IEEE SA
STANDARDS ASSOCIATION

IEEE SA WHITE PAPER

IEEE 2846

EXAMPLE APPLICATIONS OF IEEE STD 2846-2022 TO FORMAL SAFETY-RELATED MODELS

Authored by

IEEE VT/ITS/AV Decision Making Working Group

IEEE

TRADEMARKS AND DISCLAIMERS

IEEE believes the information in this publication is accurate as of its publication date; such information is subject to change without notice. IEEE is not responsible for any inadvertent errors.

The ideas and proposals in this specification are the respective author's views and do not represent the views of the affiliated organization.

ACKNOWLEDGMENTS

Special thanks are given to the following reviewers of this paper:

Mark Costin

Jack Weast

Amitai Bin-Nun

Maria Soledad Elli

John Montrym

The Institute of Electrical and Electronics Engineers, Inc. 3 Park Avenue, New York, NY 10016-5997, USA

PDF: STDVA25980 978-1-5044-9458-8

NOTICE AND DISCLAIMER OF LIABILITY CONCERNING THE USE OF IEEE SA DOCUMENTS

This IEEE Standards Association ("IEEE SA") publication ("Work") is not a consensus standard document. Specifically, this document is NOT AN IEEE STANDARD. Information contained in this Work has been created by, or obtained from, sources believed to be reliable, and reviewed by members of the activity that produced this Work. IEEE and the IEEE VT/ITS/AV Decision Making Working Group expressly disclaim all warranties (express, implied, and statutory) related to this Work, including, but not limited to, the warranties of: merchantability; fitness for a particular purpose; non-infringement; quality, accuracy, effectiveness, currency, or completeness of the Work or content within the Work. In addition, IEEE and the IEEE VT/ITS/AV Decision Making Working Group disclaim any and all conditions relating to: results; and workmanlike effort. This document is supplied "AS IS" and "WITH ALL FAULTS."

Although the IEEE VT/ITS/AV Decision Making Working Group members who have created this Work believe that the information and guidance given in this Work serve as an enhancement to users, all persons must rely upon their own skill and judgment when making use of it. IN NO EVENT SHALL IEEE SA OR IEEE VT/ITS/AV Decision Making Working Group MEMBERS BE LIABLE FOR ANY ERRORS OR OMISSIONS OR DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS WORK, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Further, information contained in this Work may be protected by intellectual property rights held by third parties or organizations, and the use of this information may require the user to negotiate with any such rights holders in order to legally acquire the rights to do so, and such rights holders may refuse to grant such rights. Attention is also called to the possibility that implementation of any or all of this Work may require use of subject matter covered by patent rights. By publication of this Work, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. The IEEE is not responsible for identifying patent rights for which a license may be required, or for conducting inquiries into the legal validity or scope of patents claims. Users are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. No commitment to grant licenses under patent rights on a reasonable or non-discriminatory basis has been sought or received from any rights holder.

This Work is published with the understanding that IEEE and the IEEE VT/ITS/AV Decision Making Working Group members are supplying information through this Work, not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought. IEEE is not responsible for the statements and opinions advanced in this Work.

TABLE OF CONTENTS

ABSTRACT 5

1. INTRODUCTION 6

2. ASSUMPTIONS IN SAFETY-RELATED MODELS FOR ADS 6

2.1. ASSUMPTIONS ON KINEMATIC PROPERTIES OF ROAD USERS 7

3. RESPONSIBILITY-SENSITIVE SAFETY 8

3.1.INTERPRETATION OF RSS RULES 8
3.2. RSS PROPER RESPONSE 9
3.3. APPLICATION OF REASONABLY FORESEEABLE ASSUMPTIONS IN RSS 10

4. RULEBOOKS 13

4.1. APPLICATION 1: OFF-LINE ASSESSMENT OF ADS-OPERATED VEHICLE BEHAVIOR 14
4.2. APPLICATION 2: MINIMUM VIOLATION PLANNING UNDER REASONABLY FORESEEABLE BEHAVIOR OF OTHER ROAD USERS 16

5. SAFETY FORCE FIELD 19

6. ANALYSIS OF SAFETY-RELATED MODELS 20

7. REFERENCES 25

EXAMPLE APPLICATIONS OF IEEE STD 2846-2022 TO FORMAL SAFETY-RELATED MODELS

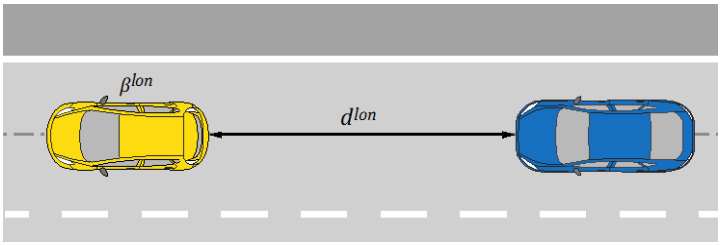
ABSTRACT

While automated driving system (ADS)-operated vehicles hold the potential for safety improvement compared to human drivers, the recognition that transportation will continue to entail some level of risk has to be considered. Human drivers rely on extensive daily experience in their interactions with other agents on the road, which helps them craft assumptions about reasonably foreseeable behavior of other road users. Similarly, ADS-operated vehicles will also need to make assumptions. Such assumptions play a role within ADS safety-related models, which provide a representation of safety-relevant aspects of driving behavior pertaining to both ADS-operated vehicles and other road users. Furthermore, formal safety-related models provide transparency and certainty in ADS decision-making contexts as they can be formally verified. Therefore, this paper introduces how several safety-related models are making use of reasonably foreseeable assumptions to help with the decision making of an ADS-operated vehicle.

1. INTRODUCTION

The safety assurance of decision making of an ADS-operated vehicle (i.e., ego vehicle) is of paramount concern to government and society. Formal models provide transparency and certainty in ADS decision-making contexts as they can be formally verified and expressed in formal notation. Formal models such as Responsibility Sensitive Safety (RSS) from Mobileye, Rulebooks from Motional, and Safety Force Field (SFF) from NVIDIA all rely on reasonably foreseeable assumptions to inform the boundaries of operation that can be assured by the formal model. This paper uses the high-level scenario depicted in FIGURE 1 to demonstrate how the reasonably foreseeable assumptions defined in IEEE Std 2846™-2022 [1] are used within these commercially deployed safety-related models.

