Arbitration Strategies in Multi-Channel Automated Driving Systems

January 31, 2024

Student: Adwaith Gopichand
Student ID: 1857304
Department: Mechanical Engineering
Section: Control Systems Technology
Master's Program: Automotive Technology

University Supervisors:
Emilia Silvas
Caspar Hanselaar
Eindhoven University of Technology

***Abstract*—As autonomous vehicles advance toward higher levels of automation, it necessitates a shift from traditional fail-degraded to fail-operational strategies. Consequently, ensuring system availability in the face of faults or functional insufficiencies (FIs) becomes increasingly safety-critical. This involves integrating diverse mission-continuing channels and sophisticated arbitration strategies. A critical challenge during this transition is balancing the safety and availability of the Automated Driving System (ADS) functionality when encountering an FI. While the Safety Shell arbitration strategy proves effective, its heavy reliance on computationally demanding predictions indicates a clear research opportunity for developing more efficient alternatives with reduced dependence on predictive models. Addressing this gap, this work introduces a novel arbitration strategy based on the Analytic Hierarchy Process (AHP) for multi-channel ADS. We benchmark this novel approach against state-of-the-art arbitration strategies through scenario-based comparisons. The results show that our proposed arbitration strategy offers a promising balance between safety, availability, and computational efficiency, serving as a viable alternative to existing strategies. This approach also acknowledges certain limitations, inviting further exploration and refinement to address these challenges in the future.**

***Index Terms*—Autonomous Vehicles, Safety, Availability, Multi-Channel Architecture, Functional Insufficiencies, Analytical Hierarchy Process**

I. Introduction

Automotive systems are undergoing a transformative shift from traditional mechanical systems to the integration of electrical systems, software, and mechatronics, paving the way for *Autonomous Vehicles* (AVs). AVs offer the promise of significant improvements in safety, comfort and mobility. Road traffic injuries stand as the foremost cause of death for children and young people aged 5–29 years, and rank as the 12th leading cause of death across all age groups as of 2019 [1]. AVs have the potential to reduce such road traffic accidents, particularly those caused by human driver errors. Beyond enhancing safety, AVs can contribute to increased comfort, improved energy efficiency, enhanced accessibility for the elderly and disabled, and reduced traffic congestion, facilitated by innovations like Mobility as a Service and platooning technologies [2], [3]. Given their potential to revolutionize people's lives and society, AVs have garnered substantial interest from companies such as Waymo, Cruise and AutoX, who are deploying and testing their robotaxi services [4]–[6].

Despite these advantages, the widespread adoption of AVs remains a challenge underscored by recent analyses that highlight significant safety concerns. Analysis of registered AV tests from 2014 - 2016 showed that AVs were 15 - 4000 times more likely to crash than Human-driven vehicles (HDVs) [7]. Another study showed AVs to be 5.2 and 1.3 times more prone to accidents in dawn/dusk or turning conditions [8]. Meanwhile, current advanced driver-assistance systems on the market, such as Tesla's Autopilot and their Full Self Driving (FSD) package, have revealed critical safety gaps. These systems, classified as SAE L2 systems [9], have shown an inability to consistently recognize and respond to hazards, as evidenced by the Tesla Autopilot's failure to detect a truck, resulting in a fatal crash [10]. Furthermore, safety tests by The Dawn Project on the FSD package revealed a range of dangerous behaviours, necessitating driver intervention in critical situations [11]. Moreover, these safety issues extend beyond Tesla, as similar concerns are observed in higher levels of automated driving, such as in SAE L4 systems [9]. Examples include GM's Cruise robotaxis failing to yield to an emergency vehicle [12] and Waymo robotaxis stopping unexpectedly in intersections [13]. These incidents underline the challenges of achieving a *"Positive Risk Balance"* [14] for AVs comparable to HDVs, indicating that there is still considerable progress to be made in ensuring the safety and reliability of these technologies.

Exploring the factors leading to hazardous AV behaviour is pivotal in addressing the challenges hindering their com-

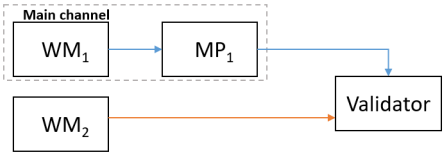

Fig. 1: Single channel Monitor-Actuator (or Doer-checker) architecture.

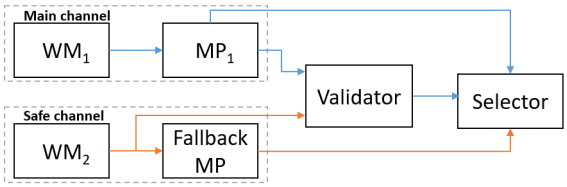

Fig. 2: Dual channel architecture with a main channel to perform the nominal function and a safe channel, as proposed by [19].

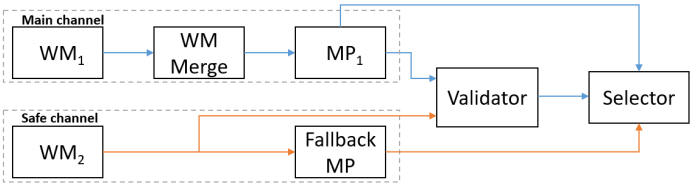

Fig. 3: Dual channel architecture with a merged WM, as proposed by [24].

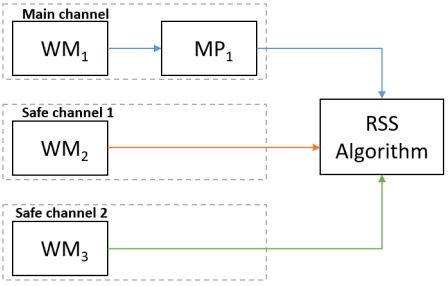

Fig. 4: Three channel architecture with a single nominal channel and 2 safety WMs, as proposed by [25], [26].

mercial deployment. The California Department of Motor Vehicles (DMV) publishes annual reports detailing *disengagements* occurring during AV test drives [15]. [16] performed a statistical analysis of the 2021 DMV report and found that about 70% of AV test disengagements are attributed to the ADS not performing their functionality as intended, leading to potentially hazardous situations. These insufficiencies, referred to as *functional insufficiencies* (FIs) [17], stem from both performance limitations of AVs' intended functionalities and insufficiencies in their specifications. These insufficiencies, if unmitigated, can lead to *output insufficiencies* (OIs) [17], impairing the ADS's ability to accurately interpret or respond to its environment. Traditional safety analysis and mitigation process, mentioned in ISO 26262 'Functional Safety' [18], provides guidelines on identifying and mitigating risks associated with system malfunctions or failures (e.g. communication failures, power system issues, software bugs, hardware faults). However, the emergence of new AV functionalities employing advanced perception modules and machine learning algorithms gives rise to novel *triggering* conditions, leading to additional safety hazards in automated driving (e.g., failure to detect road markings in adverse weather conditions, misreading traffic signals, inaccurately predicting other road users' behaviours, and generating impractical motion plans that lead to incorrect positioning or exceeding the vehicle's actuator capabilities). Addressing these hazards is crucial, as the presumption with traditional functional safety standards is that all software and hardware faults have been identified and mitigated pre-deployment [14]. The ISO 21448 'Safety of the Intended Functionality' (SOTIF)) standard [17] addresses these challenges by focusing on the safety risks associated with triggering events, ensuring a more comprehensive approach to AV safety.

Extensive research has been conducted on state-of-the-art methods to mitigate system faults. These methods focus on the redundancy and diversity of software and hardware components through N-version programming, recovery blocks, acceptance voting, sanity check patterns and watchdog patterns [20]. However, the mitigation of FIs has not received comparable attention. SOTIF [17] focuses on design time mitigation of FIs and emphasizes adapting vehicle functionality or sensor performance requirements or modifying the Operational Design Domain (ODD) [9], [21]. This adaptation occurs when FIs are identified during hazardous scenarios encountered in testing, enhancing system-level safety pre-deployment. Despite these measures, FIs continue to pose challenges in ADS post-deployment. This persistence is primarily attributed to large variations in real-world driving conditions, often characterized by edge cases. These scenarios follow a heavy tail distribution, making exhaustive testing for all rare events impractical [22], [23], highlighting the need for run-time FI mitigation strategies to maintain ADS safety.

