Encoding and Monitoring Responsibility Sensitive Safety (RSS) Rules for Automated Vehicles in Signal Temporal Logic (STL)

Mohammad Hekmatnejad, Shakiba Yaghoubi, Adel Dokhanchi, Heni Ben Amor, Aviral Shrivastava, Lina Karam, and Georgios Fainekos

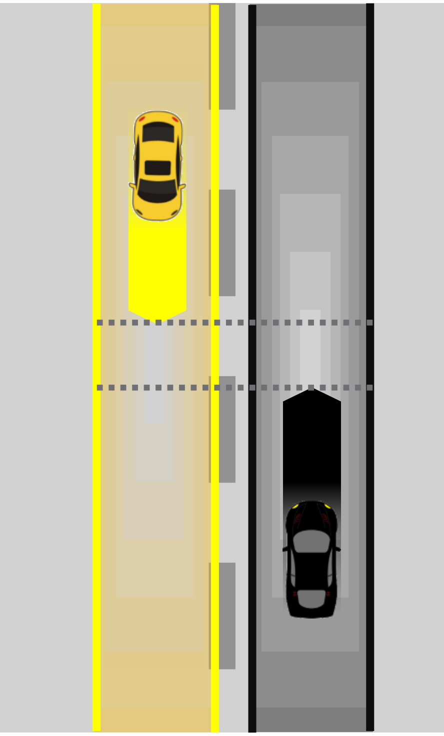
MEMOCODE 2019

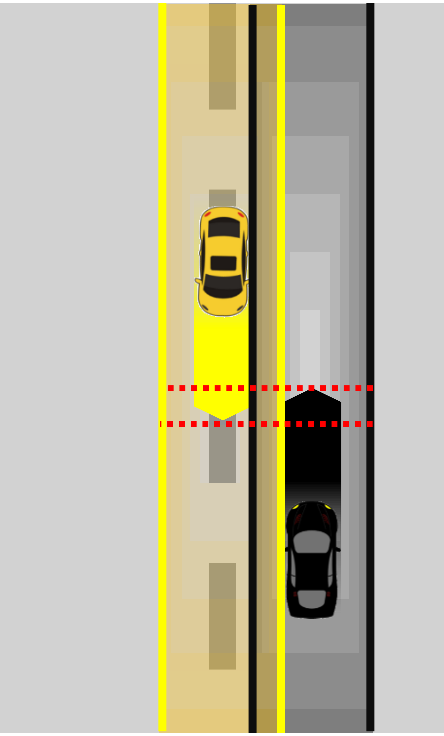
CPS Lab @ ARIZONA STATE UNIVERSITY

mhekmatn at asu edu
https://www.linkedin.com/in/mohammad-hekmatnejad-54535232

Motivation

- Responsibility Sensitive Safety (RSS) Rules
  - Developed by Intel Mobileye to capture safe driver behavior for Automated Driving Systems (ADS)
  - Alternative viewpoint: when an ADS should not be blamed for an accident



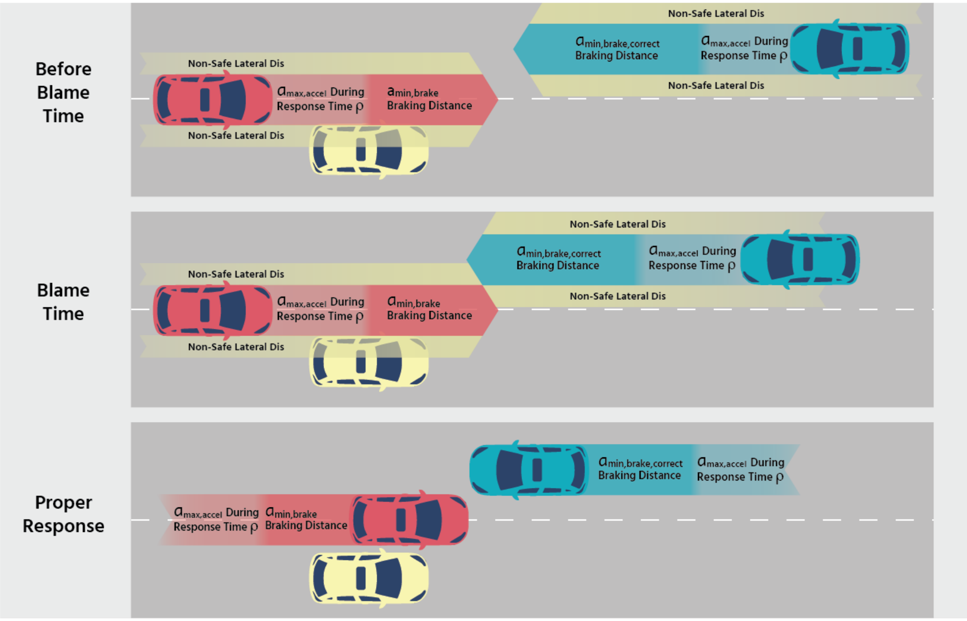


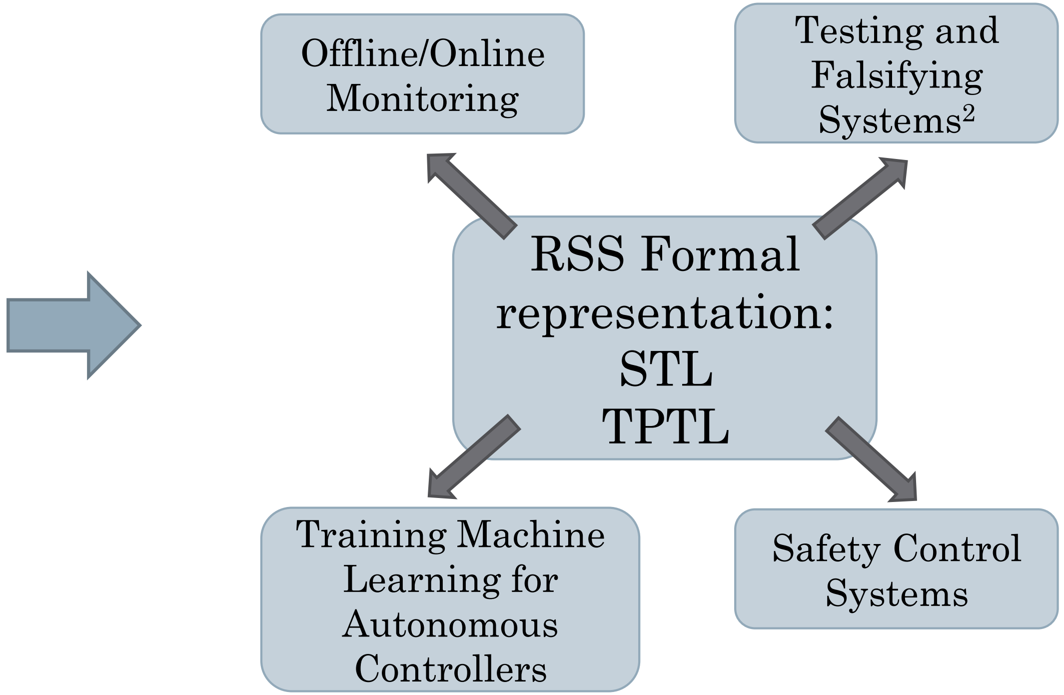
Waymo recorded video of an accident

Problem Definition & Solution Overview

Problem: How to represent and use the RSS rules in practice?

Responsible Sensitive Safety Rules[1]





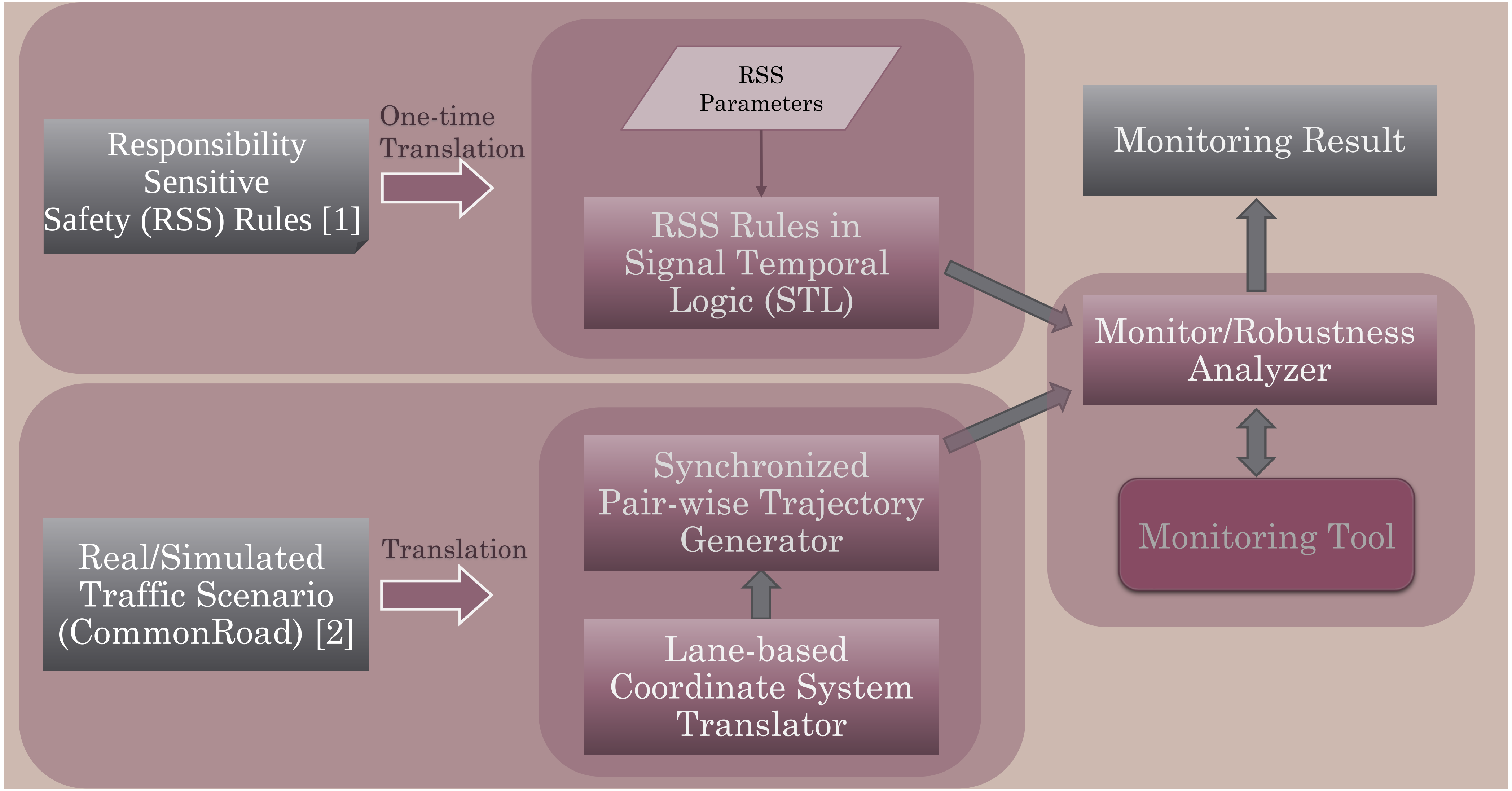
Solution: Formalizing the RSS rules in STL/TPTL

