale testing environments. Additionally, virtual testing provides greater flexibility in customizing what data is recorded. Parameters can be easily added or modified during runtime, and simulations can be paused or rerun with identical conditions, ensuring consistency in data collection. Small-scale testing, in contrast, is limited by the capabilities of physical constraints such as hardware and the fixed test environment, which may not support frequent logging.

Furthermore, virtual or simulation-based test environments allow seamless integration with data analysis tools, enabling automatic visualization, performance comparisons, and statistical evaluations. These features make virtual testing an invaluable tool for diagnosing complex system behaviours and fine-tuning CCAM systems strategies before progressing to small-scale or full-scale physical testing.

3.2.4 Additional Validation Needs

Validation is assuring that the requirements for a specific intended use have been fulfilled. Therefore, it is of utmost importance to validate the test results produced using scaled model testing to be able to argue about the usefulness of these results. Similar to the models used in a virtual test environment, the fidelity of the ones used in scaled model testing could significantly influence the usefulness of the test results. Therefore, the additional validation needs in the scaled model testing correspond to the elements and their fidelity used in this type of testing. This includes the environment, sensors, AD function, vehicle dynamics, etc.

When it comes to the environment, in the case where it is also scaled (alongside the vehicle dimensions), it is necessary to validate the representativeness of the scaled environment to be able to make good use of the test results. This includes scaled models of any objects including other vehicles, pedestrians, road infrastructure, etc. The fidelity of these models would also need to be validated for virtual testing, even though no scaling is necessary in the virtual test environment. However, it is worth noting that while virtual testing might be beneficial when testing known-unsafe [14], scale model testing is much more powerful in testing the unknown-unsafe [14].

The additional validation needs are different when it comes to sensors and AD functions. Here scaling in dimension is not as relevant as the one for the environment. However, the sensors and AD function used might not be representative of the ones used in real vehicles. This could even be the case in virtual testing as representative modelling of certain real vehicle sensors and AD functions in the virtual environment might be complex. Therefore, it is of importance to validate the sensor data and compare it with that produced by sensors in real work vehicles as well as to implement the AD function that is representative of that used in the real-world vehicle.

The validation would need to even be expanded on the different components used to model the scaled vehicle as they could significantly influence the vehicle dynamics. This validation issue is less apparent in virtual testing as in many cases the digital twin of the system under test is available.

3.2.5 Exemplary Implementation and Setup – UC 4.1

This section provides an overview of the physical setup to test the scaled truck's backing in a predefined scenario. The components for the scaled testing can be divided into the following subsystems:

- Environment – modelling the scene and operational domain where the tests take place.
- Sensors – the sensing technology on the vehicle to enable it to perceive the environment.
- Vehicle dynamics – actuators and physics responsible for moving the vehicle.
- AD function – responsible for controlling the truck to perform a sufficiently accurate reversing of the truck.

This division of subsystems closely aligns with the subsystem division presented in D4.1 [15] for the simulation system. Thus, making it viable to compare the simulation and the scaled model. The following sections provide the setup details for each of these subsystems.

Environment

To keep as close to the actual driving conditions relevant to the test of the physical truck as possible the tests are executed in an outdoor environment to create a similarity between the real truck and the virtual environment. Thus, the scenery in the simulation of a docking station is printed and placed appropriately to maintain similarity with the visual environment. A preliminary set-up of the scaled model test site is shown in Figure *10*, However, as the tests do not emphasise the accuracy of the micro details of the driving behaviour, such as road friction, these are not modelled explicitly to be of a similar nature. However, we chose a relatively standard road surface to drive on. Some of the test scenarios rely on weather parameters, specifically light, for this artificial light will be used along with a luminosity sensor mounted on the truck to measure the light intensity during the different runs.

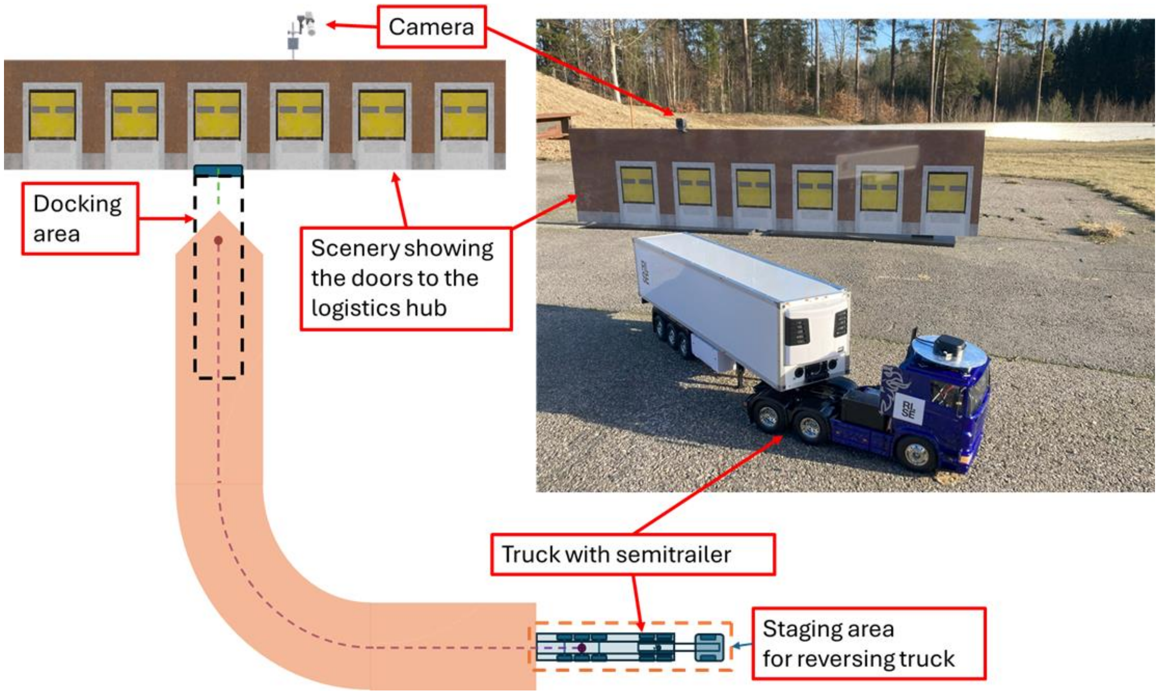

Figure 10: Preliminary set-up of for testing with the small-scale model truck used in UC4.1 testing.

AD function

The AD function to perform automated reverse docking of the truck is implemented using a prototype ecosystem powered by WayWise [16] and WayWiseR. WayWise is a rapid prototyping library developed in-house for connected, autonomous vehicles. It is designed to explore the use of autonomous vehicles, such as cars, tractors, and drones, with a focus on functional safety and cybersecurity. WayWiseR enhances the utility of WayWise by providing an abstraction layer between it and ROS2, along with necessary launch files and configuration settings. This integration facilitates the utilisation of ROS2-based open-source stacks for navigation and perception, thereby augmenting the validation process. Moreover, the abstraction layer implemented by WayWiseR promotes a high degree of code reusability between real vehicle implementations and simulations. This approach significantly enhances the validation process by allowing for the comparison of simulation results with real-world experimental data.

A pure pursuit based backing algorithm that can follow predefined points on a path is developed in WayWise to control the truck and trailer for this specific use case. This uses a Lyapunov style controller that ensures the angle between the truck and its trailer is maintained such that they are in a straight line while driving between the points. The controller takes as input GNSS data, the trailer angle and provides control outputs for the steering angle and speed for the truck to drive. The setup and the AD function used are identical in both the virtual and the scaled model testing, thus enabling the test results to be compared for validation purposes.

Sensors

Backing algorithms assume no-slip conditions and are often implemented through Proportional or Proportional Integral controllers. In recent years, growing interest has also been shown in multi-trailer vehicles, where variations of the pure pursuit algorithm are commonly applied for path tracking. For UC4.1, a similar approach employing a Lyapunov controller [17] [18] has been identified as suitable for implementing the backing function. Such a controller would require several sensors to ensure that the implementation in the simulator closely mimics its deployment in a physical truck. The sensor setup is designed to be similar in both cases. The controller relies on inputs, including the actuating angle of the steering and the angle between the trailer and the truck. Odometrical data is utilised for relative movement, while GNSS data is employed for absolute positioning and to verify speed plausibility, see Figure 11.

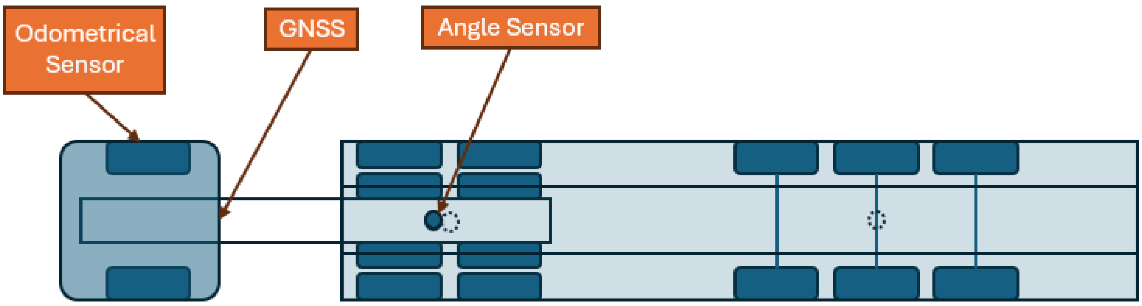

Figure 11: Schematic sensor setup for backing function under test.

Virtual sensors in simulations are often idealised, providing noise-free and perfectly accurate data that does not reflect physical sensors' imperfections. Virtual sensors should include noise, delays, and real-world data formats to ensure realistic, more readily comparable results. The level of modelling effort should align with the purpose of the simulation or testing, ensuring sufficient fidelity without unnecessary complexity.

Vehicle Dynamics

The scaled truck modelled is a modified Tamiya Scania R620 model truck in scale 1:14 with a computational unit, sensors and actuators to make it drive. One brushless motor for longitudinal control controls the actuators and a servo motor for lateral control. The truck is steered with Ackerman Steering using the servo system, and the brushless motor controls the speed using the VESC® motor controller [19], providing near to accurate speed control. The trailer is mounted on a rotary fifth wheel. The rubber tyres are chosen to provide stable rolling motion typical of such a scale model.

Validation

The primary objective of physical testing in UC4.1 using the scaled truck is to validate the accuracy and reliability of simulation-based testing. This includes verifying whether the precision of the truck's backing function, as observed in simulations, accurately translates to real-world. The validation also ensures that the truck follows a given trajectory in the physical environment according to the path-planning algorithm, thereby confirming the consistency between simulated and real-world results.

Furthermore, the physical testing serves to validate key virtual test environment parameters such as friction, fog, and sunlight, ensuring their representation aligns with real-world conditions. Another critical aspect of the evaluation is assessing whether all sensors and actuators function cohesively to execute the backing manoeuvre safely and effectively, ensuring precise docking at the docking stations.

Once the results from the physical testing are compared with those from the virtual environment, the validated findings can be extended to a full-scale truck, enabling a seamless transition from simulation to real-world implementation. This step is essential for bridging the gap between simulated and physical testing, reinforcing the credibility of simulation-based evaluations for practical deployment.

