

# Safety Validation of Autonomous Vehicles using Assertion-based Oracles

Chris Harper, Greg Chance, Abanoub Ghobrial, Saquib Alam, Kerstin Eder, Tony Pipe

**Abstract**—Safety and mission performance validation of autonomous vehicles is a major challenge. We describe a methodology for constructing and applying assertion checks, operating in simulation or in the real world. An assertion library has been derived from the UK Highway Code (UKHC), as an example of a legal code of practice. Transformation of UKHC rules into logical assertions can be achieved either by direct translation or by physical modelling, to yield logical assertion expressions. To validate the effectiveness of the assertion library we have performed assertion checking of the UKHC rules using a geospatial information system in an SQL database.

“validation suite” only occurs in abstract?

This paper presents initial simulation and realtime video analysis experiments that apply assertions related to vehicle overtaking, and a brief analysis of the safety and mission performance characteristics measured.

What are the contributions? More detail on scope (focus on overtaking to illustrate the approach as we don't yet have a “library”).

**Index Terms**—Verification and Validation (V&V), Autonomous Driving, Autonomous Vehicles, Simulation, Testing

## I. INTRODUCTION

Verification and validation are important to earning trust and gaining confidence in the safety of autonomous systems (AS) such as autonomous vehicles (AVs), which are the subject of this paper. Safety validation of any system can (and should) be performed by a variety of measures, including analysis and inspection of designs and their implementation. In this paper we focus on validation by system testing, and in particular on using monitors within simulation and at runtime for a vehicle overtaking scenario.

Manuscript received ...; revised ...; accepted.... Date of publication ...; date of current version ....

This research has in part been funded by the ROBOPILLOT and CAPRI projects. Both projects are part-funded by the Centre for Connected and Autonomous Vehicles (CCAV), delivered in partnership with Innovate UK under grant numbers 103703 (CAPRI) and 103288 (ROBOPILLOT). This research was also supported in part by the UKRI Trustworthy Autonomous Systems Node in Functionality under grant number EP/V026518/1. Also special thanks to Séverin Lemaignan.

Chris Harper (e-mail: chris.harper@brl.ac.uk), Saquib Alam (e-mail: saquib764@gmail.com), and Tony Pipe (e-mail: tony.pipe@brl.ac.uk) are with the Bristol Robotics Lab, T Block, University of the West of England, Frenchay, Coldharbour Ln, Bristol, BS34 8QZ, United Kingdom.

Greg Chance (e-mail: greg.chance@bristol.ac.uk), Abanoub Ghobrial (e-mail: abanoub.ghobrial@bristol.ac.uk), and Kerstin Eder (e-mail: kerstin.eder@bristol.ac.uk) are with the Trustworthy Systems Lab, Department of Computer Science, University of Bristol, Merchant Ventures Building, Woodland Road, Bristol, BS8 1UB, United Kingdom.

Digital Object Identifier ....

Autonomous systems are significantly more complex than their non-autonomous counterparts, and therefore more difficult to validate. Whereas a non-autonomous system is usually designed only to deliver a given servo-mechanical or information processing relationship from inputs to outputs (its function), the problem of autonomy is concerned with how that function is used to achieve one or more situated tasks within the environment, which have specified goals that must be achieved. The state of the environment is as much a part of the validation problem of an autonomous system as the internal state of the system. Hence, the task of validation is often a much higher dimensional problem to solve than for conventional systems, leading to a major expansion in the number of states of a system that may need to be tested. Field testing, often the traditional approach to validating systems in their target environments, is likely to be prohibitively expensive for autonomous systems, if a high degree of coverage or statistical confidence is required for safety assurance. This has led to much interest and research into the use of driving simulators<sup>1</sup> for safety validation. Simulation offers the possibility to exercise at least the decision-making logic of a system over a wide range of simulated conditions, and at accelerated rates compared to real-time physical testing. However, the exponential expansion of states associated with autonomy problems

autonomy problems - is this politically correct, i.e. “problems”?

can also lead to infeasible testing time and resource usage even for simulators. So, care must be taken to ensure simulators are used efficiently and effectively, obtaining as much validation information as possible and getting to the important test conditions that provide key information about the safety properties of a system as fast as possible, rather than by ‘brute force’ approaches such as statistical testing.

We are interested, therefore, in test methodologies that incorporate the classification, measurement and evaluation of a system’s situated behaviour. One such approach is based on the specification of situated behavioural properties as *assertions*, logical expressions that can be applied at some or even all steps of a simulation session

or at runtime, so I’d take this out from “at some” up to here

to monitor the behaviour of the system under test and to report any property violations. These properties (oracles) relate

<sup>1</sup>Examples include <https://www.rfpro.com> and <http://carla.org>.

to the observable driving behaviour of road vehicles. They characterise what a vehicle

really any vehicle, i.e. these could also be part of a driving simulator to be used by learner drivers

may be expected or legally required to achieve on public roads. The class of oracles we are investigating is the “rules of the road” legal driving codes for human drivers, which exist in most countries. We are focusing on our home country’s rules, the UK Highway Code [44] (UKHC), as a working example.

The UKHC is a long established document, about a century old at time of writing<sup>2</sup>, and has been subject to continual revision and evolution of the set of rules contained therein. We consider the UKHC to be a mature and valid reference source for a safe driving code of practice. Publicly available data<sup>3</sup> indicates that despite a significant increase in road users, UK road deaths have been decreasing monotonically since 1966. While there have been many causal factors contributing this trend (e.g. seat belts, airbags etc.), we argue that the advice given in the UKHC must have at least some partial contribution to this trend.

Leave out? Had any advice given at any particular time been confusing or contradictory, then it may have led to increases in accident rates, or would have been detected and corrected over time by UK motoring institutions such as driving instructors, driving test agencies, etc.

Since much of the content of the UKHC has been largely stable for many years, we argue that this is because the advice given is sound, and hence we conclude that it can be a valid reference for constructing safety validation assertions.

Why only safety? Liveness is also important. Or not?

The driving rules in the UKHC are in effect an oracle for correct behaviour on the road. In this paper we investigate how to systematically transform them into assertions which can be applied as monitoring checks. To achieve this, the driving rules must be translated from human-readable format to machine-readable expressions that can be automatically monitored during simulation-based testing and also during operation. Although we select the UKHC as the oracle for pragmatic reasons, it could equally be replaced with other metrics pertaining to functionality or even social convention. We review some related work in Section II, in which other potential sources for oracles have been identified.

Say what we do and that we do it on two examples

We present two examples for applying assertions for safety validation based on driving code of practice, within a simulation system and an online monitoring system. Both examples involve lane changes during an overtaking scenario. Assertions may be applied either in a simulation system during AV system development, or in a runtime mode during operational service. We present how the same assertion framework can be used for either case. We adopt the terms scene, scenario and situation from [45] throughout this paper.

These checks may be used as evidence to show functional safety compliance to national regulations or codes of practice,

the general safety argument being that if an AV satisfies all the rules that may be legally expected of a human driver, then its behaviour is comparable to that of human drivers and therefore should be acceptable, or at least legal. It should be noted that acceptance may therefore be somewhat context-specific, against the code of practice for each country, and what passes under one legal code may not be acceptable in another. We anticipate that there may need to be extensive harmonization of national driving codes in order to avoid having to re-certify vehicles for each new national market and improve the cost-effectiveness of AV certification processes, but this is outside the scope of this paper.

The need to measure the compliance of an AV’s behaviour with legal codes of practice may go beyond the initial design validation or certification stage as performed in simulation and (to some extent) road tests. It may become necessary to measure such compliance during vehicle operation, for example to establish that the vehicle’s behaviour is correct (or otherwise) in the event of an incident that requires subsequent investigation. If an AV is involved in an incident, but it can be shown by assertion-based assessment during operation that it was adhering to all relevant legal standards of driving behaviour, then this may have an impact on any liability or insurance-based compensation that may be due as a consequence of the incident. A detailed discussion of the rationale and requirements for explainability of AV decision-making in post-incident investigation is beyond the scope of this paper.

find some literature?

This is where we ought to say how the paper is structured

We present a novel technique for deriving, systematically formalising and encoding assertions from the UK Highway Code

Always write out or use the UKHC abbreviation introduced earlier.

and applying them in two modes: (i) within a Geospatial Information System (GIS) database linked to a vehicle simulator, and (ii) within a real-time monitoring system during vehicle operation.

In the first mode of use, we encode the UKHC assertions [44] as SQL queries and run them within a PostgreSQL database with a PostGIS extension.<sup>4</sup> In this mode, the assertions can make use of any data stream available, passed into the database records.

In the second mode of use, the principal idea is that data about the AV, other road users and the adjacent environment from the test vehicle is sufficient to evaluate the assertions during operation. The same PostgreSQL-based database as used in the simulator can therefore be installed as embedded software, but receiving its data directly from vehicle sensors instead of a simulator. This captured data need not be of a high fidelity, but rather should have “just enough” detail [27] to allow the assertions to be evaluated with good confidence.

This paper is structured as follows: Related work is reviewed in Section II. Section III explains the structure of the

<sup>2</sup>[https://en.wikipedia.org/wiki/The\\_Highway\\_Code#History](https://en.wikipedia.org/wiki/The_Highway_Code#History)

<sup>3</sup>[https://en.wikipedia.org/wiki/Reported\\_Road\\_Casualties\\_Great\\_Britain](https://en.wikipedia.org/wiki/Reported_Road_Casualties_Great_Britain)

<sup>4</sup>Refer to <https://www.postgresql.org/> and <http://postgis.net>.

PostgreSQL database and how it integrates with the underlying simulator or runtime monitoring environment. Section IV discusses the methodology of developing assertions derived from the UKHC, and the principles of using assertions as an oracle both in simulations and for operational analysis. Section V presents the scenario we have used to demonstrate the application of assertion-based oracles and describes two case studies which we use to illustrate the two options, simulation testing and operational monitoring. The results from the case studies are presented and analysed. In Section VI we discuss some important aspects of assertion-based monitoring, which have emerged from the experimental work, in particular the use of assertions for performance monitoring as well as for safety, and the observation that in many assertions the intention of ego-vehicle agents may also need to be monitored.

say why

Finally, in Section VII we draw conclusions and suggest directions for future work.

## II. RELATED WORK

Over the past few years, several researchers have investigated the classification and analysis of autonomous driving safety. However, most early works focus on autonomous driving robots in an industrial environment. Park et al. [34] uses braking distance to calculate so-called risk distance, and then plan a safe trajectory to avoid obstacles. While the resulting method has a good performance in a rule-free environment, it is highly unlikely that this method alone would be sufficient to accommodate any human-defined safety rules in its path planning algorithm.

But that is to inform the vehicle, not for external assessment, or?

In [6] Bourbakis developed a priority language to formalise human-defined traffic rules in a machine-understandable fashion. Their priority language, called KYKLOFORIA, can be used by autonomous robots to navigate dynamic environments by defining a navigation corridor for every robot in the environment and then generate a navigation strategy using the language. While the language appears very promising for navigation tasks in a free dynamic environment, it is unclear how to use it in a constrained setting of the road traffic to verify the correctness of the vehicle's manoeuvre.

So, both of these are for the robots to use to improve their behaviour, not for external assessment - yes?

There has been some recent work to make autonomous vehicles abide by highway traffic laws by formalizing the rules in a language that robots can understand.

This sentence fits better into our introduction? This is challenging not only because the traffic rules are often vague and conflicting, but also due to rule changes and the need for common sense knowledge.

Prakken et al. [35] discusses some of these issues and argue in favour of the need for formalization of traffic rules. Rizaldi et al. [37] use the formalization as a means to resolve accountability issues in an event of an accident. To the best of

our knowledge, most work [37, 16, 38, 4] advocates the use of Higher-Order Logic (HOL) to codify a wide range of traffic rules into a formal language. The authors of [37, 16, 38, 4] use Linear Temporal Logics (LTL) to construct a set of equations for various manoeuvres to ensure road safety. To construct these equations, a formal model of each lane and lane marking is required. The authors of [38] used *lanelets* [5] to build these lane models. While many works use the German or Vienna convention [15] on road traffic, in [4] a road junction rule from the UK Highway Code [44] is used to demonstrate their methodology. Some researchers [22, 18] have used temporal logic queries to formulate the HOL. These queries can be formulated using a variety of query languages like SQL, OQL, RQL, etc. In this paper, we use SQL to build the assertions due to their high expressiveness [30]. Encoding every rule by hand is, however, an extensive and potentially error-prone task.

DeCastro et al. [10] guide an agent to learn probabilistic models, via a learning-based algorithm, constrained by safety contracts which are in essence traffic rules. However, the major challenges for using learning algorithms are completeness and verification of the solutions. Such systems are essentially a black box, and it is extremely difficult to validate their correctness. Another major problem is what reward or cost function should be used to train the model. One approach can be to have an oracle providing rewards for correct behaviour, but this leads back to developing traditional assertion-based rules using the relevant highway code.