- use formalized RSS rules in standardizing, designing, training, testing and controlling ADSs.

[1] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "**On a formal model of safe and scalable self-driving cars**," arXiv:1708.06374v6, 2018.
[2] Cumhur Erkan Tuncali, Georgios Fainekos, Hisahiro Ito, James Kapinski, "**Sim-ATAV: Simulation-based Adversarial Test generation framework for Autonomous Vehicles (AV)**", HSCC 2018

* Figure is taken from Mobileye "**Implementing the RSS Model on NHTSA Pre-Crash Scenarios**"

Solution Architecture



[1] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "**On a formal model of safe and scalable self-driving cars**," arXiv:1708.06374v6, 2018.
[2] Matthias Althoff, Markus Koschi, and Stefanie Manzinger, "**CommonRoad: Composable Benchmarks for Motion Planning on Roads**", 2017 IEEE Intelligent Vehicles Symposium (IV)

Summary of Our Contribution

- We demonstrate that the RSS model can be encoded in assume-guarantee STL requirements.
- To motivate how the resulting STL requirements could be used in practice, we monitor multiple real driving data scenarios* offline over some of the RSS rules written in STL [1].
- Finally, we have released our case-study and experiments publicly available as part of S-TALIRO available at: https://cpslab.assembla.com/spaces/s-taliro_public/ .

[1] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "**On a formal model of safe and scalable self-driving cars**," arXiv:1708.06374v6, 2018.
* Matthias Althoff, Markus Koschi, and Stefanie Manzinger, "**CommonRoad: Composable Benchmarks for Motion Planning on Roads**", 2017 IEEE Intelligent Vehicles Symposium (IV)

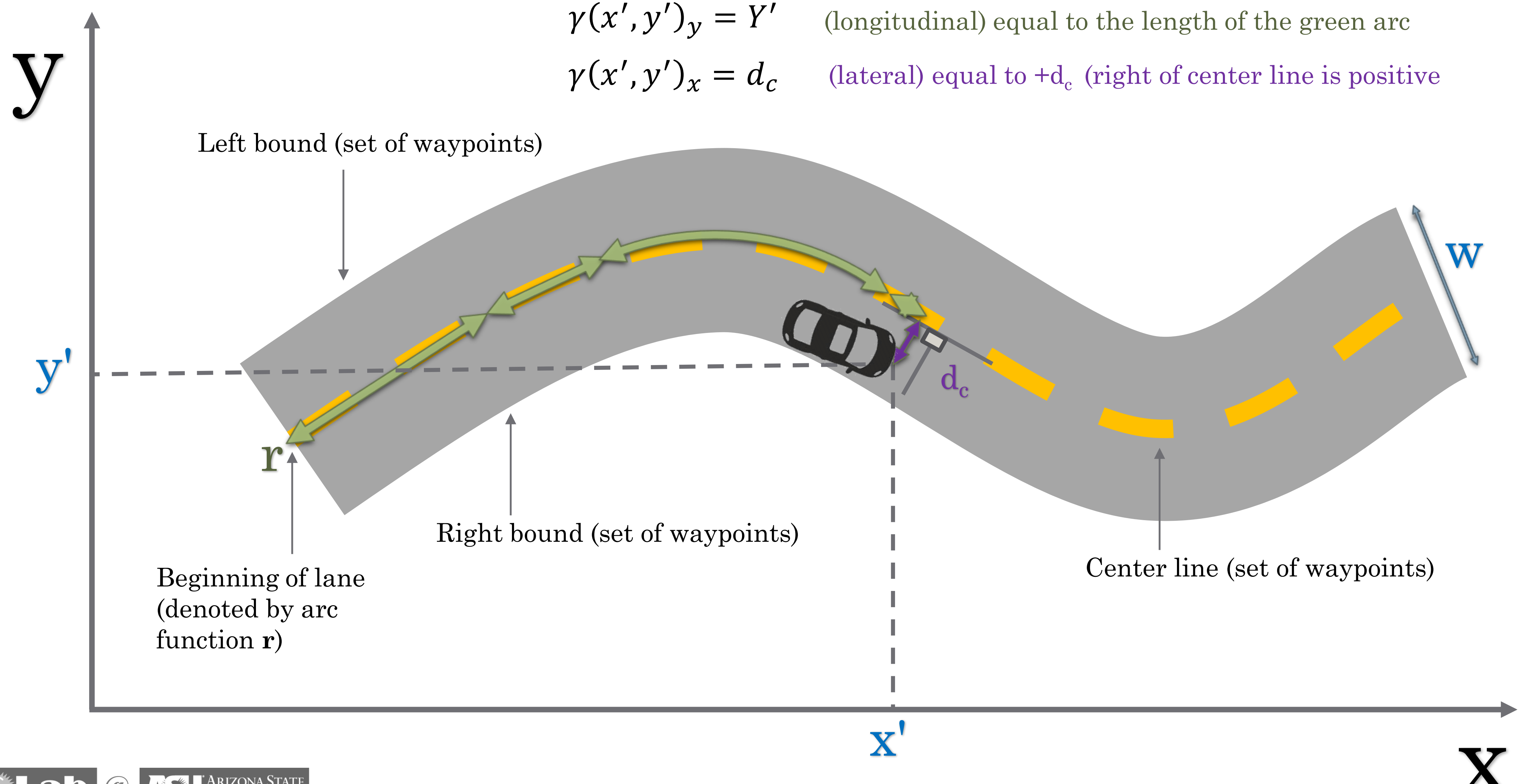
Outline

- Preliminaries
  - Lane-based Coordinate System
  - RSS Safe Distances
  - Metric/Signal Temporal Logic
- RSS Translation into STL
- Monitoring RSS Rules in DP-TALIRO
- Experimental Results
- Conclusion

Lane-based Coordinate System

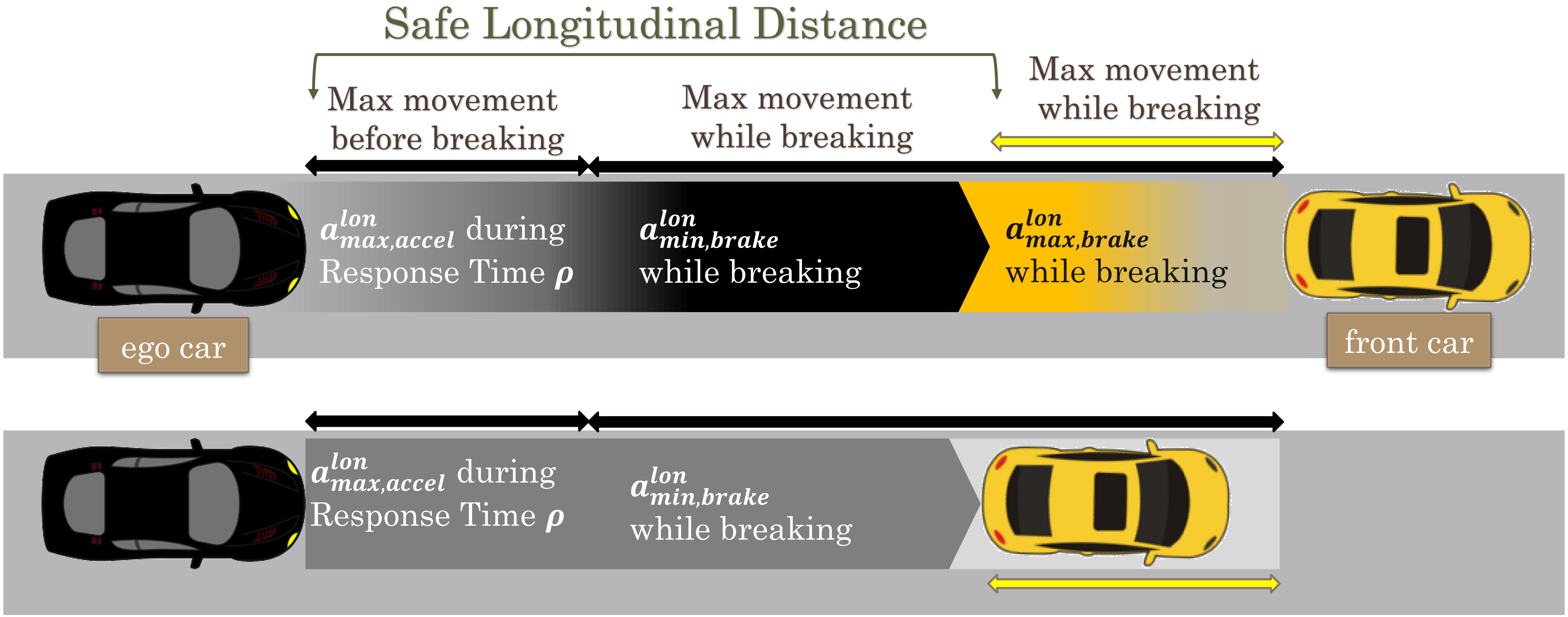
$\gamma(x',y')_y = Y'$ (longitudinal) equal to the length of the green arc