FIGURE 1 Ego vehicle driving longitudinally behind another road user



The level of safety ultimately delivered by the ADS-equipped vehicle may, additionally, be dependent on what values are used for the reasonably foreseeable assumptions considered in safety-related models. The whitepaper "Literature Review On Kinematic Properties of Road Users for Use on Safety-Related Models for Automated Driving Systems" [2] presents a summary of peer-reviewed scientific publications, related standard documents, and active industry documents that studied road users' behavior, identifying values (and/or distributions) for the kinematic properties of road users using data-driven analyses.

2. ASSUMPTIONS IN SAFETY-RELATED MODELS FOR ADS

This section provides a high-level introduction of the assumptions about reasonably foreseeable behavior of road users that are defined in IEEE Std 2846-2022.

2.1. ASSUMPTIONS ON KINEMATIC PROPERTIES OF ROAD USERS

The reasonably foreseeable assumptions defined in IEEE Std 2846-2022 are primarily based on kinematic properties of other road users that include, among others, longitudinal and lateral velocities, accelerations, and decelerations,[1] and response time. The response time of a road user, while not a kinematic property, is relevant in the context of a safety-related model and should be understood as the time it takes a road user to perceive a specific stimulus and start executing a response (e.g., braking, steering, etc.). TABLE 1 presents a summary list of the kinematic properties considered in the standard. See ITS IEEE VT [2] for the related literature on values for kinematic properties.

Within the scope of IEEE Std 2846-2022, the assumptions about kinematic properties are based on the classification of different road user types, namely, pedestrians, bicyclists, vehicles, and other vulnerable road users (VRUs), such as a person riding an electric scooter, or a person using a wheelchair. Assumptions about these kinematic properties can take the form of bounding limits, such as reasonably foreseeable minimum and maximum boundaries (e.g., $v^{lon} \leq v^{lon}_{max}$), and their applicability depends on the driving scenario and the safety-relevant road users to be considered.

TABLE 1 List of road user properties considered in IEEE Std 2846-2022

Notation	Description
v^{lat}, v^{lon}	Lateral and longitudinal velocity of a road user
$\alpha^{lat}, \alpha^{lon}$	Lateral and longitudinal acceleration of a road user in its direction of travel
β^{lat}, β^{lon}	Lateral and longitudinal deceleration of a road user in its direction of travel
h	Heading angle (yaw) of a road user
h'	Heading angle rate of change (yaw rate) of a road user
λ	Lateral margin for small lateral fluctuation performed by road user moving in forward motion
ρ	Response Time of a road user

[1] The lateral deceleration of a road user, β^{lat}, can be understood as the decrease in the absolute value of the lateral speed of a road user, whereas a lateral acceleration, α^{lat}, can be understood as the increase of the absolute value of the lateral speed of a road user.

3. RESPONSIBILITY-SENSITIVE SAFETY

Mobileye first published the Responsibility-Sensitive Safety (RSS) model in 2017 [3]. RSS is an open and transparent technology-neutral formal model for safety that provides complete coverage for any driving scenario that an ADS-operated vehicle may encounter within the bounds of assumptions about reasonably foreseeable behaviors of other road users.

RSS is based on common-sense human notions of what it means to drive safely. RSS formalizes common-sense human driving rules as the following:

1) Do not hit someone from behind
2) Do not cut in recklessly
3) Right-of-way is given, not taken
4) Be careful in areas of limited visibility
5) If you can avoid an accident without causing another, you must do it

3.1. INTERPRETATION OF RSS RULES

The first two rules formally define a longitudinal and lateral safety envelope that is the foundation of determining what constitutes safe distances with respect to other road users in all driving situations. These safe distances are physics-based calculations that account for the capabilities of the ADS-operated vehicle while also incorporating reasonably foreseeable assumptions about the behavior of other road users.

Rule 1 states that the minimum longitudinal distance for a trailing road user is to have an adequate stopping distance to prevent a collision due to the sudden hard braking of a lead road user (e.g., see FIGURE 1).

$$d_{min}^{lon} = \left[v_r \rho + \frac{1}{2} \alpha_{max}^{lon} \rho^2 + \frac{\left(v_r + \rho \alpha_{max}^{lon}\right)^2}{2\beta_{min}^{lon}} - \frac{v_f^2}{2\beta_{max}^{lon}} \right]_+$$

where

v_r is the following vehicle's longitudinal speed [m/s]

ρ is the following vehicle's response time [s]

α_{max}^{lon} is the following vehicle's maximum longitudinal acceleration during the response time [m/s^2]

β_{min}^{lon} is the following vehicle's minimum longitudinal deceleration after response time [m/s^2]

v_f is the longitudinal speed of the leading vehicle [m/s]

β_{max}^{lon} is the reasonably foreseeable assumed maximum longitudinal deceleration of the leading vehicle [m/s^2]

$[x]_+$ denotes the $\max(0, x)$

Rule 2 states that the minimum lateral distance to prevent a side collision between two road users side-by-side, given the sudden lateral motion of one of them, is defined as:

$$d_{min}^{lat} = \lambda + \left[\frac{v_1 + v_{1,\rho}}{2}\rho_1 + \frac{v_{1,\rho}^2}{2\beta_{1,min}^{lat}} - \left(\frac{v_2 + v_{2,\rho}}{2}\rho_2 - \frac{v_{2,\rho}^2}{2\beta_{2,min}^{lat}}\right)\right]_+$$

with $v_{1,\rho} = v_1 + \rho_1\alpha_{1,max}^{lat}$, $v_{2,\rho} = v_2 - \rho_2\alpha_{2,max}^{lat}$

where

v_1 is the left vehicle's lateral speed [m/s]

ρ_1 is the left vehicle's response time [s]

$\alpha_{1,max}^{lat}$ is the left vehicle's maximum lateral acceleration during the response time [m/s^2]

$\beta_{1,min}^{lat}$ is the left vehicle's minimum lateral deceleration after response time [m/s^2]

v_2 is the right vehicle's lateral speed [m/s]

ρ_2 is the right vehicle's response time [s]

$\alpha_{2,max}^{lat}$ is the right vehicle's maximum lateral acceleration during the response time [m/s^2]

$\beta_{2,min}^{lat}$ is the right vehicle's minimum lateral deceleration after response time [m/s^2]

λ is the lateral fluctuation margin [m]

3.2. RSS PROPER RESPONSE

A collision between two road users becomes possible when they are at both an unsafe longitudinal distance and an unsafe lateral distance. This is considered as a *Dangerous Situation* by RSS. Once a *Dangerous*

Situation occurs, any involved road user can perform a *Proper Response* to mitigate the *Dangerous Situation*.