The Responsibility Sensitive Safety (RSS) framework developed by MobilEye [40, 25] proposes mathematical models for mimicking the human subjective decision making as an effort to provide safe driving behaviour by AVs. On the other hand, legal codes such as the UKHC often define the interactions required with particular features of their respective road environments (for example, junctions, roundabouts, or crossings). Our work is intended to address this latter domain, although a complete database will need to incorporate sets of assertions of both categories.

All the works discussed earlier are limited to detecting the occurrence of some undesirable events while driving. However, such events are not equal in severity [41]. Myers et al. [32] propose a framework to access autonomous driving systems using two types of scoring rules: prescriptive and risk-based. To the best of our understanding, the prescriptive scoring mechanism captures the capability of the AV to follow traffic rules while the risk-based scoring system assesses the behaviour of the AV in rare events where undesirable outcomes are unavoidable, and it is expected of the AV to exhibit behaviour that would minimise the risk. Developing a generic system to score risks in different scenarios is challenging and requires a separate study of its own. Therefore, risk-based assessment is not part of this paper.

Conclude with how our work differs and contributes to the existing body of knowledge.

### III. SYSTEM ARCHITECTURE FOR ASSERTION-BASED MONITORING

The evaluation of assertions requires an environment to generate vehicle data, and apply the safety assertions to interpret the driving situation. This section presents the generic components and software architecture we have deployed to support assertion-based monitoring either in simulation or operational modes.

#### A. Simulation Environment

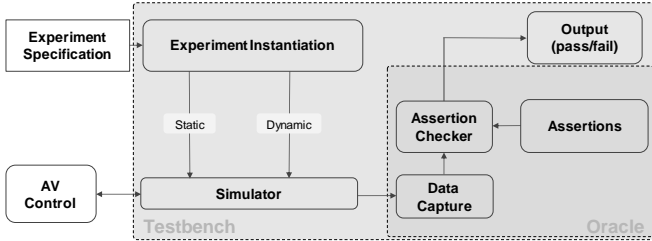


Fig. 1: Simulation testbench architecture

Figure 1 shows a layout for a simulation testbench that indicates the context in which the assertions database subsystem is used:

Should the figure not also include assertion coverage?

Why subsystem? Can we not just use the name of the respective box(es)?

The *Experiment Instantiation* subsystem generates test scenarios derived from the *Experiment Specification* which may include test objectives or requirements (defined in natural language). This subsystem identifies the *Static* information about the environment (e.g. maps, road features) and *Dynamic* descriptions of the behaviour required of simulated actors (vehicles, other road users, traffic lights, any third-party agents necessary to the test requirement). The experimental configuration determines the test case that will contribute coverage towards the assertion.

Nothing said yet about coverage.

The *Simulator* subsystem interacts with the *AV Control* subsystem - although for the purposes of this paper a dummy controller is used that can be set to pass or fail assertions to demonstrate the full extent of the assertion-based safety validation approach. The simulator used is the CARLA simulator<sup>5</sup>, but in principle this could be exchanged for other simulators with only moderate effort. The *Data Capture* subsystem records relevant simulation data about the driving situation of the AV, including actor positions and road network information, and at each time step of the simulation writes this information into the records of the database to be used for assertion check-

ing. Should there not be a DB symbol (looks like a tin) to show the database we are sourcing?

The *Assertion Checker* subsystem together with the *Assertions* library constitute the test *Oracle*. These subsystems evaluate all assertions relevant to the driving situation defined at each step of the simulation. The *Outputs* are a set of Pass/Fail declarations which can be used to annotate the simulation log or be recorded in a separate file if desired.

Should there not be a DB symbol to represent the assertion library, i.e. a set of SQL queries?

We will focus on describing the assertion oracle in greater detail in the remainder of this paper.

add the relevant section

Really, the “Simulator” interfaces with the “Data Collector”, which writes to a “Simulation Log” DB. There is also an Assertion DB. The Assertion Checker sources both the “Simulation Log” DB and the “Assertion” DB. There is enough space to do this, as follows: Experimental Instantiation -change to- Experimental Instantiation and make box smaller, this means the LHS occupies less space. Move “Data Collector” into that space. Add DB to RHS of Data Collector - this results in two DB icons on RHS or one with two parts, one for the log data and one for the assertions/queries.

#### B. Operational Monitoring

The basic argument underlying our approach to assertion-based monitoring during operation is that the assertions we have developed for checking in the geospatial database can also be utilised to act as a reference model and monitor the vehicle during operation. Thus, the same Assertion Checker and Assertions contained in Figure 1 can also be used in operational monitoring mode. The external system architecture required to achieve this is illustrated in Figure 2. In this architecture, the external subsystems comprise:

- The *AV Sensors* acquire real-time information from the vehicle’s sensors.
- During *Feature Extraction*, the data obtained from AV sensors are processed to provide a trace equivalent to that obtained from simulations, so that the *Assertion Checker* can also be applied to operational monitoring.

Ensure same style as prev. figure, i.e. “Feature Extraction” - same for remaining items.

- *Data Capture*: similarly to the simulator architecture in Figure 1, the data describing the vehicle’s environment is captured and stored in a database for subsequent assertion checking.

Check with Greg re calling this a DB and the use of a DB symbol in prev. figure.

- The *Assertion Checker* and *Assertions* library: this subsystem is essentially identical to that of the simulation testbench architecture in Figure 1, evaluating the relevant assertions and generating a pass or a fail verdict [29] as an output. In some applications, there may be differences in the deployment (as shown in Figure 2) compared to the simulation testbench, where the system would

<sup>5</sup><https://carla.org/>



be deployed on embedded hardware. These applications include operational data recording for further processing or possibly its use as a real-time reward for online learning.

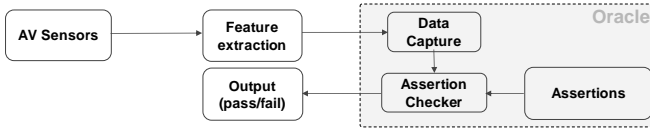


Fig. 2: Operational monitoring architecture

Revise to match new version of Fig. 1

### C. Geospatial Database for Assertion Checking

To develop assertions derived from UKHC rules, we need a formal language into which they can be translated for automatic checking. Two aspects are important to the choice. First, as we discuss later, the rules and guidelines in the UKHC are written for human readers not automatic systems. They consist of sets of discrete rules applicable to specific aspects of road user behaviour, which must be taken collectively, and not necessarily always in a particular order, to form the legally required profile of behaviour. This suggests that a declarative programming language, such as Prolog or SQL, may provide an easier and more easily validated means of encoding rules than more familiar procedural languages such as the C family. Second, we are interested in processing the data produced during driving, so we seek languages that handle large data sets naturally, where the assertions we develop constitute formal operations on datasets corresponding to the logical properties we extract from the UKHC rules. Taking these two requirements into consideration, SQL emerges as the most suitable commonly available programming language, having all the properties mentioned as well as being supported by many publicly available databases and software development environments. SQL has been demonstrated formally [30] to be sufficiently expressive to be able to encode useful problems such as state or graph reachability as well as all standard problems in relational algebra, albeit with the use of some of its later extensions such as recursive operators. Safety properties can be defined in terms of reachability [19, 31] or (in-)stability [21, 50] of desirable (goal) and undesirable (hazardous) situated states, so the ability of SQL to express such problems in a general sense is another important capability. For these reasons, we have selected SQL-based database technology as the framework for our assertion-based oracle.

The technical challenge then becomes how to compare the state of the vehicle against the assertion or safety property. We argue that a geospatial relational database is well suited to the task for formal reasoning about the vehicle properties against a set of assertions. The vehicle state can be easily stored and accessed and additional insights, such as dynamic properties, can also be derived from this base information. A database system, such as PostgreSQL, can record, access and perform assertion checks upon this state information quickly and reliably and is especially useful considering libraries that make use of geographical shape objects, e.g. PostGIS.

We have implemented the Assertion Checking subsystem as a PostgreSQL database with a PostGIS geospatial information system extension, which provides an extensive library of SQL-native spatial measurement functions. We chose these particular software products because (a) they are open-source, (b) PostgreSQL implements powerful table searching algorithms that have high performance - in the order of hundreds of thousands to millions of records searched per second. Real applications of this kind of simulator system will involve the use of large data sets, so the performance of the underlying database engine is an essential factor in achieving a practicable simulation-based safety validation toolset.

The database management system (DBMS) receives data from the simulator (or other data source) containing an actor state vector of variables such as position, velocity and yaw angle of the vehicles and other road users (e.g. pedestrians). This data is then entered into a set of tables, and the safety assertions are applied to the data as one or more database queries (written in SQL) acting on the relation.

Figure 3 is a UML-style deployment diagram of the Data Capture and Assertion Checking functions using a PostgreSQL DBMS server. In fact, the server hosts both elements; the Data Capture function is implemented as a set of records that can be updated externally by the Simulator or Experiment Instantiation functions; the Assertions Checking function is performed by the DBMS query engine by execution of SQL queries encoded with the assertion logic, potentially as a stored procedure, that will typically run all relevant assertions sequentially at each time step. The assertions are stored in the database server as a set of SQL queries or procedural functions to be invoked by the query engine whenever required.

The use of PostGIS introduces an extensive library of data types and method functions suited to measurement of basic physical (geometric) relationships between the object data received from the simulator. For the particular scenario considered in this paper (overtaking), functions measuring distance between geometric objects (the bounding-box shapes of vehicles on the road), and partial or total containment of the AV within road lane geometries, are of particular use, as these are used to identify initial and final conditions of overtaking manoeuvres.

PostGIS functionality can also be used to generate more complex geometric dynamic data derived from the basic state of the AV or other simulated agents. An example of this is shown in Figure 4, in which two polygons for *thinking distance* (inset, green hashed area) and *braking distance* (red hashed area), which are parameters defined in the UKHC (Rule 126), can be generated as a function of the basic forward velocity data of the AV. Figure 4 shows a projection of these derived geometries onto the associated street map. In this way new abstractions can be developed by building new derived data sets from the original (captured) simulation data. Where any such abstraction of the simulation data can be generated it can be manipulated effectively as a new variable (SQL attribute) appended to the original data set, and referenced by queries in the same way.

This is an essential computational mechanism for the methodology we present in this paper. We can develop pro-

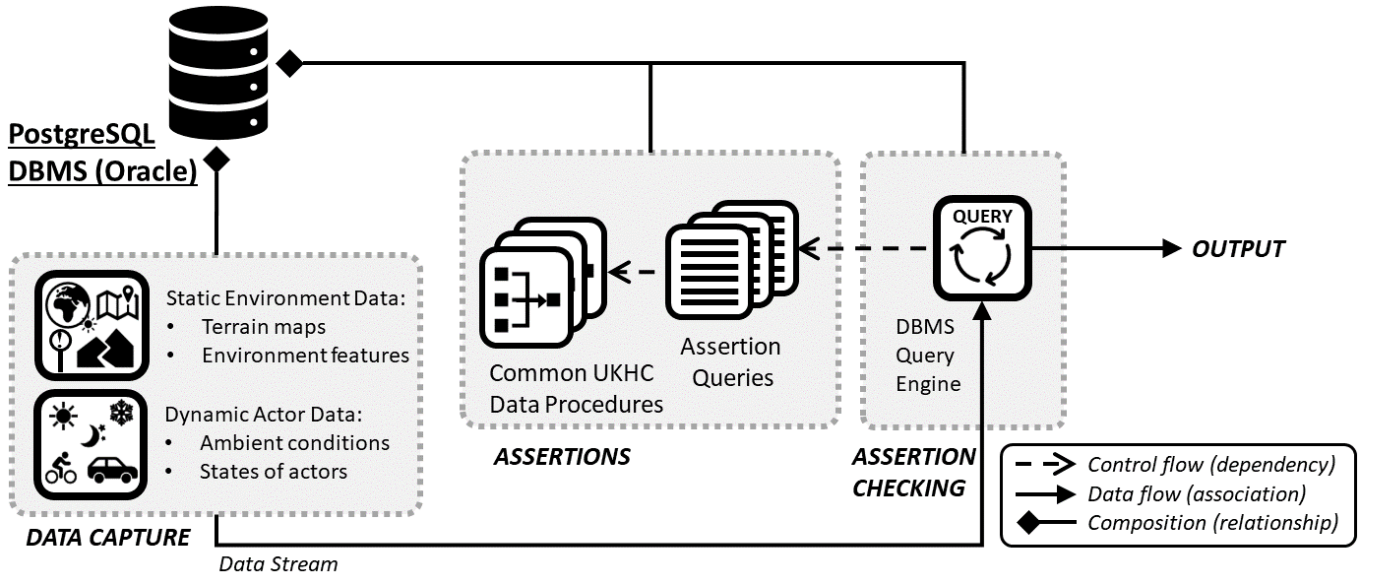


Fig. 3: Deployment Diagram of the Oracle as a PostgreSQL DBMS, showing how the Data Capture and Assertion Checking functions are allocated.