$\gamma(x',y')_x = d_c$ (lateral) equal to $+d_c$ (right of center line is positive



Safe Longitudinal Distance in One-Way Traffic

All cars move at the same direction from left to the right



Longitudinal Minimum Safe Distances

- **Based on Lemma 2 of RSS [1]:**
- Ego vehicle ***b*** is always behind the Front ***f***

$$d_{min,lon} = \max\left(d_{b,preBrake} + d_{b,brake} - d_{f,brake}, 0\right),$$

- Maximum frontal movement by accelerating as maximally allowed (before taking any action w.r.t reaction time)

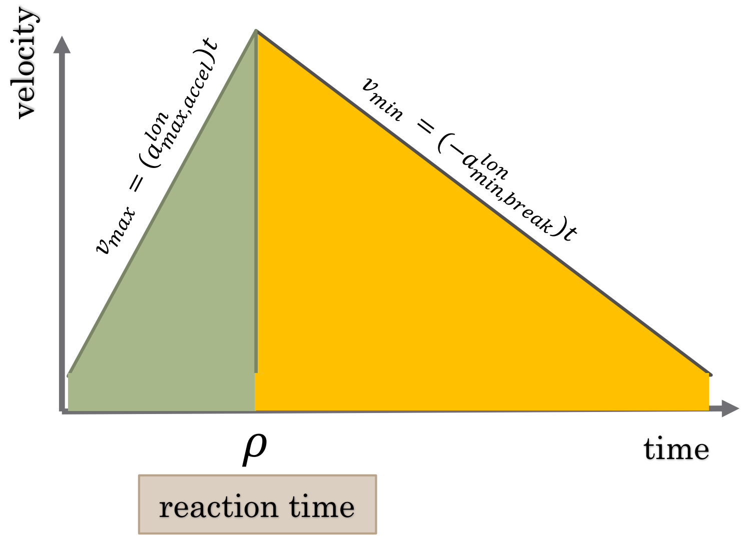
$$d_{b,preBrake} = v_b^{lon}\rho + \frac{1}{2}a_{max,accel}^{lon}\rho^2$$

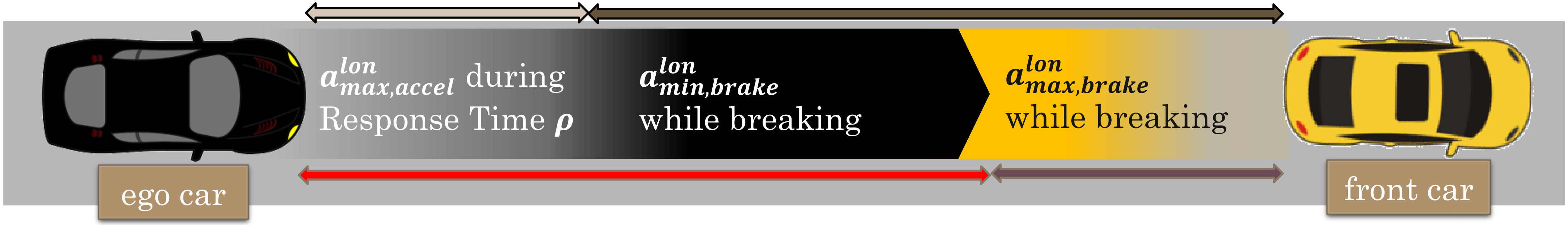
- Maximum frontal movement after braking as minimally required

$$d_{b,brake} = \frac{\left(v_b^{lon} + \rho a_{max,accel}^{lon}\right)^2}{2a_{min,brake}^{lon}}$$

- Minimum frontal movement by braking as maximally allowed

$$d_{f,brake} = \frac{{v_f^{lon}}^2}{2a_{max,brake}^{lon}}$$





[1] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "**On a formal model of safe and scalable self-driving cars**," arXiv:1708.06374v6, 2018.

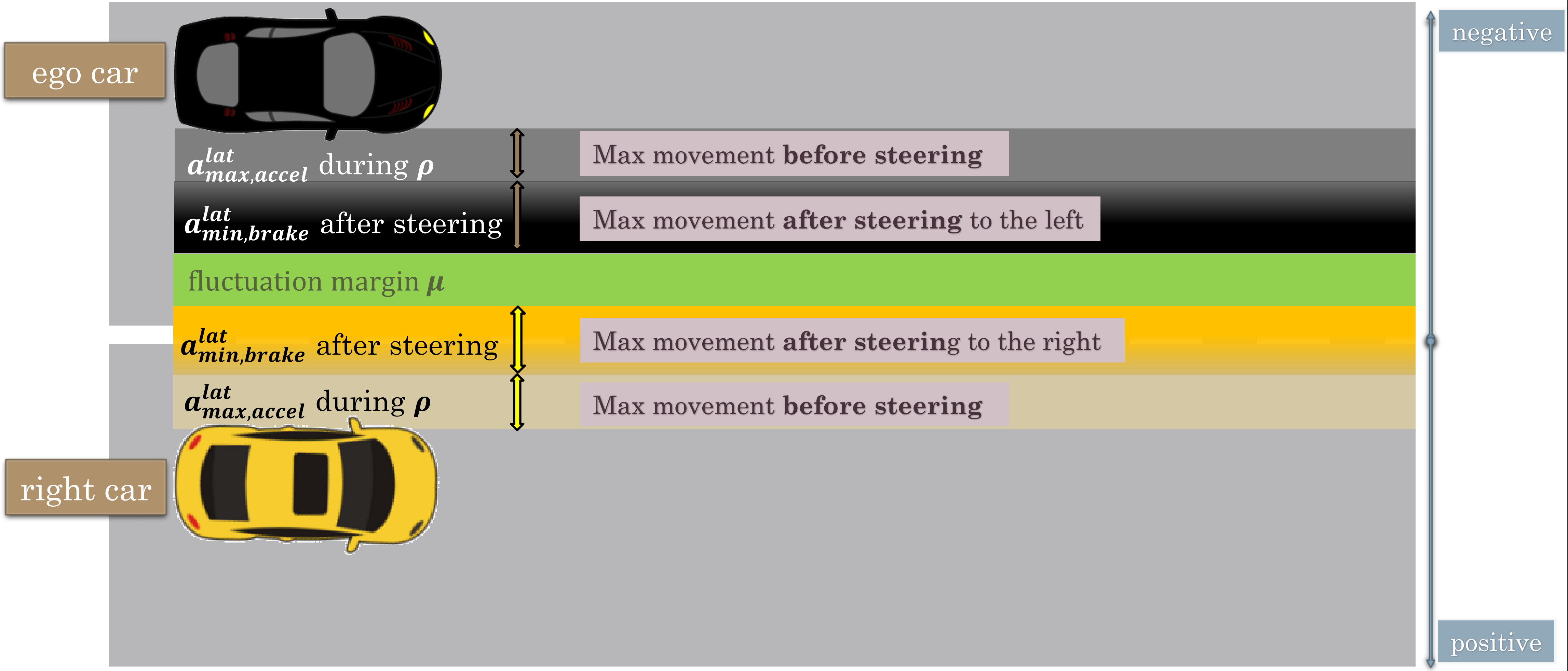
Longitudinal Minimum Safe Distances (cont')

- $D_{f,b} = longitudinal\ disntance\ - d_{min,lon}$
- $D_{f,b} > 0$ is **safe**
- $D_{f,b} \leq 0$ is **unsafe**

- Longitudinal **dangerous threshold time** is as follows:
- was $(D_{f,b} > 0), and\ now\ (D_{f,b} < 0)$

Safe Lateral Distance in One-Way Traffic

All cars move at the same direction from left to the right

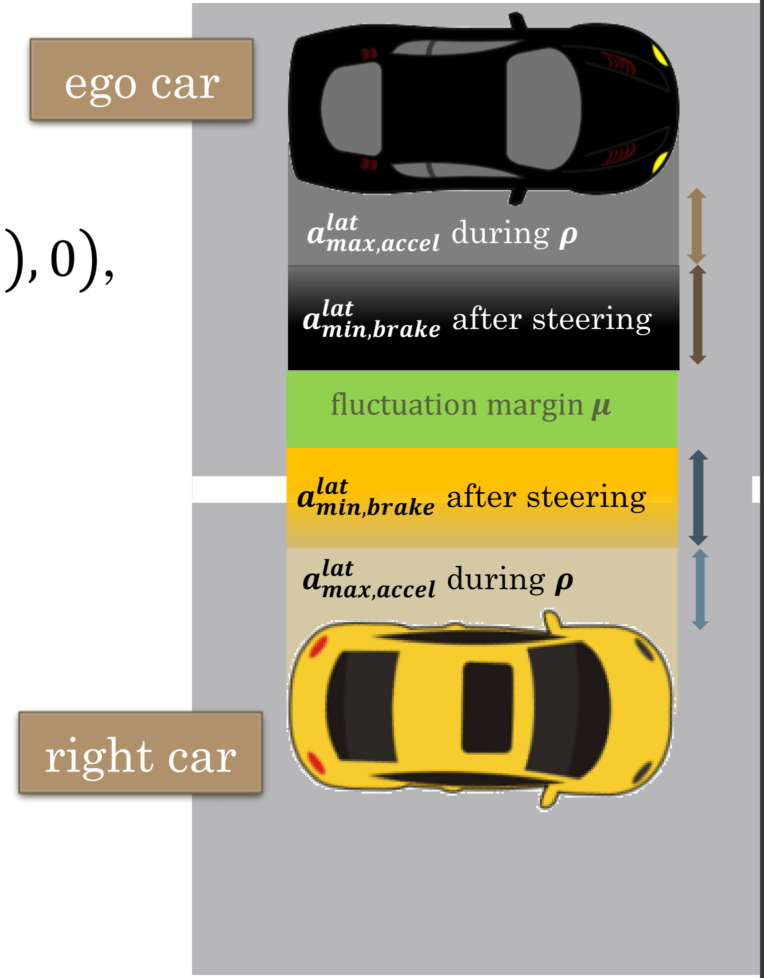


Lateral Minimum Safe Distances

- **Based on Lemma 4 of RSS [1]:**
- If Ego vehicle l is on the left of any car in the Front r

$$d_{min,lat} = \boldsymbol{\mu} + \max\left(d_{l,preBrake} + d_{l,brake} - \left(d_{r,preBrake} - d_{r,brake}\right), 0\right),$$