From [3] definitions, if a vehicle is driving behind another, they are already at an unsafe lateral distance. If they get closer longitudinally until the longitudinal distance, d_{min}^{lon}, becomes unsafe, then the situation becomes dangerous. Therefore, it makes sense that the rear vehicle brakes longitudinally with at least β_{min}^{lon}, to restore a safe headway space ahead (i.e., apply a longitudinal *RSS Proper Response*).

Similarly, when two road users are driving side-by-side (e.g., in adjacent lanes), they are already at an unsafe longitudinal distance. If the road users get closer laterally such that the lateral distance, d_{min}^{lat}, becomes unsafe according to Rule 2, this is considered a *Dangerous Situation*. Therefore, in this case, the *RSS Proper Response* would be to decelerate laterally, with at least β_{min}^{lat}, (i.e., steer away from the other road user) until the Dangerous Situation has been mitigated or until reaching a zero lateral velocity again.

3.3. APPLICATION OF REASONABLY FORESEEABLE ASSUMPTIONS IN RSS

The RSS model makes use of the reasonably foreseeable assumptions defined for the kinematic properties described in TABLE 1. These assumptions are used for the calculation of the safety envelope along with the determination of the Proper Response needed to avoid crashes. Examples of how RSS makes use of reasonably foreseeable assumptions are presented below.

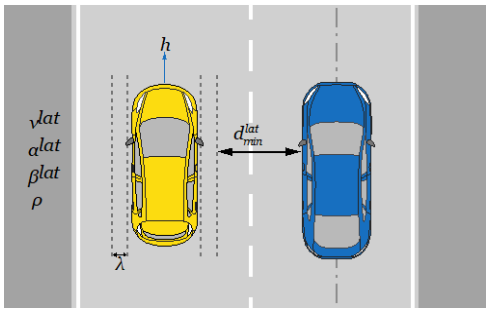
The RSS minimum longitudinal distance, d_{min}^{lon}, takes into account the kinematics between road users by utilizing the reasonably foreseeable assumptions in a manner consistent with IEEE Std 2846-2022 [1]. Considering the car following scenario illustrated in FIGURE 1, the RSS minimum longitudinal distance calculation between a follower vehicle that drives behind another in the same direction with longitudinal velocities v_r and v_f, respectively, assumes that the leading vehicle could brake up to a reasonably foreseeable deceleration, denoted as β_{max}^{lon}. At the same time, the formulation considers that the following vehicle could accelerate up to α_{max}^{lon} during the response time ρ, after which it would start decelerating with at least a deceleration of β_{min}^{lon}.

The RSS minimum lateral distance formula, d_{min}^{lat}, accounts for the ADS-operated vehicle driving next to another road user, as depicted in FIGURE 2, both with lateral speed v_1 and v_2, respectively, which during the response time ρ_1 and ρ_2 both apply a lateral acceleration of at least $\alpha_{1,max}^{lat}$ and $\alpha_{2,max}^{lat}$, respectively,

towards each other and after that both apply lateral braking of at least $\beta_{1,min}^{lat}$ and $\beta_{2,min}^{lat}$ until reaching zero lateral velocity. In this way, RSS calculates a minimum lateral distance based on a consideration of the measured lateral velocity of the laterally adjacent road user, assuming reasonably foreseeable maximum lateral acceleration $\alpha_{2,max}^{lat}$, maximum lateral braking $\beta_{2,min}^{lat}$ (i.e., steering away from the other road user), and response time ρ_2, as defined in IEEE Std 2846-2022 [1].

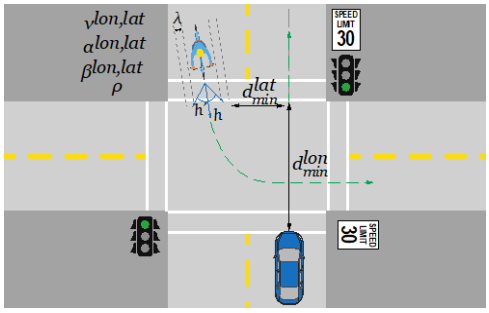
NOTE—For simplicity, these equations assume instantaneous max acceleration or deceleration, but more dynamic jerk-bounded braking and acceleration profiles can easily be added [4].

FIGURE 2 Ego vehicle driving next to other road users



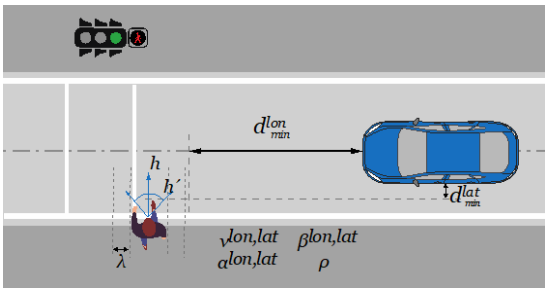
In the case of the ADS-operated vehicle negotiating an intersection with other road users, as depicted in FIGURE 3, in addition to considering the right-of-way rules, the RSS model's Rule 3 states that right-of-way is given, not taken. The reasonably foreseeable assumptions from IEEE Std 2846-2022 are also useful in this context. The approach is to perform minimum distances calculations that are based on the equations defined in Rule 1, which provide insight into whether the unprioritized road user would be able to stop in time in order to avoid a collision, assuming that the unprioritized vehicle could accelerate up to α_{max}^{lon} during its response time ρ, and after that, start braking with a minimum reasonably foreseeable braking of at least β_{min}^{lon}, and give way or not.