4 VEHICLE-IN-THE-LOOP TESTING FOR CCAM VALIDATION

4.1 Vehicle-In-the-Loop on Proving Grounds

Vehicle-In-the-Loop (VIL) testing on proving grounds is very similar to Blackbox Testing on Proving Grounds (see section 3.1). The setup of the testing is nearly the same. The main difference between both testing concepts is that in VIL testing at least one element of the environment and sensors components is virtual and not real, see Table 1. The Vehicle Dynamics Component is fully real, otherwise it is VIL testing on Testbenches (see section 4.2). Also, in general the AD function is real since the real vehicle is used containing the ECU on which the AD function is running. However, there might be VIL tests using a real vehicle where the AD function only exists in virtual world as a model which is running on external hardware, means hardware not directly belonging to the vehicle itself but connected to it. Such constellations make sense where the AD function is only planned for the concept phase and not planned to go in series production (e.g., SUNRISE UCs). The type of AD function in these setups could be called "virtual AD function". However, the combination of external hardware and AD function must be "real enough" to control the real VUT subjected to physical vehicle dynamics. Therefore, as indicated in Table 1, the term "real" AD function will be used also in such cases.

In general, the VUT and AD function are real, which means, that for the system to be called VIL, at least one element of the “Environment” or “Sensors” components is virtual. If all sensors are real (real “Sensors” component), then the Environment component contains some virtual elements such as precipitation, surrounding traffic, infrastructure or communication. Real sensors and a virtual environment implicate that at least one stimulator is required for the VUT to perceive the virtual environment. But there might be rare Vehicle-In-the-Loop tests on Proving Grounds where this virtual environment information goes only to a Human-Machine-Interface (HMI) where no stimulator is used, e.g. some take-over test cases.

Usually, at least one element of the “Sensors” component is also virtual as counterpart to the virtual “Environment” element if the virtual environment information must be provided to the AD function. First, only a virtual sensor model can perceive virtual environment information, otherwise a stimulator must be used for real sensors to perceive such virtual information. Second, it is in general unfeasible to use only real sensors on Proving Grounds, some of them equipped with stimulators without any virtual sensor, because of the stimulators' large dimensions (e.g., LiDAR or radar stimulators, or even big screens for cameras; see Figure 17 or Figure 19). An exception to this can be the stimulation of cameras (including 3D cameras), which can be stimulated by small high-resolution screens, as known from virtual reality (VR) or augmented reality (AR) glasses. However, usually different types of sensors are used with similar Field of View for redundancy reasons combined with sensor fusion where the use of multiple stimulators is not feasible on Proving Grounds or the usage of less stimulators than needed would generate bias.0

On the other hand, at least one virtual “Sensors" element implicates at least one virtual “Environment" element because a virtual sensor model cannot perceive data from real environment.

4.1.1 Physical and Virtual Components on Subsystem Level

Environment

At least one element of the Environment is virtual, such as precipitation, surrounding traffic, infrastructure, or communication.

Sensors

Usually at least partly virtual to provide virtual Environment elements to AD function, but there might be rare tests just using sensor stimulators with real sensors or provide the virtual Environment information only to the driver using an HMI.

AD function

In general, a real AD function as part of a real prototype or series production vehicle running on one of its ECUs. There might be rare cases, where a virtual AD function is running on external hardware (e.g., at concept phase). In this case, the combination must be able to control a real VUT subjected to physical vehicle dynamics.

Vehicle Dynamics

A real prototype or series production vehicle.

4.1.2 Advantages in Comparison to Virtual Testing

The main advantage of VIL testing on Proving Grounds (PGs) compared to (pure) virtual testing such as Model-In-the-Loop (MIL) or SIL is the strong focus on a few models (e.g., one sensor and its counterpart from environment) and having all other components real. Having the focus on just one model for CCAM validation is usually not feasible, because this one model would be in case of VIL testing a sensor model and at least the counterpart of the environment would have to be modelled, too. But, if this one sensor model would be a camera and a sensor stimulator (small high-resolution screen such as in VR glasses) playing a video with real-world data, then also one model would be realisable. Nowadays, Tesla vehicles for example work just with cameras and don’t make use of additional sensors such as radars or LiDARs and therefore no sensor fusion where the field of view of the cameras is not overlapping [20].

Other advantages are the same as for black box testing on proving grounds which is the strong reduction in usage of models (such as vehicle dynamics model or most elements of the environment) which always represent a kind of simplification of the real world (virtual testing might lack fidelity or realism). An additional advantage is the combination of purposes for the testing that require physical tests. For example, the same test could be used for the safety assurance of the SUT, but also to test the performance of the system (real-time behaviour) or any kind of driving study (e.g., Human-Machine Interface, take-over scenarios). But keep in mind that the last one might not require physical testing and could maybe be realised by Driver-In-the-Loop (DIL) testing (see chapter 5).

4.1.3 Limitations in Comparison to Virtual Testing

The limitations of VIL testing compared to (pure) virtual testing such as MIL or SIL don't differ to the limitations of black box testing on proving grounds. The disadvantages are therefore the low test throughput and the high costs (including test setup, equipment, equipment maintenance, staff). The strongest limitation is the limited parameter or scenario space. There are only a few proving grounds at a few locations which are limiting the environment such as weather conditions, the road network (e.g., kind of junctions) and road layout (number of lanes, lane widths, lane markings) or the infrastructure (e.g., buildings and trees). The number and variety of agents are also strongly limited due to available resources for equipment (e.g., test vehicles or robots/targets) and staff controlling these. But also, the communication is limited due to the limited number of agents or communication network (e.g. only 5G but no WIFI).

4.1.4 Additional Validation Needs

Since VIL testing includes at least one virtual model it must be ensured that the test is still representable for the real case. Therefore, all of the used virtual models must be validated beforehand and test cases at the ODD boundary of the models might need to be repeated via black box testing on proving grounds or on real roads.

4.1.5 Exemplary Implementation and Setup

Applus+ IDIADA has developed an innovative approach to testing and validating Advanced Driver Assistance Systems (ADAS) and Automated Driving functions through the integration of virtual and physical testing methodologies. This approach bridges the gap between simulation and physical tests by **synchronizing a virtual environment with a real vehicle** on a simplified track.

At the core of this methodology is the use of **virtual traffic agents and sensor models**. These components are part of the environment and sensors components, respectively, and allow for **comprehensive testing scenarios on proving grounds**. By incorporating these virtual elements, IDIADA can test safety-critical scenarios that would be dangerous or impractical to recreate in purely physical environments.

The **fidelity** of the virtual traffic agents and sensor models **can be adjusted** based on the specific requirements of each test case and the capabilities of the simulation solver. These can range from simple 3D bounding boxes with object-list based sensor models (either idealized or stochastic) to more complex real shapes with physics-based sensor models running on GPUs.

Digital maps of the proving ground play a crucial role in this process, allowing for the **validation** of virtual traffic agent trajectories and ensuring **accurate synchronization** between the physical and virtual worlds. The geo-localization of the vehicle, combined with these digital maps, creates an **immersive scenario** for the driver while enabling the **precise injection** of virtual sensor information into the vehicle's Electronic Control Unit (ECU).

A diagram of the applied methodology can be found in the Figure 12 below.

This VIL methodology offers several key advantages:

1. **Broader testing coverage:** It allows for a wider range of complex traffic scenarios to be tested, including those that are too dangerous or complex for purely physical testing.
2. **Cost-effectiveness:** By reducing the requirements for specialized proving grounds and complex physical targets, VIL offers a more economical solution for ADAS and Automated Driving system development and validation.
3. **Increased safety:** Complex or dangerous scenarios can be tested without risk to drivers or equipment.
4. **High accuracy and repeatability:** The virtual components ensure consistent and repeatable test conditions.
5. **Flexibility:** IDIADA's versatile and modular virtual platform can be tailored to meet specific client requirements.
6. **Future-proofing:** The system is designed to accommodate both existing and future standards for vehicle safety testing.

IDIADA's VIL services are particularly valuable for **pre-calibration** and **evaluation** of ADAS and Automated Driving systems. Moreover, this methodology is being assessed by **Euro NCAP** for potential inclusion in their vehicle safety rating process, which could significantly expand the test matrix coverage for safety evaluations.

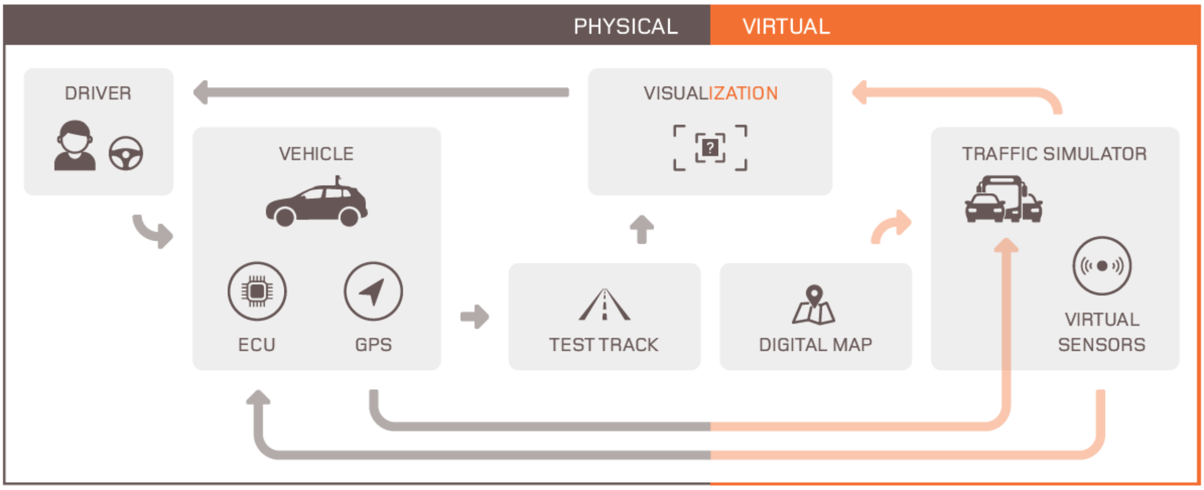

Figure 12: IDIADA VIL approach

4.1.6 Exemplary Implementation and Setup – UC 1.3

As part of the SUNRISE project, VED and ICCS have designed and developed two alternate Vehicle-In-the-Loop testing systems to validate the proposed Safety Assurance Framework for hybrid testing of UC 1.3 (Collective Perception) scenarios. Both systems support testing of ETSI-compliant Cooperative Awareness Message (CAM) or Cooperative Perception Message (CPM) V2X message exchanges, assuming collective perception is performed off-board on a smart RSU. The primary technical difference between the two VIL systems lies in the virtual network-traffic co-simulation sub-modules and the interfaces connecting the real agent and the virtual world running on a server simulation. The first system employs an OMNeT++/Veins/SUMO co-simulation environment and communicates via MQTT, supporting realistic traffic scenarios. In contrast, the second system utilizes a CARLA-ns3 co-simulation environment and communicates through gRPC interfaces, enabling the simulation of sensor data. A significant challenge in these implementations is ensuring that the V2X messaging stack used in the real world operates similarly in the network co-simulation virtual counterpart. Below, we present the VED VIL setup as an exemplary implementation.