- Maximum to the right movement by accelerating as maximally allowed (before taking any action w.r.t reaction time) $d_{l,preBrake} = \frac{v_l^{lat} + v_{l,\rho}^{lat}}{2}\rho$
- Maximum to the right movement after braking as minimally required $d_{l,brake} = \frac{{v_{l,\rho}^{lat}}^2}{2a_{min,brake}^{lat}}$
- Maximum to the left movement by accelerating as maximally allowed (before taking any action w.r.t reaction time) $d_{r,preBrake} = \frac{v_r^{lat} + v_{r,\rho}^{lat}}{2}\rho$
- Maximum to the left movement after braking as minimally required $d_{r,brake} = \frac{{v_{r,\rho}^{lat}}^2}{2a_{min,brake}^{lat}}$



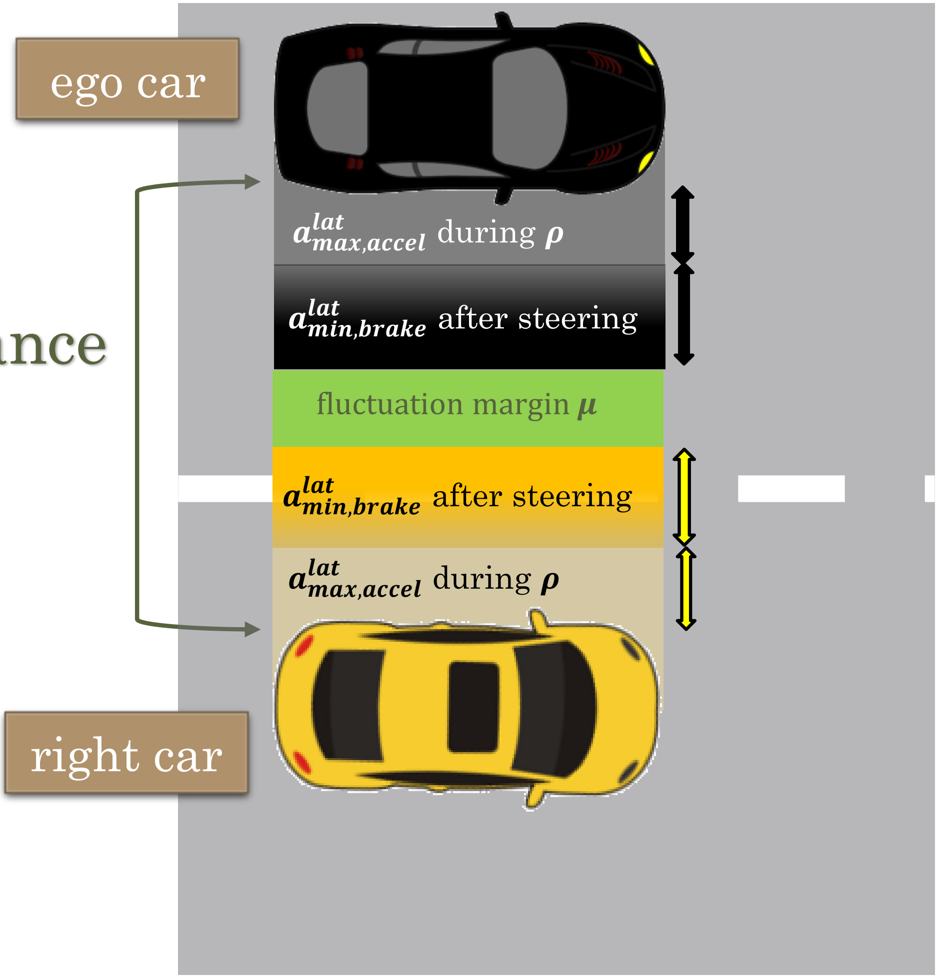
$$v_{l,\rho}^{lat} = v_l^{lat} + \rho a_{max,accel}^{lat}, \quad v_{r,\rho}^{lat} = v_r^{lat} - \rho a_{max,accel}^{lat}$$

[1] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "**On a formal model of safe and scalable self-driving cars**," arXiv:1708.06374v6, 2018.

Lateral Minimum Safe Distances (cont')

- $D_{l,r} = lateral\ disntance\ - d_{min,lat}$
- $D_{l,r} > 0$ is **safe**
- $D_{l,r} \leq 0$ is **unsafe**

Safe Lateral Distance



- Lateral **dangerous threshold time** is as follows:
- was $(D_{l,r} > 0)$, *and now* $(D_{l,r} < 0)$

Metric Temporal Logic* (MTL)

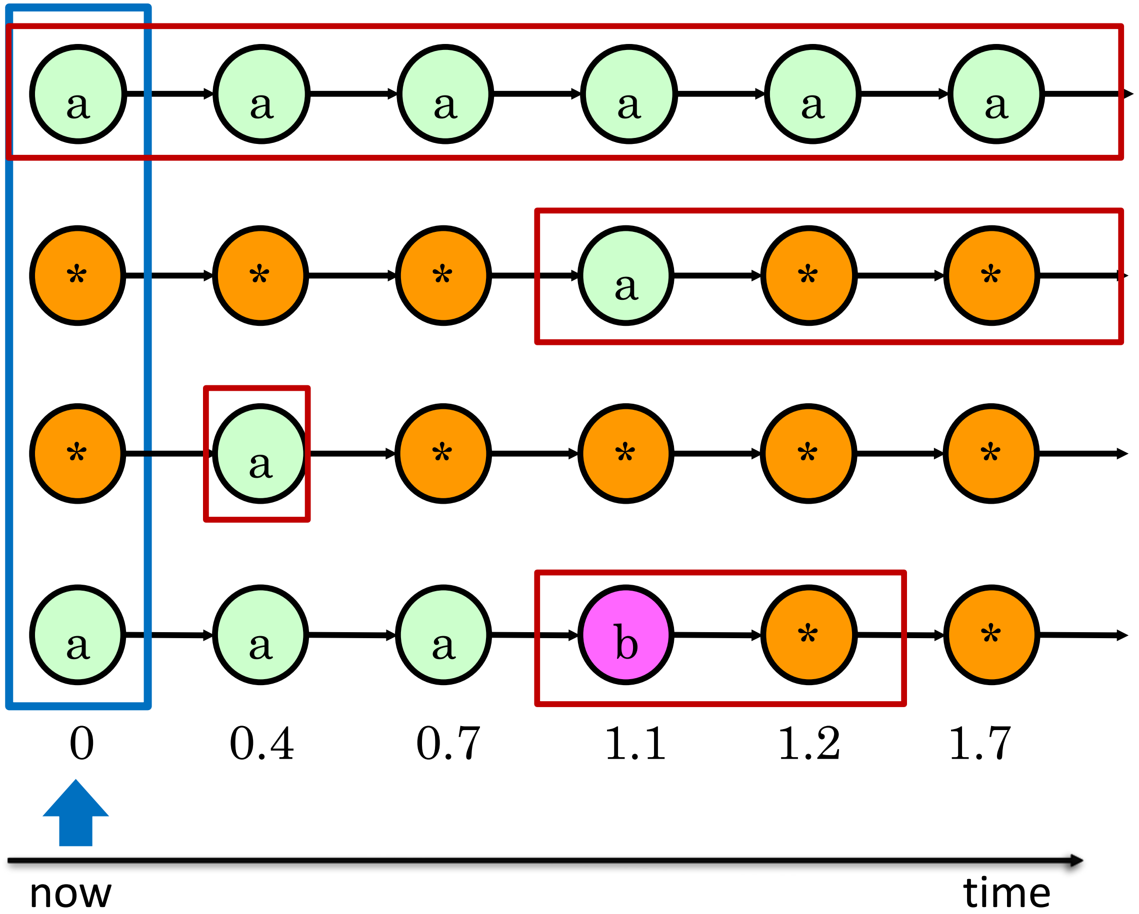
- Syntax: $\phi \ ::= \top \ |p| \ \neg\phi \ | \ \phi_1 \vee \phi_2 | \Box_I \phi \ | \Diamond_I \phi | \bigcirc\phi | \ \phi_1 U_I \phi_2 | \ \phi_1 R_I \phi_2$
- Semantics:

$G_{[0,\infty)} a \equiv \Box_{[0,\infty)} a$ - Always a

$F_{[1,3]} a \equiv \Diamond_{[1,3]} a$ - Eventually a

$Xa \equiv \bigcirc a$ - Next a

$a \ U_{[1,1.5]} \ b$ -a until b



* R. Koymans "Specifying real-time properties with metric temporal logic" Real-Time Systems, 2(4):255–299, 1990