A straightforward approach to run time mitigation of FIs is by using a single-channel architecture (see Section II for the definition of a channel). In this architecture, a single channel is tasked with processing sensor data to construct an environmental model, formulating driving trajectories, and subsequently sending setpoints to the actuators. This is referred to as a single-channel Monitor-Actuator design pattern (MA) [20], [22], shown in Fig 1. This architecture is composed of two heterogeneous subsystems: the 'doer or actuator', responsible for executing a function, and the 'monitor or checker', tasked with monitoring and validating the doer's output. The doer performs specific functionalities like processing sensor data for environment modelling to generating trajectories and actuator setpoints. In scenarios where the doer malfunctions, for instance, by generating hazardous trajectories, the monitor intervenes by shutting down the entire function, thereby transitioning the system into a fail-silent mode. The implementation

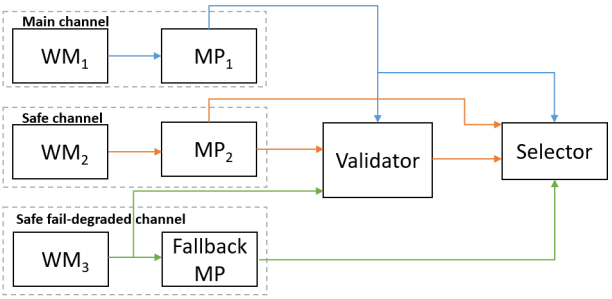

Fig. 5: Scalable fail-operational multi-channel architecture concept in the BMW ADS [27].

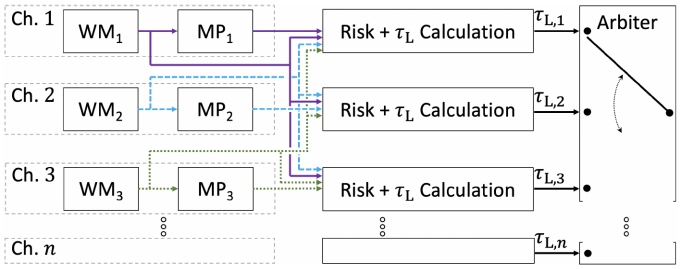

Fig. 6: Scalable multi-channel architecture, with each channel having mission continuing capabilities, as proposed by [32]. This figure is reused from [32] with permission.

of heterogeneous redundancy aims to mitigate risks associated with faulty doer behaviour; however, it also means a complete loss of functionality if something goes wrong, posing a significant challenge to maintaining consistent ADS availability (probability that the system keeps operating in the advent of a fault or FI). Ideally, a fail-silent system should not only cease its regular operation but also ensure that it transitions to a *safe state* before becoming silent. To ensure that the ADS remains operational for a minimal yet crucial duration when its standard functionality is impaired, a dual-channel architecture is proposed by [19]. This architecture introduces an additional channel, enabling the system to maintain fail-degraded functionality. This allows for fallback operations and enables the vehicle to reach a safe state. A safe state, as defined in [28], is an operating mode where the level of risk is not unreasonably high, known as a *Minimal Risk Condition* (MRC). A safe state is attained when the ADS executes a *Minimal Risk Manoeuvre* (MRM) [9], such as slowing down or pulling over to the side of the road. This architecture, shown in Fig 2, compares the nominal channel's trajectory with the world models to initiate MRMs when necessary. As described in [22], the architectural design pattern employing multiple heterogeneous ADS channels requires sufficient independence or diversity between the channels to prevent common cause and cascading failures. Diversity between the two channels can be achieved, for instance, by employing independent sensor modalities for each channel or by utilizing different software modules that maintain a sufficient level of independence [29]. However, this can lead to inconsistent outputs due to replica indeterminism [30], potentially causing false positives, leading to ghost object-triggered fallback MP, affecting ADS availability. [24] presents an architectural design pattern for systematic false positive FI mitigation, shown in Fig 3. This pattern involves merging the drivable space as determined by the main channel's WM with that identified by the safe channel's WM, resulting in a more conservative trajectory plan. Even if one of the ADS channels provides faulty data due to the presence of FIs, the merged output from the dual-channel architecture is generally considered safe, provided that at least one channel is functioning without a FI. However, it is important to note that this safety is contingent on the assumption that a more conservative trajectory plan resulting from the merger does not inadvertently introduce any additional risks or dangers. This architecture, while prioritizing safety, can result in more conservative motion planner behaviour, potentially degrading the overall driving experience. Mobileye's True Redundancy [25], [26] introduces an architecture employing diverse sensor modalities and integrated with the Responsibility-Sensitive Safety (RSS) algorithm, as depicted in Fig 4. The driving policy generated by the nominal channel is evaluated against the 2 safety-oriented WMs to ensure that the actuator setpoints from the MP adhere to the safe driving parameters defined by the RSS model. The RSS model creates a safe operational space, considering both the dynamics of the ego vehicle and the predicted dynamics of other road agents [26]. If the setpoints deviate from this RSS-defined safe operational space, they are replaced with actions generated by RSS to re-establish safety. The RSS model's notable limitation lies in its sensitivity to environmental variability and uncertainty, where handling unpredictable elements such as road conditions and the status of other vehicles can significantly constrain the vehicle's operational permissiveness [31].

The state-of-the-art ADS architectures mentioned above, [19], [20], [24]–[26], offer mechanisms to mitigate FIs during runtime, primarily through shutdown processes or fallback manoeuvres. While these strategies enhance system safety, they often compromise the availability of the ADS functionality, which is crucial for SAE L4 and L5 systems. To achieve fail-operational capability in AVs, multiple redundant mission-continuing channels with design diversity are essential to prevent common mode, cascading failures, and FIs [22]. BMW's implementation of a scalable fail-operational multi-channel architecture for its SAE L3 ADS is shown in Fig 5. This system comprises a 'main channel' and a 'safe channel', each tasked with generating mission-continuing trajectory plans. Operational control alternates between these channels based on their safety assessments, ensuring the availability of ADS functionality in the advent of a FI rather than just going to a degraded state. In the case of conflicting trajectories, the system defaults to a 'safe fail-degraded channel', facilitating an MRM [27], [33]. However, specifics regarding the arbitration mechanism, which governs how the ADS alternates

between the two mission-continuing channels in the event of a FI are not clearly defined. The Safety Shell [16], [32], a scalable architecture consisting of multiple heterogeneous mission-continuing ADS channels is shown in Fig 6. Each channel in this architecture has its own WM and MP functions. The diverse nature of the channels in the architecture decreases the risk of unhandled OIs and enhances ADS's availability, allowing a switch to a safe journey continuation channel in the event of a fault or a FI without compromising safety. The safety shell employs an arbitration strategy based on risk-based calculations, evaluating the MP generated by one channel against all available WMs. In its decision-making process, the Safety Shell arbiter accumulates risk and preference indicators for each channel. The risk indicator identifies potential trajectory hazards to maintain safety, while the preference indicator evaluates the suitability of each channel. It then uses the last safe intervention time to identify a last-time moment to either switch to an alternative channel or execute a timely evasive manoeuvre, enhancing overall vehicle safety and operational efficiency. Additionally, the Safety Shell includes a fallback MP to generate a safe trajectory if no channel provides a safe motion plan. However, the Safety Shell depends heavily on prediction and incurs substantial computational overhead in its arbitration process.

Research on alternate arbitration strategies for multi-channel mission continuing systems are limited. To address this gap, we introduce a novel arbitration strategy, coined the 'AA', employing the Analytic Hierarchy Process (AHP). The AHP's versatility and effectiveness in various complex decision environments are extensively detailed in [34], with its applications spanning critical areas such as decision-making in healthcare, urban planning resource allocation, strategic planning prioritisation, and software selection evaluation. AHP methodology has been used in this paper to develop a computationally efficient arbitration strategy and to balance multiple conflicting criteria. The strategy focuses on choosing a channel that minimises negative impact on safety and maximises the availability of mission-continuing channel choices. Additionally, we benchmark this novel arbitration strategy against state-of-the-art arbitration strategies using numerical simulations.

The rest of this paper is organized as follows. Section II provides a preliminary overview of key components within AVs, exploring the concept of an autonomous driving channel and detailing the general architecture of an AV. Section III identifies the existing gaps in the field and deals with the problem formulation. The novel arbitration strategy is introduced in Section IV. Section V elaborates on the methodology, including the simulation framework and test setup, and also covers test cases for benchmarking multi-channel ADS architectures. The benchmarking results obtained from the scenario-based comparison of the proposed arbitration algorithm against the arbitration strategies employed in the Safety Shell and RSS architectures are discussed and analyzed in Section VI. Conclusions of the paper with an outlook on future research directions are highlighted in Section VII.