VED VIL System Implementation

The system combines a real Connected Automated Vehicle (CAV) on a test track with a virtual simulation environment, enabling seamless interaction between real and virtual entities. VIL testing offers higher reliability compared to HIL testing. Below is a detailed description of the VIL testing system.

Real Environment Setup

The real-world testing will be conducted at the Satory test track in Versailles, France, a facility equipped to emulate urban driving conditions. The test track includes a fully developed road network featuring road markings, pedestrian crossings, intersections, and traffic controls. A Roadside Unit (RSU) is installed to enable Vehicle-to-Infrastructure (V2I) communications. Additionally, the testing involves a real subject vehicle shown in Figure 13, an automated vehicle equipped with an Onboard Unit (OBU) that facilitates V2X communication systems. The real ego vehicle has a screen that displays the received CPM information on it, shown in Figure 14.

Figure 13: Real ego of CAV of VED

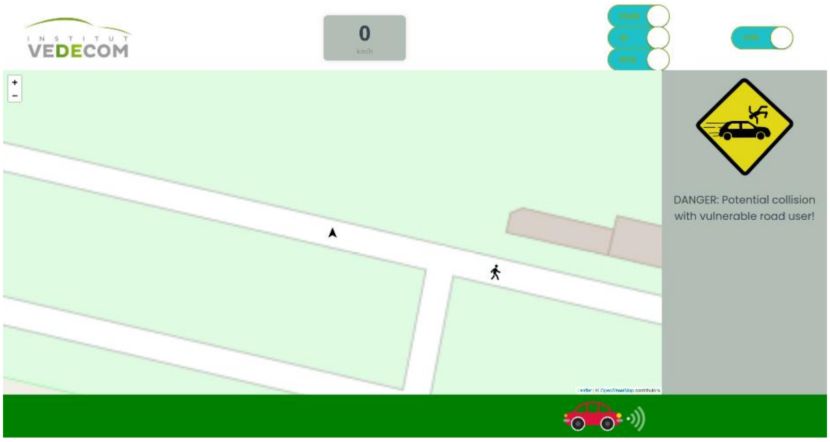

Figure 14: Visualisation screen inside the ego CAV showing received CPM information on the map.

Virtual Environment Setup

The virtual environment is constructed on a simulation platform integrating OMNeT++, Veins, and SUMO. OMNeT++ serves as an open-source C++ platform that manages simulation control, data exchange, and synchronization between real and virtual components. Veins operates as a V2X simulation framework, tailored to meet UC 1.3 requirements. It implements ETSI-standard Vehicle-to-Vehicle V2V and V2I communication protocols, including Cooperative Awareness Messages for situational awareness and Cooperative Perception Messages for sharing local perception data. A two-dimensional (2D) perception model within Veins simulates camera-based object detection in a vehicle's local field of view, accounting for occlusions caused by other objects and incorporating this information into CPMs.

SUMO, a traffic simulator for urban mobility scenarios, defines UC 1.3 scenarios by modelling road networks, traffic flows, and pedestrian behaviours. Operating in server mode, SUMO dynamically interacts with Veins to ensure seamless integration. Figure 15 below displays the SUMO Graphical User Interface (GUI), showcasing the scenario's road network, complete with road markings, pedestrian crossings, and vehicles. In the GUI, the virtual ego vehicle is depicted in white, virtual non-ego vehicles in yellow, and pedestrians in red. Additionally, it highlights the camera perception cones of the vehicles and illustrates how a yellow block occludes the ego vehicle's view of a pedestrian.

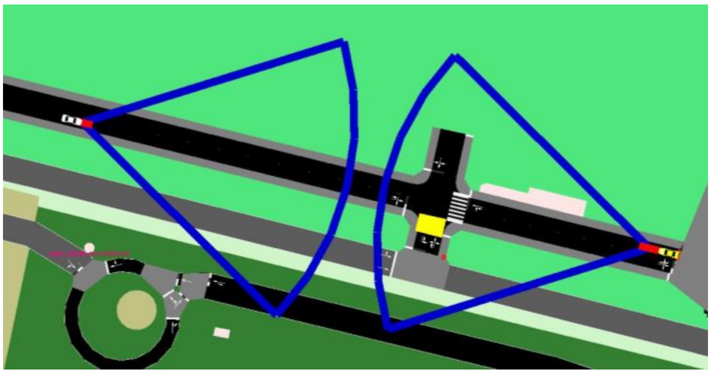

Figure 15: SUMO GUI of the XIL simulator.

Hybrid Communication Framework

The hybrid communication system enables interaction between the real ego-vehicle and virtual entities. The real ego-vehicle, equipped with VED's V2X stack software, transmits real CAMs to virtual vehicles in the simulator and receives virtual CPMs from the simulator to inform its control systems. Virtual vehicles in the simulation utilize the same ETSI-standard V2X communication protocols implemented in Veins, ensuring consistency in messaging.

Sensors and Perception

The system incorporates both virtual and real-world perception. The 2D perception model in Veins simulates camera-based detection of objects in the local environment, including occlusion handling. The real ego-vehicle leverages its V2X connectivity to obtain perception objects, integrating this data into its control systems.

Automated Driving (AD) Functionality

The ego-vehicle operates as a fully functional automated vehicle. It employs real-world AD systems for decision-making and control, V2X communication capabilities for interaction with real and virtual entities, and real-time reception of virtual CPMs that influence its behaviour during testing scenarios.

Vehicle Dynamics

The real ego-vehicle operates with authentic vehicle dynamics on the test track, providing realistic responses to the environment. In parallel, virtual vehicles simulate realistic dynamics using SUMO and Veins, ensuring coherent interaction between real and simulated components.

System Integration

The VIL testing system ensures seamless integration between real and virtual components by enabling the real ego vehicle to transmit live CAM data to the simulator. The simulator processes this data, simulates virtual interactions, and generates appropriate CPM responses. These CPMs are then transmitted back to the real ego vehicle, which uses the information to make informed decisions. For instance, in the UC 1.3A pedestrian darting-out scenario, the real ego vehicle reacts by slowing down or stopping to avoid a collision with the pedestrian. This dynamic interaction between the real and virtual environments significantly enhances the fidelity and reliability of testing for UC 1.3 scenarios.

Validation

VED will validate the execute block of the SAF for Use Cases UC 1.3A and UC 1.3B, focusing on hybrid testing of urban perception validation. In UC 1.3A, the perception functionality of the subject vehicle and other vehicles will be tested using a virtual simulator, where virtual vehicles implement a virtual camera sensor to detect objects in their perception cone. The perceived information is transmitted via CPM over an MQTT interface to a real vehicle, which receives and processes the data. Conversely, the real vehicle sends CAM via MQTT, allowing the simulator to update the subject vehicle's location and speed accordingly. The correct transmission and reception of pedestrian information between the real and virtual environments will validate system connectivity, and the real vehicle will use the received CPM data to execute automated driving (AD) functions such as slowing down or stopping. Simulations will be recorded, showcasing vehicles and pedestrians on a road network, with the received CPM data displayed on a map interface to confirm proper information exchange.

For UC 1.3B, the Vehicle-to-Infrastructure (V2I) communication will be validated. The virtual vehicle will send a CPM containing pedestrian information to a Roadside Unit (RSU), which will aggregate this data with its own local sensor information before transmitting it to the real subject vehicle. The vehicle will then process the received CPM from the RSU and apply the appropriate AD functions, such as braking or stopping. This process confirms the RSU’s role in perception fusion and decision-making. Additionally, other blocks of the SAF will be validated accordingly, either within the virtual simulator or on a real vehicle on the test track.

4.2 Vehicle-In-the-Loop on Testbenches

Vehicle-In-the-Loop (VIL) testing on Testbenches bridges the gap between physical and virtual validation by integrating real vehicles with virtual environments in a controlled testbench setup. This approach enables the testing of ADAS and AD functionalities with high precision, allowing engineers to evaluate how subsystems interact under a wide range of reproducible conditions. VIL testing enhances the development process by blending physical components, such as vehicles and sensors, with virtual elements, such as traffic scenarios and environmental conditions.

Placing a fully operational vehicle on a chassis dynamometer or powertrain testbed, VIL on Testbenches enables precise testing under controlled conditions while simulating diverse virtual driving scenarios. By incorporating advanced technologies such as highly dynamic wheel dynamometers and novel steering force emulators, along with validated vehicle models and detailed environment simulations, VIL setups allow for comprehensive testing across a wide range of scenarios, including vehicle dynamics, highway, and urban driving. Figure 16 shows an example component of a VIL Testbench setup including a sensor stimulator and wheel-steering force emulation etc.. The integration of Powertrain-, Steering-, Brake-, and Sensor-In-the-Loop facilitates highly realistic testing that mirrors real-world operations, even at the dynamic limits of vehicle performance [21].

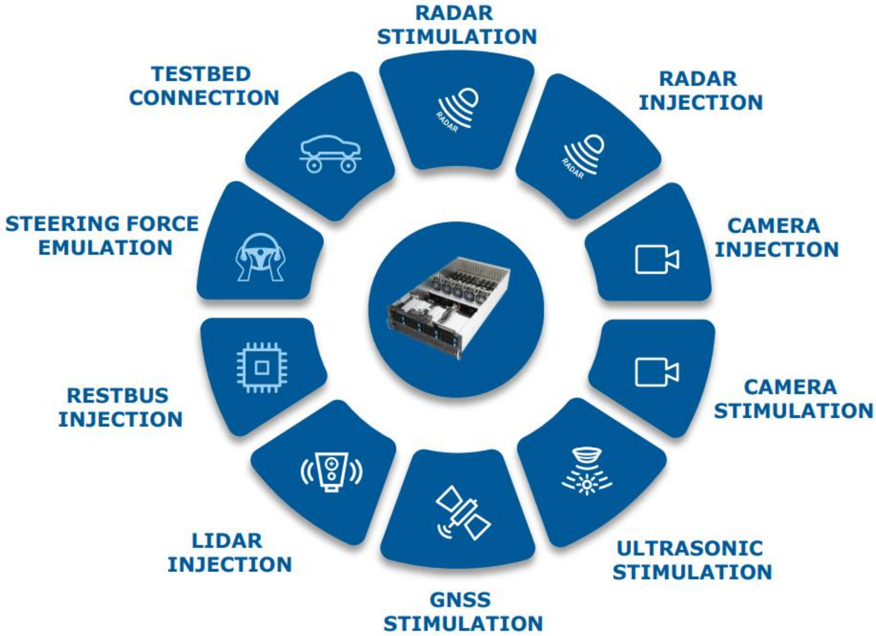

Figure 16: Components of a VIL Testbench Setup

In addition to bridging the gap between virtual and full-scale physical testing, VIL is particularly effective for validating CCAM systems in the development phase. By uncovering integration issues between hardware and software, it minimizes costly design iterations later in the development process. Furthermore, the controlled laboratory environment enhances safety by eliminating the risks associated with on-road testing, especially during the evaluation of critical or edge-case scenarios.

4.2.1 Physical and Virtual Components on Subsystem Level

Environment

In a VIL setup, the environment is virtually simulated while interacting with the physical vehicle in real-time. Elements such as road configurations, traffic scenarios, lighting conditions, and weather effects are generated using sophisticated simulation platforms. These simulations provide the flexibility to test various scenarios, from urban driving to highway navigation, under consistent and controlled conditions. However, the physical test environment must also account for accurate coupling with virtual inputs, ensuring seamless interaction between the real vehicle and the simulated world.