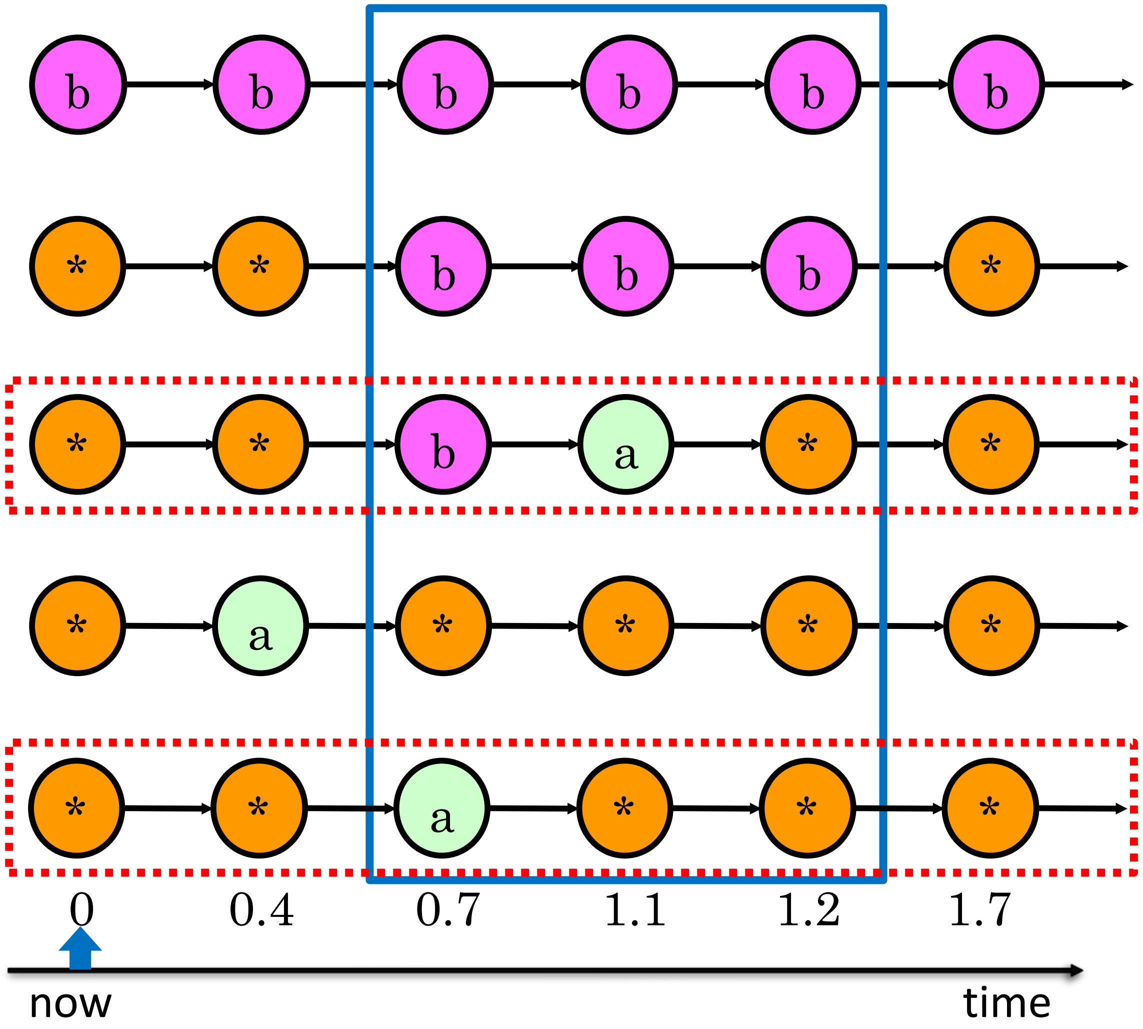
Metric Temporal Logic* (MTL)

- Syntax: $\phi ::= \top \mid p \mid \neg\phi \mid \phi_1 \vee \phi_2 \mid \Box_I \phi \mid \Diamond_I \phi \mid \bigcirc \phi \mid \phi_1 U_I \phi_2 \mid \phi_1 R_I \phi_2$
- Semantics:

$a\,\overline{R}_{[0.5,1.5]}\,b$ - a release b

Satisfy b in the interval [0.5,1.5] unless a has happened in the past.

The requirement to satisfy b in the interval [0.5,1.5] is released when a was true in the past.



* R. Koymans "Specifying real-time properties with metric temporal logic" Real-Time Systems, 2(4):255–299, 1990

Longitudinal Safety Requirements

- **Longitudinal Safety Requirement for Ego vehicle:**

$$\varphi^{lon}_{resp} \equiv \square\left(\left(S^{lon}_{b,f} \wedge \bigcirc \neg S^{lon}_{b,f}\right) \rightarrow \bigcirc\, \boldsymbol{P}^{lon}\right)$$

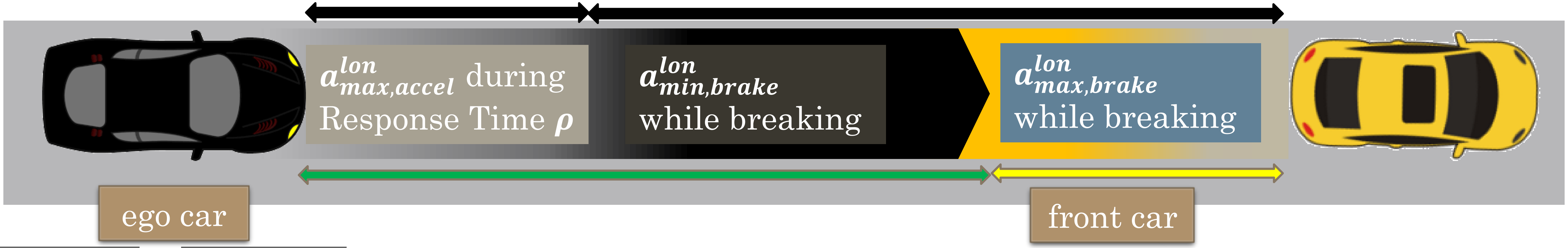
$$\boldsymbol{P}^{lon} \equiv \left(S^{lon}_{b,f}\, \bar{\mathcal{R}}_{[0,\rho)} \left(A^{lon}_{b,maxAcc} \wedge A^{lon}_{f,maxBr}\right)\right) \wedge \left(S^{lon}_{b,f}\, \bar{\mathcal{R}}_{[\rho,+\infty)} \left(A^{lon}_{b,minBr} \wedge A^{lon}_{f,maxBr}\right)\right)$$

$$S^{lon}_{b,f} \equiv \gamma\left(y_f, x_f\right)_y - \gamma(y_b, x_b)_y - d_{min,lon} > 0$$

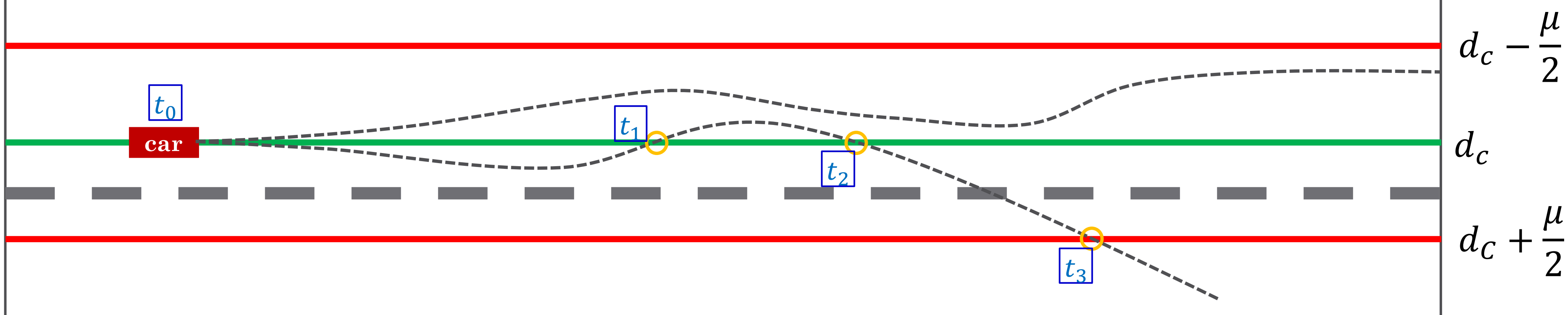
$A^{lon}_{b,maxAcc} \equiv a^{lon}_{b} \leq a^{lon}_{max,accel}$

$A^{lon}_{b,minBr} \equiv a^{lon}_{b} \leq -a^{lon}_{min,brake}$

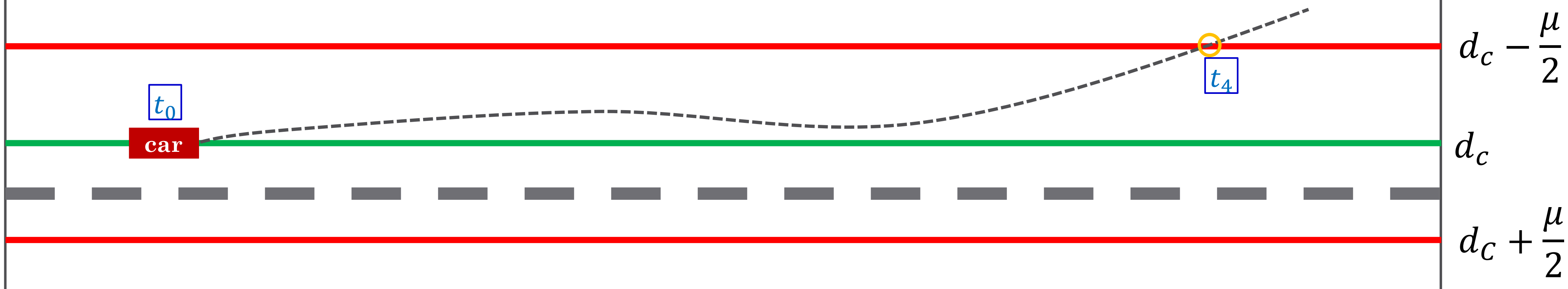
$A^{lon}_{f,maxBr} \equiv a^{lon}_{f} \geq -a^{lon}_{max,brake}$