FIGURE 3 Ego vehicle negotiating an intersection



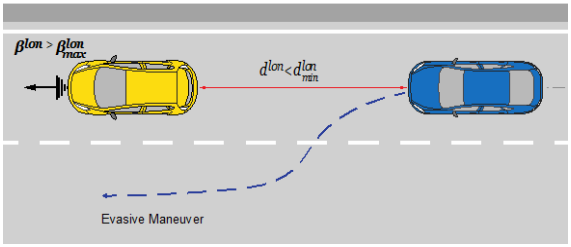
In the case of the ADS-operated vehicle negotiating an intersection with other road users, as depicted in In the case of the ADS-operated vehicle interacting with pedestrians, as depicted in FIGURE 4, the RSS model takes a similar approach, consistent with the assumptions defined in IEEE Std 2846-2022. RSS assumes that a pedestrian with a heading of h could exhibit any reasonable trajectory within the bounds of a rate of change of the heading angle of at most h'_{max} that considers a reasonably foreseeable acceleration of at most α^{lon}_{max} during its response time ρ, after which the pedestrian would start stopping with at least β^{lon}_{min} until reaching a full stop. Assumptions about the reasonably foreseeable behavior of the pedestrian on α^{lon}_{max} , ρ, β^{lon}_{min}, and h'_{max} are at the core of minimum distance calculations for this case.

FIGURE 4 Ego vehicle interacting with pedestrians



In the case that the other road user exceeds the reasonably foreseeable assumptions considered by RSS, as depicted in FIGURE 5, RSS specifies that the ADS performs an evasive maneuver (one example of a possible evasive maneuver is shown by the dashed blue line in FIGURE 5) to avoid an accident without creating another, by maintaining a safety envelope around all other road users, as defined by RSS rules 1–4. IEEE Std 2846-2022 does not suggest any specific ADS evasive maneuvers and the situation depicted in FIGURE 5 is not addressed by the standard.

FIGURE 5 Ego vehicle evasive maneuver



4. RULEBOOKS

Rulebooks is a technology-neutral framework to formally specify and assess the desired behavior of an ADS-operated vehicle [5]. The Rulebooks framework along with certain extensions and applications are spelled out in a series of publications listed in the references of this white paper [5] [6] [7]. The Rulebooks framework can be used to develop various ADS applications but should be considered distinct from the ADS technology itself. While this section offers several potential application areas in the context of the Rulebooks framework, the approaches are subject to further evaluation and may or may not prove suitable for use for a commercial product.

The Rulebooks framework allows developers to create specific instantiations of behavioral specifications or rulebooks. A rulebook has two main ingredients:

- A set of formal driving rules that a safe, lawful, and natural driver should follow
- A priority structure that specifies the relative importance of the rules

One application of a rulebook is to rank a set of trajectories under consideration in a given scenario.

Formal driving rules can capture safety considerations (e.g., do not collide with other vehicles), legal rules of the road (e.g., yield to pedestrians on a crosswalk), good driving practices (e.g., do not make unnecessary lane changes), and product performance (e.g., drive comfortably, reach the destination). Formal driving rules include a violation metric that specifies the degree to which a given ADS-operated vehicle trajectory violates a rule, with zero indicating no violation and a positive value indicating a violation. The Rulebooks framework acknowledges that in complex driving scenarios, the rules may represent competing objectives, resulting in trade-offs. The priority structure provides a transparent way to specify which driving rules should take precedence over others. For example, if faced with a choice between braking hard (i.e., violating the rule to drive comfortably) and failing to yield to a pedestrian on a crosswalk, the priority structure would specify that it is preferable to brake hard. The original Rulebooks framework envisioned a priority structure as a hierarchical graph mathematically known as a pre-order [5], but other types of priority structures can be used as well (e.g., a set of weights for the formal rules).

A key feature of the Rulebooks framework is that the set of formal driving rules and the priority structure remain the same for all scenarios, allowing reasoning over broadly applicable principles to define preferred outcomes [8]. This makes it possible to use the same internally consistent rulebook to specify the desired ADS-operated vehicle behavior in any scenario within the ODD and to systematically assess the actual

behavior against the rulebook at scale [8].

The Rulebooks framework takes the perspective of an outside, independent observer with access to perfect information: it evaluates an ADS-operated vehicle trajectory as it happened (after-the-fact) against the rulebook. This perspective is both necessary and advantageous. It is necessary because driving behavior is assessed based on what actually happened rather than what a driver (or ADS) thought might happen in real time. It is advantageous because it makes the approach technology-neutral: one can use the same rulebook to evaluate different ADS versions or even different ADS architectures (if operating in the same ODD). The after-the-fact perspective makes the Rulebooks framework particularly suitable for off-line assessment of ADS-operated vehicle behavior at scale. Section 4.1 gives an example of a potential application of the scenarios and assumptions of IEEE Std 2846-2022, providing a set of reasonably foreseeable scenarios for off-line, rule-based assessment of the ADS-operated vehicle behavior.

A second potential application of the scenarios and assumptions is to account for uncertainty about what other road users might do when an ADS uses a rulebook in real time. IEEE Std 2846-2022 recognizes that there are limits on what can be considered reasonably foreseeable behaviors of other road users. Section 4.2 describes how one may apply assumptions about other road users to evaluate ADS-operated vehicle trajectories in real time using a rulebook.

To illustrate the use of a rulebook under assumptions about reasonably foreseeable behaviors of other road users, we consider the scenario depicted in FIGURE 1 which, for simplicity, characterizes the reasonably foreseeable behavior of the lead vehicle using a single parameter, β_{max}^{lon}.

4.1. APPLICATION 1: OFF-LINE ASSESSMENT OF ADS-OPERATED VEHICLE BEHAVIOR

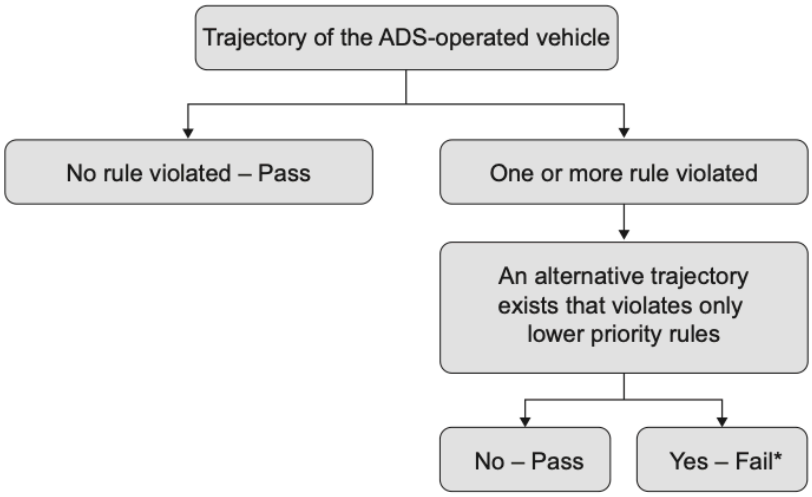
This application evaluates an ADS-operated vehicle against the rulebook behavior specification by testing it within the limits of reasonably foreseeable behaviors of other road users. The tests may be performed in simulation or, if possible to conduct safely, on a closed course. For the scenario in FIGURE 1, the test would have the ego vehicle driving at a distance d^{lon} behind the lead vehicle when the lead vehicle suddenly starts braking at a deceleration β_{max}^{lon} until it comes to a full stop.