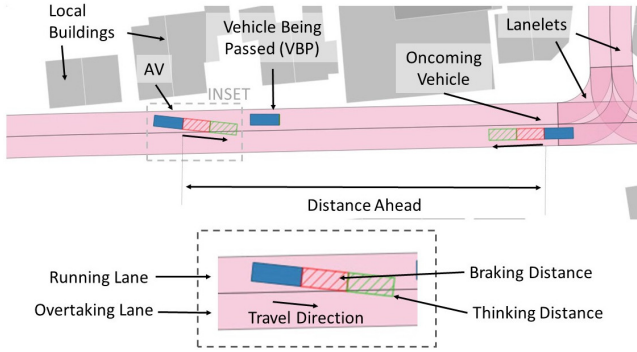


Fig. 4: Database view of a driving scenario

cedures and queries to generate abstractions directly related to the concepts expressed in the rules of the UKHC. The methodology for identifying assertions and key abstractions is described in Section IV. Where such concepts are used in multiple rules the corresponding abstractions form a library of common queries or operations on data, which can be incorporated into the PostgreSQL DBMS as a set of stored procedures within the general database schema for the Test-bench Architecture. Figure 3 shows how the procedure library is included in the overall deployment and its dependencies to other elements of the system. This library will be developed incrementally as each assertion is produced. We anticipate that eventually the library of procedures may be packaged and made available as a PostgreSQL extension. Data and derived abstractions can be combined within queries for assertion checking and coverage collection. For example, if an assertion is required to determine whether the braking zone of the AV has intersected another road user, such as a vehicle travelling in the opposite direction, then the PostgreSQL query can be formulated as:

```
SELECT ST_Overlap(g1.geom::geography,
                  g2.geom::geography)
FROM sim_log.data g1
WHERE g1.agent_id = AV_braking_zone
AND g2.agent_id = Oncoming_Vehicle;
```

where a PostGIS library function, `ST_Distance`, is used to measure the shortest Euclidean distance between the two physical objects or geographical areas `g1` and `g2`, which are a vehicle braking zone and an oncoming vehicle.

The main illustration in Figure 4 shows the operating situation associated with the overtaking scenario described in Section V, and gives an example of a derived variable called *Safe Distance Ahead* (SDA), whose name is derived from terminology in Rule 162 of the UKHC. The figure shows an urban scene depicting a section of road network divided into *lanelets* denoting legal division between driving lanes and any local buildings in the area. In this example the AV is on the left of the figure overtaking a parked vehicle requiring a lane change. The distance to the oncoming vehicle is termed the *Distance Ahead*. The inset shows annotation denoting the braking and thinking distances of the vehicle which can be treated as geographical object entities in the database and hence easily used for assertion checking.

The assertion to be tested performs a check (as defined in UKHC Rule 162) that a safe distance ahead exists between an AV and the nearest vehicle in an oncoming lane, labelled the ‘Oncoming Vehicle’ (OV) as the AV starts to overtake the ‘Vehicle Being Passed’ (VBP). The start time of the overtake is defined as the time that the AV first begins to enter the adjacent lane (Overtaking Lane) of the road, at which point it must commit to the manoeuvre or else abort and return to driving in the initial lane (the Running Lane).

In the case of assertions related to overtaking, it is necessary to determine whether an AV wholly or partly occupies a given

road lane, since overtaking manoeuvres require it to move at least partially into an adjacent lane and back. Therefore, it is necessary for the database to include lane geometries (polygons) in order to allow checking of vehicle occupancy of lanes by testing for the overlap between vehicle and lane objects. Figure 4 shows the lane polygons projected onto the map section (an extract from the map used for our initial assertion checking experiments described in Section V). The start of an overtaking manoeuvre is identified as the first time step in which the AV overlaps the overtaking lane (and stops being wholly contained within the running lane), and the overtake manoeuvre ends when the AV is entirely within the running lane once more after having passed the VBP. Lane geometry objects (shape or bounding box polygons) are therefore an essential element of the assertions related to overtaking.

#### IV. METHODOLOGICAL PRINCIPLES FOR DEVELOPING AND APPLYING ASSERTIONS

In this section we report how we develop and apply assertion checks, and present our methodology in its current form. We identify the types of assertions that may exist. We discuss how to develop assertions from the natural language statements of the UK Highway Code, either by direct translation or by modelling. We also discuss how we plan to use assertions, both in simulation and for runtime monitoring.

##### A. Categories of Assertions

We have found that assertions applied to simulation or runtime monitoring tend to fall into one of four principal types (some of which have sub-types). In general, assertions are defined with respect to some *assertion reference point* - a particular step of a captured data trace, defined by some reference condition which holds at a specified time step. The assertion is then defined as some spatio-temporal condition relative to the reference point, as shown in Fig. 5.

The four categories of assertions are:

- **Invariant Condition:**
  - An condition that must be satisfied at all time steps within the captured data trace (i.e. there is no specific assertion reference point, or the reference point is in effect the entire trace).
- **Execution Condition:**
  - A condition that must be satisfied at the assertion reference point.
- **Pre-condition:**

A spatio-temporal phenomenon that must exist in the steps preceding the assertion reference point. Pre-conditions can be one of two sub-categories:

  - *Temporal*: A time (step or period) prior to the assertion reference point, within which a specified condition holds true. Essentially, this is the temporal component of the pre-condition.

*Example: an assertion for the behaviour of an AV at a highway exit lane may check that the vehicle's indicators have been set correctly in a preceding period*

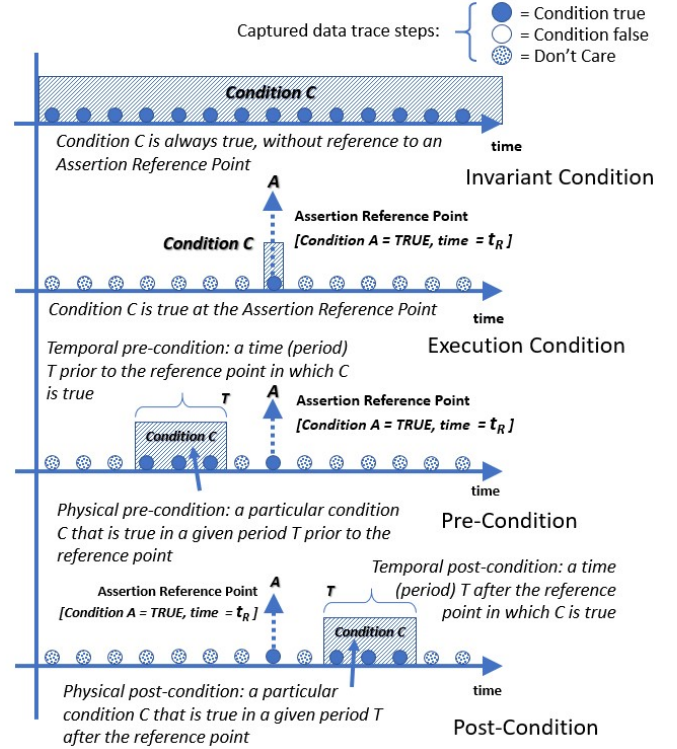


Fig. 5: Categories of assertions

(e.g. 5s) before it crosses the exit lane boundary (i.e. checking for unduly late or intermittent indications).

- **Physical**: A physical (e.g. spatial or state) condition that must be satisfied at a specified time (or within a time period) before the assertion reference point. This is the physical component of the overall pre-condition. *Example: a similar highway exit assertion may check, as the AV passes the the exit lane boundary, that the vehicle's indicators were set correctly at the last highway exit sign before the exit (i.e. checking for omitted indications).*
- **Post-condition**:
 

A spatio-temporal phenomenon that must exist in the steps following the assertion reference point. Post-conditions can be one of two sub-categories:

  - *Temporal*: A time (period) after the assertion reference point, within which a specified condition holds true. This is the temporal component of the post-condition. *Example: a highway exit assertion may check that if the vehicle's starts indicating that it intends to leave at the exit lane, that it does enter the exit lane within a given time period (i.e. checking for unnecessarily early indications).*
  - *Physical*: A condition that must be satisfied within a specified time after the assertion reference point, essentially the physical component of the post-condition. *Example: a highway exit assertion may check that if the vehicle is indicating to leave at the exit lane, that it does in fact do so (i.e. checking for false or spurious indications).*

### B. Selecting relevant assertions

We have also found that it may be useful at times to identify whether an assertion is relevant at a given time step of the simulation, in order to improve the performance of the assertion checking system. For example, two particular reasons for pre-selecting assertions are:

- Some assertions may be applicable to mutually exclusive situations, and therefore should only be selected for their own relevant scenarios if one wishes to avoid confusing reporting of assertion results. For example, an assertion which checks that an AV is braking smoothly (for example when approaching a T-junction or traffic light indicating 'stop') will inherently fail in a situation where it is required to accelerate (for example when moving off after having been stationary). If assertions of this kind are not filtered out in scenarios where their checks are not pertinent to what may be identified as correct behaviour, they will generate nuisance failure results. Similarly, assertions related to motorway (highway) driving are not applicable to urban or rural road driving, so need not be applied (and may generate inappropriate or confusing results) while an AV is on any such road.
- Where assertions are being applied in a more time-critical mode of operation - for example for simulation control (see Section IV-D), online warnings, or for self-adaptation (see Section IV-E) - the speed of their execution can be increased by filtering out those assertions that are not relevant to the situation concerned at a given time (step). If the category type can be embedded in the assertion (for example, as a special field in its script), then the Assertion Checking process in the Testbench (see Figures 1 or 2) can be set to check only assertions of a given type, which can improve runtime performance.

Discrimination or pre-selection of assertions can also be achieved by specifying 'trigger conditions' based on the type of *Operating Design Domain (ODD)* in which the assertion is relevant.<sup>6</sup>

### C. Deriving machine-oriented logical assertions from human-oriented Highway Code rules

In this section we discuss how to develop assertions from the Highway Code. We discuss issues of interpreting the natural language of the Code in a consistent manner, and two approaches to constructing logical assertions from the text.

1) *Interpretation of Natural Language UKHC Rules:* The UKHC comprises several hundred rules and guidelines, applicable to different categories of road user (including those on foot - pedestrians, cyclists, horse riders - as well as the usual cars, vans, or trucks and buses) and covering a number of different aspects of road use (including general driving advice, instructions for maintenance and repair of

vehicles, and instructions for management of situations such as emergencies). The UKHC rules are written entirely for human readers, and seek to offer guidance in common sense terms that any reader should understand clearly. For example, two rules of central interest to this paper relate to overtaking manoeuvres:

#### Rule 162

**Before overtaking** you should make sure

- the road is sufficiently clear ahead
- road users are not beginning to overtake you
- there is a suitable gap in front of the road user you plan to overtake.

#### Rule 163

**Overtake only** when it is safe and legal to do so.

You should

- not get too close to the vehicle you intend to overtake
- use your mirrors, signal when it is safe to do so, take a quick sideways glance if necessary into the blind spot area and then start to move out
- not assume that you can simply follow a vehicle ahead which is overtaking; there may only be enough room for one vehicle
- *move quickly past the vehicle* you are overtaking, once you have started to overtake. *Allow plenty of room. Move back to the left as soon as you can but do not cut in*
- take extra care at night and in poor visibility when it is harder to judge speed and distance
- give way to oncoming vehicles before passing parked vehicles or other obstructions on your side of the road
- only overtake on the left if the vehicle in front is signalling to turn right, and there is room to do so
- stay in your lane if traffic is moving slowly in queues. If the queue on your right is moving more slowly than you are, you may pass on the left
- give motorcyclists, cyclists and horse riders at least as much room as you would when overtaking a car (see Rules 211 to 215).

There are several problems to resolve in converting rules of this kind into structured logical assertions of a kind usable by computer:

- 1) As these examples show, the rules are expressed in a second-person tense, offering advice directly to the reader. To convert them into assertions that are in effect performed (measured) from an external third-person perspective requires changes of perspective that may transform the logic of the rule to some extent.
- 2) The logical sense of the safety property expression (i.e. whether the rule expresses a pass or a fail condition) varies depending on the concept being expressed. Since the general guidance for direct conversion of rules is to keep as close as possible to the top level natural language, the logical sense of the assertions will follow suit. This contrasts with typical practice for assertions

<sup>6</sup>This is similar to typical practice in computer hardware verification [42], where an assertion may take the form "under <environment conditions> if <trigger> then <expectation>", where <environment conditions> may be "we are not in RESET mode" and <trigger> may be "we receive a request" and <expectation> may be "an acknowledge signal is driven high sometime in the X next cycles".



as used in conventional computer programming, where the standard practice is to encode an assertion such that a pass condition permits a program to continue running normally, and a fail causes an exception to be triggered.