μ –lateral-velocity



From time t_0 to t_2 predicate $\boldsymbol{V^{lat}_{l,stop}}$ is true (zero-mu-lateral velocity)



From time t_0 to t_4 predicate $\neg\boldsymbol{V^{lat}_{l,nneg}}$ is true (non-positive mu-lateral velocity)

Lateral Safety Requirements

- **Lateral Safety Requirement for Ego vehicle:**

$$\varphi_{resp}^{lat} \equiv \Box\left(\left(S_{l,r}^{lat} \wedge \bigcirc \neg S_{l,r}^{lat}\right) \rightarrow \bigcirc\, \boldsymbol{P^{lat}}\right)$$

$$\boldsymbol{P^{lat}} \equiv \left(P_{o,\rho}^{lat} \wedge P_{\rho,\infty}^{lat,1} \wedge P_{\rho,\infty}^{lat,2}\right)$$

$$P_{o,\rho}^{lat} \equiv S_{l,r}^{lat}\bar{\mathcal{R}}_{[0,\rho)}\left(A_{l,maxAccel}^{lat} \wedge A_{r,maxAccel}^{lat}\right)$$

$$P_{\rho,\infty}^{lat,1} \equiv \left(\left(S_{l,r}^{lat} \vee V_{l,stop}^{lat}\right)\bar{\mathcal{R}}_{[\rho,+\infty)}A_{l,minBrake}^{lat}\right) \wedge \left(\left(S_{l,r}^{lat} \vee V_{r,stop}^{lat}\right)\bar{\mathcal{R}}_{[\rho,+\infty)}A_{r,minBrake}^{lat}\right)$$

$$P_{\rho,\infty}^{lat,2} \equiv \left(S_{l,r}^{lat}\bar{\mathcal{R}}_{[\rho,+\infty)}\left(V_{l,stop}^{lat} \rightarrow \bigcirc\, V_{l,npos}^{lat}\right)\right) \wedge \left(S_{l,r}^{lat}\bar{\mathcal{R}}_{[\rho,+\infty)}\left(V_{r,stop}^{lat} \rightarrow \bigcirc\, \Box\left(V_{r,nneg}^{lat}\right)\right)\right)$$

$$S_{l,r}^{lat} \equiv \gamma(y_r, x_r)_\alpha - \gamma(y_l, x_l)_\alpha - d_{min,lat} > 0$$

$$V_{l,stop}^{lat} \equiv v_l^{\mu-lat} = 0,\ V_{r,stop}^{lat} \equiv v_r^{\mu-lat} = 0$$

$$V_{l,npos}^{lat} \equiv v_l^{\mu-lat} \leq 0,\ V_{r,nneg}^{lat} \equiv v_r^{\mu-lat} \geq 0$$

(i) Computed at signal level
(ii) Formalized as TPTL formula

ego car

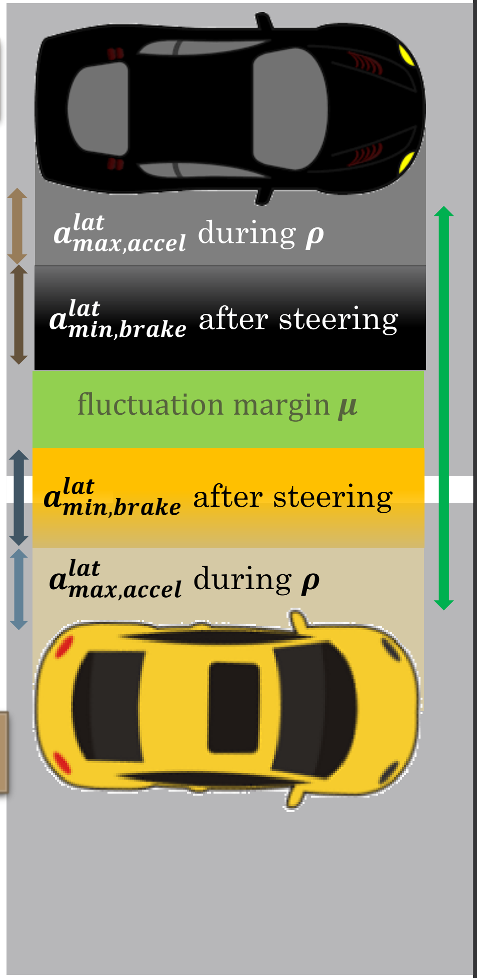
$$A_{l,maxAccel}^{lat} \equiv |a_l^{lat}| \leq a_{max,accel}^{lat}$$

$$A_{l,minBrake}^{lat} \equiv a_l^{lat} \leq -a_{min,brake}^{lat}$$

$$A_{r,minBrake}^{lat} \equiv a_r^{lat} \geq a_{min,brake}^{lat}$$

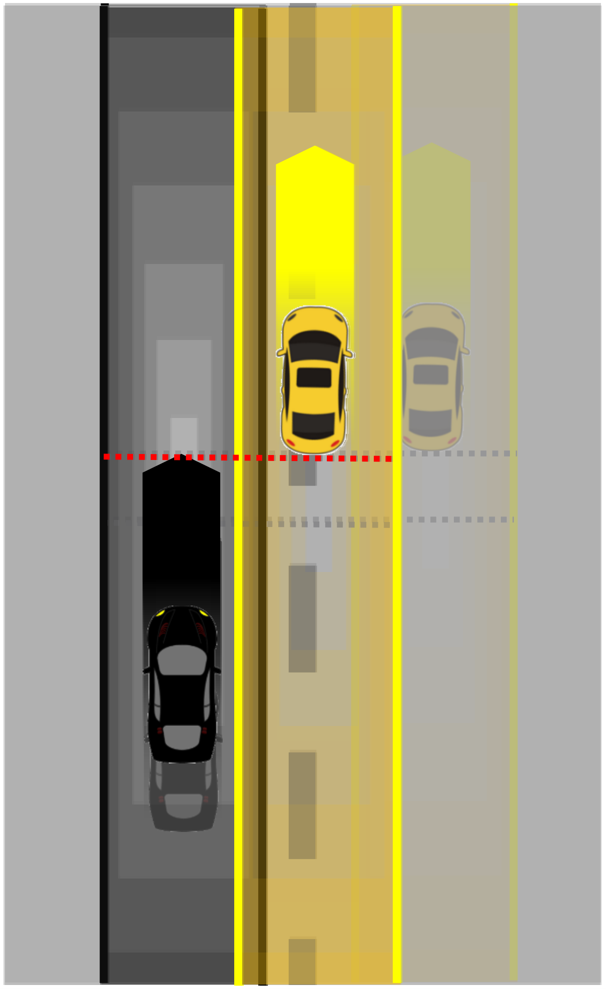
$$A_{r,maxAccel}^{lat} \equiv |a_r^{lat}| \leq a_{max,accel}^{lat}$$

right car



Basic Proper Response: From Laterally Unsafe to Unsafe

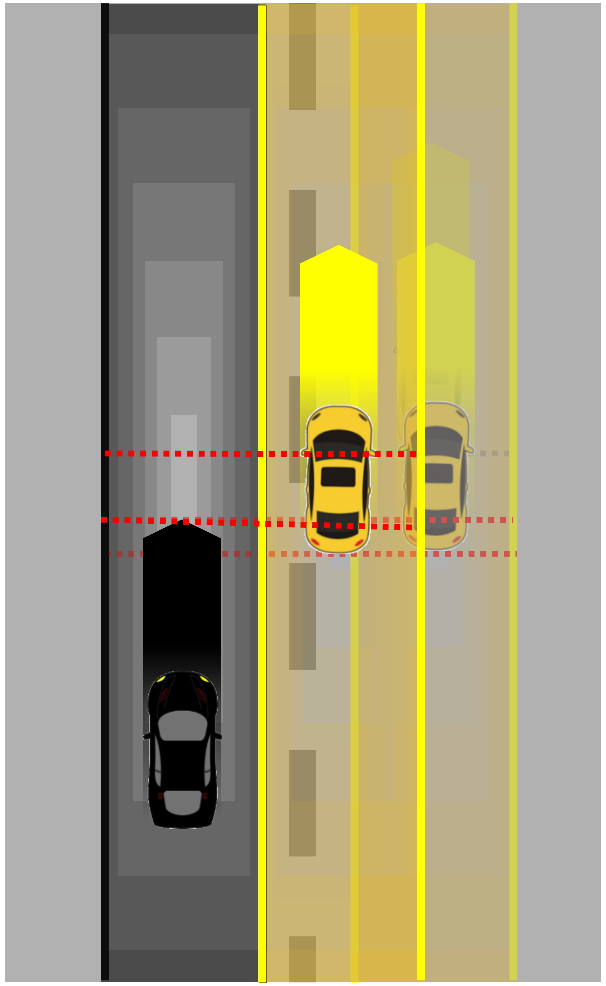
Laterally Unsafe



$$\varphi^{lon} \equiv \square\left(\left(\neg S^{lat}_{l,r} \wedge S^{lon}_{b,f} \wedge \circ\left(\neg S^{lat}_{l,r} \wedge \neg S^{lon}_{b,f}\right)\right) \rightarrow \circ\, P^{lon}\right)$$