II. Preliminaries

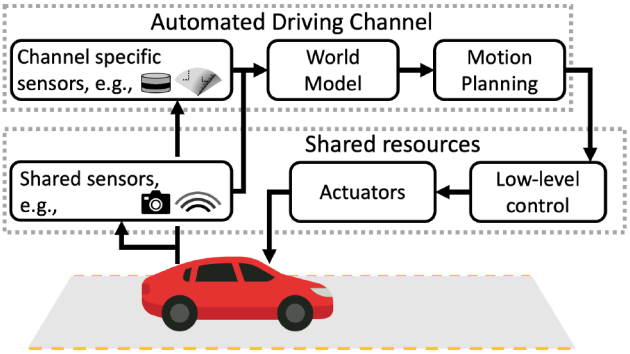

Fig. 7: An automated vehicle system architecture. A set of sensors captures environmental data, which is then synthesized by the WM to create an internal representation of the environment. The output from the WM is fed to the MP, which then selects a suitable driving strategy, ultimately executed by the vehicle's low-level controllers and actuators. This figure is reused from [32] with permission.

A simplified architecture of an ADS is shown in Fig 7, detailing its key components and structural layout. A multi-channel ADS is comprised of a series of parallel operating subsystems, commonly referred to as channels. Each channel combines a variety of sensor modalities—such as LIDAR, radar, cameras, and ultrasonic sensors tailored to specific functionalities within the channel. The data acquired by these sensors are further processed and utilized by two components within each channel: the World Model (WM) and the Motion Planning (MP) functions. The WM in an ADS represents the state of the external environment as perceived by the ego vehicle and integrates multiple subfunctions: sensor fusion, which combines data from various sensors to create a detailed environment model, identifying both static and dynamic road elements; localization, which is responsible for determining precise ego vehicle position; free space detection, which describes driveable areas; and state estimation and prediction, which involves anticipating future locations of the other road users and estimating their states over time. The data from the WM is fed to the MP, which comprises three subfunctions: mission planner, responsible for determining an optimal route to the destination, taking into account factors like distance, traffic conditions and road types; behavioural planner, responsible for decision-making, which generates manoeuvre specifications; and trajectory planner, which formulates an optimal driving strategy. Low-level controllers take the trajectory plans from the MP and translate them into specific control commands that are sent to the actuators (e.g., motor torque, brake pressure) to correctly execute the planned trajectory.

III. Problem Statement

For higher levels of automated driving (SAE L4 and L5), the availability of the system is also safety-critical, necessitating a shift from the traditional fail-degraded to fail-operational

strategies. This requires the integration of heterogeneous mission-continuing channels. However, switching the driving modes between multiple mission-continuing channels necessitate an effective arbitration strategy, as the overall measures of safety and availability need to be balanced. Relying solely on the most conservative channel can hinder progress in traffic, affecting availability, while prioritizing availability might compromise safety. Therefore, there is a need for research into advanced channel selection strategies that extend beyond the conservative 'safest' approach. A notable issue in the arbitration strategy used in the Safety Shell architecture [32] is its reliance on prediction, where the risk analysis is evaluated over a prediction time horizon into the future. Additionally, Safety Shell struggles with demanding computational requirements. However, research on other suitable arbitration strategies for multi-channel mission-continuing systems is limited. This gap in research leads to the formulation of the following problem statement for this paper: Develop an arbitration strategy that reduces reliance on computationally expensive predictions while still minimising safety risks and maximising the availability of mission-continuing channel choices.

IV. Novel Arbitration Strategy

To address the problem identified in the previous section, we propose the AA arbitration strategy in this paper. In heterogeneous multi-channel ADS systems, where there may be a divergence between the outputs (e.g., different motion plans generated by two channels), the arbiter's role is to select the most suitable channel that balances safety with system availability. This decision-making scenario, due to the requirement to balance multiple criteria is inherently a Multi-Criteria Decision Making (MCDM) problem [35]. Within this context, MCDM aims to identify the most appropriate channel by considering multiple criteria (e.g., safety and comfort) and assigning weights to each based on their relevance in the specific context. To tackle this MCDM challenge, the Analytic Hierarchy Process (AHP) [36] has been utilized in the development of the arbitration logic. AHP effectively handles the challenge of evaluating channel alternatives, encompassing both tangible and intangible criteria. Its capability to assign priorities to these criteria and then rank channel alternatives based on the assigned priorities, makes AHP an effective methodology for arbitrating between the various channels in an ADS.

The AHP methodology involves the following steps: structuring the decision problem into a hierarchy of subproblems, gathering empirical information and data, evaluating these hierarchies through pair-wise comparisons to compute priority scores, ensuring consistency in the pair-wise comparisons and synthesizing these comparisons to determine the overall ranking of the alternatives. Each step contributes to a decision framework that accommodates quantitative and qualitative data. These steps are elaborated in detail in the subsequent subsections.

A. Structuring the AHP decision hierarchy

In this study, the decision problem is defined as selecting the most suitable ADS channel that ensures both the continued functionality of the ADS and the maximization of safety. There are two decision alternatives for achieving this goal: Channel 1 and 2. To systematically address this decision problem, the primary criteria identified for the analysis are *safety* and *comfort*, reflecting the core focus of the decision-making process. These criteria are divided into a cluster of sub-criteria. Under the safety criterion, the sub-criteria include *time to collision* (TTC) and *post-encroachment time* (PET), which are vital for assessing collision risks. The comfort criterion encompasses sub-criteria such as *following distance*, *average speed-to-intended-speed ratio*, and measures of *braking*, *lateral*, and *forward accelerations*. After establishing these criteria and sub-criteria, the decision problem is modelled into a hierarchy, as illustrated in Fig 8, with the overall goal at the top, followed by layers of criteria and subcriteria, with the channel alternatives at the bottom.

B. Performing pair-wise criteria evaluations

Once the hierarchy is constructed, pair-wise comparisons are conducted to assess the criteria in terms of their importance relative to the goal. Similarly, within each cluster of sub-criteria, the significance of each sub-criterion in relation to others in the same cluster is evaluated. To quantify how much more important or dominant one element is over another, AHP employs a numerical scale, as proposed by Saaty [36] and depicted in Table I. For instance, if criteria are evaluated as equally significant, each is assigned an equal weight of 1. However, if one criterion is deemed more important than another, the more significant criterion receives a weight ranging between 2 and 9, reflecting its relative importance in the decision-making hierarchy. In the AHP methodology, selecting relevant criteria and sub-criteria (see Section IV-A) and conducting pair-wise preference evaluations are typically done by a group of stakeholders. This approach is favoured as it leverages collective wisdom and helps to mitigate individual biases. However, it is essential to note that in the context of this study, the setting of the criteria, sub-criteria, and preference evaluations derive from the judgements of a single individual under the guidance of a moderator. In this exploratory study, the precise accuracy of the pairwise comparison evaluations is not of paramount concern. The primary objective is to explore the feasibility and potential implications of the proposed strategy.

To facilitate these comparisons, the following pairwise comparison questions are formulated with the relative scores assigned according to the numerical scale (Table I):

- Safety vs. comfort: When considering the overall objective of selecting the most appropriate ADS channel, to what degree is safety more important than comfort (or vice versa) in ensuring the system's functionality?
- Safety sub-criteria analysis: When evaluating the safety aspects of the ADS channels, what is the relative im-

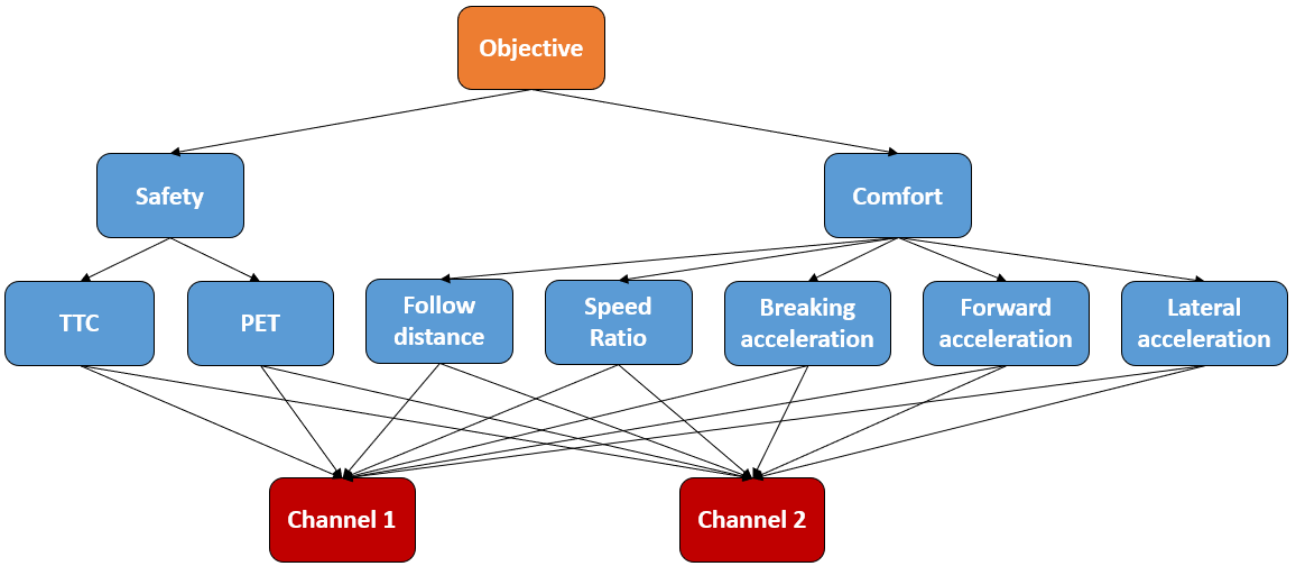

Fig. 8: Hierarchical decision structure for optimal channel selection, outlining criteria and sub-criteria for evaluating ADS channels.