- 3) Being written at a natural human-readable level of abstraction, the rules and guidelines rely greatly on the reader having sufficient background knowledge and capability to resolve their generalized constraints to the point where actions can be selected that satisfy them.

This abstraction also interferes with the ability to validate the assertions by any means other than manual design review; formalization of the assertions may require all the hidden complexity to be reintroduced.

- 4) Many clauses of UKHC rules offer advice to drivers about their internal decision making, often in the form of constraints. The third bullet-item of Rule 163 is a typical example. These clauses cannot be measured externally without some form of communication by the driving agent (for example, an AV) of the results of its decision-making processes. To date we have not attempted to establish assertions of this kind, as there are no standards or conventions by which an AV might explain and communicate its actions. However, we note that this is an interesting direction of future work (see Section VII-A).

Two methods have evolved for developing assertions from the Highway Code: direct translation into logic, and the use of modelling. These are discussed in the next subsections.

2) *Method 1: Direct translation into logic:* The basic approach involves simply converting the natural language text into a logical predicate that captures the entities described and their relationship to one another, by the following steps:

- 1) Identify and extract the natural language clause, for which an assertion is to be developed. A separate assertion will be required for every distinct subject-verb-object clause in the text (meaning that a given UKHC rule could easily require on the order of 5-10 separate assertions to be fully covered). Note that the clause must define a *testable* requirement - some UKHC rules are (at least at present) not testable. Refer to Section VI for further discussion of the testability of UKHC rules.
- 2) Write a natural language hypothesis that reflects the safety property intended by the UKHC rule. One useful approach to doing this is to phrase the natural language hypothesis interrogatively, i.e. as a question. Many rules are written imperatively, as orders or advice for road users to follow. The hypothesis is in effect an interrogative statement, asking whether or not a condition is true (this condition being the safety property to be satisfied).
- 3) Write a logical statement that reflects as closely as possible the relationships between concepts in natural language of the rule (albeit that changes of perspective may need to be applied). Define each high-level concept as a predicate and develop expressions for each in terms of variables in captured or derived data, or in terms of functions/ predicates of the PostGIS library. The concept-predicates will be implemented as the library functions discussed in Section III-C and Figure 3.

- 4) Verify the correctness of the above steps by manual design review. (As previously mentioned, this is pragmatically the only method available.)

As an example, consider the following extract from UKHC Rule 258:

- UKHC rule (Imperative instruction or advice):

The natural language statement is as follows:

**Rule 258: Red flashing lights.** *If red lights flash on a signal and a red 'X' is showing, you MUST NOT drive in the lane shown as closed beyond the signal.*

- Assertion Hypothesis (Interrogative question):

The assertion hypothesis (test condition) embodied in the rule can be captured as the following question:

*Does the AV pass a 'Lane Closed' overhead gantry signal?*

- Assertion Statement (Predicate logic):

Since the UKHC rule applies to roads with an overhead sign gantry, such as Automatic Traffic Management (ATM) motorways, the assertion is an Execution Condition that need only be applied when the AV is on such a road. The assertion question concept of passing a sign then becomes a test of whether the AV bounding-box geometry overlaps with a boundary line between the posts of the overhead gantry object, which must lie across the roadway. This could be converted into the following essential SQL query: <sup>7</sup>

```
SELECT
( time,
  ST_Overlap(
    AV.geom::geography,
    CurrentGantryLine::geography)
) ASC LIMIT 1
FROM simulation_data
WHERE
  simulation_data.agent_id
  = /* the AV's ID */ AND
  CurrentRoadType(agent_id)
  = 'ATM_MOTORWAY' AND
  CurrentLaneStatus(agent_id)
  = 'LANE_CLOSED' AND
  CurrentGantryLine
  = FindCurrentGantryLine(agent_id);
```

The key concepts of the assertion question become predicates in the assertion logic, and will be implemented as SQL procedural functions in the PostgreSQL database (refer back to Section III-C and Figure 3). It should be

<sup>7</sup>It should be noted the skeleton SQL schemas presented in this section are illustrative examples only and have not yet been tested experimentally. The experimental results presented in section V are based on a clause from UKHC Rule 162 on overtaking.

noted that there are library functions available to generate (or retrieve) the boundary line of the overhead gantry immediately ahead of the current position of the AV. A skeleton SQL schema to create such a function could be:

```
CREATE FUNCTION
  FindCurrentGantryLine (INT agent_id)
RETURNS
  geometry CurrentGantryLine
AS $$
  SELECT
    ST_Distance (agent_id::geography,
      GantryLines::geography)
  ASC LIMIT 1
FROM simulation_data
WHERE gantry line location is ahead
      of the vehicle
$$ LANGUAGE SQL;
```

We anticipate that typically assertion queries will make use of PostGIS relational operators and functions for object geometries, whereas the library functions will typically use PostGIS geometry constructors to set up the geometric objects for the assertions as variables derived from the captured simulation data (as discussed in Section III-C). A hierarchical organization of queries and procedural functions (from Assertions to CAV DBMS library functions to PostGIS functions to captured simulation data) emerges from this approach.

3) *Method 2: Model-based Analysis:* While many UKHC rules can be translated directly into logical conditions or constraints expressible as simple Boolean predicates, some describe more complex situations that may require more explicit modelling to identify the logical basis of the assertion hypothesis. We require no particular constraints on the modelling methodology used, except that a well formed logical expression deriving the essential parameter of the assertion hypothesis from the captured data stream should be produced as the result.

As a worked example, we present an assertion we developed from Rule 162, which we applied experimentally as discussed later in Section V. The particular clause within this rule that we studied is:

*Before overtaking you should make sure the road is sufficiently clear ahead.*

While this seems an apparently simple, innocuous requirement, the natural language actually encapsulates a complex judgement that a vehicle driver must make, at the moment the overtake manoeuvre commences. The driver must judge whether the distance between their vehicle and any oncoming vehicle in the adjacent road lane is sufficient to allow the manoeuvre to be completed safely, which is contingent on the (relative) speeds of the driver's vehicle (the AV), the VBP and the OV, the line-of-sight distances to limits of view or obscuring obstacles such as bends in the road, blind summits/dips, or road-side buildings, as well as the style of manoeuvre that the driver intends to perform - an 'aggressive' overtake manoeuvre

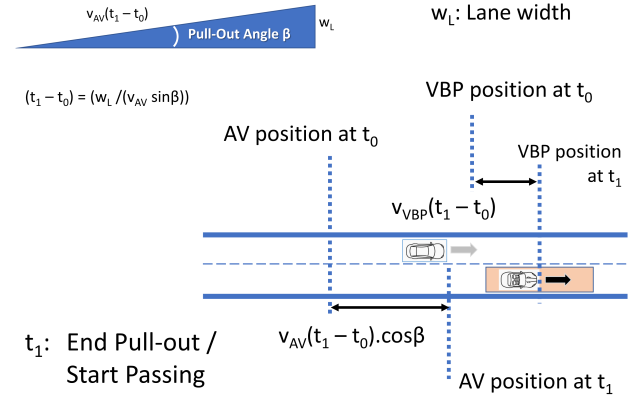


Fig. 6: Scene as the AV begins to pass the VBP, time =  $t_1$ .

can probably be completed in less distance than a 'relaxed' one [11, 43].<sup>8</sup>

In order to produce a measurable assertion it was necessary to develop a dynamic model of the overtaking manoeuvre to derive formulae for the road distance consumed by performing various stages of the manoeuvre. The model divides the whole overtaking manoeuvre into three distinct phases: Pulling Out (into the oncoming lane), Passing the VBP, and Cutting In (to the original running lane), and defines four time points (way points) at which the distances covered by the vehicles involved are measured. Ultimately, a formula was developed for the overall distance required to complete the manoeuvre without violating safety constraints (such as impingement upon vehicle thinking or braking distance (see Figure 4) or clearance distances around the VBP).

A diagram of the full analysis model, and the derivation of the formula for Safe Distance Ahead, is provided in Figure 11 in Appendix A. For the purposes of describing the general approach we describe the analysis of one of the way points as a typical example.

The second waypoint is set at the transition from pull-out to passing the VBP, as shown in Figure 6. The figure shows the expressions for the distance travelled by all vehicles in the  $[t_0, t_1]$  time period (taking into account aspects such as the movement of the AV into the oncoming traffic lane):

The overtaking manoeuvre analysis in Figure 11 derives expressions for distance covered by each vehicle in the time periods between each waypoint, and hence a formula for the Safe Distance Ahead at time  $t_0$  can be constructed by summing all the distances covered by the AV and the OV

The analysis identifies several key characteristic parameters that characterise the overtaking manoeuvre. These include:

<sup>8</sup>The internal plans and intentions of an AV may not be available to the external validation system, so the definition of the assertions may need to make default assumptions about its intentions, or make a 'worst case' test that checks against the most conservative case that can be assumed. For example, using the case of Overtaking, the assertion could assume that the AV was making a 'relaxed' overtake manoeuvre, in which case an assertion might register a safety violation for anything less than large separation distances between the AV and oncoming vehicle at the start of the manoeuvre, even though the manoeuvre could probably be completed safely if a more urgent profile were performed.

- *Stopping distance* (SD) is defined as the sum of the ‘thinking distance’ and ‘braking distance’ of the vehicle (assuming it has a human driver) as defined in UKHC Rule 126 and illustrated in Figure 4. A table of the typical stopping distances for travelling speeds ( $v$ ) between 20 mph - 70 mph is provided with Rule 126 in the UKHC. Equation 1 is a regression formula derived from this table to estimate the stopping distance ( $m$ ) given the vehicle speed ( $mph$ ).<sup>9</sup> The coefficients are:  $a = 0.300$ ,  $b = 0.058$ ,  $c = -0.011$  and  $d = 0.015$ .

$$SD = \underbrace{av}_{\text{Thinking Distance}} + \underbrace{b + cv + dv^2}_{\text{Braking Distance}} \quad (1)$$

- *Danger Space* (DS) is an area projected forward in the direction of travel with a length equal to the stopping distance and a width equal to that of the vehicle. SD varies with vehicle speed, making the DS a function of speed, therefore intuitively as the speed increases the DS ahead of the vehicle will project further forward accordingly.
- *Safe Distance Ahead* (SDA) is the result of the formula, as identified in the first item of UKHC Rule 162, which calculates how far away the Oncoming Vehicle needs to be from the AV for an overtaking manoeuvre to be conducted without risk of collision. The formula for SDA incorporates the DS of the OV.
- *Pull-out Clearance* distance is the minimum safe separation that must be maintained as the AV pulls out into the oncoming lane (and carries its own constraints (and assertions) as defined in the first bullet-point of Rule 163 - see Section IV-C1).
- *Pull-out Angle* ( $\beta$ ) is the steering angle taken by the AV as it moves into the oncoming lane.
- The *Cut-in Clearance* distance is the minimum required separation distance between AV and VBP when the AV begins to cut back into the running lane.
- *Cut-in Angle* ( $\theta$ ) is the steering angle taken during the cut-in move.

The ensemble of parameter values can therefore define a *driving profile* for the manoeuvre, which can informally be thought of as a level of urgency or aggression. An ‘aggressive’ profile might be taken as one which has a low pull-out clearance, high pull-out angle, low cut-in clearance and high cut-in angle, indicating a ‘sharp and close’ manoeuvre taken by the AV around the VBP. A more ‘relaxed’ driving profile may have larger clearances and lower steering angles. Various profiles of manoeuvre with different characteristic values are assessed in the experimental work discussed later. It should be noted that other UKHC rules place (at least qualitative) restrictions on these parameters to ensure safe driving behaviour (see the fourth bullet-point of Rule 163 in Section IV-C1, and also Section V-B).

As with the problem of direct translation of the UKHC rules, the approach aimed to develop a model using the

concepts of the natural language of the rule rather than by other parameters, for example time-to-collision, which has been found in human factors studies to be the most significant parameter characterising overtaking manoeuvres [28, 8]. Hence the model developed for the Safe Distance Ahead measurement differs to many that have been developed for analysis of overtaking manoeuvres, which aim to capture the decision making of the driver rather than the characteristics of the legal code of behaviour that is essentially an independent observer’s perspective on the scenario.

4) *Geospatial Database for Assertion Checking*: Analysis of the overtaking manoeuvre problem fits neatly into the geospatial database framework for assertion checking. The vehicles can be represented as shape objects using the postgis system extension which allows for logical tests and measurement metrics to be efficiently calculated between shapes. An Entity Relation Diagram (ERD), Figure 7, shows the variables within the database structure and the relation between them.

For each log file there exists an environment table containing map information, which in this case is based on the openDrive<sup>10</sup> and lanelet formats. From the lanelet data we derive a table containing useful entries such as lane width and orientation for the environment. Captured data from the simulation (or real-time monitor), contains a state vector for the actors (vehicles in this case) in *data\_capture.actor\_state* which contains position and orientation information. From this captured data table we can calculate dynamic state information, that is information that is not fixed but may change over time (e.g. velocity, acceleration) or with vehicle speed (e.g. braking distance).