Basic Proper Response: From Longitudinally Unsafe to Unsafe

Longitudinally UnSafe

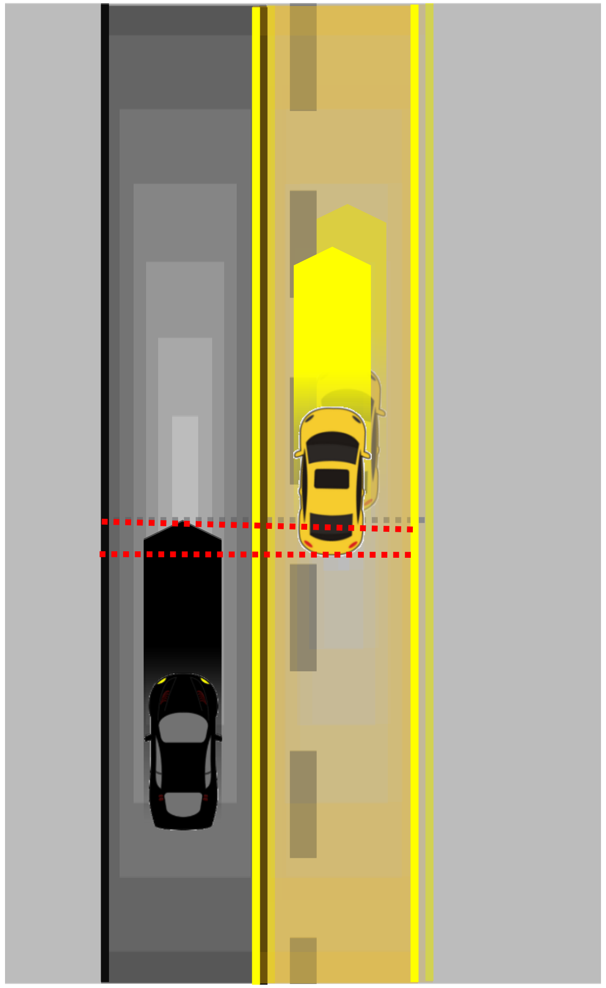


$$\varphi^{lat} \equiv \square\left(\left(\neg S_{b,f}^{lon} \wedge S_{l,r}^{lat} \wedge \circ\left(\neg S_{b,f}^{lon} \wedge \neg S_{l,r}^{lat}\right)\right) \rightarrow \circ\, P^{lat}\right)$$

Basic Proper Response: From Safe to Unsafe

Unsafe

Safe



$$\varphi^{lat,lon} \equiv \square\left(\left(S^{lat}_{l,r} \wedge S^{lon}_{b,f} \wedge \circ\left(\neg S^{lat}_{l,r} \wedge \neg S^{lon}_{b,f}\right)\right) \rightarrow \circ\left(P^{lon} \wedge P^{lat}\right)\right)$$

Basic Proper Response Specification

- $\boldsymbol{\varphi_{resp}^{lat,lon}} \equiv \varphi^{lon} \wedge \varphi^{lat} \wedge \varphi^{lat,lon}$
- $\varphi^{lon} \equiv \square\left(\left(\neg S_{l,r}^{lat} \wedge S_{b,f}^{lon} \wedge \circ\left(\neg S_{l,r}^{lat} \wedge \neg S_{b,f}^{lon}\right)\right) \rightarrow \circ\, P_{lat}^{lon}\right)$
- $\varphi^{lat} \equiv \square\left(\left(\neg S_{b,f}^{lon} \wedge S_{l,r}^{lat} \wedge \circ\left(\neg S_{b,f}^{lon} \wedge \neg S_{l,r}^{lat}\right)\right) \rightarrow \circ\, P_{lon}^{lat}\right)$
- $\varphi^{lat,lon} \equiv \square\left(\left(S_{l,r}^{lat} \wedge S_{b,f}^{lon} \wedge \circ\left(\neg S_{l,r}^{lat} \wedge \neg S_{b,f}^{lon}\right)\right) \rightarrow \circ\left(P_{lat}^{lon} \vee P_{lon}^{lat}\right)\right)$
- P_{lat}^{lon} and P_{lon}^{lat} are modified versions of P^{lon} and P^{lat} where the propositions $S_{l,r}^{lat}$ and $S_{b,f}^{lon}$ are replaced with the formula $(S_{l,r}^{lat} \vee S_{b,f}^{lon})$.

Remarks on $\boldsymbol{\varphi_{resp}^{lat,lon}} \equiv \varphi^{lon} \wedge \varphi^{lat} \wedge \varphi^{lat,lon}$

- (1)
  - $\varphi^{lat,lon} \equiv \square\left(\left(S_{l,r}^{lat} \wedge S_{b,f}^{lon} \wedge \circ\left(\neg S_{l,r}^{lat} \wedge \neg S_{b,f}^{lon}\right)\right) \rightarrow \circ\left(P_{lat}^{lon} \vee P_{lon}^{lat}\right)\right)$

$\varphi^{lat,lon}$ is implicitly defined in Def. 10.

Def 10 implies conjunction; however this is too conservative.

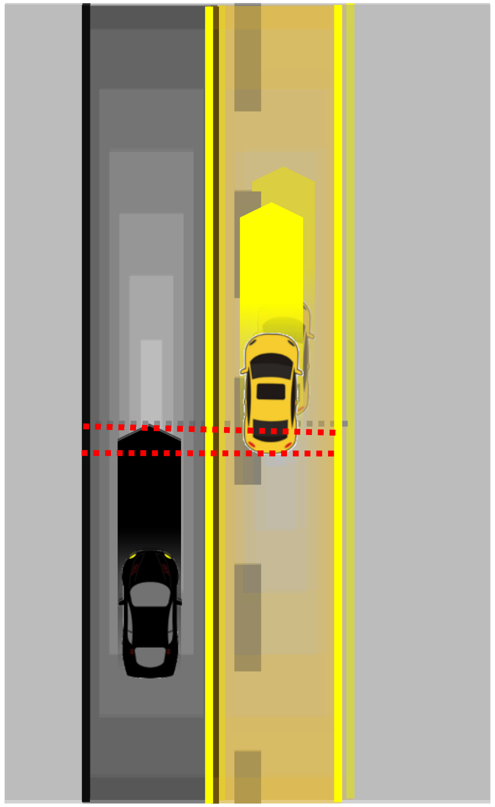
$P^{lon} \wedge P^{lat}$

- (2)
  - $S_{l,r}^{lat}\bar{\mathcal{R}}_I A^{\mathrm{lat}}$ rewritten as: $\left(S_{l,r}^{lat} \vee S_{b,f}^{lon}\right)\bar{\mathcal{R}}_I A^{\mathrm{lat}}$
  - $S_{b,f}^{lon}\bar{\mathcal{R}}_I A^{\mathrm{lon}}$ rewritten as: $\left(S_{l,r}^{lat} \vee S_{b,f}^{lon}\right)\bar{\mathcal{R}}_I A^{\mathrm{lon}}$

How a situation became dangerous does not imply it must become safe the same way

- (3)
  - $\boldsymbol{\varphi_{resp}^{lat,lon}} \equiv \varphi^{lon} \wedge \varphi^{lat} \wedge \varphi^{lat,lon} \wedge \varphi^{\neg lat,\neg lon}$
  - $\varphi^{\neg lat,\neg lon} \equiv \left(\neg S_{l,r}^{lat} \wedge \neg S_{b,f}^{lon}\right) \rightarrow \circ\left(P_{lat}^{lon} \vee P_{lon}^{lat}\right)$

What if a situation is unsafe from the beginning



Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2018). **On a formal model of safe and scalable self-driving cars**. *arXiv preprint arXiv:1708.06374 v6*.

CommonRoad Real Scenarios

- A composable framework for benchmarking motion planning on roads.
- Highway scenarios without intersection
- Vehicles in the same lane move the same direction
- Longitudinal Distance: Front-Rear Safety Requirement
- Lateral Distance: Left-Right Safety Requirement