If the ADS-operated vehicle responds to the scenario by executing a trajectory that does not violate any rule in the rulebook, then it passes the test. However, it is possible that collision avoidance may necessitate a

violation of some other rule (e.g., an evasive maneuver may violate a rule to stay in the drivable area). Despite the rule violation, this behavior might be entirely consistent with a well-designed rulebook that prioritizes collision avoidance over staying in the drivable area. However, to examine whether the behavior complies with the rulebook, we need to know whether it was feasible to do even better than the evasive maneuver that leaves the drivable area, for example, by violating a less important rule such as a rule to keep longitudinal deceleration within comfortable limits.

To determine whether the ADS-operated vehicle performance passes the test, rather than looking at individual rule violations, FIGURE 6 proposes a rule-based pass/fail process for trajectories that considers whether a materially better trajectory than the one executed was feasible. This process defines a materially better trajectory as one that violates only lower priority rules than the executed trajectory (or that violates no rules at all). By evaluating the trajectory against materially better trajectories, this process creates an incentive for the ADS to try to minimize rule violation in situations that mandate a rule violation. This process requires after-the-fact consideration of alternative trajectories. These alternative trajectories may come from manual driving, human driving models, expert review, or algorithmic approaches like rule-based optimal control [6].

FIGURE 6 **Rule-based pass/fail evaluation process for after-the-fact off-line evaluation of ADS-operated vehicle behavior**



*NOTE - If an alternative trajectory exists that violates the same rule as the ADS-operated vehicle trajectory but to a lesser degree, then the tester may use further review or performance criteria to determine pass/fail

4.2. APPLICATION 2: MINIMUM VIOLATION PLANNING UNDER REASONABLY FORESEEABLE BEHAVIOR OF OTHER ROAD USERS

This application illustrates how the dynamic driving task (DDT) of the ADS may use a rulebook while considering the standard's assumptions about reasonably foreseeable behaviors of other road users.

Let $T(t) = \{\tau_1, \tau_2, \ldots, \tau_n\}$ be a set of proposed trajectories in arbitrary order considered by the DDT at a given point t in time. Each trajectory τ_j is a list of ego vehicle positions at subsequent time points in the scenario over some planning horizon t_H. Trajectories may come from any ADS planner or controller, including rule-aware planners [9], [10] , rule-based optimal control algorithms [8], or rule-agnostic planning algorithms based on machine learning or other approaches.

The Rulebooks framework evaluates the proposed ego vehicle trajectories based on a set of rules $R = \{r_1, r_2, \ldots, r_m\}$ and a priority structure $P(R)$. $P(R)$ is a function that takes as input the degrees of violation $\rho_{i,j} = \rho_{r_i}(\tau_j)$ of each trajectory τ_j with respect to each rule r_i under perfect information and outputs the set of trajectories in order of preference T_{pref} according to the rulebook behavior specification, assuming perfect information. The ordering can be based on a pre-order, a weighted sum of violations for all rules, or any other method to encode a priority structure. If T contains all possible trajectories, then the trajectory preferred by the rulebook is the minimum (rule) violation trajectory.

As mentioned, while performing the DDT, an ADS does not have perfect information and may need to consider uncertainty about the behavior of other road users. The assumptions state that it is reasonably foreseeable that the lead vehicle's deceleration β^{lon} at time t is in the range $[0, \beta^{lon}_{max}]$. Using this assumption, we determine the worst possible violation of each trajectory with respect to each rule as:

$$\rho^{RF}_{ij} = \max_{\beta^{lon} \in [0, \beta^{lon}_{max}]} \rho_{ij}(\beta^{lon})$$

where

$\rho_{ij}(\beta^{lon})$ is the rulebooks' violation score of trajectory τ_j with respect to rule r_i for a given value of β^{lon} at time t

Now the rulebook priority structure can be applied to rank the proposed trajectories according to minimum violation under all reasonably foreseeable behaviors of the lead vehicle (hence the notation ρ_{ij}^{RF} where the superscript RF stands for reasonably foreseeable).

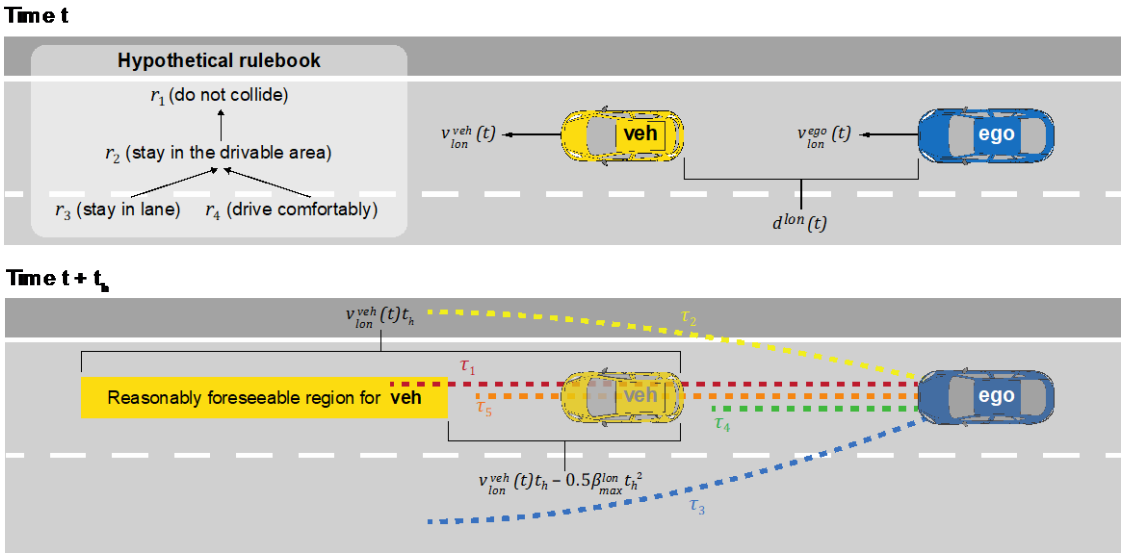
FIGURE 7 illustrates the concept of priority structure. The top figure shows the situation at time t and the bottom figure shows the situation at time $t + t_H$. FIGURE 7 also shows:

- A hypothetical rulebook with four rules. This pre-order prioritizes collision avoidance (r_1) over staying in the drivable area (r_2), while staying in lane (r_3) and driving comfortably (r_4) are both incomparable to each other and less important than all other rules. Incomparability of rules here means that the rulebook expresses no preference between a trajectory that fails to stay in lane versus one that drives uncomfortably.