Further insight to the driving situation can be derived when combining actor state information with information about the environment. For example, driving lane orientation (*lanelets.lane\_orientation*) can be used to calculate the angle of the vehicle relative to the driving lane to infer the cut-in or pull-out angles (*state\_dynamics.pull\_out\_angle*) required for the Safe Distance Ahead calculation. The final *assertions* table is where the Safe Distance Ahead test output can be written and stored as a Boolean pass/fail result.

#### D. Use of Assertions at Simulation Time

When using an assertion database as a test oracle in a simulation system (see Section III), we have identified four different modes by which assertions can be used:

##### 1) Simulation Control(Halting)

Assertions can be checked as the simulation runs, and if a major safety violation occurs (for example, a vehicle collision) then the simulation can be halted, as there is little value in continuing with a simulation beyond that point (unless the post-incident behaviour of an AV is being investigated). It may be a more efficient management of the testing process to stop and review the test failure, or proceed to the next test case, than to waste time (simulated and real) by letting a simulation continue.

##### 2) Operator Warnings

Similar to the above, the assertions can be checked as the

<sup>9</sup>Refer to <https://www.gov.uk/guidance/the-highway-code/general-rules-techniques-and-advice-for-all-drivers-and-riders-103-to-158>

<sup>10</sup><https://www.asam.net/standards/detail/opendrive/>

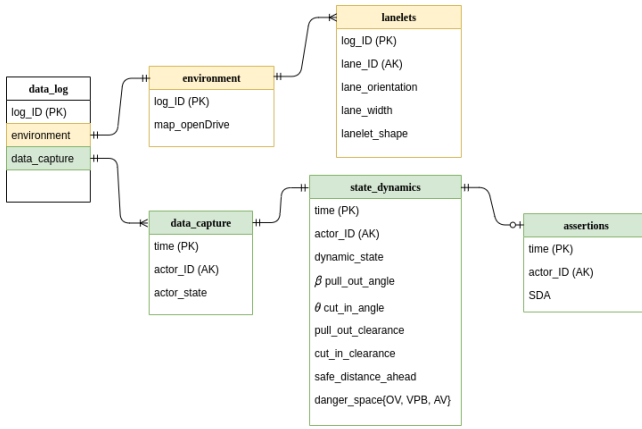


Fig. 7: Entity Relationship Diagram for the PostgreSQL database, for cardinality see crow's foot notation (See URL: [https://en.wikipedia.org/wiki/Entity-relationship\\_model](https://en.wikipedia.org/wiki/Entity-relationship_model) Accessed: 2021-08-17.)

simulation runs, but instead of halting the simulation if a safety assertion check fails, one can issue warnings to the simulation operator (if present) to draw their attention to the current situation of the simulation, to allow them to exercise their judgement on whether to continue. We anticipate that this may be the best policy for minor safety violations or events such as collisions which may lead to secondary effects like non-deterministic behaviour in the simulator [7], where the log file data might be degraded or unrepeatable following the event. The simulation may need to be halted, but pragmatically we believe this decision is probably best left for test engineers to decide. However, the use of assertions to provide operator warnings could be a beneficial mode of use.

### 3) Simulation Log Annotation

For negligible safety assertions failures, or for non-safety checks such as service performance monitoring, assertions could be checked as the simulation runs. But, instead of providing any automatic process control or operator warnings, any assertion failures could simply result in annotations being added to the simulation log data, indicating that an assertion failed at a given time step but otherwise having no other externally observable effect. The log can then be analysed retrospectively after the simulation run has finished.

### 4) Retrospective analysis

Instead of checking assertions as the simulation runs, the simulation captured data can be recorded and passed to the assertion checking database after simulation has finished, to analyse the AV's performance retrospectively. Assertion failures would probably then be presented in a display on screen to the analyst, showing the time steps in which they occurred and the dynamic state of the system at the (simulated) time.

The simulation experiments with assertions presented in Section V were all run retrospectively, as they were a simple demonstration of concept rather than full implementation of the above modes of use. In the ongoing development of the

simulation pipeline at Bristol Robotics Laboratory, we are focusing on the retrospective mode of use initially, as this represents the least complex software development option, and will then proceed to enabling dynamic checking during simulation time once this basic mode of operation has been demonstrated.

## E. Use of Assertions at Runtime

As for the use of assertions as an oracle at runtime monitoring, three modes were identified by which assertions can be used:

### 1) Warning

As discussed in the uses of assertions at simulation testing, assertions can be used to raise warnings for the human operator responsible of monitoring the vehicle during runtime to draw their attention or request them to take control over the vehicle when minor safety violations occur. For runtime, these warning messages importance may also extend for major safety violations, if handing the control to the human operator is the AV's risk mitigation plan for major violations. Additionally these warnings can be used to indirectly assess the performance of subsystems in an AV during operation to provide feedback for further development in future versions of the system.

### 2) Execution of risk mitigation plans

Often safety critical systems have automated pre-planned risk mitigation plans encoded within the system to be executed in the case of major violations or where high risk is possibly detected [29]. However, having a ground truth reference to judge on monitored data at runtime can be challenging. Assertions failures can provide a reliable verdict to help AVs execute appropriate risk mitigation plans.

### 3) Self-adaptation

The use of assertions during runtime may also be extended beyond the execution of pre-planned risk mitigation plans by combining it with key performance indicators (KPIs), safety performance indicators (SPIs) [26] or some scoring functions, as was shown in [23], to help the AV and autonomous systems more generally to actively learn and reliably adapt their autonomous functionalities during operation.

### 4) Accidents Investigation

In case of catastrophic accidents or just violations of explicit highway code rules, such as surpassing statutory speed limits, assertions can be used to annotate the internal status data logs of an AV to help identify why and how an accident or a violation occurred. This mode links to the application of *ethical black boxes* introduced in [48], which is a similar concept to flight data recorders, but for autonomous systems.

## V. EXPERIMENTAL CASE STUDIES

As an aid to developing the methods and techniques described previously, we created some worked examples of assertions and measurements applied both to simulation data and real-time video footage. The experiments were early



stages of work in developing the methodology, and were not developed by all the methods described in this paper, nor possess all the features described herein, but they did guide the development of the methodology and we believe that they do prove the general concept and utility of the approach. The use of methods such as modelling, and the need for features such as the use of standard procedures/library functions, all emerged as direct consequences of these experiments.

We chose two case studies inspired from a challenging overtaking scenario; the first aims at demonstrating the use of assertions in simulation testing, and the second shows how assertions can be used as SPIs or monitors to assess the safety performance of an AV or other road users during runtime.

#### A. Experimental Scenario

To demonstrate the functionality of the assertion based safety validation process, we chose to consider an overtaking scenario based on the driving situations from the NHTSA<sup>11</sup> pre-crash typology [46] that were used in the CARLA Challenge<sup>12</sup>. The corresponding UKHC rules [44] that apply to this scenario are Rules 162-169.

Specifically, the scenario involves overtaking on a single-lane road where the test vehicle must perform a lane change to avoid a parked or broken down vehicle, Figure 8. The vehicle must leave sufficient gap to the vehicle it is overtaking and also not endanger any vehicles that are oncoming.



Fig. 8: Overtaking scenario from NHTSA [46]

This scenario is difficult from the side of the autonomous vehicle as there are competing goals firstly to ensure safety and avoid accidents but also to make reasonable progress along the road network given the conditions, i.e. maintain *liveness*. There is also significant challenge in validating the safety of the vehicle in such a scenario, whether in real-world trials or in a simulated environment. For the overtaking scenario several parameters need to be monitored in order to check each assertion.

In the simulation and the real-time video analysis experiments, we chose to assess clauses in two different UKHC overtaking rules. The simulation study developed an assertion for the first bullet-item of Rule 162 (as discussed in Section IV-C1 and also in Section IV-C3), developing an assertion

check to determine whether a Safe Distance Ahead existed for an AV performing an overtake manoeuvre on a single-lane road, as it began to cross the road centre line at the start of the manoeuvre. The real-time safety validation study developed video analysis measurements that would support assertion checks against the fourth bullet-item of Rule 163 (see Section IV-C1; the relevant clauses are in *italics* for ease of identification) - while the measurements were not developed fully into assertions in SQL code, the study does capture the information essential to the assertion hypothesis for that UKHC clause (the Cut-in Clearance distance as discussed in Section IV-C3), and would be generated as a derived variable in the database using a standard procedure function as discussed previously.

#### B. Simulation case study

In the simulation case study we developed an assertion check for Safe Distance Ahead at the start of overtaking, as discussed earlier in Section IV-C3. We ran the simulation using a map segment of road network of Bristol, UK (part of an AV test route being considered for the ROBOPILOT project<sup>13</sup>).

1) *Environment model*: The map segment used for the simulation case study is shown in Figure 9 (and also Figure 4). We ran the simulation on one of the straight sections of road, for the sake of simplicity, i.e. avoiding having to take considerations such as obscuring objects or limited lines of sight into account. The road layout model was developed using the Road Runner tool from an OpenDrive environmental model developed from Open Street Map (OSM) data, obtainable freely from UK government sources (Ordnance Survey).<sup>14</sup>

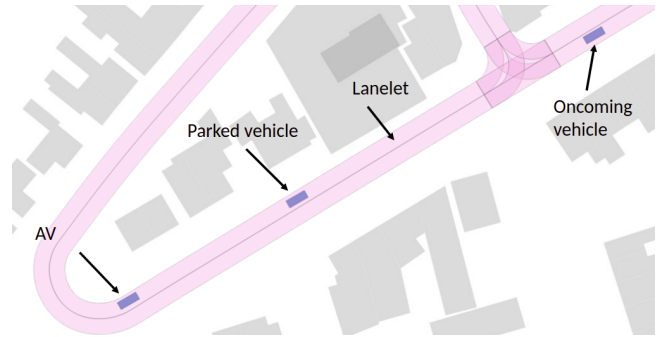


Fig. 9: Lanelets overlaid onto road network.

As discussed in Section III-C, lane polygons are required for efficient measurement of vehicle lateral movement within, and partial occupancy of, road lanes; in this simulator case study we used a lane partitioning concept termed *lanelets* [5], which allows exactly this level of description of the scene. We used an open-source software conversion package [3] to generate the lanelet data from the openDrive [14] description of a road network and divides roads into individual sub-lanes. This data was imported into the PostgreSQL database as a table of lane segment polygon objects (*lanelets.lanelet\_shape*) covering the whole of the road layout used for the study. Additionally,

<sup>11</sup><https://www.nhtsa.gov/>

<sup>12</sup><https://carlachallenge.org/>

<sup>13</sup>See URL: <http://www.bristol.ac.uk/tsl>

<sup>14</sup><https://os.openstreetmap.org/>

information about the lanes boundaries and orientations, required to identify road orientations and running lane directions, was appended to this data. Figure 9 shows an example of the lanelet shape object overlaid onto a road network. The blue boxes represent the dynamic actors, which in our case are the vehicles involved in the overtaking scenario. All of the above world map information was pre-loaded statically into the database, for use by the SQL query engine running the assertions queries, as discussed in Section III-C and shown in Figure 3.

2) *Overtaking Profiles*: The design intent of the assertion we applied to the simulation data was to assess whether or not the distance ahead of the simulated AV was sufficient for it to complete an overtaking manoeuvre safely. But in practice, the sufficiency of the distance ahead is dependent on the driving style (see Section IV-C3) that the AV controller has selected to perform. Also, it is always possible for the AV to change its profile in mid-manoevre, thereby shortening the remaining distance required, which can be applied as an emergency action if the risk of the situation should increase for any reason. So, the assessment whether there is a safe distance ahead in any given situation at the start of an overtake manoeuvre is dependent in part on the internal plans of the AV at the time, and must be made without full knowledge of exactly how the manoeuvre will play out after the Safe Distance Ahead assessment has been made. To reduce any such risks as far as reasonably practicable in this situation, the assessment must therefore make conservative assumptions about the manoeuvre, at least if it is to be applied as an execution condition (as discussed in Section IV-A; other post-condition assertions could be developed in principle to identify errors retrospectively, although this is not the intent of the first bullet-item of UKHC Rule 162).

In our experiment, we developed three different variations of the Safe Distance Ahead assertion, each calculating the safe distance ahead using the formula developed from the modelling analysis (see Section IV-C3), but which assumed three different driving profiles corresponding to three different levels of urgency of overtake manoeuvre (called ‘relaxed’, ‘hurried’ and ‘aggressive’). The levels of urgency were based loosely on similar concepts developed in other Human Factors research [8, 47], which had equated level or degree of urgency with *Time To Collision* (TTC). These sources identified that a TTC of approximately two seconds corresponds approximately to a medium degree of urgency, which we have labelled ‘hurried’.