Sensors

VIL Testbench setups replicate real-world sensor interactions by feeding simulated data to the physical vehicle's sensors. Technologies like radar stimulation and camera feed generation ensure that sensors perceive the virtual environment as realistic.

GNSS Stimulator: Customizable and GNSS Real-time satellite model (GPS, Galileo, GLONASS, Beidou). GNSS Stimulator allows the simulation of all possible positions on earth, while the vehicle or truck is driving on a test bed.

Radar Stimulator: A radar sensor stimulator is a device that simulates radar reflections from objects in a virtual environment, enabling the testing and validation of radar-based ADAS/AD Functions. It takes input data such as object positions, velocities, sizes, and material properties, as well as environmental factors like weather conditions from a 3D Virtual Driving Environment. The output includes raw radar signals with realistic delays, Doppler shifts, and amplitudes.

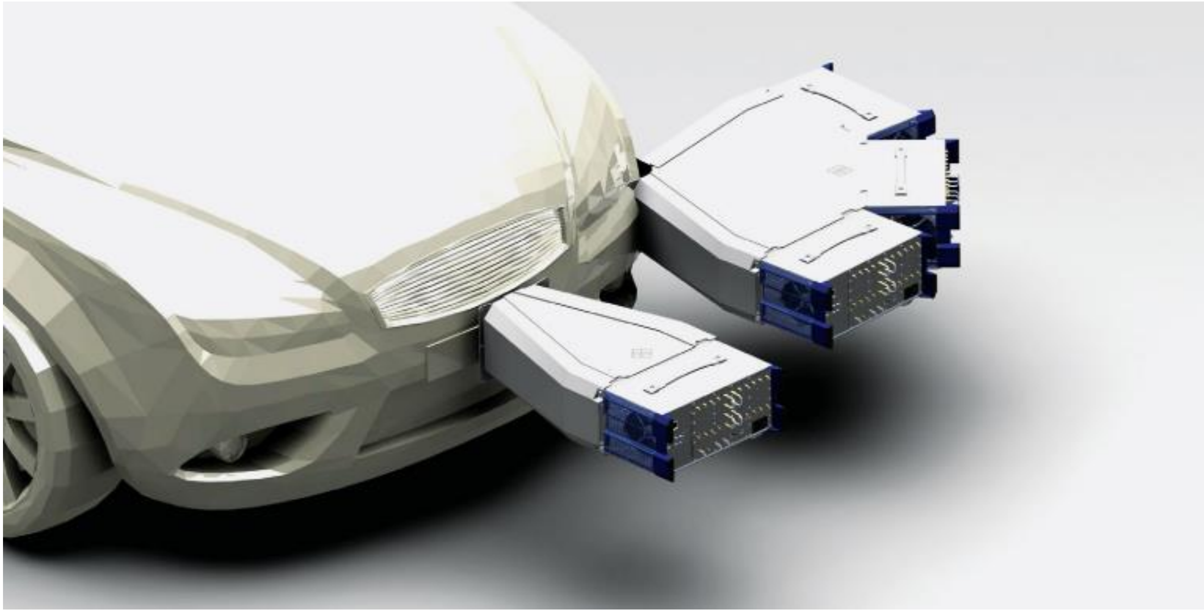

Figure 17: Radar Stimulator configuration with modular absorber system with variable airgap

Camera Stimulator: A camera stimulator setup integrates a physical vehicle on a powertrain testbed with a virtual environment to test vision-based functions like Lane Keeping Assist (LKA). The virtual environment, including lane markings and objects, is captured by a virtual camera, and the images are projected onto a screen or sent to the vehicle’s mono camera after calibration.

LiDAR Target Stimulator: A LiDAR Target Stimulator is a system designed to simulate target distance and reflectivity for LiDAR sensors, enabling controlled and repeatable testing of LiDAR-based perception systems. There is still development going on for finalizing a state-of-the-art device. LiDAR Target Simulator can be used where a LiDAR sensor is placed on the vehicle on the testbenches.

AD function

The AD function, typically bundled in an ECU, processes sensor data and executes decision-making algorithms to control the vehicle. In VIL Testbenches, the AD functions are tested using a combination of real sensor outputs and virtual environmental inputs.

Vehicle Dynamics

The physical vehicle on the testbench operates with its real-world dynamics, such as steering, acceleration, and braking. The testbench itself is able to represent realistic driving dynamic conditions with a Dynamic Steering Force Emulator and a chassis dynamometer which ensures that these dynamics interact naturally with the simulated environment. By testing with real vehicle components, VIL allows precise validation of control systems and their responses to dynamic scenarios.

4.2.2 Advantages in Comparison to Virtual Testing

Vehicle-in-the-Loop testing on Testbenches offer distinct advantages over purely virtual testing by incorporating real hardware components into the validation process. While virtual testing relies entirely on simulated models, VIL enables the direct evaluation of physical systems, including sensors, actuators, and vehicle dynamics, capturing real-world behaviours and constraints with higher fidelity. Although virtual testing is highly effective in terms of scalability and flexibility, it may lack the realism required to fully assess the interactions between hardware and software. VIL complements this by providing dynamic responses from physical components, such as drivetrain and steering mechanisms, in combination with virtual environments. This hybrid approach enables the identification of hardware-specific issues that might remain undetected in purely virtual tests, ensuring more thorough validation of ADAS/AD functionalities under realistic conditions.

4.2.3 Limitations in Comparison to Virtual Testing

Applications of VIL on Testbenches are subject to certain inherent limitations due to its reliance on physical systems. One significant challenge is scalability, as VIL setups are designed to test individual vehicles, requiring substantial infrastructure and effort for each configuration. Additionally, the cost of hardware, such as chassis dynamometers, sensor stimulators, and the associated testbench components, makes VIL more resource-intensive than other testing methods. Flexibility is another constraint, as physical setups require time and effort to adapt to new scenarios, unlike virtual environments, which can be modified with minimal adjustments. These limitations highlight the practical challenges associated with implementing VIL on Testbench systems.

4.2.4 Additional Validation Needs

In Vehicle-In-the-Loop (VIL) testing on testbenches, additional validation efforts are essential to ensure the accuracy and representativeness of the results. First, the vehicle models, including tire, steering, and suspension dynamics, must be validated against real-world behaviour to ensure the fidelity of the simulated responses. Similarly, the sensor stimulators (e.g., radar, camera, and LiDAR) require strict calibration and validation to align virtual inputs with real-world sensor outputs, accounting for effects such as noise, latency, and physical distortions.

The virtual environment must also be validated for realism, ensuring accurate representations of road geometry, lane markings, traffic scenarios, and environmental conditions (e.g., lighting and weather). This validation is critical for assessing perception and control systems reliably.

Finally, the integration of physical and virtual components, including the Real-time synchronization of steering forces, torque feedback, and sensor data streams, requires thorough validation to avoid inconsistency between real and simulated interactions.

4.2.5 Exemplary Implementation and Setup

The Vehicle-In-the-Loop (VIL) on Testbench solution, implemented using the **AVL DRIVINGCUBE™**, was designed to evaluate its potential for ADAS validation and its correlation with real-world testing results. This project focused on the Adaptive Cruise Assist (ACA) function of an Audi A8, aiming to test the integrated driving function with vehicle dynamics while minimizing interference with internal interfaces. The system architecture of the A catalogue of ten highway driving scenarios, including NCAP AEB tests, was used as the evaluation basis.

Figure 18 illustrates the system architecture of the VIL Testbench setup, highlighting the interaction between physical and virtual components. It showcases the integration of the powertrain testbed, dynamic steering force emulator, and sensor stimulation systems with the virtual environment.

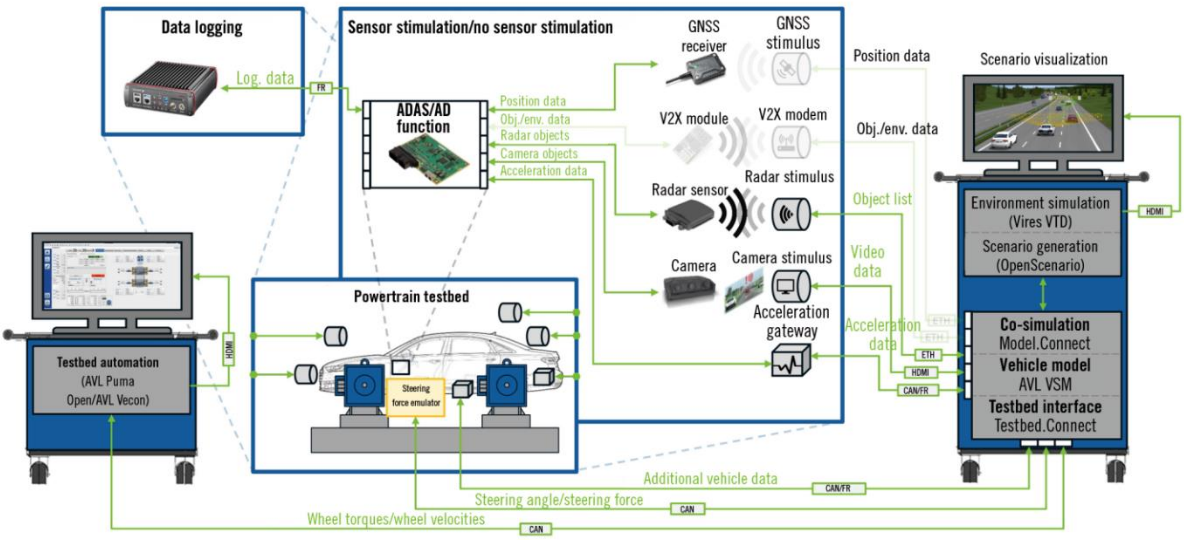

Figure 18: System Architecture for VIL on Testbench for Adaptive Cruise Assist

The VIL setup employed various advanced systems and components, as shown in Figure 19 to ensure accurate simulation and validation of the ACA function.

Powertrain Testbed

Equipped with highly dynamic dynamometers to simulate longitudinal forces, including realistic tire slip and dynamic limits.

Coupled with AVL VSM (vehicle dynamics simulation) to accurately reproduce driving resistances and reaction forces in both longitudinal and lateral dynamics.

Dynamic Steering Force Emulator

Applies realistic steering forces directly to the steering rack, up to 5 kN, with steering speeds of 250 mm/s, ensuring the vehicle experiences the same reaction forces as in real-world tests, critical for validating lateral dynamics and steering interventions.

Braking System

Operated in two modes:

- Real braking, replicating road test behaviour.
  Decoupled braking, where brake pressure is measured and mapped to the simulation to avoid wear and cleanup of brake dust.

Radar Sensor Stimulation

The RTS100 radar target stimulator by Rohde & Schwarz created artificial radar objects with configurable parameters such as distance, velocity, azimuth angle, and size.

Camera Sensor Stimulation

- A calibrated projector-screen system displayed virtual images of lane markings and objects, ensuring proper alignment with the mono camera's field of view.

Virtual Environment Simulation

- Generated using the Vires Virtual Test Drive (VTD) platform, replicating traffic scenarios, lane geometries, and dynamic objects.
- Provided a controlled, closed-loop environment for testing the ACA function.