portance of time to collision (TTC) compared to post-encroachment time (PET) for assessing collision risks?

- Comfort sub-criteria analysis: When evaluating the comfort aspects, for instance, what is the relative importance of following distance in comparison to braking acceleration for driver comfort?

The results of the pairwise comparison questions are systematically organized into three tables: Table II, Table III and Table IV. Each table catalogues the comparative assessments for its designated category. Specifically, in Table IV, abbreviations are used for clarity: Speed ratio, braking acceleration, lateral acceleration, forward acceleration and following distance are abbreviated as Spd Ratio, Brake acc, Lat acc, Fwd acc and Follow Dist, respectively.

In addition to the evaluations outlined in Tables II, III and IV, pairwise evaluations of channels (alternatives) relative to each sub-criterion must be performed. This process involves a critical step: mapping the simulation outputs obtained during online operations onto the AHP scale. Such mapping is necessary for ensuring that the pairwise evaluations accurately reflect the relative performance of each alternative under the defined sub-criteria. This methodology is described in Section IV-F.

TABLE I: AHP SCALE

Scale	Judgments of preferences
1	Equal Importance
3	Moderate Importance
5	Strong Importance
7	Very strong Importance
9	Extreme Importance
2,4,6,8	Intermediate values

C. Computations required to obtain the criteria and sub-criteria scores

After performing the pairwise evaluations and assigning the corresponding weights, reciprocal matrices for the pairwise comparisons are constructed [36]. This process begins by defining a set of weights, $W = \{w_1, w_2, \ldots, w_n : w_i \in \mathbb{R},\ i = 1, 2, \ldots, n\}$, where each weight w_i is a positive real number reflecting the relative importance or preference of each element compared to others at the same level of the hierarchy. These weights have been previously determined, as detailed in subsection IV-B. The structure of a pairwise comparison matrix A of order n is as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} \tag{1}$$

Total $n(n-1)/2$ pairwise comparisons contribute to form a pairwise comparison matrix. The diagonal entries of the pairwise comparison matrix are equal to 1, indicating the equivalence of each element to itself. The off-diagonal entries are the reciprocals of the direct comparisons, thus ensuring the matrix's reciprocal nature. Mathematically, the matrix A is represented as follows:

$$A = [a_{ij}], \text{ where } a_{ij} = \begin{cases} 1 & \text{if } i = j \\ \frac{1}{a_{ji}} & \text{if } i < j \end{cases} \tag{2}$$

Where a_{ij} represents the preference of the i^{th} element over the j^{th} element.

TABLE II: PAIRWISE COMPARISON OF CRITERIA

Criteria A	Criteria B	More Important	Score	Justification for pairwise AHP score
Safety	Comfort	A	7	Comfort has no significance without safety.

TABLE III: PAIRWISE COMPARISON OF SAFETY SUB-CRITERIA

Criteria A	Criteria B	More Important	Score	Justification for pairwise AHP score
TTC	PET	A	2	TTC is a direct measure of the imminent risk of collision.

TABLE IV: PAIRWISE COMPARISON OF COMFORT SUB-CRITERIA

Criteria A	Criteria B	More Important	Score	Justification for pairwise AHP score
Spd ratio	Brake acc	B	3	Braking can cause sudden discomfort, hence more critical than maintaining speed ratio.
Spd ratio	Lat acc	B	3	Lateral movements are felt more intensely by passengers, affecting comfort more than speed variations.
Spd ratio	Fwd acc	B	3	Forward acc's immediate physical effect on passengers is more discomforting than speed ratio variations.
Spd ratio	Follow dist	B	3	Affects comfort by reducing the need for sudden braking or steering.
Brake acc	Lat acc	B	3	Side-to-side motion of lateral acc is more discomforting than the linear deceleration of braking.
Brake acc	Fwd acc	A	2	Sudden braking causes more discomfort.
Brake acc	Follow dist	A	3	Immediate discomfort from abrupt braking affects comfort more than maintaining a following distance.
Lat acc	Fwd acc	A	3	Side-to-side motion of lateral acc is more discomforting than the linear acceleration.
Lat acc	Follow dist	A	5	Discomfort from lateral movement far outweighs the unease caused by a diminished sense of perceived safety.
Fwd acc	Follow dist	A	3	Forward acc can cause sudden discomfort, hence more critical than maintaining speed ratio.

For clarity, it should be noted that in these pairwise comparisons, the more preferred option is assigned a value up to 9, reflecting its relative importance, and the less preferred option receives the reciprocal of that value. This scoring system ensures the reciprocal nature of the matrix and accurately reflects the relative importance of each option.

This structural approach ensures that for any entry a_{ij}, where i is not equal to j, it can be approximated by the ratio of the corresponding weights w_i/w_j [36]. Consequently, matrix A can be expressed in terms of these weight ratios, providing a representation of the elements' comparative evaluations. The matrix is articulated as follows:

$$A = \begin{bmatrix} 1 & w_1/w_2 & w_1/w_3 & \cdots & w_1/w_n \\ w_2/w_1 & 1 & w_2/w_3 & \cdots & w_2/w_n \\ w_3/w_1 & w_3/w_2 & 1 & \cdots & w_3/w_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & w_n/w_3 & \cdots & 1 \end{bmatrix} \quad (3)$$

The next step in the AHP methodology is the calculation of the priority scores or priority vectors. Saaty recommends the principle eigenvector method [36] to determine these scores. The importance of this method is encapsulated in Theorem 1.

Theorem 1 (Principle Eigenvector and eigenvalue, [36])**.** *For a given positive matrix A, the only positive vector w and only positive constant c that satisfy $A \cdot w = c \cdot w$ is a vector w_{max} that is a positive multiple of the Perron vector (Principal eigenvector) of A, and the only such λ_{max} is the Perron value (Principle eigenvalue) of A.*

The theorem shows that for a positive reciprocal matrix A, which we construct from pairwise comparison data in the AHP, there will be a consistent and unique set of weights that can be derived. Mathematically, if A is our pairwise comparison matrix, Theorem 1 guarantees that we can find a vector w_{max} that reflects the relative weights of the criteria, sub-criteria, or alternatives being compared. To extract this vector, the largest eigenvalue, λ_{max} and its corresponding eigenvector w_{max} is identified. This principal eigenvector is normalized so that the sum of its components equals 1. This normalized vector is the priority vector or priority score for the criteria, sub-criteria, or alternatives being compared.

$$w_{\text{normalized}} = \frac{w_{\text{max}}}{\sum_{i=1}^{n} w_i}, \forall i = 1, 2.....n \quad (4)$$

Building upon this, 3 pair-wise comparison matrices are constructed: one detailing the criteria relative to the goal, shown in Table V; two for the sub-criteria relative to the safety and the comfort criterion, shown in Table VI and Table VII respectively. It should be noted that the bottom row of each table contains a 'CR' value. This refers to the Consistency Ratio (CR), a measure of the consistency of the pairwise comparisons, which is discussed in the following section, Section IV-D.

D. Consistency analysis

The integrity of the AHP is significantly influenced by the consistency of pairwise judgments. While absolute consistency in decision matrices is ideal, it is often unachievable due to the subjective nature of human judgment [36]. Inconsistencies can arise when judgments on relative importance are not perfectly transitive [36]. Therefore, the AHP methodology aims to minimize inconsistency to a level that does not significantly skew the decision-making process.