3) *Simulator system details*: The simulator used for the case study was a minimal version of the generic simulator testbench architecture described in Section III-A. The system used the CARLA simulation software [13], an open-source autonomous driving simulation environment. Dynamic state data generated by the simulator was passed to a PostgreSQL database as shown in Figure 3, with the SQL assertion script being applied in retrospective analysis mode as discussed in Section IV-D.

The paths followed by each simulated vehicle (AV, VBP, OV - see Figure 4) were manually designed to follow a series of way-points. Using fixed vehicle paths facilitates evaluation

of the assertion, as it made the test repeatable<sup>15</sup> and allowed us to see how different characteristic parameters affected the assertion checking (see Section V-B5 for details).

CARLA generates dynamic state information about the simulation actors (vehicles), which is made available via an API, and Python scripts were written to transfer data from the simulator into the PostgreSQL database on a step-by-step basis.

4) *Simulations*: Three simulation runs were executed to generate simulation data traces for assessment by assertion checking. In each case, the trajectory of the AV remained unchanged, overtaking the stationary VBP and cutting back into the running lane with the same pull-out and cut-in clearance distances each time. The OV trajectory simply moved along the oncoming lane at constant speed, but the starting position of the Oncoming Vehicle was varied so as to generate three different scenarios of i) a safe overtake manoeuvre, ii) a near miss and iii) a dangerous overtake manoeuvre resulting in a collision. The simulation data was logged and transferred into the assertion database as three distinct subsets of samples within the main captured data tables. These were then analysed retrospectively after the end of the simulation run, using assertions designed to assess whether (and how well) the AV performed overtake manoeuvres of differing profiles (as discussed in Section IV-C3). These results are discussed below.

5) *Simulation Assertion Checking Results*: In this section we discuss the outcomes of the simulation safety validation analysis, and the implications on the methodology of developing and using assertions.

The three simulation runs described previously produced simulation data logs that were subjected to assertion checking to ensure that a safe distance ahead existed as the AV began to cross the road centre line into the oncoming lane at the start of the overtake manoeuvre. However, there are several factors conditioning the test results that need to be taken into account.

Comparison of the three variations of test against the simulated scenarios (safe, near miss and collision) produces nine distinct tests.

The fixed overtaking manoeuvre trajectory that had been pre-set for the AV vehicle approximated to a medium level of urgency, in terms of the pull-out and cut-in clearance distances achieved (about 2s time to collision at the simulated speeds of the vehicles). This meant that for a set of tests having a range of urgency levels centred on the medium urgency case, one should get a sliding scale of test case passes and failures as the test case variants are applied to decreasing initial separation distances between AV and OV in the three simulation logs. The results of the tests are shown in Table I, which shows the required separation distances at the start of each manoeuvre (determined by the modelling analysis formula described in Section IV-C3), the Distance Ahead that was achieved in each simulation run, and thence the pass/fail assertion results of comparing the two distances in each case.

<sup>15</sup>Test repeatability is needed to ensure trust in the simulated assertion results and also to ensure software bugs are found and fixed efficiently. Repeatability relies on deterministic simulation code and hardware. If the *simulation variance* is within tolerance then verification coverage results are stable and if an assertion fails the test can be re-run to provide the same output as described in [7].

Scenario	Achieved Distance Ahead (m)	Distance Ahead required by profile (m)		
		'Relaxed'	'Hurried'	'Aggressive'
		101.39	63.73	40.02
Safe	76.43	FAIL	Pass	Pass
Near miss	58.33	FAIL	FAIL	Pass
Collision	35.63	FAIL	FAIL	FAIL

TABLE I: Pass/Fail Assertion results.

The table shows that it would have been inadvisable to perform a 'relaxed' overtake in any of the three scenarios presented, as this would have resulted in a near miss or a collision. If a 'hurried' overtake is attempted (which actually corresponds to the behaviour of the AV in the scenarios), this policy was acceptable for the 'Safe' scenario, but inadvisable for the 'Near Miss' and 'Collision' scenarios (as their descriptive names imply). And if the AV had been programmed to overtake aggressively in these scenarios, then only in the case of the Collision scenario circumstances would the Distance Ahead have been short enough for the manoeuvre to be inadvisable; in the other two cases the overtake could have been achieved successfully (although aggressive overtaking may result in other safety assertions being violated).

### C. Runtime case study

In the runtime case study we developed an assertion monitoring for checking safe stopping distance during the overtake. To demonstrate the use of assertions at runtime monitoring of an AV, in the absence of a fully functioning AV controller, we used manual video analysis in combination with a pre-trained object classifier to extract data from a dashboard video camera of a dangerous overtaking vehicle.<sup>16</sup> The aim of the assertions applied to runtime data was to assess whether the AV risks colliding with any of the other vehicles involved in the scenario during the overtake and in turn take an informative decision on aborting or continuing the overtake. Such situations are very likely to happen if road layouts are more complex, vehicles change speed or if for any other reason the SDA calculation at the start of the manoeuvre gives misleading estimates.

1) *Danger Space (DS) Assessment*: The overtake manoeuvre can be split into three distinct stages as was described in the modelling analysis (Section IV-C3): pulling out, passing VBP and cutting in. During the different stages of the overtake, none of the vehicles should be present in the DS of another vehicle nor should the DS of one of the vehicles overlap with the DS of another. To satisfy this requirement, six assertions are needed to be checked at the different stages of the overtake. The subscript DS will be used to refer to a vehicle's danger space.

The first four assertions are applicable throughout the three stages of the overtake, these are:

- i) the VBP should not be in the AV danger space at anytime during the overtake,
- ii) the OV should not be in the AV<sub>DS</sub>,
- iii) the AV should not be in the OV<sub>DS</sub>,

iv) there should be no intersection between the AV<sub>DS</sub> and OV<sub>DS</sub>.

There are two key SPIs applicable for the last stage of the overtake (cutting in):

v) the AV should not enter the VBP<sub>DS</sub> and

vi) there should be no intersection between the AV<sub>DS</sub> and the VBP<sub>DS</sub>.

2) *Video analysis*: Usually well trained machine learning models are used with stereo vision and information from other sensors, like LIDAR, odometers etc, to estimate accurately the position, orientation and velocities of the AV and surrounding static and dynamic objects. However, due to the absence of a fully functioning AV, we used the following approach to approximate the processing of the real-time data stream for the runtime case study:

First, a pre-trained object classifier, YOLOv3 [36], was used to provide locations of vehicles in each frame and how much area each vehicle occupies of the frame. The separation distances between the AV and the other vehicles are the SPIs at the various stages of the overtaking manoeuvre. The longitudinal distance ( $s$ ) of a vehicle along the road from the AV can be estimated using Equation 2:

$$s = \frac{c \times W}{w} \quad (2)$$

where  $c$  is a constant capturing the focal length of the camera and is found empirically,  $W$  is the actual width of a vehicle in meters, and  $w$  is the width of the box detected by YOLOv3 in pixels.

Second, the lateral distance  $d$  of the AV from the middle white line is estimated using Equation 3:

$$d = \frac{w_{L(real)}}{w_{L(pixels)}} \times d_{pixels} \quad (3)$$

where  $w_{L(real)}$  is the real lane width measured in meters,  $w_{L(pixels)}$  is the lane width measured in pixels and  $d_{pixels}$  is the lateral distance of the AV from the middle white line in pixels.  $d_{pixels}$  is determined by simply finding where the middle white line is compared to the middle of the frame. The VBP and OV are assumed to have a fixed lateral distance from the white line separating lanes.

Third, the pull out  $\beta$  and cut in  $\theta$  angles of the AV are calculated using the change in lateral distance and change in the longitudinal distance measured.

Fourth, as we were not able to measure vehicle speed directly from the video, we based the DS calculation on the worst case scenario; that is all vehicles were traveling at the speed limit of the road. The speed limit of the single carriage way in the video was found to be the national speed limit. According to Rule 124 in the UKHC the national speed limit for cars (AV and OV) is 60mph, and 50mph for goods vehicles (VBP).

Using the above steps the real time video was converted to a 2D driving scenario, where the danger space assertions listed in section V-C1 can be checked.

3) *Runtime Validation Results*: Table II shows the results for the six assertions introduced in section V-C1 at the three different stages of the overtake. At pulling out, the assertion checks suggest that it was safe to continue the overtake

<sup>16</sup>Link to video used in analysis: <https://youtu.be/Gzi4X7WZkRI>.

manoeuvre as non of the vehicles were present in another vehicles danger space and there was no overlap in any of the danger spaces.

When passing the VBP the assertion checks indicate that the overtake should have been aborted or if it was aborted then it should have been aborted earlier during passing the VBP stage as both the AV and the OV have entered each others danger spaces.

Finally at the last stage of the overtake, the assertion checks suggest that the AV has attempted to cut in before it has completely overtaken the VBP, whilst still failing the assertions relating to the OV. However, the AV has passed the assertion of not being in the  $VBP_{DS}$ , which suggests, given all of the other assertions have failed, that it has not overtaken the VBP before cutting in.

Assertion	Overtaking manoeuvre stages		
	Pulling out	Passing VBP	Cutting in
VBP not in $AV_{DS}$	PASS	PASS	FAIL
OV not in $AV_{DS}$	PASS	FAIL	FAIL
AV not in $OV_{DS}$	PASS	FAIL	FAIL
No $AV_{DS} \cap OV_{DS}$	PASS	FAIL	FAIL
AV not in $VBP_{DS}$	N/A	N/A	PASS
No $AV_{DS} \cap VBP_{DS}$	N/A	N/A	FAIL

TABLE II: Pass/Fail danger space assertion checks results.

## VI. DISCUSSION

### A. Assertion Checking to measure Performance

While we have applied the assertion checking methodology to measure the safety properties of vehicles in simulation and at runtime, the methodology can be extended to other important system engineering aspects as well. This may include performance based metrics such as smooth driving (minimal jerk), accuracy of control (such as parking or stopping precisely at a bus stop), and decisive behaviour at junctions or roundabouts, sometimes termed *liveness* [24]. In particular, the mission performance of a vehicle, for example, can be measured in journey time or efficiency of movement through traffic, which may affect not only passenger expectations but also the commercial viability of such an autonomous system.

Performance (achievement of task goals) and safety (avoidance of task hazards) are distinct behavioural properties of a system; achievement of one does not entail achievement of the other. A task controller must achieve satisfactory compliance with both properties if a dependable system is to be developed. In some scenarios, such as during overtaking, they may exist a competition within the controller to satisfy both the achievement of goals and the avoidance of hazards.

For example, an autonomous taxi may need to perform a minimum number of fare-paying journeys in a day in order to remain commercially viable. Performance assertions can measure whether suitable opportunities for manoeuvres such as overtaking were taken efficiently, without incurring delays or driving sub-optimally for the given road network i.e. using all available space. However, it must still drive safely as it conducts each journey, so in situations that may induce journey

delays, such as high traffic congestion, optimum performance and safety may not be mutually achievable. Assertions can measure how the AV trades off these two properties in the execution of its task(s).

An analysis for overtaking performance can be observed in the simulation results presented in Section V-B by comparing the required SDA to the actual distance ahead of the Oncoming Vehicle as listed in Table I. In this previous example, a failure of the SDA assertion indicates a violation of the UKHC and this happens when the actual distance ahead is less than the required Safe Distance Ahead, i.e. intersecting the danger space of the Oncoming Vehicle. Considering this example from the point of view of performance, the AV must not violate the safety property but must overtake within a reasonable window of opportunity ( $\delta$ ), i.e. taking any presented opportunity to overtake if safe to do so and if appropriate to the road network. We define  $\delta$  as the ratio of ‘achieved distance ahead’ to the ‘SDA required by profile’ from Table I. The size of this *opportunity window* will be dependent on a myriad of different contextual and preferential conditions such as the driving style, vehicle capability, passenger comfort, etc.

Table III shows the classification for the overtaking scenario but from the perspective of performance rather than safety. There are three conditions in Table III describing the distance to the Oncoming Vehicle with respect to required SDA and the opportunity window.

- 1) Unsafe:  $d_{OV} < SDA$
- 2) Higher Performance:  $SDA \leq d_{OV} \leq SDA.\delta$
- 3) Lower Performance:  $d_{OV} > SDA.\delta$

Here,  $d_{OV}$  is the distance to the Oncoming Vehicle, SDA is the required Safe Distance Ahead based on the model given in Appendix VIII and the ratio  $\delta$  is the optimal window of opportunity for the vehicle to perform the manoeuvre. For condition 1, the distance to the Oncoming Vehicle is less than SDA, so to not overtake is the only safe action and indeed to overtake in this condition would result in the SDA safety assertion failure. Safe driving behaviour whilst making timely use of the overtaking opportunities presented to the vehicle will see the vehicle progress through traffic most efficiently. This sentiment is captured in condition 2, where the vehicle is taking opportunities to overtake whilst maintaining a safe distance to the oncoming vehicle. In condition 3, the distance to the Oncoming Vehicle is greater than SDA plus the window of opportunity, where the vehicle has unhindered opportunity, i.e. no currently oncoming traffic. In this case not overtaking, although completely safe, will incur unnecessary delay to the journey and may be given a lower performance score. Overtaking in this condition would be the expected behaviour of the AV and as such could be assigned a neutral or small positive performance score.