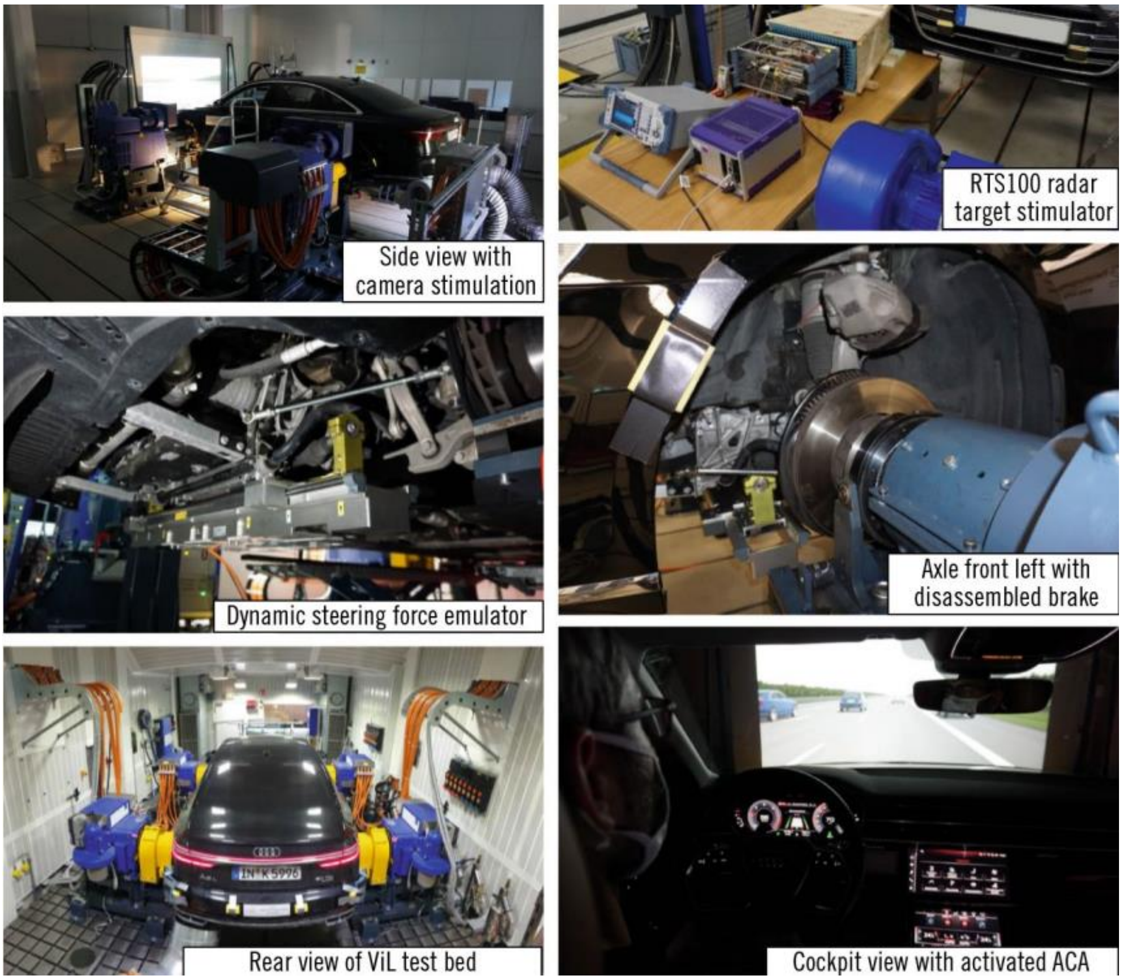

Figure 19: Real setup of the VIL Testbench

RESULTS AND CORRELATION

The basic correlation between a powertrain testbed and the real world, that is, proving ground and road, has already been demonstrated in several papers [22] [23]. Therefore, the focus of this project was on the correlation of sensor stimulation with real sensor behaviour. As described above, ten different scenarios were executed, each with multiple repetitions. All tests were performed on the proving ground as well as on the AVL DRIVINGCUBE. Additionally, to over-the-air stimulation of the sensors, an over-the-cable injection was applied for comparative purposes. The scenarios were executed completely closed-loop. Meaning it is not a simple replay of recorded data. The ACA function was active and fully functional at the testbed. Based on the sensor information, it actuated the vehicle in the longitudinal direction (acceleration and deceleration) as well as in the lateral direction (steering).

The validation results showed strong agreement between the VIL setup and real-world tests. Key observations included:

- Object Detection: Radar and camera sensor outputs closely matched their real-world counterparts, with deviations below 1 m for target distance (Figure 19).

- Lane Following: Lane detection by the mono camera exhibited less than 15 cm deviation from the actual lane position, indicating precise sensor stimulation and system performance (Figure 20).
- Consistent System Behaviour: The ACA function demonstrated reliable control of acceleration, braking, and steering in a closed-loop environment, successfully handling all ten highway scenarios, including cut-ins and emergency braking.

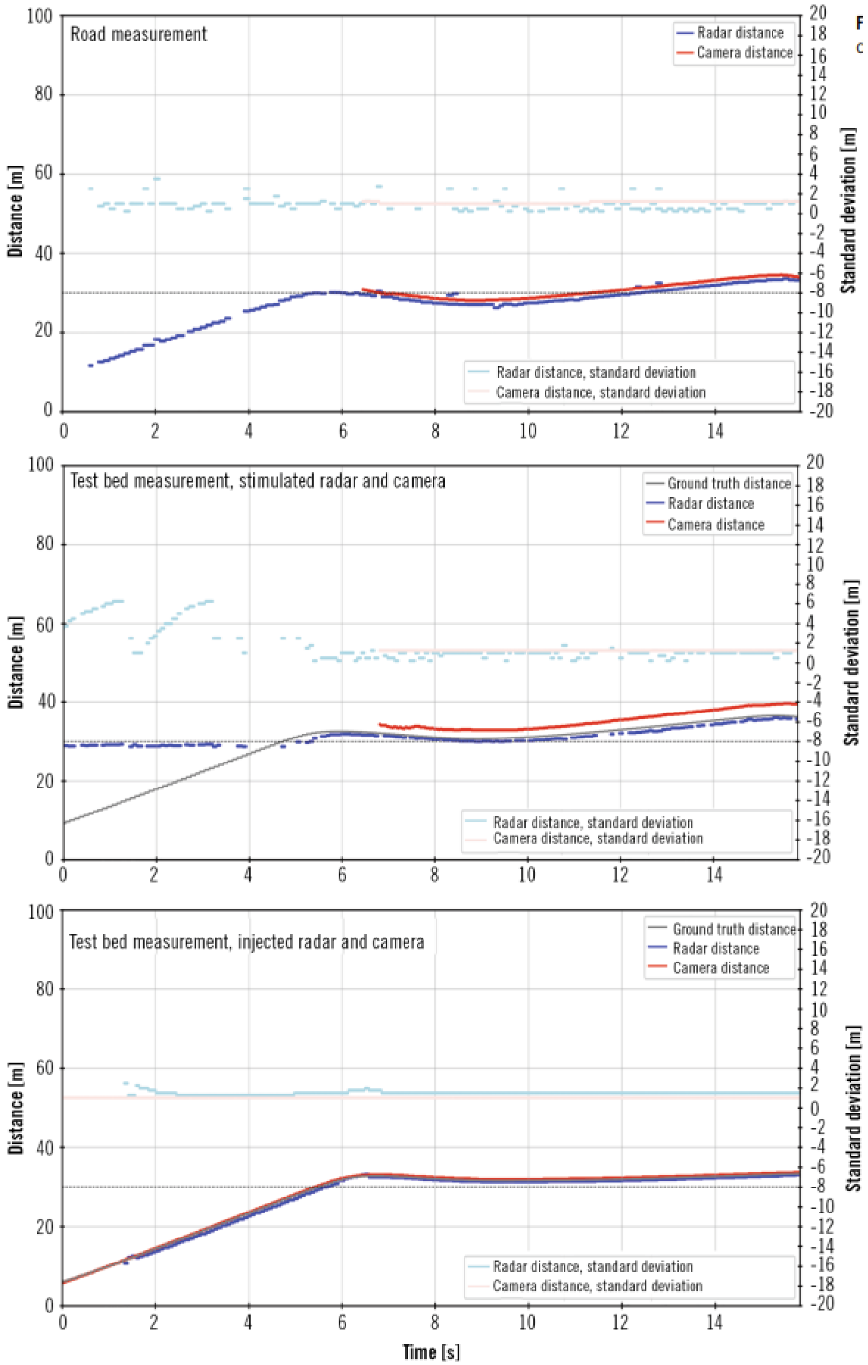

Figure 20: Comparison of the ground truth distance and the measured distance

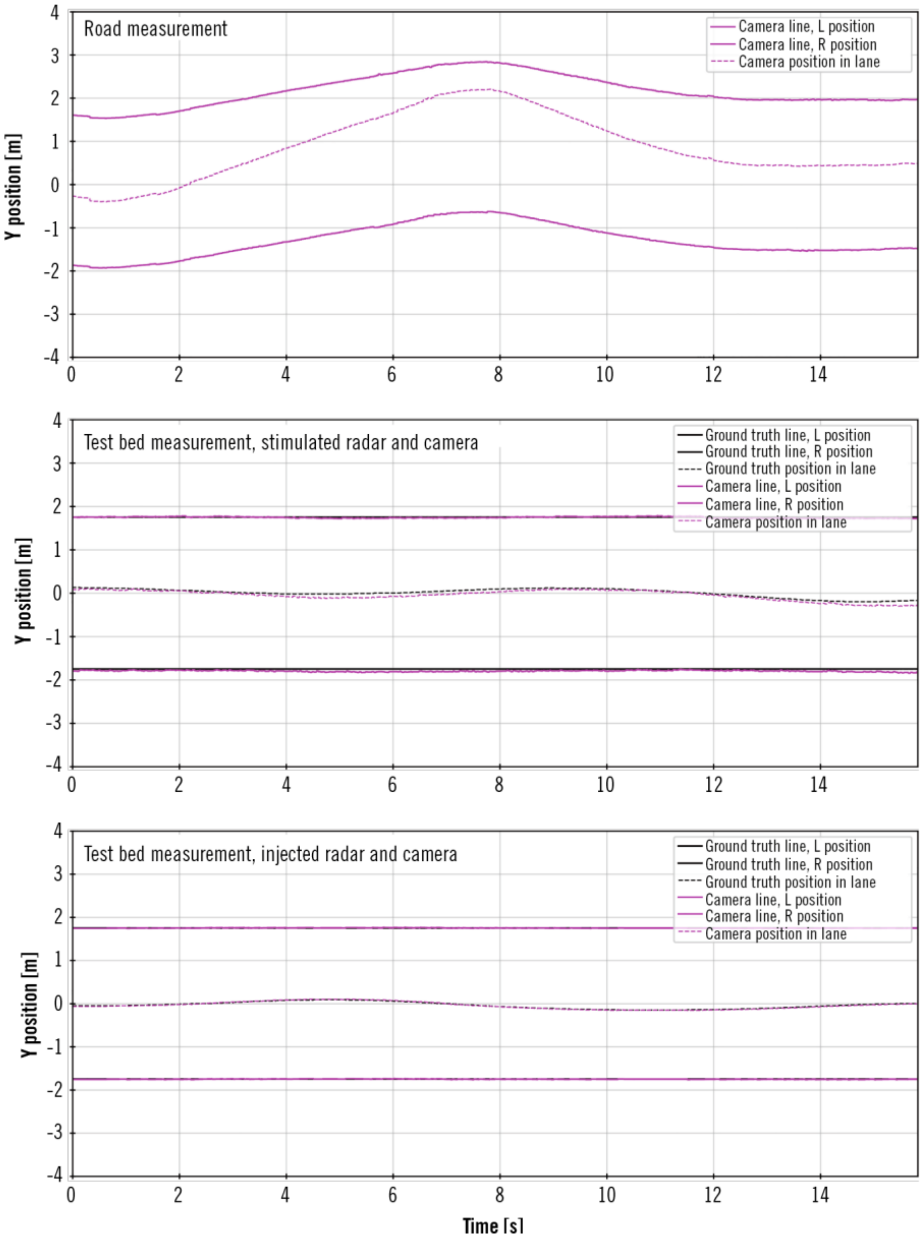

Figure 21: Comparison of the ground truth lane and line position and the measured position