- The reasonably foreseeable region where the lead vehicle might be at time $t + t_H$ for $\beta^{lon} \in [0, \beta_{max}^{lon}]$.

- The realizations through time $t + t_H$, of five hypothetical candidate trajectories that the ADS may propose at time t.

FIGURE 7 Minimum violation planning using a rulebook under reasonably foreseeable behavior of other road users



In trajectory τ_1, the ego vehicle does not decelerate. For most assumptions about reasonably foreseeable behaviors of the lead vehicle, a collision does not occur. However, if β^{lon} is close to β^{lon}_{max}, then a collision would occur, leading to a violation of r_1. The violation score with respect to rule r_1 is the highest if $\beta^{lon} = \beta^{lon}_{max}$ and therefore $\rho^{RF}_{11} = \rho_{11}(\beta^{lon}_{max}) > 0$. Since none of the other trajectories violate r_1 under any assumptions about reasonably foreseeable behaviors of other road users about the lead vehicle's behavior, trajectory τ_1 is the least preferred trajectory. Regardless of the behavior of the lead vehicle, trajectory τ_2 violates r_2, while the remaining candidate trajectories do not, so trajectory τ_2 is the next least preferred trajectory. Assuming trajectory τ_3 does not violate r_4 while trajectory τ_4 does, the rulebook would be indifferent between trajectories τ_3 and τ_4 under reasonably foreseeable behavior of the lead vehicle since both violate different incomparable rules. If trajectory τ_5 does not violate r_4, then it would be the preferred trajectory under reasonably foreseeable behavior of the lead vehicle. Thus, in that case, the ranking would be (from best to worst, where "<>" indicates incomparable): $\{\tau_5, \tau_4 <> \tau_3, \tau_2, \tau_1\}$.

NOTE—If τ_5 does violate r_4, then presumably it violates r_4 less than τ_4 does, and the ranking would be: $\{\{\tau_5, \tau_4\}<>\tau_3, \tau_2, \tau_1\}$.

TABLE 2 summarizes the maximum foreseeable rule violations of each trajectory with each rule. Note that this scenario assumes no presence of other road users. If there was a vehicle behind the ego vehicle in the left lane, then the ego vehicle would have to consider whether a collision with that vehicle is reasonably foreseeable and possibly other rules about cutting in front of other vehicles.

TABLE 2 Matrix ρ^{RF}_{ij} of the worst violation of the *i-th* rule by the *j-th* trajectory within the bounds of reasonably foreseeable behavior by the lead vehicle (assuming τ_5 does not violate r_4)

	r_1	r_2	r_3	r_4
τ_1	$\rho^{RF}_{11} > 0$	—	—	—
τ_2	0	$\rho^{RF}_{22} > 0$	—	—
τ_3	0	0	$\rho^{RF}_{33} > 0$	0
τ_4	0	0	0	$\rho^{RF}_{44} > 0$
τ_5	0	0	0	0

NOTE— If a trajectory violates a rule, then its performance on lower priority rules does not impact trajectory selection, denoted as "—".

Please remember that alternative methods exist and may be appropriate for dealing with uncertainty about the behavior of other road users. For example, instead of considering the bounds on assumptions of what is reasonably foreseeable (i.e., β_{max}^{lon}), the DDT may consider the entire probability distribution of β^{lon} to make an informed decision. This approach may require using a different priority structure that is less hierarchical than a pre-order.

5. SAFETY FORCE FIELD

The NVIDIA Safety Force Field (SFF) [11] is built on a simple core concept: Actors in traffic should apply a safety procedure or equivalent action before it is too late. It maps world perception into constraints on control that, if obeyed, prevents all collisions. It is a physics-based model that takes into account lateral and longitudinal separations within the same calculation.

The negative gradient $F_{AB} = -\frac{\partial \rho_{AB}}{\partial x_A}$ of the SFF safety potential ρ_{AB} is called the safety force field on actor A from actor B where x_A is the state of actor A.

The car-following scenario from FIGURE 1 is used to illustrated SFF. In FIGURE 1, the blue vehicle represents the ego vehicle, and the lead vehicle is assumed to have the maximum deceleration capability β_{max}^{lon}. Two cases are considered: (1) the lead vehicle changes lanes into the ego vehicle lane and d_{lon} does not allow for sufficient ego vehicle stopping distance for the assumed maximum deceleration, and (2) the ego vehicle is approaching the lead vehicle from behind with greater speed than the lead vehicle and starts to slow down to maintain the safe stopping distance given the assumed maximum deceleration.

The normative requirement in IEEE Std 2846-2022 is for the ADS-operated vehicle to consider whether $\beta^{lon} \leq \beta_{max}^{lon}$, and $\beta^{lon}$is the observed longitudinal deceleration of the leading vehicle, where β_{max}^{lon} is the assumed maximum reasonably foreseeable deceleration of the leading vehicle.