The previous simulation results from the perspective of performance are shown in Figure 10. Here  $\delta$  is set to be equivalent to  $TTC=2.5s$  [39, 8, 47] for the purposes of illustrating this concept. The required SDA is shown as a function of the actual distance ahead for the three overtaking profiles. Unsafe test cases are shown as cross markers, all being within *Zone A* which is an area where the SDA assertion will be violated. At the edge of this area defining unsafe driving behaviour



Distance Ahead Condition	AV Does Not Overtake	AV Overtakes
1: Unsafe	safe	unsafe (assertion failure)
2: Higher Performance	safe but not taking opportunities	safe and taking opportunities
3: Lower Performance	safe but incurring significant delay	safe & expected behaviour

TABLE III: Performance classifications for deciding to overtake based on the distance ahead to the Oncoming Vehicle.

is a *safety boundary* (dashed line) being the edge of Zone A. Beyond the safety boundary exists the opportunity window that extends to the limit of the *performance boundary* (dotted line), and covers Zone B and Zone C. Zone B represents a conceptual *safety margin* or *buffer* between optimal performance and the safety boundary, and in reality this may be the limit of a near-miss condition, e.g.  $TTC=0.5s$ . Reward should be given for an adaptive controller that makes use of Zone C but does not enter Zone B, i.e. executes optimal overtaking given opportunity  $\delta$  but does not cause a near-miss event. Zone D can be considered safer than the other zones, i.e. the vehicle is giving lots of distance to the Oncoming Vehicle, but may be unnecessarily incurring delays and hence considered lower performance. An improvement could be made by moving from Zone D to Zone C without violating any safety property.

There is therefore a trade-off between safety and performance for optimal driving behaviour, where performance is improved by moving towards Zone B (but not passing the safety boundary), and safety is improved by moving towards Zone D (see large arrows in Figure 10). While we have observed this conceptual trade-off between safety and performance in our analysis, we also see evidence of this in the UKHC rules themselves. An example of this is exemplified in the 4th bullet point in Rule 163 (see Section IV-C1) where the overtaking rule states “Move back to the left as soon as you can but do not cut in”, which clearly suggests that performance will play a part in correct driving behaviour. There is also evidence of this trade-off between performance and safety, while the AV should complete the overtake maneuver “as soon as you can” (higher performance) but “do not cut in” (higher safety) we see these competing demands resolve in optimal driving behaviour analogous to that described by Zone C in Figure 10. This analysis considers only the overtaking scenario and in reality there will be many competing demands on the decision making process of a real controller.

### B. The Significance of Agent Intentions during Assertion Checking

The initial work presented in this paper has raised a number of consequent issues that require further investigation.

One significant topic is the extent to which assertions checking the UKHC require access to states of internal decision making within the autonomous vehicle being checked monitored or validated. Many UKHC rules are advisories to drivers or other road users to be aware of potentially hazardous situations, and to actively perceive, deliberate, and take special

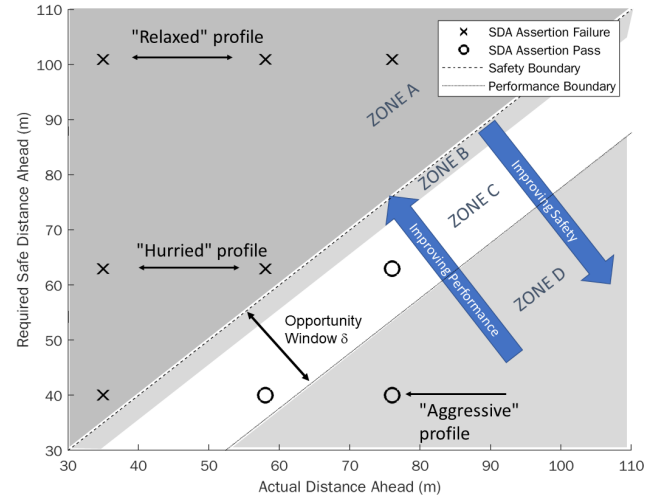


Fig. 10: Simulation results for performance.

or particular action to avoid harmful events. The following extracted rule is an example:

**Rule 146** *Adapt your driving to the appropriate type and condition of road you are on. In particular*

- *do not treat speed limits as a target. It is often not appropriate or safe to drive at the maximum speed limit*
- *take the road and traffic conditions into account. Be prepared for unexpected or difficult situations, for example, the road being blocked beyond a blind bend. Be prepared to adjust your speed as a precaution*
- *where there are junctions, be prepared for road users emerging*
- *in side roads and country lanes look out for unmarked junctions where nobody has priority*
- *be prepared to stop at traffic control systems, road works, pedestrian crossings or traffic lights as necessary*
- *try to anticipate what pedestrians and cyclists might do. If pedestrians, particularly children, are looking the other way, they may step out into the road without seeing you.*

Several bullet-items listed above ask the road-user to consider particular situations explicitly, advising that increased attention or effort be put into the perception of particular conditions, features of, or objects in the environment (e.g. blind summits, pedestrians, unmarked junctions). If we want to validate the correctness of an AV’s behaviour in respect of these rules, either in simulation or in runtime monitoring, we must have the capability to observe the decisions made by the AV or its control system as they are made.

This case highlights the need for *explainable AI*, a topic gaining increasing attention in the AI, robotics and autonomous systems community. The opinion is widely held [12, 33, 49] that it may be essential for any such system employed in a safety related application to be able to explain its decision making. The complexity and opacity of internal functional

representations in many AI technologies, combined with the non-deterministic nature of complex operational situations in real world environments means that they may not be repeatable in post-incident analysis without the system being able to report the precise situation that it perceived, and the planning or action-selection decisions that it made as a consequence. This requires some form of communication protocol or interface from the decision-making centres of the system to the outside world, so that the decisions can be observed and recorded for later analysis. This is a significant motivation behind the use of assertion checking oracle subsystems as a runtime analysis facility, discussed in Section IV-E.

At present, we are not in a position to perform assertion checking for this kind of safety property by means of an independent assertion monitoring oracle. There is no externally objective way to observe whether an AV agent is actually performing deliberative reasoning, planning or action selection consistent with rules such as Rule 146, unless it communicates explicitly its planning and decision-making through an interface, whose information can be read by the oracle for checking. A standard protocol for reporting decision-making information would need to be developed and at time of writing no such standard for doing this has been agreed. And given the commercially sensitive nature of such information, we do not anticipate that manufacturers would be willing to design systems that volunteer such information without legal requirements to do so and a standard protocol to use.

In the absence of any such explanation capabilities or facilities, it is still possible to achieve some degree of validation by performing checks derived from the text of the rule, which attempt to challenge the hypothesis expressed or implied by the UKHC rule or rule clause (e.g. *falsification testing* of the safety property). For example, in the testing phase of an AV, to exercise the last bullet item of Rule 146 above one could construct a test case, perhaps similar in nature to those defined in [17], where simulated ‘children agents’ (that model the physical appearance of children as well as some aspects of their behaviour) step out into a road in front of the AV agent, having appeared either to be looking away from an AV agent, towards it, or some mix of the two focuses of attention. Such a test suite could then test the capability of the AV control system to evaluate the potential behaviour of children to determine what driving action can be taken successfully and safely. This can test the ‘end-to-end’ situated behaviour of the vehicle and thereby evaluate the cumulative effect of decision making, but as evidence to support a safety case this is less effective than explicit confirmation (by recording) of the AV’s decision process that would come from Explainable AI.

## VII. CONCLUSION

The results obtained from both the simulation and the runtime video analysis case studies show the general feasibility and utility of assertion-based safety validation testing, with the assertions being derived from regulatory documentation written in natural language and intended for human drivers rather than written specifically to be usable by computer. The results provide our first indication that human-oriented natural

language rules of driving are an effective validation oracle for assessment of artificial systems.

The case studies were carried out at an early stage of development so we did not employ every element of the methodology that has emerged. Nevertheless, we believe that enough of the process has been demonstrated to show that the full methodology is viable. We also present the idea of using UKHC-based assertion checking methodology as a method to infer autonomous vehicle performance and have demonstrated this with an overtaking scenario using simulation results. We have demonstrated that a common UKHC-based assertion oracle is equally usable in runtime monitoring as it is in simulation.

## A. Future Work

In addition to the ongoing work of development of the BRL Simulator Testbench, we are considering numerous other directions in which to continue this work:

- The assertions database can be extended to other sources of information regarding safe driving behaviour of vehicles. This can include the rules and regulations of countries other than the UK (such as the German study described in [37]), or models such as the RSS model [40, 25].
- We are studying how the simulation testbench can be used to perform scenario-based validation methodologies such as falsification testing [9, 1] or *situation coverage testing* [2]. We are exploring how other scenario-based safety assurance techniques such as *Environmental Survey Hazard Analysis* [20] can be used to generate systematically comprehensive test objectives, scenarios and specific test cases, to produce a comprehensive safety validation methodology and supporting tool chain framework.
- We aim to investigate and extend the use of assertion checking to provide scores as a critic or reward function in reinforcement learning processes. Allowing for continuous improvement of the AVs balance between safety and liveness during operation.

## VIII. ACKNOWLEDGMENTS

This research has in part been funded by the ROBOPILOT and CAPRI projects. Both projects are part-funded by the Centre for Connected and Autonomous Vehicles (CCAV), delivered in partnership with Innovate UK under grant numbers 103703 (CAPRI) and 103288 (ROBOPILOT). This research was also supported in part by the UKRI Trustworthy Autonomous Systems Node in Functionality under grant number EP/V026518/1. Also special thanks to Séverin Lemaignan.

## REFERENCES

- [1] T. Akazaki, Y. Kumazawa, and I. Hasuo. “Causality-Aided Falsification”. In: *Electronic Proceedings in Theoretical Computer Science* 257 (Sept. 2017), pp. 3–18. URL: <http://dx.doi.org/10.4204/EPTCS.257.2>.