5 DRIVER IN THE LOOP TESTING FOR CCAM VALIDATION

5.1 Introduction to Driver-in-the-Loop Testing

Driver-in-the-Loop simulations are an essential component in the development and validation of AD and ADAS systems. Although they are not explicitly part of the Safety Assurance Framework (SAF), driving simulations continue to play a significant role in the CCAM safety assurance process by:

- Gathering data to develop human driver models as a reference for automated driving functions
- Assessing the impact of mixed traffic on human driver behaviour and performance
- Testing driver interaction with systems of SAE Level 1-3, including the analysis of take-over situations

Within the context of the SUNRISE Project, Driver-in-the-Loop testing is a tool to close research and validation gaps for SAE-Level 1-3 systems featuring driver interaction.

The following two chapters present two related but fundamentally different approaches to integrating the driver into the simulation loop. Section 5.2 shows how the driver can be involved in real-world test drives on a test track. Section 5.3 driver testing in virtual environments, a method commonly used in driving simulators.

5.2 Driver-In-The-Loop on Proving Grounds

Test track-based driver-in-the-loop testing remains largely an academic and research-driven topic. On the market there are no complete, ready-to-run commercial solutions available for driver in the loop testing on test tracks. Several initiatives have shown the feasibility of driving simulators based on test tracks, using virtual and augmented reality (VR/AR) technologies to enrich the driving experience. These systems combine real-world driving with virtual elements, creating immersive simulation environments for the driver.

In his 2008 Ph.D. thesis, Thomas Bock [24] presented an augmented reality (AR) headset integrated into a real vehicle. This setup projected virtual objects into the real world, allowing the driver to experience simulated scenarios while driving on a test track. Vehicle localization was achieved using a high-precision GPS system, while the driver's head position was tracked using a laser-based method.

Building on Bock's work, Berg's Ph.D. thesis at the University of the Bundeswehr Munich [25] in 2014 extended the concept by introducing a head-mounted display (HMD)-based simulation environment. This system synchronized virtual scenarios with real-world driving, allowing the driver to experience realistic accelerations while interacting with virtual elements. The simulation used Vires' Virtual Test Drive software, which tracked the position of the VR headset using a combination of optical methods and inertial sensors. GPS data was used to determine the vehicle's position in the real world, which was then translated into the virtual environment.

In 2018, David Goedicke et al. [26] presented a different approach in their paper VR-OOM: Virtual Reality On-road driving siMulation. They used an Oculus VR headset, external IMUs, and a LEAP motion controller for hand tracking. Unlike previous systems, there was no precise external tracking of the VR headset; instead, vehicle motion was calculated using IMU data and the vehicle's diagnostic interface. A study with six participants showed that VR headsets in real vehicles have a promising potential for driving simulation, although further studies and technical advancements are needed.

In his Ph.D. thesis in 2024, Nicolas Wagener [27] developed an urban driving simulator that combines real-world vehicle dynamics with visual simulation of the environment and other traffic participants. The work focuses on the challenges of simulating urban driving, in particular complex manoeuvres such as turning, and the effects of real-world dynamics on immersion and motion sickness. Both VR and 2D screen-based visualizations are discussed. Wagener shows that using real-world dynamics in combination with VR improves immersion and reduces simulator sickness compared to static simulations.

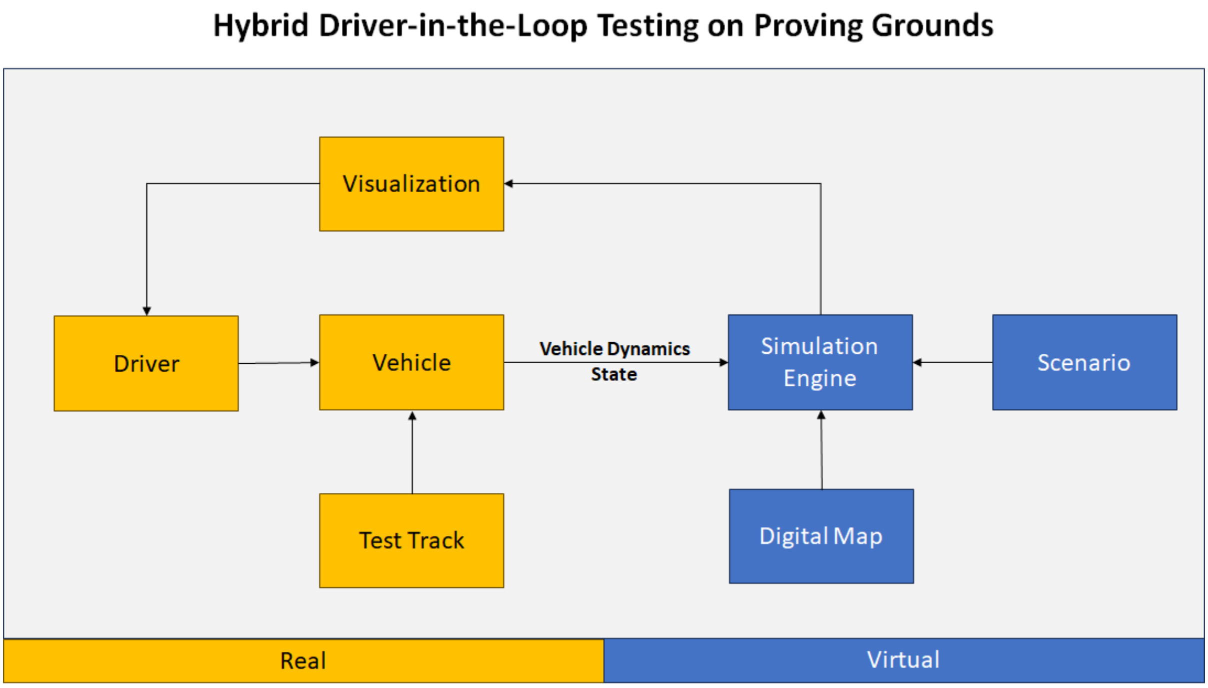

Figure 22: Driver in the Loop on Proving Grounds

These studies demonstrate the potential of combining VR/AR technology with real-world driving to create advanced simulation experiences and provide valuable insights for future developments in driving simulation systems.

5.2.1 Physical and Virtual Components on Subsystem Level

Environment

Physical: The drivable surface of the test track is a physical part of the environment that influences the vehicle being tested and the driver. The vehicle and driving behaviour are influenced by the ground properties, which interact with the tires and by vibration excitation. The driver also perceives the test track surface haptically and acoustically. The size of the drivable surface also plays a role in the test design.

Digital: When using virtual reality environments, then the digital elevation model needs to match the real-world terrain, such that the vehicle feedback resulting from reality matches the virtual environment. In virtual reality application it is also recommended that the ground texture closely resembles reality, so that the visualization is a plausible match for tire forces and acoustic feedback. Static objects, such as buildings and props, can be placed virtually to either enhance realism or to serve specific purposes within the scenario, such as occluding or distracting the driver. Traffic rules and signs are also introduced into the digital environment, allowing precise control of the scenario situation. Additionally, the digital implementation of other road users, such as cars, bicycles and pedestrians, enables flexible and dynamic test scenario design.

Sensors

Virtual: While the vehicle under test requires an inertial navigation system to synchronize the virtual and physical vehicle, sensor data for ADAS or AD functions is typically simulated virtually. This approach allows the vehicle to react to virtual traffic participants.

AD function

Virtual: Prototypical ADAS and AD functions can be executed entirely in the virtual environment and are typically not implemented on the final hardware.

Vehicle Dynamics

Physical: All parts of the vehicle exist in the physical real world. This means the vehicle-road interaction occurs in reality and the driver can interact with real interfaces to the car, such as the steering wheel, gas pedal, and brake pedal. The resulting kinematic state of the vehicle is synchronized with the virtual environment.

5.2.2 Advantages in Comparison to Virtual Testing

When researching partially automated driving and handover situations for automated systems, the driver is the element, that is the hardest to model virtually. Therefore, this approach involves a real driver, that interacts with a system. This approach retains the key benefits of traditional track testing, such as real physics, authentic steering feedback, and natural noise, vibration, and harshness (NVH) feedback. In addition, the virtual components allow for flexible adaptation and customization of test scenarios. Furthermore, this method allows for safe testing of critical situations, such as near misses, that would be difficult to replicate in a purely physical study.

5.2.3 Limitations in Comparison to Virtual Testing

Despite its advantages, this testing method has several practical limitations. Safety considerations often restrict the range of vehicle speeds, and strict adherence to safety margins relative to the test track boundaries is essential. Moreover, the harsh vehicle environment demands significant calibration efforts to ensure precise car and VR tracking. Another notable drawback is that testing with a real driver and a real vehicle inherently limits the process to real-time execution models, which may constrain the scope of the tests. Addressing these challenges requires a careful balance between realism and safety, alongside meticulous calibration procedures.

5.2.4 Additional Validation Needs

While vehicle behaviour does not need to be validated in this test environment, it is essential to validate the accuracy and precision of the tracking system to ensure valid results. If the focus is on the driver, the fidelity of the static environment and the behaviour of other road users must also be sufficiently realistic to produce valid study results. Tracking requirements and simulation fidelity requirements depend on the specific use case and study objectives [28].

5.2.5 Exemplary Implementation and Setup

In this section, we present an exemplary implementation of how ika conducts human subject studies using a Virtual Reality (VR) headset in a vehicle on a test track. This innovative setup, based on a PhD thesis by Nicolas Wagener [27], integrates virtual reality with real-world vehicle dynamics, allowing for highly immersive simulation.