Theorem 2 (Consistency of a reciprocal matrix, [37])**.** *Let $A = a_{ij}$ be an nxn matrix of positive coefficients with $a_{ij} = \frac{1}{a_{ji}}$ then A is consistent if and only if $\lambda_{max} = n$.*

TABLE V: PAIRWISE COMPARISON MATRIX FOR CRITERIA

	Safety	Comfort	Priorities(Weights)
Safety	1	7	0.875
Comfort	0.142	1	0.125
CR = 0			

TABLE VI: PAIRWISE COMPARISON MATRIX FOR SAFETY SUB-CRITERIA

	TTC	PET	Priorities(Weights)
TTC	1	2	0.667
PET	0.5	1	0.333
CR = 0			

TABLE VII: PAIRWISE COMPARISON MATRIX FOR COMFORT SUB-CRITERIA

	Spd ratio	Brake acc	Lat acc	Fwd acc	Follow dist	Priorities(Weights)
Spd ratio	1	0.333	0.333	0.333	0.333	0.0715
Brake acc	3	1	0.5	2	3	0.2563
Lat acc	3	2	1	3	3	0.3743
Fwd acc	3	0.5	0.333	1	3	0.1857
Follow dist	3	0.333	0.333	0.333	1	0.1123
CR = 0.074						

To measure the degree of inconsistency, we calculate the Consistency Index (CI) using the formula:

$$CI = \frac{\lambda_{\text{max}} - n}{n - 1} \tag{5}$$

A non-zero CI indicates some inconsistency. To interpret the CI, we compare it to the Random Index (RI), an average CI obtained from a large sample of random pairwise comparison matrices [36]. This is shown in Table VIII. The Consistency Ratio (CR) is then computed as:

$$CR = \frac{CI}{RI} \tag{6}$$

Tables V, VI and VII include the computed CR for each pairwise comparison matrices. Saaty prescribes a CR threshold of 0.10 [36]. If CR values exceed 0.10, it indicates potential inconsistencies in the evaluations. In such cases, the stakeholders should revisit and adjust the pairwise evaluations until the CR falls below the threshold, ensuring the consistency of the decision-making process.

E. Calculating global weights of sub-criteria

Once the inconsistency in the pairwise comparisons is reduced to an acceptable level, the next step in the AHP methodology is to calculate the global weights for the sub-criteria. This begins with determining the local weights of each sub-criterion and the weight of their corresponding main criteria. The local weights for each sub-criterion are derived from the pairwise comparison matrices, as discussed in Section IV-C and shown in Tables VI and VII. Similarly, the weights of the main criteria are determined from their respective pairwise comparison matrices, detailed in Table V.

To calculate the global weight of a sub-criterion, its local weight is multiplied by the weight of its parent main criterion. This is represented mathematically as follows: for the i^{th} sub-criterion under the j^{th} main criterion, the global weight ($\mathcal{K}_{ij}$) is calculated as follows:

$$\mathcal{K}_{ij} = \kappa_{ij} \times \omega_j \qquad \forall i = 1, \ldots, n_j \text{ and } \forall j = 1, \ldots, m \tag{7}$$

where κ_{ij} is the local weight of the sub-criterion, ω_j is the weight of the parent criterion, n_j is the number of sub-criteria under the j^{th} criterion and m is the number of criteria.

The global weights are integral to the final decision-making process as they reflect the combined influence of each sub-criterion's local importance and the overarching priority of their corresponding main criteria.

F. Evaluating alternatives (channels) during online operation and synthesizing channel priorities

This section addresses the evaluation of channel alternatives, focusing on their performance relative to each sub-criterion identified in the AHP framework. During the simulation, each channel receives a value for the sub-criterion at every simulation time step, provided the data is available. For the purpose of this analysis, let's denote the outputs for channels i and j as γ_i and γ_j, respectively. The aim is to assess the relative significance of these values and accurately map them onto the AHP scale. This translation is accomplished using the following formula:

$$\Gamma_{ij} = \begin{cases} \min\left(\frac{\gamma_i}{\gamma_j} \times \frac{9}{6}, 9\right), & \text{if } \gamma_i > \gamma_j \text{ and } i \neq j \\ 1 \qquad\qquad\qquad\quad , & \text{if } \gamma_i \leq \gamma_j \text{ and } i \neq j \end{cases} \tag{8}$$

$$\Gamma_{ji} = \begin{cases} \min\left(\frac{\gamma_j}{\gamma_i} \times \frac{9}{6}, 9\right), & \text{if } \gamma_j > \gamma_i \text{ and } i \neq j \\ 1 \qquad\qquad\qquad\quad , & \text{if } \gamma_j \leq \gamma_i \text{ and } i \neq j \end{cases} \tag{9}$$

TABLE VIII: AVERAGE RANDOM CONSISTENCY INDEX (R.I.)

N	1	2	3	4	5	6	7	8	9	10
Random consistency index (R.I.)	0	0	0.52	0.89	1.11	1.25	1.35	1.40	1.45	1.49

where Γ_{ij} represents the preference of γ_i when evaluated against γ_j. Through trial and error, a ratio score of 6 is identified as indicating a significant difference, where the channel with the higher score gets a 9 on the AHP scale, and the other gets 1/9. For intermediate values, the ratio scores are scaled from the 1-6 range to the 1-9 AHP scale using the formula:

The pairwise comparison matrix is then constructed based on these scores, providing a structured approach to assess and compare the performance of the channels. This matrix, denoted as Γ, is of order 2 to represent the two-channel alternatives. It is structured as follows:

$$\Gamma = \begin{bmatrix} 1 & \Gamma_{ij} \\ \Gamma_{ji} & 1 \end{bmatrix} \tag{10}$$

In this matrix, Γ_{ij} represents the preference score of channel i when evaluated against channel j, while Γ_{ji} represents the preference score of channel j when evaluated against channel i. The diagonal elements are set to 1, indicating the equivalence of each channel to itself, and the off-diagonal elements represent the relative preference scores derived from the equation 8 and 9. Following this, the weights of the alternatives are calculated according to the steps outlined in Section IV-C.

Finally, to synthesize the overall priority of each channel, the weights of the alternatives with respect to each sub-criterion are aggregated. This aggregation is performed by multiplying the weights of each alternative with the global weight of the corresponding sub-criterion (as calculated in Section IV-E) and then summing these products. The channel with the highest aggregated score is deemed the most suitable for performing the driving task. Thus, the overall priorities, which encompass both the specifics of each sub-criterion and the overarching importance of each main criterion, are used to determine the most preferred alternative. This holistic approach ensures that the final decision in selecting the most appropriate ADS channel is well-informed and balanced, taking into account all pertinent factors and criteria.

V. METHODOLOGY

This section outlines the methodology for validating the proposed arbitration strategy, including benchmarking against selected architectures referenced in Section I. It begins with an overview of the simulation framework, designed to evaluate the performance of various arbitration strategies. Next, the architectures chosen for scenario-based comparison are listed. This is followed by an elaboration of the test cases and the FIs introduced in the channels. The section then provides an overview of the simulation settings applied in these test cases. Finally, it underscores the key metrics essential for assessing the quality of an arbitration strategy.

A. Simulation environment

The numerical simulations performed in this study were executed using a Frenet-based testing framework in Matlab 2023a. This framework, originally developed by Hanselaar et al. [32], utilizes the Automated Driving Toolbox [38]. Within this framework, a toolchain was developed to compile and analyze the experimental data to evaluate the performance of the arbitration strategies against the performance metrics mentioned in Section V-C. For further insights into the simulation framework, readers are directed to Section IV, as detailed in [32].

B. Designing the experiment

In the benchmarking process, the *AA* arbitration strategy was compared against the state-of-the-art arbitration strategies used in selected architectures, as referenced in Section I. The architectures selected for this comparison include the Safety Shell architecture with two channels (*SaS2*) [32], its expanded version with three channels (*SaS3)* [32] and Mobileye's three-channel architecture, which we refer to as the *RSS* [26].