- [2] R. Alexander, H. R. Hawkins, and A. J. Rae. “Situation coverage—a coverage criterion for testing autonomous robots”. In: (2015).
- [3] M. Althoff, S. Urban, and M. Koschi. “Automatic conversion of road networks from opendrive to lanelets”. In: *2018 IEEE International Conference on Service Operations and Logistics, and Informatics SOLI*. IEEE. 2018, pp. 157–162.
- [4] G. Alves, L. Dennis, et al. “Formalisation of the Rules of the Road for embedding into an Autonomous Vehicle Agent”. In: (2018).
- [5] P. Bender, J. Ziegler, and C. Stiller. “Lanelets: Efficient map representation for autonomous driving”. In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE. 2014, pp. 420–425.
- [6] N. G. Bourbakis. “A traffic priority language for collision-free navigation of autonomous mobile robots in dynamic environments”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B Cybernetics* 27.4 (1997), pp. 573–587.
- [7] G. Chance, A. Ghobrial, K. McAreavey, S. Lemaignan, T. Pipe, and K. Eder. “On Determinism of Game Engines used for Simulation-based Autonomous Vehicle Verification”. In: *arXiv preprint arXiv:2104.06262* (2021).
- [8] R. Chen, K. D. Kusano, and H. C. Gabler. “Driver behavior during overtaking maneuvers from the 100-car naturalistic driving study”. In: *Traffic injury prevention* 16.sup2 (2015), S176–S181.
- [9] A. Corso, R. J. Moss, M. Koren, R. Lee, and M. J. Kochenderfer. “A survey of algorithms for black-box safety validation”. In: *arXiv preprint arXiv:2005.02979* (2020).
- [10] J. DeCastro, L. Liebenwein, C.-I. Vasile, R. L. Tedrake, S. Karaman, and D. L. Rus. “Counterexample-guided safety contracts for autonomous driving”. In: (2018).
- [11] J. DeCastro, L. Liebenwein, C.-I. Vasile, R. L. Tedrake, S. Karaman, and D. L. Rus. “Counterexample-guided safety contracts for autonomous driving”. In: (2018).
- [12] A. Deeks. “The judicial demand for explainable artificial intelligence”. In: *Columbia Law Review* 119.7 (2019), pp. 1829–1850.
- [13] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun. “CARLA: An open urban driving simulator”. In: *Conference on robot learning*. PMLR. 2017, pp. 1–16.
- [14] M. Dupuis, M. Strobl, and H. Grezlikowski. “Open-DRIVE 2010 and Beyond—Status and Future of the de facto Standard for the Description of Road Networks”. In: *Proc. of the Driving Simulation Conference Europe*. 2010, pp. 231–242.
- [15] Economic Commission for Europe Inland Transport Committee. *Convention On Road Traffic*. 1968.
- [16] K. Esterle, L. Gressenbuch, and A. Knoll. “Formalizing traffic rules for machine interpretability”. In: *2020 IEEE 3rd Connected and Automated Vehicles Symposium CAVS*. IEEE. 2020, pp. 1–7.
- [17] European New Car Assessment Programme. *Euro NCAP Vulnerable Road User VRU Test Protocol*, v-3.0.3. Accessed: 2021-10-20.
- [18] M. Gueffaz, S. Rampacek, and C. Nicolle. “Temporal logic to query semantic graphs using the model checking method”. In: *Journal of Software* 7.7 (2012), <http://www>.
- [19] J. Guiochet. “Trusting robots: Contributions to dependable autonomous collaborative robotic systems”. PhD thesis. Université de Toulouse 3 Paul Sabatier, 2015.
- [20] C. Harper and P. Caleb-Solly. “Towards an Ontological Framework for Environmental Survey Hazard Analysis of Autonomous Systems.” In: *SafeAI AAAI*. 2021.
- [21] C. J. Harper and A. F. Winfield. “A methodology for provably stable behaviour-based intelligent control”. In: *Robotics and Autonomous Systems* 54.1 (2006), pp. 52–73.
- [22] S. Huang and R. Cleaveland. “Temporal-logic query checking over finite data streams”. In: *International Conference on Formal Methods for Industrial Critical Systems*. Springer. 2020, pp. 252–271.
- [23] D. Kang, D. Raghavan, P. Bailis, and M. Zaharia. “Model assertions for monitoring and improving ML models”. In: *arXiv preprint arXiv:2003.01668* (2020).
- [24] K.-D. Kim and P. R. Kumar. “An MPC-based approach to provable system-wide safety and liveness of autonomous ground traffic”. In: *IEEE Transactions on Automatic Control* 59.12 (2014), pp. 3341–3356.
- [25] P. Koopman, B. Osyk, and J. Weast. “Autonomous vehicles meet the physical world: Rss, variability, uncertainty, and proving safety”. In: *arXiv preprint arXiv:1911.01207* (2019).
- [26] P. Koopman and M. Wagner. “Positive Trust Balance for Self-driving Car Deployment”. In: *International Conference on Computer Safety, Reliability, and Security*. Springer. 2020, pp. 351–357.
- [27] P. Koopman and M. Wagner. *Toward a framework for highly automated vehicle safety validation*. Tech. rep. SAE Technical Paper, 2018.
- [28] J. Lenard, R. Welsh, and R. Danton. “Time-to-collision analysis of pedestrian and pedal-cycle accidents for the development of autonomous emergency braking systems”. In: *Accident Analysis & Prevention* 115 (2018), pp. 128–136.
- [29] M. Leucker and C. Schallhart. “A brief account of runtime verification”. In: *The Journal of Logic and Algebraic Programming* 78.5 (2009), pp. 293–303.
- [30] L. Libkin. “Expressive power of SQL”. In: *Theoretical Computer Science* 296.3 (2003), pp. 379–404.
- [31] L. Masson. “Safety monitoring for autonomous systems: interactive elicitation of safety rules”. PhD thesis. Université Paul Sabatier-Toulouse III, 2019.
- [32] R. Myers and Z. Saigol. “Pass-fail criteria for scenario-based testing of automated driving systems”. In: *arXiv preprint arXiv:2005.09417* (2020).
- [33] S. OSullivan et al. “Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence AI and autonomous robotic surgery”. In:

- The International Journal of Medical Robotics and Computer Assisted Surgery* 15.1 (2019), e1968.
- [34] J.-H. Park and U.-Y. Huh. "Path planning for autonomous mobile robot based on safe space". In: *Journal of Electrical Engineering and Technology* 11.5 (2016), pp. 1441–1448.
- [35] H. Prakken. "On the problem of making autonomous vehicles conform to traffic law". In: *Artificial Intelligence and Law* 25.3 (2017), pp. 341–363.
- [36] J. Redmon and A. Farhadi. "Yolov3: An incremental improvement". In: *arXiv preprint arXiv:1804.02767* (2018).
- [37] A. Rizaldi and M. Althoff. "Formalising traffic rules for accountability of autonomous vehicles". In: *2015 IEEE 18th international conference on intelligent transportation systems*. IEEE. 2015, pp. 1658–1665.
- [38] A. Rizaldi et al. "Formalising and monitoring traffic rules for autonomous vehicles in Isabelle/HOL". In: *International conference on integrated formal methods*. Springer. 2017, pp. 50–66.
- [39] C. J. Robbins, H. A. Allen, and P. Chapman. "Comparing drivers' gap acceptance for cars and motorcycles at junctions using an adaptive staircase methodology". In: *Transportation research part F: traffic psychology and behaviour* 58 (2018), pp. 944–954.
- [40] S. Shalev-Shwartz, S. Shammah, and A. Shashua. "On a formal model of safe and scalable self-driving cars". In: *arXiv preprint arXiv:1708.06374* (2017).
- [41] A. Sinha, V. Vu, S. Chand, K. Wijayaratna, and V. Dixit. "A Crash Injury Model Involving Autonomous Vehicle: Investigating of Crash and Disengagement Reports". In: *Sustainability* 13.14 (2021), p. 7938.
- [42] Y. Tao. "An introduction to assertion-based verification". In: *2009 IEEE 8th International Conference on ASIC*. IEEE. 2009, pp. 1318–1323.
- [43] P. Tkachenko, J. Zhou, D. Gagliardi, and L. del Re. "On-line maneuver identification in highway traffic using elastic template matching". In: *IFAC-PapersOnLine* 51.15 (2018), pp. 557–562.
- [44] UK Driving Standards Agency. *The Official Highway Code*. Her Majesty's Stationery Office, 2012.
- [45] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer. "Defining and substantiating the terms scene, situation, and scenario for automated driving". In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE. 2015, pp. 982–988.
- [46] US Department of Transportation. *PreCrash Scenario Typology for Crash Avoidance Research*. <https://www.nhtsa.gov/sites/nhtsa.gov>. Accessed: 2021-10-20.
- [47] X. Wang, M. Yang, and D. Hurwitz. "Analysis of cut-in behavior based on naturalistic driving data". In: *Accident Analysis & Prevention* 124 (2019), pp. 127–137.
- [48] A. F. Winfield and M. Jirotko. "The case for an ethical black box". In: *Annual Conference Towards Autonomous Robotic Systems*. Springer. 2017, pp. 262–273.
- [49] R. H. Wortham. *Transparency for Robots and Autonomous Systems: Fundamentals, technologies and applications*. Institution of Engineering and Technology, 2020.
- [50] B. Xue, M. Fränzle, N. Zhan, S. Bogomolov, and B. Xia. "Safety Verification for Random Ordinary Differential Equations". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 39.11 (2020), pp. 4090–4101.



**Chris Harper** is a Research Fellow in Robotics Safety & Control at Bristol Robotics Laboratory, researching methods and technologies for safety assurance of autonomous systems, machine learning and AI. For over 20 years Dr Harper has worked in both industry and academia as a safety assurance specialist with experience in system/software hazard and safety analysis, safety cases and safety management, and has also been involved in development of international industry safety standards for robotics. His research interests include assurance methods, high integrity design, safety validation, and safety arguments/cases of autonomous systems. He has recently been working on the CCAV-funded R&D projects ROBOPLOT (autonomous parcel delivery), CAV Forth (autonomous bus service) and SafeMRX (requirements for CAV minimum risk manoeuvres) and on a project for the Assuring Autonomy International Programme (AAIP) on assistive robotics in healthcare.



**Greg Chance** received the B.Sc. (Hons.) and Ph.D. degrees in physics from the University of Bath, Bath, U.K., in 2001 and 2005, respectively. He was previously at the Bristol Robotics Lab and is currently a Senior Research Associate with the Trustworthy Systems Laboratory, University of Bristol, Bristol, U.K. His current research interests are simulation-based verification for autonomous systems and cybersecurity. He has ten years industrial experience researching for Oxford Instruments, U.K., and BAE Systems, U.K. Dr Chance is a Chartered Engineer and a member of the Institute of Physics and holds a patent for a novel plasma control method.

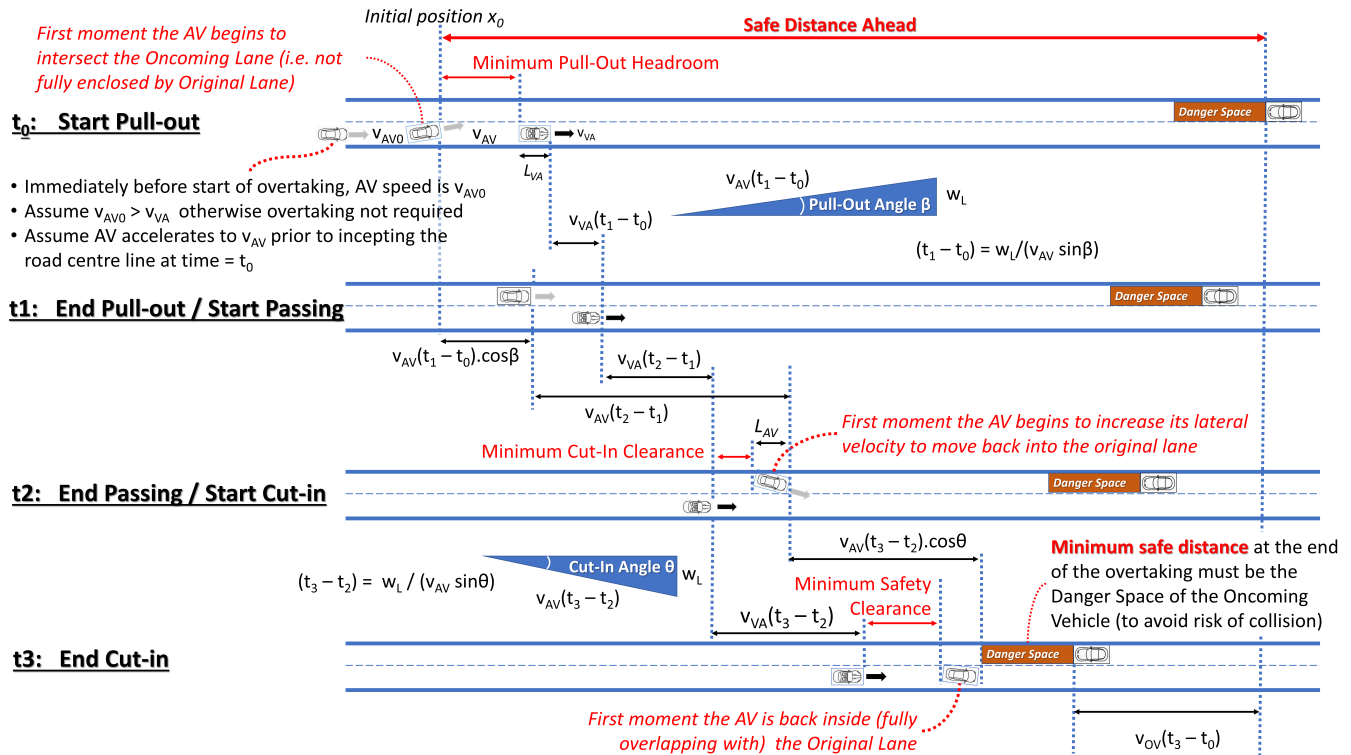


**Abanoub Ghobrial** received the MEng degree in mechanical engineering from the University of Bristol, Bristol, U.K., in 2018. He is currently pursuing the Ph.D. degree in computer science at the University of Bristol and is a part-time Research Associate with the Trustworthy Systems Lab, Bristol, U.K. From 2018 to 2020, he was a full-time Research Associate with the Trustworthy Systems Lab. His current research interests are techniques to allow self-managing of autonomous safety-critical systems via continual learning during operation; and the development of simulation-based verification techniques for autonomous systems.



## APPENDIX A

Figure 11 presents a model of an overtaking manoeuvre, showing the derivation of a formula for calculating Safe Distance Ahead, which was used in the simulation study results presented in Section V-B.



$$\text{Safe Distance Ahead} = \frac{w_L}{\tan \beta} + v_{AV} \left[ \frac{k + (v_{VBP} - v_{AV} \cos \beta) \left( \frac{w_L}{v_{AV} \sin \beta} \right)}{(v_{AV} - v_{VBP})} \right] + \frac{w_L}{\tan \theta} + DS(v_{OV}) + v_{OV}(t_3 - t_0)$$

where DS = Danger Space ( Thinking + Braking distances from UKHC Rule 126)

$$k = \text{Minimum Pull-Out Headroom} + L_{VA} + \text{Minimum Cut-In Clearance} + L_{AV}$$

Fig. 11: Schematic of Model-based Analysis of Overtaking Manoeuvre.