The SFF safety potential ρ_{AB} is defined to be strictly positive on the unsafe set and non-negative elsewhere and defined as:

$$\rho_{AB} = \max\left(t_{LEADstop} - p_t, t_{EGOstop} - p_t\right)$$

where

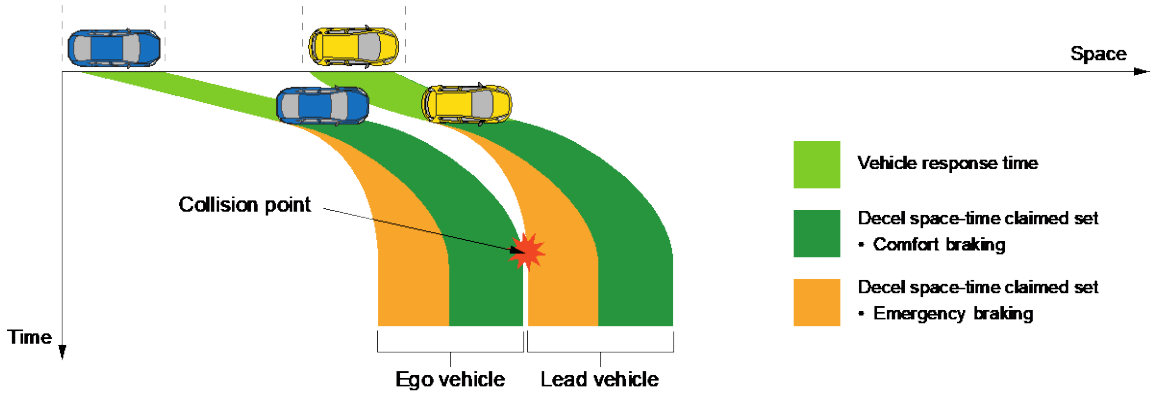
$t_{LEADstop}$ is the stopping time for the lead vehicle

$t_{EGOstop}$ is the stopping time for the ego vehicle

p_t is the collision time (p is the collision point in space-time, and then p_t is the time in seconds) between the two vehicles based on β_{max}^{lon}

This is illustrated in FIGURE 8.

FIGURE 8 Safety Force Field ego vehicle following lead vehicle



The objective of SFF is to define the acceptable actions for the ego vehicle. For case (1), the SFF would require that the ego vehicle reduce speed (apply its safety procedure) to allow for sufficient braking time to prevent a collision. For case (2), the SFF will allow the ego vehicle to approach the lead vehicle until the point where their claim sets overlap. The "Decel" space-time claimed set of the lead vehicle can be calculated using the assumed maximum reasonably foreseeable deceleration β_{max}^{lon}.

6. ANALYSIS OF SAFETY-RELATED MODELS

In addition to defining a minimum set of reasonably foreseeable assumptions that shall be considered, IEEE Std 2846-2022 [1] also defines a set of attributes common among safety-related models. This section analyzes a subset of the models that were contributed to the standard against the defined attributes. Some models may be complete embodiments of the attributes, while others may provide a framework for creating a model that implement the attributes. The intention for the reader is to understand how the defined attributes can be embodied within a safety-related model. Attributes that can only be validated by demonstration or through empirical examination and tests are omitted for this document. Please refer to the Verification and Validation methods defined in IEEE Std 2846-2022 [1].

TABLE 3 Analysis of safety-related models attributes

Attribute	RSS	Rulebooks	SFF
Incorporates the laws of physics	Minimum safe distances in all situations are derived from kinematics of the road users	Individual rules can incorporate the laws of physics	The control model and safety procedure are both determined based on physical properties of the vehicles.
Accommodates acceptable risk	RSS is a parameterized model that accommodates for behavior that can map to a desired level of risk	Rulebooks provides a customizable framework to compare alternative trajectories leading to different outcomes	Uncertainty in driving environment is handled by providing confidence intervals for all the metrics needed to calculate the Safety Force Field constraints.
Supports reasonably foreseeable scenarios	Considers longitudinal and lateral conflicts with vehicles and VRUs on roads with multiple geometries, merges, cross traffic, signalized and unsignalized right-of-way violations, occlusions and unstructured roads and their derivations	A rulebook can address any scenario within its ODD, foreseeable, or not	Considers safe trajectories, given all actors (static and dynamic) in the scenario with visibility limitations, latency, and discretized wait conditions.
Focused on motion control	Derives control restrictions (e.g., longitudinal, and lateral accelerations) based on minimum safe distance violations	Rulebooks can be used to assess planned or executed trajectories	Calculates a sound control policy.
Incorporates assumptions	RSS is a parameterized model that considers assumptions about reasonably foreseeable behaviors of road users (e.g., speed and acceleration capabilities)	Rulebooks does not explicitly incorporate assumptions but defines a proper response to assumed behavior of other actors. Rulebooks supports rules that incorporate assumptions	SFF takes the world state into account so various assumptions about ego and other actors can be incorporated.

Attribute	RSS	Rulebooks	SFF
Based on current position, heading and velocity of other safety-relevant objects	RSS defines minimum safe longitudinal and lateral distances and proper responses based on current position, heading and velocity of other safety-relevant objects	Rules can be based only on specific attributes of other safety-relevant objects	Models based on current time measurements.
Supports prioritization of safety goals	The proper response definition is by design free of contradictions at a global level as it is performed in a pairwise structure considering the full complexity of the driving situation, including traffic rules	Rules and priority structure designed to generate explainable, predictable behavior	Includes caveat to "allow an actor to not engage in their safety procedure if it does not help." Can add extensions for additional safety goals.
Is sensitive to adjustment in parameter values	RSS is a parameterized model that adjusts the minimum safe distances and proper response calculations based on its inputs	All rules are parameter based; does not use machine learning	Parameter based.
Supports diverse safety-relevant objects	Allows for the adjustment of safe minimum distances and proper response for VRUs (e.g., pedestrians, motorcycles, etc.) and vehicles (e.g., four-wheel vehicles, trucks)	Rulebooks supports formalization of different behaviors for different objects	Actors and claimed sets are very general. They can be from any road user (e.g. truck, pedestrian, bicycle, commercial vehicle, bike, bus) with all possible configurations of vehicle model/margin etc.
Includes emergency maneuvers	Defines evasive maneuver actions to take if longitudinal and lateral minimum safe distances are violated due to improper behavior of other road users (e.g., other road users' behavior outside the assumptions)	Emergency maneuver emerges from rules	Defines action to take if SFF is violated. Emergency maneuvers are defined as a safety procedure or safe control policies.
Defines a hazardous situation	Hazardous situations are defined as "dangerous situations," when the longitudinal and lateral minimum safe distances are violated	Rule violations can be considered definition of hazardous situation	Assume that a hazardous situation arises when SFF is violated when the claimed sets overlap.