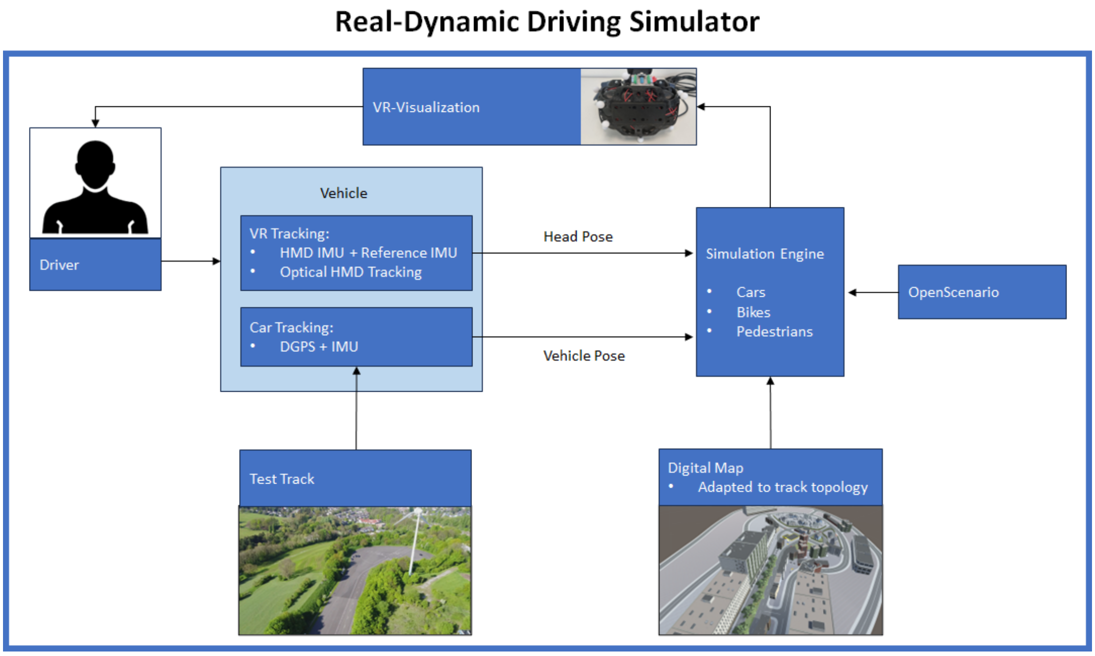

Figure 23: Structure of the Real-Dynamic Driving Simulator

VR-based approach on the test track

The approach is to integrate the vehicle into the simulation while immersing the driver through VR visualization. To achieve this, two main challenges must be overcome:

- **Tracking the VR headset in the vehicle:** Conventional tracking methods often fail in vehicles due to interference by the vehicle motion. To overcome this, an optical tracking system based on multiple stereo camera pairs is used. This optical tracking determines the precise pose of the VR headset and fuses it with data from the headset's Inertial Measurement Unit (IMU) and the vehicle's IMU. This fusion combines the drift-free characteristics of optical tracking with the high update rate and dynamic accuracy of IMUs to provide reliable and accurate tracking.
- **Tracking the vehicle state:** Vehicle motion is tracked using an Inertial Navigation System (INS) that combines high-precision global navigation satellite system (GNSS) signals with inertial data from an IMU fixed in the vehicle. This system allows accurate tracking of the vehicle's speed, rate of rotation, and position. The simulation dynamically places the vehicle according to these measurements.

Hardware Setup

The hardware used in the vehicle includes

- a VR headset,
- a VR reference IMU,
- an infrared optical tracking system,
- a simulation computer,
- a high precision vehicle INS.

In addition, an independent power supply system has been integrated to ensure the uninterrupted operation of all components during the drive.

Safety Concept

To ensure safety during the studies, where the driver is fully immersed in the VR simulation, a robust safety concept has been implemented. The test vehicle is equipped with a secondary brake pedal on the passenger side. A safety driver on the passenger side can intervene at any time to bring the vehicle to a safe stop. This configuration ensures that the test vehicle remains safely on the test track in the event of a failure.

This setup demonstrates a seamless blend of cutting-edge VR technology with real-world vehicle integration, enabling ika to conduct safe, immersive and highly accurate studies on the test track.

5.3 Driver-In-The-Loop in Simulators

A driver-in-the-loop (DIL) simulator integrates human decision-making processes within the broader testing ecosystem, providing an essential bridge between purely virtual simulation and real-world human interactions. This approach is particularly valuable for validating transitional automation scenarios, where control shifts between human operators and the ADS.

DIL simulations offer a controlled environment where human interactions can be evaluated alongside ADS functionalities. By interfacing with driver interaction tools, such as steering wheels, pedals, and other human-machine interfaces (HMI), DIL simulators emulate real-world driving dynamics. These simulations support various operational configurations, including real-time feedback loops, to replicate dynamic scenarios influenced by human actions. The result is a detailed analysis of ADS performance under diverse conditions.

5.3.1 Physical and Virtual Components on Subsystem Level

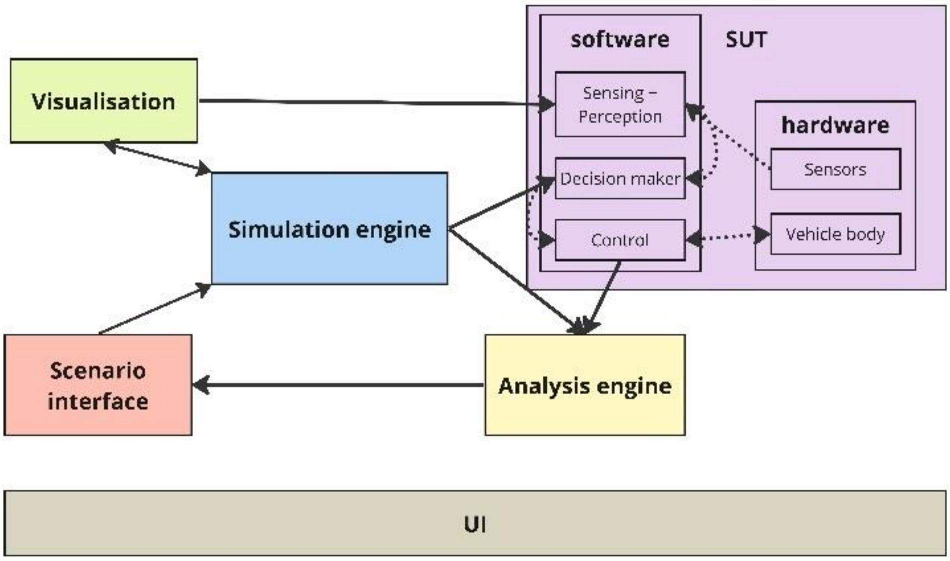

Figure 24: ADS simulation reference architecture

The modular simulation framework used for DIL simulation provides the flexibility and scalability required for driver-in-the-loop testing applications. This framework includes both physical and virtual components that interact to create a robust simulation environment. Key elements of the framework are:

1. **Sensor Simulation**: Provides high-fidelity sensing data to emulate sensor inputs, supporting both ADS testing and driver feedback loops.
2. **Simulation Engine**: Executes scenario descriptions, supporting object-based and sensor-based simulations for automated and driver-involved contexts.
3. **Scenario Interface**: Ensures accurate execution of scenarios, translating scenario semantics into actionable execution logic.
4. **Analysis Engine**: Iterates through logical and concrete scenarios, optimizing parameters to identify potential system vulnerabilities.
5. **System Under Test (SUT)**: Represents the ADS and driver interaction modules to ensure a comprehensive evaluation of the combined system.
6. **User Interface (UI)**: Allows users to configure and interact with the framework, essential for human-in-the-loop scenarios.

The framework's modular nature ensures it can interface with driver interaction hardware and software, integrating human behaviour into simulation workflows.

5.3.2 Advantages in Comparison to Virtual Testing

- **Realistic Human Interaction:** Captures driver responses, including reaction times, decision-making tendencies, and behavioural variances under different conditions.
- **Enhanced Transition Testing:** Enables testing of transitional automation phases where control is exchanged between the ADS and the human driver.
- **Improved Safety and Usability Assessments:** Helps evaluate how humans interact with automation features, improving system usability and safety measures.
- **Validation of Complex Scenarios:** Allows testing of edge cases that require human intervention, such as unpredictable pedestrian movements or unexpected road obstacles.

5.3.3 Limitations in Comparison to Virtual Testing

- **Higher Setup Complexity:** Requires additional hardware and software integration to accommodate real-time driver interactions.
- **Increased Cost and Maintenance:** Physical components such as steering interfaces, HMI devices, and motion platforms add to the cost and maintenance overhead.
- **Variability in Human Input:** Unlike purely virtual tests, which are deterministic, human-in-the-loop tests introduce variability, making reproducibility more challenging.

5.3.4 Additional Validation Needs

The hybrid test environment described here can tackle multiple abstraction levels, including planning, sensor simulation, and human-in-the-loop testing. The validation needs depend on the test objective. From the driver-in-the-loop perspective, ensuring real-time operation is crucial, along with maintaining a visual fidelity level that meets human perception requirements.

5.3.5 Exemplary Implementation and Setup

WMG's 3xD simulator incorporates the simulation framework described above and has been used for research focused on human driver studies and automated driving system testing. Example projects relevant to CCAM Safety Assurance include:

- **Motorway Gantry Demonstration**: This project used the 3xD simulator to study the impact of variable gantry signage on motorways. The demonstration explored how drivers react to changes in road signage within a controlled virtual environment, incorporating adverse weather conditions such as heavy rain and fog.
- **SAVVY Project**: This project focused on testing Advanced Driver Assistance Systems (ADAS) in a controlled environment. The study addressed challenges such as decreased driver engagement, limited access to prototype vehicles, and test time constraints. The project developed a simulation-based validation and verification process for ADAS technologies, leveraging deep learning and Convolutional Neural Network (CNN) algorithms on FPGA architectures.

Figure 25: The 3xD immersive driving simulator in WMG, University of Warwick

Figure 26: WMG's 3xD used for gantry study

Figure 27: WMG's 3xD used for the SAVVY project

6 SUMMARY OF ADVANTAGES AND DISADVANTAGES OF HYBRID AND REAL-WORLD TEST METHODS

The investigation of various hybrid and real-world test methods revealed that each method has its own advantages and disadvantages. No single solution outperforms all other test methods or can be universally recommended for all use cases. This chapter summarizes the advantages and disadvantages of hybrid and real-world test methods to support the selection of suited testing methods. Table 2 compiles the information that has been discussed for

- Section 3.1 Black Box Testing on Proving Grounds and Public Roads,
- Section 3.2 Scaled Model Testing,
- Section 4.1 Vehicle-in-the-Loop on Proving Grounds,
- Section 4.2 Vehicle-in-the-Loop on Test Benches,
- Section 5.2 Driver-in-the-Loop on Proving Grounds,
- Section 5.3 Driver-in-the-Loop in Simulators.

The table presents a comparison of the advantages and disadvantages in comparison to pure virtual testing.