The scenarios selected for evaluating the arbitration strategies, as illustrated in Fig 9, 10, and 11, reflect common driving situations encountered in real-world conditions. Fig 9 focuses on testing the arbitration strategy's response to insufficiencies in standard driving scenarios, including challenges such as identifying and navigating around Vulnerable Road Users (VRUs). Additionally, as shown in Fig 10, the focus extends to manoeuvring in complex traffic situations, specifically T-junctions. The simultaneous missed objects scenario, depicted in Fig 11, was selected to assess the arbitration strategy's

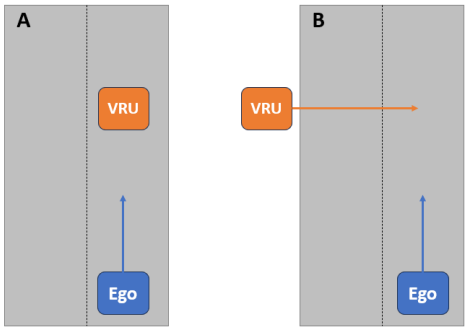

Fig. 9: Two straight road scenarios: A. Vulnerable Road User (VRU) in the ego lane, B. VRU approaching the ego lane to cross the road.

capability to handle multiple potential hazards, thereby testing the robustness and reliability of the arbitration strategies.

Our study specifically targets FIs in the World Model (WM) and Motion Plan (MP) categories. This emphasis stems from the findings in [16], which identify these categories as having the most significant prevalence of OIs. Table IX presents a detailed description of these FIs and the respective simulation scenarios in which they were tested.

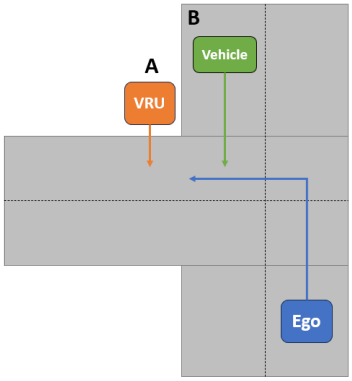

Fig. 10: Two T-junction scenarios: A. VRU approaching the lane to cross the road, B. Vehicle approaching from the opposite direction lane.

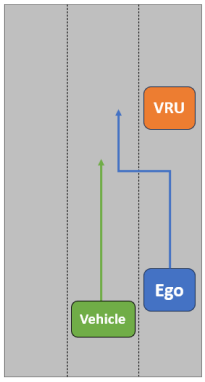

Fig. 11: Three-lane scenario with a VRU in the ego lane and a vehicle in the adjacent lane with the same initial speed as the ego vehicle.

A parameter variation study was conducted to assess how changes in specific simulation parameters impact the performance of different arbitration strategies. The settings relevant to the test cases are shown in Table X. Key parameters varied in this study include the *ego vehicle speed* and the *object disappearing time*. The ego vehicle's speed range in the simulation was set from 8 m/s to 20 m/s (29 km/h to 72 km/h). The lower limit of 29 km/h is based on the observation that collisions occurring at speeds below 30 km/h infrequently lead to fatal outcomes [39].

Test cases 1 through 4 are designed to evaluate the performance of the arbitration strategies when the current driving channel encounters false negative FIs. Test cases 5 and 6 are run with the second channel's WM detecting the ghost object for a limited time by defining an object's disappearing time relative to the expected impact time if the car never responds to the object. In Test case 7, both channels consistently fail to detect their respective missed objects for the entire scenario duration.

C. Criteria for evaluating the quality of arbitration strategies

Safety, *comfort*, and *availability* criteria are considered to assess the quality of an arbitration strategy. These criteria are crucial in determining the effectiveness of the strategy in balancing critical operational dynamics. This assessment includes ensuring the physical safety of passengers, enhancing the overall comfort of the travel experience and maintaining the availability of the ADS functionality.

Comfort is evaluated by the frequency of channel switches within multi-channel systems. Excessive switching may cause discomfort for passengers, as frequent changes in trajectories or setpoints may result in erratic vehicle movements [29]. Additionally, comfort is assessed by monitoring lateral and braking accelerations, focusing on peak accelerations in various scenarios, with lower peaks indicating a smoother ride.

Availability is measured by the average to intended-speed ratio, reflecting the vehicle's efficiency in maintaining the planned speed. This aspect is crucial as it affects both the smoothness of the ride and the overall travel duration.

For safety, the primary metric is the number of collisions, directly evaluating the arbitration strategy's effectiveness to avoid crashes.

VI. Results

A. Test cases 1 to 4 with a false negative FI affecting channel 1

For the given test cases, all the arbitration strategies successfully identified the risk posed by the nominal trajectory plan and were able to avoid the collision, ensuring safety. Upon the detection of a predicted collision, SaS2, SaS3, and AA strategies triggered a switch to a secondary mission-continuing channel that took over the Dynamic Driving Task (DDT) [9], thereby maintaining the ADS's availability. On the other hand, the RSS strategy responds to predicted collisions by implementing motion restrictions, such as breaking or slowing down. While the RSS strategy effectively prevented collisions in these scenarios, it was unable to maintain system availability in test cases 1 and 3, where an object remains in front of the vehicle. This limitation stems from the RSS architecture's lack of a motion planner other than the nominal planner (see Fig 4), limiting its ability to generate alternative trajectories when the primary plan is inadequate.

Due to RSS-prescribed motion restrictions in response to the detected FIs, the RSS strategy significantly impacts passenger comfort. This is evident in 12, which shows that RSS interventions necessitate severe braking at 8 m/s^2. The peak braking and lateral acceleration distributions (see Fig. 12 and 13) of SaS2 and SaS3 architectures are not identical due to the implementation of a noise level in the measured

TABLE IX: DESCRIPTION OF FIS THAT ARE TESTED IN THE SIMULATION SCENARIOS

Test case	Type of FI	Scenario	Description of the FI
1	Missed object	Fig 9.A	The VRU on the ego lane is not detected.
2	Missed object	Fig 10.B	The vehicle approaching the T-junction is not detected.
3	Incorrect location	Fig 9.B	The VRU crossing the road is observed to be further away from the ego lane.
4	Incorrect location	Fig 10.A	The VRU crossing the T-junction is observed to be further away from the T-junction.
5	Ghost object detection	Fig 9.A	Non-existent VRU is falsely detected on the ego vehicle's path.
6	Ghost object detection	Fig 10.A	Non-existent VRU is falsely detected attempting to cross the T-junction.
7	Simultaneous missed objects	Fig 11	Channel 1 does not detect the VRU and channel 2 does not detect the vehicle,

TABLE X: OVERVIEW OF THE SIMULATION SETTINGS FOR THE TEST CASES

Test case	Ego vehicle speed	Object disappearing time
1 - 4	[8, 10, ... , 20] [ms^{-1}]	n.a.
5 & 6	[8, 10, ... , 20] [ms^{-1}]	[2, 1.75, ... , 0] [s]
7	[8, 10, ... , 20] [ms^{-1}]	n.a.

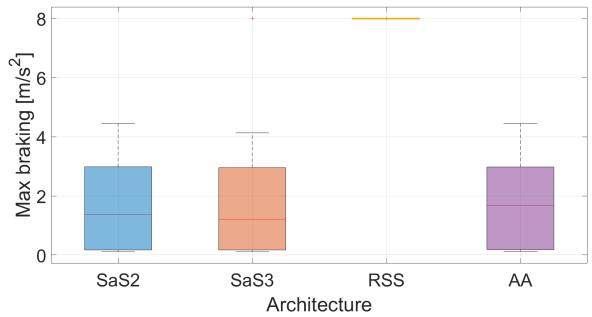

Fig. 12: Maximum braking deceleration for test cases 1 to 4.

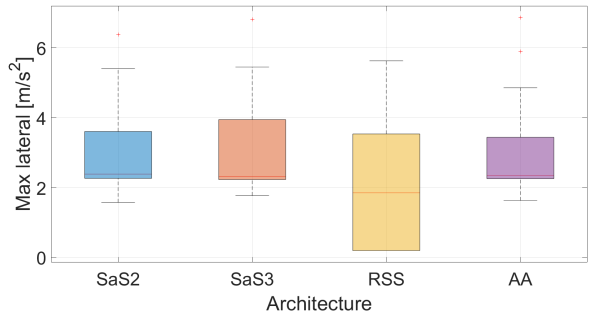

Fig. 13: Maximum lateral acceleration for test cases 1 to 4.

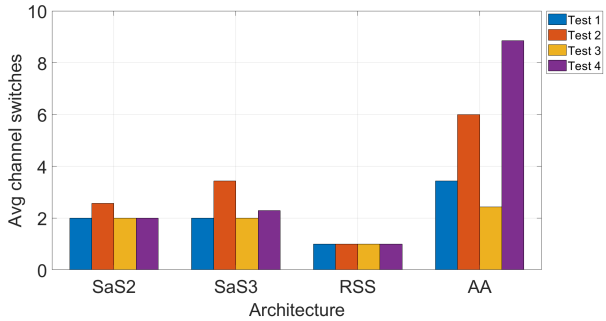

Fig. 14: Averaged channel switches for test cases 1 to 4, averaged over all tested speeds.