Attribute	RSS	Rulebooks	SFF
Defines proper responses	Defines a proper response in all cases that specify actions to be taken when there is a violation of the minimum safe distances	Response to hazardous situation emerges from rules and priority structure, but is not explicitly defined	The ego vehicle adjusts to maintain SFF.
Differentiates between initiator and responder	Defines a proper response in all cases that specify actions to be taken by the initiator and/or the responder. This allows responsibility to be assigned in cases where a crash happens	Rules differentiate between the behavioral expectations of the initiator and responder; rules can be formulated to be only violated by one actor (initiator)	Has concept of Out-of-Policy Detection. Can have a symmetric case where both are required to act. The one who fails to act would be the initiator, but if both fail to act then both are initiators.
Supports directional flexibility	Considers road users driving in opposite direction and in unstructured scenes with no clear path or lanes	Rules apply on unstructured roads	SFF can use fixed properties of the world (e.g., road shape or map) but is not constrained if these are unavailable.
Supports occlusion scenarios	Assumes reasonable behavior for occluded safety-relevant objects	Supports rules addressing occlusion scenarios, but occlusion is not explicitly addressed	SFF uses more general term of visibility; assumes worst case after "excluding extreme states of actors that may be physically possible but would force us to behave too conservatively for practical use if we have to take them into account."
Defines a safety envelope	Defines minimum safe lateral and longitudinal distances to be maintained in all cases	Rules support multiple safety envelope definitions	Defines safety procedures and safe control policies.
Considers reasonably foreseeable events regarding right-of-way	Defines Rule 3 based on axiom: "Right-of-way is given, not taken." The model performs calculations using the assumptions about reasonably foreseeable behaviors of other road users to determine whether the other safety-relevant object will respect right-of-way or not.	Rules support specification of ADS-operated vehicle responsibilities both with legal right-of-way and without legal right-of-way	"The safety force field should be obeyed regardless of right-of-way. On the other hand, if we are expected to yield, additional requirements are on us beyond the safety force field. The recurring theme is that we should strive to behave in such a way that we do not cause a safety force field on the actor to whom we should yield."

Attribute	RSS	Rulebooks	SFF
Supports a theoretical guarantee of no collision upon universal adoption	Model provides inductive proof that if all safety relevant agents will adhere to the model and behave within the assumptions, responding properly to dangerous situations, then Utopia is possible, in the sense that there will be no collision	Supports combinations of rules and priority structures which satisfy no collision upon universal adoption	SFF is based on this assumption.
Is formally verifiable	Formal verification can be done based on model's assumptions about reasonably foreseeable behaviors of other road users and parameters	Rules can be implemented as behavioral requirements	SFF is based on one core concept that is formally verifiable.
Supports creation of performance indicators	Not covered in the paper but model can be used to calculate safe distances and keep track of violations as a safety performance indicator	Rules serve as behavioral metrics; to be fully specified, rules come with violation metrics that specify the degree of violation	Not covered in the paper but can support via keeping track of SFF violations; can track safety potential (changes sign when violated).
Can be expressed in formal notation	Formal notation not included in the original RSS paper, but the model's definitions could be converted to a formal notation	Rules are expressed in formal logic	SFF expressed in formal notation.
Is transparent	Full RSS model definitions published in 2017 paper and extra supporting material was also released	Framework and applications published. Small set of model rules published.	SFF is disclosed in two papers as well as blog posts and videos.
Considers weather-related environmental conditions and road surface conditions	RSS is a parametrized model where the values of the parameters represent the different assumptions and/or operating conditions of the ADS-operated vehicle and other road users, including weather conditions	Rules support specification of weather-related or other environmental conditions	Handled, for example, as loss of visibility and reduced allowable deceleration and cornering capabilities.

7. REFERENCES

The following sources either have been referenced within this paper or may be useful for additional reading:

[1] IEEE SA, IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems, https://standards.ieee.org/ieee/2846/10831/, 2022.

[2] ITS IEEE VT, "Literature Review on Kinematic Properties of Road Users for Use on Safety-Related Models for Automated Driving Systems," pp. 1-35, 2022.

[3] S. Shalev-Shwartz, S. Shammah and A. Shashua, "On a Formal Model of Safe and Scalable Self-driving Cars," arXiv:1708.06374, 2017.

[4] S. Shalev-Shwartz, S. Shammah and A. Shashua, "Vision zero: on a provable method for eliminating roadway accidents without compromising traffic throughput," arXiv preprint arXiv:1901.05022, 2018.

[5] A. Censi, K. Slutsky, T. Wongpiromsarn, D. Yershov, S. Pendleton, J. Fu and E. Frazzoli, "Liability, Ethics, and Culture-Aware Behavior Specification using Rulebooks," 2019. [Online]. Available: https://www.aptiv.com/docs/default-source/white-papers/aptiv-rulebooks.pdf.

[6] A. Collin, A. Bilka, S. Pendleton and R. D. Tebbens, "Safety of the intended driving behavior using rulebooks," in *IEEE Intelligent Vehicles Symposium (IV)*, 2020.

[7] J. De Freitas, A. Censi, B. W. Smith, L. D. Lillo, S. E. Anthony and E. Frazzoli, "From driverless dilemmas to more practical commonsense tests for automated vehicles.," in *National Academy of Sciences 118, no. 11*, 2021.

[8] W. Xiao, N. Mehdipour, A. Collin, A. Y. Bin-Nun, E. Frazzoli and a. C. B. Radboud Duintjer Tebbens, "Rule-based optimal control for autonomous driving," in *ACM/IEEE 12th International Conference on Cyber-Physical Systems*, 2021.

[9] J. Tůmová, L. I. R. Castro, S. Karaman, E. Frazzoli and D. Rus, "Minimum-violation LTL planning with conflicting specifications," in *IEEE American Control Conference*, 2013.

[10] K. Slutsky, D. Yershov, T. Wongpiromsarn and E. Frazzoli, "Hierarchical Multiobjective Shortest Path Problems," in *Workshop on the Algorithmic Foundations of Robotics*, 2020.

[11] D. Nistér, H.-L. Lee, J. Ng and Y. Wang, "The Safety Force Field," 2019. [Online]. Available: https://www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/the-safety-force-field.pdf.

RAISING THE WORLD'S STANDARDS

3 Park Avenue, New York, NY 10016-5997 USA http://standards.ieee.org

Tel.+1732-981-0060 Fax+1732-562-1571