Table 2: Advantages and Disadvantages of Hybrid and Real-World Testing

Method	Advantages	Disadvantages
Black Box Testing on Proving Grounds and Public Roads	• Unbiased and focusing only on system outputs • Testing fully implemented ADS function (real perception) • Independent from manufacturer • No virtual models needed	• Can not uncover architecture flaws • Limited resources (test track, test equipment) • ADS possible limited by geofencing • Possible permits needed for testing on public roads
Scaled Model Testing	• Cost effective test • Partial validation possible • Minimized risk compared to full scale real-world testing	• Hardware Limitations for physical down scaling
Vehicle-in-the-Loop on Proving Grounds	• Few models needed • Real components do not have to be validated	• Low test through put • Expensive test setup • Limited scenario space • Limited availability of test setups

	• Many elements, such as powertrain, can be treated as black boxes	• Safety limitations due to potential of physical crashing into targets • Limited number and types of targets
Vehicle-in-the-Loop on Test Benches	• Hil evaluation of real hardware components (sensors, actuators, vehicle dynamics) • Increased realism • Dynamic responses for drivetrain and steering • Identification of hardware issues possible	• Limited scalability • Costs of test architecture such as dynos and sensor stimulators • Limited flexibility towards new scenarios
Driver-in-the-Loop on Proving Grounds	• Validation of human interactions for example in takeover situations possible • Retaining real physics and feedback • Dynamic objects can be virtually altered for the driver	• Limited scenario space due to safety concerns • Limited scalability due to limited available test setups and high preparation efforts • Limited repeatability due to human input
Driver-in-the-Loop in Simulators	• Validation of human interactions for example in takeover situations possible • Validation of complex scenarios with human interaction possible	• High setup complexity • Increased test setup costs • Limited repeatability due to human input

To further elaborate on the advantages of hybrid and real-world test methods in CCAM validation, we examine how they address the limitations of pure virtual testing and help bridge the gaps identified in **Limitations of Virtual Testing (2.3)**.

One key limitation of pure virtual testing is the reliance on validated simulation models, which may not always be available, as highlighted in **Limitations for the Subject Vehicle – AD Function (2.3.3)**. Current simulation tools often lack direct support for interfacing embedded AD stacks, making their validation challenging. **Black Box Testing on Proving Grounds and Public Roads (3.1)** overcomes this by allowing the validation of series or prototype vehicles without requiring detailed models, making it particularly useful when such models are missing.

Similarly, pure virtual testing often struggles to uncover implementation flaws and misunderstandings in early development stages, as noted in **Limitations for the Subject Vehicle – AD Function (2.3.3)**. Software implementation errors can persist undetected due to the abstraction level of the simulation. **Scaled Model Testing (3.2)** mitigates this by enabling early CCAM validation in a physical environment, helping to detect such issues earlier in the development process.

Another significant limitation of pure virtual testing is the difficulty in accurately representing real vehicle behavior, especially in dynamic maneuvers, as discussed in **Limitations for the Subject Vehicle – Vehicle Dynamics (2.3.4)**. Tire-road interactions, traction estimation, and environmental factors create gaps between simulated and real behavior. **Vehicle-in-the-Loop on Proving Grounds (4.1)** addresses this by integrating real vehicle behavior into a virtual traffic environment, ensuring that physical vehicle dynamics are inherently valid.

Likewise, virtual testing can struggle with realistic sensor interactions, as sensor models are difficult to validate and may not fully capture integration issues, as outlined in **Limitations for the Subject Vehicle – Sensors (2.3.2)**. Even with high-fidelity physics-based rendering, sensor models struggle with weather effects, lens distortions, and environmental reflections. **Vehicle-in-the-Loop on Test Benches (4.2)** helps overcome this limitation by using real sensors in a controlled virtual environment, providing a more accurate assessment of sensor performance and integration.

A major shortcoming of purely virtual approaches is their inability to fully capture human behavior and interactions, as driver models are inherently limited and human variability is difficult to simulate. **Driver-in-the-Loop on Proving Grounds (5.2)** and **Driver-in-the-Loop in Simulators (5.3)** directly address this by incorporating real human drivers into the test process, making it possible to study driver interactions in takeover scenarios and other complex situations that are not sufficiently covered in simulation platforms.

These hybrid and real-world methods serve as execution platforms within the SAF when no validated virtual execution platform is available. They can be selected immediately through the allocation process or used as a fallback if a virtual execution run is deemed invalid and requires **reallocation**. By integrating physical and hybrid testing where virtual methods fall short, SAF ensures a more robust, adaptable, and comprehensive validation framework.

7 CONCLUSIONS

This deliverable significantly contributes to the overall project objectives by developing a clear understanding of the importance of **hybrid and real-world testing** within the **Safety Assurance Framework (SAF)**. It defines and describes hybrid and real-world CCAM safety assessment methods that serve as a **test execution** environment within the SAF, providing a valuable alternative for structured safety assessment both within and beyond the limits of virtual simulation.

This deliverable motivates the need for hybrid and real-world testing methods by discussing **limitations of the harmonized V&V simulation framework** in **Chapter 2.3**. While virtual testing is efficient and safe, several limitations remain, including **latency timing discrepancies**, the gap between simulation and reality (**sim2real gap**), and challenges in accurately **modeling sensor behavior**. Additionally, virtual testing omits challenges arising from **target hardware constraints**, which can impact system performance in real-world deployment.

Consortium experts have contributed novel and state-of-the-art **hybrid and real-world testing methods** to this deliverable that address these limitations. For detailed discussions, see **Chapters 3, 4 and 5**. The analysis identifies specific **advantages and disadvantages** associated with each testing method, and their role in overcoming limitations of pure virtual CCAM safety assessment is summarized in **Chapter 6**.

Hybrid and real-world safety assessment techniques can be widely employed, from early development to production vehicles, helping to **solve CCAM safety assessment problems and reduce testing costs**. This deliverable focuses on real-world testing methods, which offer the benefit of exposing systems to **real-world conditions** and **capturing complex interactions** that may be difficult to simulate virtually. However, physical testing is resource-intensive and may have limitations in scalability and repeatability. **Hybrid methods**, such as Vehicle-in-the-Loop (VIL) testing, combine real and simulated components, **balancing realism and controllability**. This deliverable also explores how humans can be incorporated into the safety assessment process through **Driver-in-the-Loop** testing, addressing the challenge of modeling **human behavior**.

By providing a structured overview of hybrid and real-world methods, this deliverable **strengthens the SAF's allocation** process to appropriate test execution environment. AD developers and regulators benefit from awareness of existing physical and hybrid methods, enabling informed decisions on execution environments. The insights gained in Task T4.6 contribute to the SAF demonstration through various use cases in Task T7.3.

The hybrid and real-world methods employed in Task T7.3 serve as exemplary implementations, showcasing their practical application of the SAF rather than providing fully validated toolchains. At the time of this task, no toolchain ready for validation exists, hindering a complete validation of the testing methods. Instead of explicit validation results, this deliverable discusses **general validation concerns** for each presented method. For the testing methods demonstrated in T7.3, **potential validation strategies** with respect to the requirements of T7.1 are proposed, but their full verification and validation will require further toolchain development and research efforts beyond this task.

In conclusion, D4.6 serves as a central resource for advancing the understanding of hybrid and real-world safety assessment techniques for CCAM systems. It strengthens the SAF by providing insights into efficient and effective test execution, enabling stakeholders to make informed decisions about CCAM safety assessment methods.

8 REFERENCES

[1] SUNRISE, "D7.1 - CCAM Use-cases-validation-requirements," 2024.

[2] SUNRISE, "D3.3 - Report on the Initial Allocation of Scenarios to Test Instances," 2024.

[3] SUNRISE, "D4.4 - Report on the Harmonised V&V Simulation Framework," 2024.

[4] SUNRISE, "D4.5 - Report on Validation of the harmonized V&V simulation framework," 2024.

[5] H. Moosbrugger and A. Kelava, "Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien")," in *Testtheorie und Fragebogenkonstruktion*, 2020, pp. 13-38.

[6] J. Gäde, K. Schermelleh-Engel and C. Werner, "Klassische Methoden der Reliabilitätsschätzung," in *Testtheorie und Fragebogenkonstruktion*, 2020, pp. 305-334.

[7] Messring, "SprayMaker: Regensimulation für Active Safety Testing," 21 01 2025. [Online]. Available: https://www.messring.de/de/blog/spraymaker-regensimulation-fuer-active-safety-testing/.

[8] AVL, "Mobility and Sensor Test Center," 21 01 2025. [Online]. Available: https://www.avl.com/en/engineering/automated-and-connected-mobility-engineering/calibration-verification-and-validation/mobility-and-sensor-test-center.

[9] UNECE, "UN Regulation No. 157 - Automated Lane Keeping Systems (ALKS)," United Nations, 2021.

[10] SUNRISE, "D7.2 - Safety assurance framework demonstration instances design," 2024.

[11] G. Özbilgin, A. Kurt and Ü. Özgüner, "Using scaled down testing to improve full scale intelligent transportation," in *IEEE Intelligent Vehicles Symposium Proceedings*, Dearborn, MI, USA, 2014.

[12] H. M. La, R. S. Lim, J. Du, S. Zhang, G. Yan and W. Sheng, "Development of a Small-Scale Research Platform for Intelligent Transportation Systems," *IEEE Transactions on Intelligent Transportation Systems,* pp. 1753-1762, 2012.

[13] V. Kumar and e. al., "iTETRIS: Adaptation of ITS Technologies for Large Scale Integrated Simulation," in *IEEE 71st Vehicular Technology Conference*, Taipei, Taiwan, 2010.

[14] I. O. f. S. ISO, *ISO 21448:2022 Road vehicles — Safety of the intended functionality,* 2022.

[15] SUNRISE, "D4.1 - Report on relevant subsystems to validate CCAM systems," 2024.

[16] M. H. R. a. M. A. Damschen, "WayWise: A rapid prototyping library for connected, autonomous vehicles," *Software Impacts,* p. 100682, 2024.

[17] K. Kvarnfors, "Motion Planning for Parking a Truck and Trailer System," KTH, School of Electrical Engineering and Computer Science (EECS), Stockholm, 2019.

[18] A. Dosovitskiy, G. Ros, F. Codevilla, A. M. López and V. Koltun, "CARLA: An Open Urban Driving Simulator," in *Conference on Robot Learning*, 2017.

[19] B. Vedder, "VESC project," [Online]. Available: https://vesc-project.com/. [Accessed 20 February 2025].

[20] Tesla, 2025. [Online]. Available: https://www.tesla.com/en_gb/support/transitioning-tesla-vision.

[21] C. Schyr, H. Inoue and Y. Nakaoka, "Vehicle-in-the-Loop Testing - a Comparative Study for Efficient Validation of ADAS/AD Functions," in *International Conference on Connected Vehicle and Expo (ICCVE)*, Lakeland, FL, USA, 2022.

[22] T. Mockeridge, H. Dohman and Phiilips, "HIL Driveline Dyno.," *SAE Technical Paper,* 2014.

[23] T. Weck and T. Sugiura, "Funktionale Prüfung des Allradantriebsstrangs im neuen XV Hybrid. CTI Symposium Fahrzeuggetriebe," in *HEV- und EV-Antrieb Berlin*, 2013.

[24] T. Bock, *Vehicle in the Loop - Test- und Simulationsumgebung für Fahrerassistenzsysteme,* Cuvillier Verlag, 2008.

[25] G. Berg, *Das Vehicle in the Loop,* Luxemburg: Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 2014.

[26] J. L. V. E. W. J. David Goedicke, "Vr-oom: Virtual reality on-road driving simulation," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, Montreal, 2018.

[27] N. Wagener, *Aufbau und Bewertung eines realdynamischen urbanen Fahrsimulators,* Aachen: Rheinisch-Westfälische Technische Hochschule Aachen, 2024.

[28] M. Fischer, A. Labusch, T. Bellmann and C. Seehof, "A Task-oriented Catalogue of Criteria for Driving Simulator Evaluation," in *Driving Simulator Conference*, Tübingen, 2015.