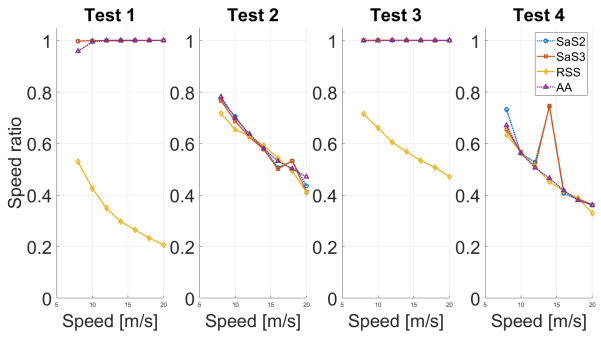

Fig. 15: Average speed-to-intended-speed ratio for test cases 1 to 4.

positions. AA, SaS2 and SaS3 exhibit similar distributions in peak braking acceleration, with AA showing slightly high average values. Nonetheless, this difference is not significant, indicating comparable performance among these strategies. For peak lateral accelerations, AA, SaS2 and SaS3, on average, exhibit comparable levels of performance.

The bar chart in Fig 14 illustrates the average frequency of channel switches across various arbitration strategies. The SaS2 and SaS3 architectures exhibit a relatively low frequency of channel switching. The RSS strategy, which lacks an alternative MP, resorts to motion restrictions in response to a FI, counted as a channel switch in this analysis. Notably, the AA strategy shows the highest frequency of channel switching. Such frequent switching could impact ride comfort and raise safety concerns due to continuous transitions between heterogeneous channels. The AA strategy's tendency for frequent channel switching can be attributed to its underlying arbitration logic. This logic is designed to initiate a switch between channels whenever there are any differences in the calculated global priorities of the channels.

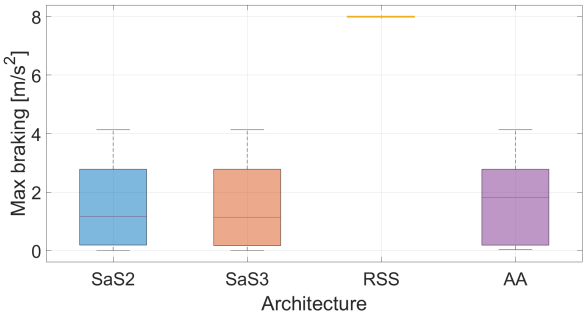

Fig. 16: Maximum braking deceleration for test cases 5 and 6.

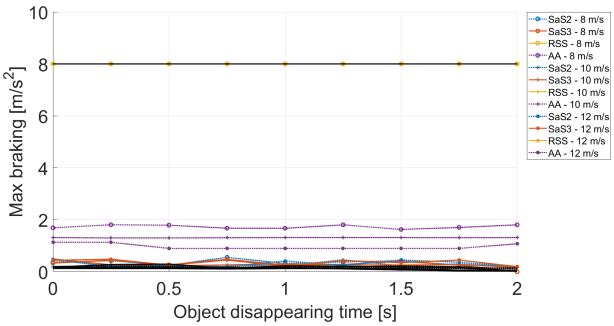

Fig. 17: Peak braking deceleration for test case 5, as a function of disappearing time of the object by the 2nd channel for various vehicle speeds.

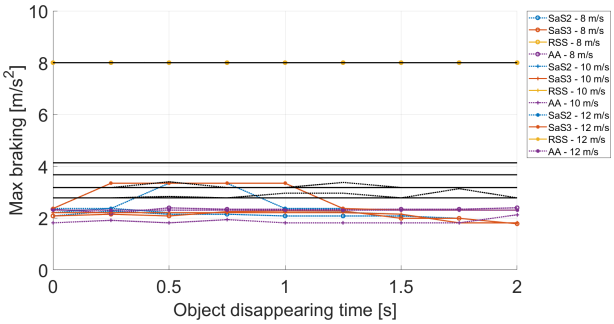

Fig. 18: Peak braking deceleration for test case 6, as a function of disappearing time of the object by the 2nd channel for various vehicle speeds.

The line graphs in Fig 15 illustrate the average speed-to-intended-speed ratio across four test cases. For test cases 1 and 3, SaS2, SaS3, and AA arbitration strategies maintained a ratio close to one, indicating seamless availability of the ADS functionality in the advent of a FI. By contrast, the RSS strategy shows lower speed ratios for the same test cases. This issue stems from RSS architecture's lack of flexible alternative MPs, and it relies on motion restrictions such as breaking or slowing down rather than rerouting around obstacles. Conversely, the other three strategies successfully planned a path around the VRU, maintaining system availability. In the T-junction scenarios of test cases 2 and 4, all strategies exhibit speed ratios below one as the ego vehicle decelerates to navigate the junction. This behaviour includes waiting for passing traffic in test case 2 and waiting for the pedestrian to cross the ego lane in test case 4.

B. Test cases 5 and 6 with a false positive FI affecting channel 2

In these test cases, collisions are not relevant because the FI affecting channel 2 is a false positive detection.

RSS initiates motion restrictions and necessitates an emergency braking of 8m/sec^2 as depicted in Fig 16) and is consistent with observations from previous test cases. The AA strategy exhibits a braking distribution similar to that of the SaS2 and SaS3 strategies, albeit with slightly higher average maximum braking values. However, this difference is not significant.

At higher speeds ranging from 14 to 20 m/sec, AA, SaS2, and SaS3 strategies demonstrate almost identical behaviours and are hence given the same colour for clarity in Figures 17 and 18. However, at lower speeds, specifically from 8 to 12 m/sec, a notable difference emerges when comparing the AA strategy with SaS2 and SaS3, with AA having higher peak braking deceleration as a function of object disappearing time for tested speeds. This distinction is evident in both test cases. The cause of this divergence in behaviour at lower speeds remains unclear and highlights an area for further investigation

C. Test Cases 7 with false negative FIs in both the channels affecting different objects

All the arbitration strategies successfully avoided a collision for the given test case. The RSS and SaS2 strategies achieved this by activating their escape trajectories, leading to severe discomfort due to emergency braking at 8m/s^2 as seen in Fig 19. SaS3 and AA initiated high peak braking at low speeds due to the poorly tuned motion planner used in these simulations [32]. Among the tested strategies, the AA maintained a balance between braking and steering to prevent collisions, showing better performance than the other strategies. However, at higher speeds, specifically at 16 and 18 m/s, the AA strategy tended to rely on severe lateral acceleration for collision avoidance (see Fig 20). This may be attributed to AA's frequent channel switching, which may cause delayed responses to FIs in the ego lane. Consequently, the limited braking time available in these situations forces the AA strategy to employ aggressive lateral movements to avert collisions.

VII. Discussion and Conclusion

In this paper, we have proposed a novel arbitration strategy, referred to as the AA, for arbitrating multi-channel mission continuing ADS channels. Numerical simulations show that the proposed arbitration strategy is a good alternative to the current state-of-the-art arbitration strategies employed in multi-channel architectures. This strategy's key advantage lies in its simplicity, which effectively balances functionality and efficiency, contrasting favourably with the more intricate

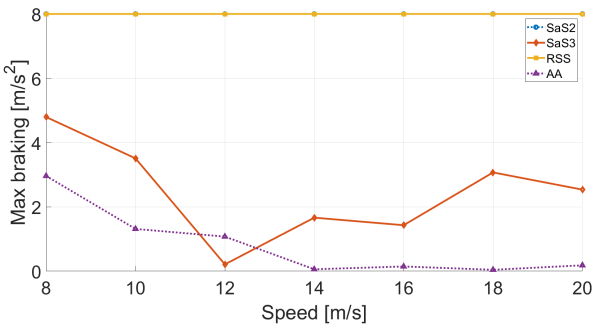

Fig. 19: Maximum braking deceleration for test case 7.

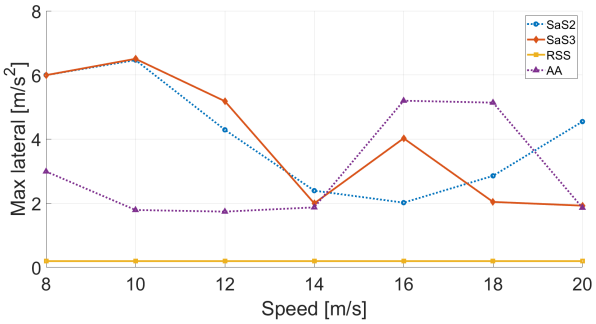

Fig. 20: Maximum lateral acceleration for test case 7.

implementations of other existing strategies. The AA strategy demonstrated its efficacy by successfully avoiding collisions and maintaining ADS functionality across all the test scenarios. However, it is observed that frequent channel switching (see Fig 14) and high-risk lateral manoeuvres at certain speeds (see Fig 20) are a cause for concern. The high frequency of channel switching can be eliminated to a certain extent by the implementation of 'dead zones'. This concept involves establishing thresholds within which minor performance variations between channels do not prompt a switch. Such a mechanism aims to reduce unnecessary transitions between channels.

Future work will focus on enhancing the evaluation framework by integrating additional criteria and sub-criteria into the AHP model. For instance, this expansion can include incorporating *Safety Performance Indicators* under the safety criterion and *Jerk* under comfort, allowing for a more comprehensive evaluation of the channel alternatives. Moreover, the transition to the Analytic Network Process (ANP) will enable a more detailed exploration of the complex interrelations and feedback mechanisms among various elements in the hierarchy. This enhanced approach will allow for a more thorough understanding of how different aspects of the ADS channels interact and impact the overall system performance.

References

[1] World Health Organization, "Global status report on road safety," 2023.

[2] T. S. Stephens, J. Gonder, Y. Chen, Z. Lin, C. Liu, and D. Gohlke, "Estimated bounds and important factors for fuel use and consumer costs of connected and automated vehicles," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2016.

[3] P. Hogeveen, M. Steinbuch, G. Verbong, and A. Hoekstra, "The energy consumption of passenger vehicles in a transformed mobility system with autonomous, shared and fit-for-purpose electric vehicles in the netherlands," *The Open Transportation Journal*, vol. 15, no. 1, 2021.

[4] The Conversation, "Robot, take the wheel: waymo has launched a self-driving taxi service," 2020. [Online]. Available: https://theconversation.com/robot-take-the-wheel-waymo-has-launched-a-self-driving-taxi-service-147908

[5] Engadget, "Gm's cruise is now offering driverless taxi rides in san francisco," 2021. [Online]. Available: https://www.engadget.com/g-ms-cruise-is-now-offering-driverless-taxi-rides-in-san-francisco-103542426.html

[6] The Robot Report, "Autox deploys its first fully driverless robotaxi fleet in shenzhen, china," 2020, accessed online Jan 30th, 2024. [Online]. Available: https://www.therobotreport.com/autox-deploys-first-fully-driverless-robotaxis-shenzhen-china/

[7] S. S. Banerjee, S. Jha, J. Cyriac, Z. T. Kalbarczyk, and R. K. Iyer, "Hands off the wheel in autonomous vehicles?: A systems perspective on over a million miles of field data," in *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 2018, pp. 586–597.

[8] M. Abdel-Aty, "A matched case-control analysis of autonomous vs human-driven vehicle accidents," 2023. [Online]. Available: https://www.researchsquare.com/article/rs-3401212/v1

[9] SAE, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," 2021.

[10] The Washington Post, "The final 11 seconds of a fatal tesla autopilot crash," 2023. [Online]. Available: https://www.washingtonpost.com/technology/interactive/2023/tesla-autopilot-crash-analysis/

[11] The Dawn Project, "Critical safety issues revealed by the dawn project's testing of tesla full self-driving," 2022. [Online]. Available: https://dawnproject.com/critical-safety-issues-revealed-by-the-dawn-projects-testing-of-tesla-full-self-driving/

[12] Fox Business, "Cruise robotaxi crashes with fire truck in san francisco," 2023. [Online]. Available: https://www.foxbusiness.com/technology/cruise-robotaxi-crashes-fire-truck-san-francisco

[13] Jalopnik, "Two waymo cars block san francisco traffic again as robotaxi stalling incidents rise 300 percent," 2023. [Online]. Available: https://jalopnik.com/two-waymo-cars-block-san-francisco-traffic-again-as-rob-1850581369

[14] P. Koopman and M. Wagner, "Positive trust balance for self-driving car deployment," in *Computer Safety, Reliability, and Security. SAFECOMP 2020 Workshops: DECSoS 2020, DepDevOps 2020, USDAI 2020, and WAISE 2020, Lisbon, Portugal, September 15, 2020, Proceedings 39*. Springer, 2020, pp. 351–357.

[15] Department of Motor Vehicles, "2021 autonomous vehicle disengagement reports," 2021.

[16] Y. Fu, J. Seemann, C. Hanselaar, T. Beurskens, A. Terechko, E. Silvas, and M. Heemels, "Characterization and mitigation of insufficiencies in automated driving systems," in *27th ESV Conference*, 2023.

[17] "Road vehicles – Safety of the intended functionality (ISO Standard No. 21448:2022)."

[18] "Road vehicles – Functional safety (ISO Standard No. 26262:2018)."

[19] A. Mehmed, W. Steiner, M. Antlanger, and S. Punnekkat, "System architecture and application-specific verification method for fault-tolerant automated driving systems," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 39–44.

[20] A. Armoush, "Design patterns for safety-critical embedded systems." Ph.D. dissertation, RWTH Aachen University Aachen, Germany, 2010.

[21] B. PAS, "Operational design domain (odd) taxonomy for an automated driving system (ads)," *Specification, standard by BSI Group*, vol. 31, 2020.

[22] P. Koopman and M. Wagner, "Challenges in autonomous vehicle testing and validation," *SAE International Journal of Transportation Safety*, vol. 4, no. 1, pp. 15–24, 2016.

[23] P. Koopman, *How Safe is Safe Enough?: Measuring and Predicting Autonomous Vehicle Safety*. Amazon Digital Services LLC - Kdp, 2022. [Online]. Available: https://books.google.nl/books?id=lNVozwEACAAJ
[24] A. Mehmed, W. Steiner, and A. Čaušević, "Systematic false positive mitigation in safe automated driving systems," in *2020 International Symposium on Industrial Electronics and Applications (INDEL)*. IEEE, 2020, pp. 1–8.
[25] Mobileye, "True redundancy." [Online]. Available: https://www.mobileye.com/technology/true-redundancy/
[26] J. Weast, "Sensors, safety models and a system-level approach to safe and scalable automated vehicles," *arXiv preprint arXiv:2009.03301*, 2020. [Online]. Available: https://arxiv.org/abs/2009.03301
[27] S. Furst, "Scalable architecture for autonomous driving," in *9th Vector Congress*, 2018.
[28] Mercedes-Benz Group, "Safety first for automated driving," 2019. [Online]. Available: https://group.mercedes-benz.com/documents/innovation/other/safety-first-for-automated-driving.pdf
[29] The Autonomous, "Safe automated driving: requirements and architectures," 2023. [Online]. Available: https://www.the-autonomous.com/news/the-autonomous-safety-architecture-full-report-unlocking-key-findings/
[30] S. Poledna, *Fault-tolerant real-time systems: The problem of replica determinism*. Springer Science & Business Media, 2007, vol. 345.
[31] P. Koopman, B. Osyk, and J. Weast, "Autonomous vehicles meet the physical world: Rss, variability, uncertainty, and proving safety," in *Computer Safety, Reliability, and Security: 38th International Conference, SAFECOMP 2019, Turku, Finland, September 11–13, 2019, Proceedings 38*. Springer, 2019, pp. 245–253.
[32] C. A. J. Hanselaar, E. Silvas, A. Terechko, and W. Heemels, "The safety shell: An architecture to handle functional insufficiencies in automated driving," *IEEE Transactions on Intelligent Transportation Systems*, 2024.
[33] BMW Group, "Safety assessment report," 2020. [Online]. Available: https://lindseyresearch.com/wp-content/uploads/2020/06/BMW.pdf
[34] O. S. Vaidya and S. Kumar, "Analytic hierarchy process: An overview of applications," *European Journal of operational research*, vol. 169, no. 1, pp. 1–29, 2006.
[35] E. Triantaphyllou and E. Triantaphyllou, *Multi-criteria decision making methods*. Springer, 2000.
[36] T. Saaty and L. Vargas, *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*, ser. International Series in Operations Research & Management Science. Springer, 2012. [Online]. Available: https://books.google.nl/books?id=6J9XI8I1qjwC
[37] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *Journal of mathematical psychology*, vol. 15, no. 3, pp. 234–281, 1977.
[38] The MathWorks, Inc., "Automated driving toolbox version: 3.7 (r2023a)," 2023.
[39] D. Richards, "Relationship between speed and risk of fatal injury: pedestrians and car occupants," 2010.