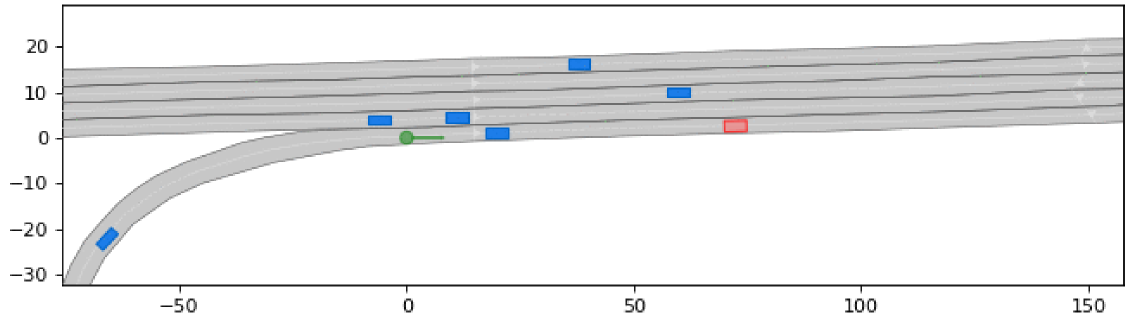
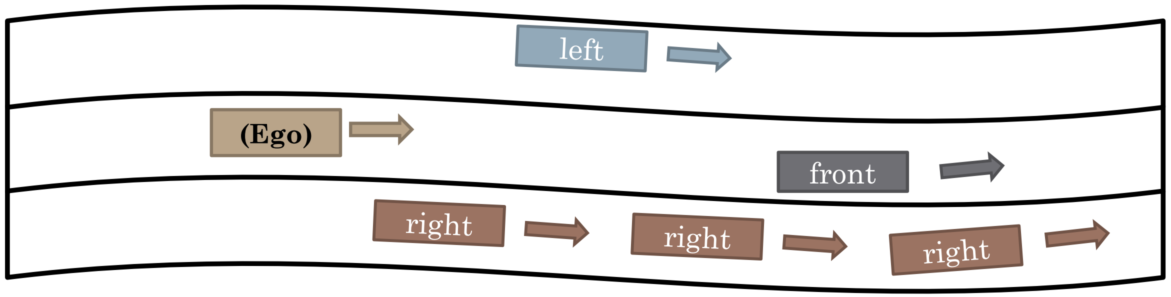
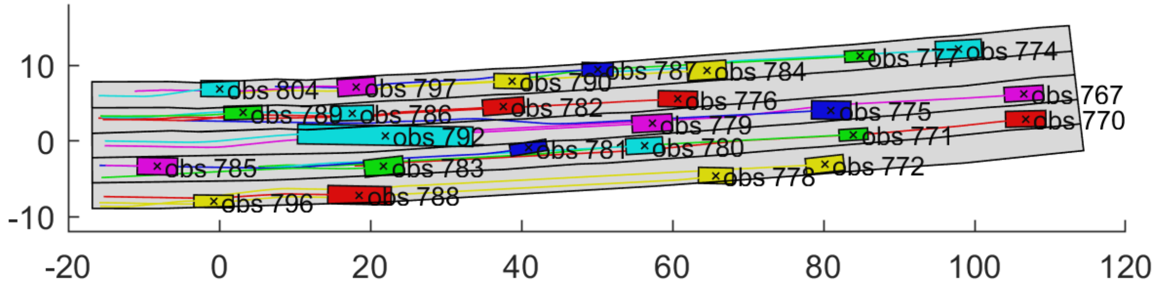
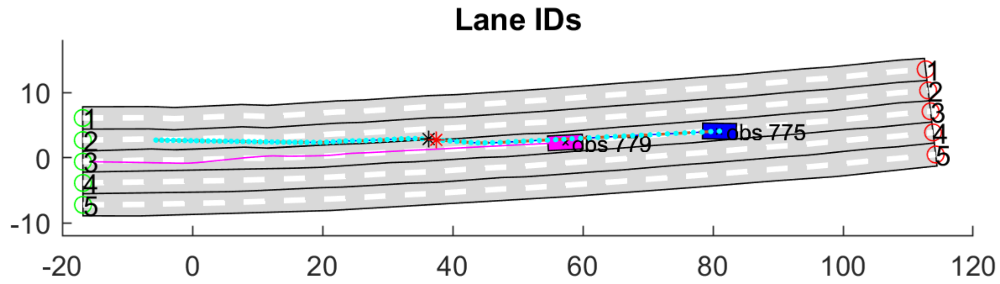


Image taken from: https://commonroad.in.tum.de



Case Study



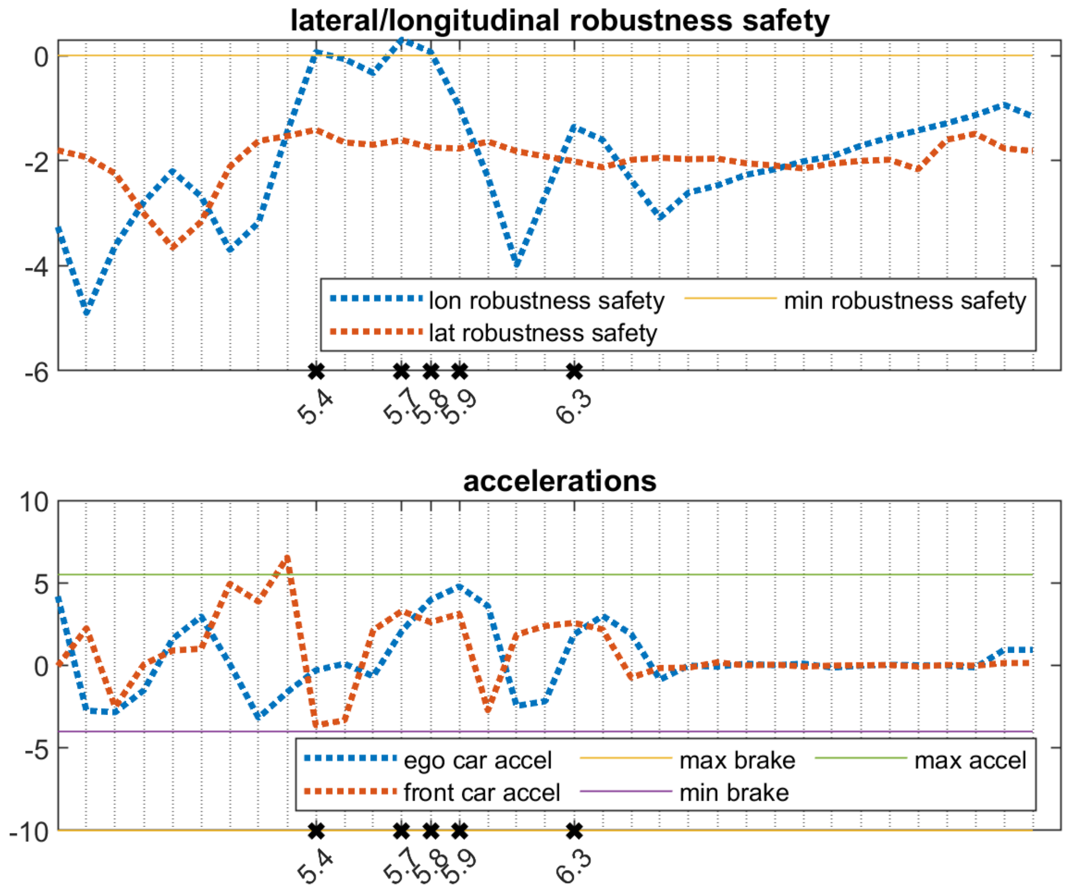


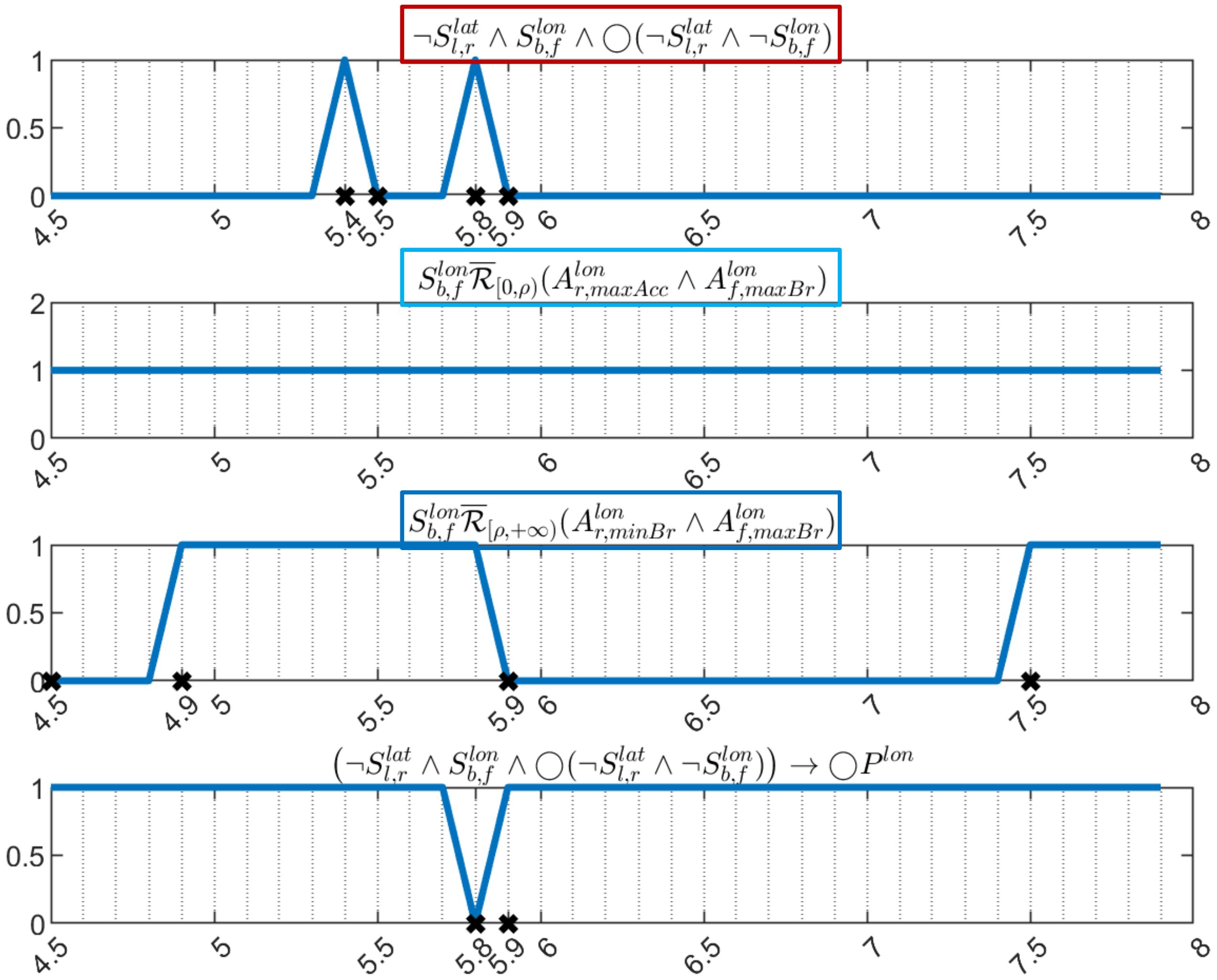
- $a^{lon}_{max,acc} = 5.5\ m/s^2$
- $a^{lat}_{max,acc} = 3\ m/s^2$
- $a^{lon}_{min,brake} = 4\ m/s^2$
- $a^{lon}_{max,brake} = 10\ m/s^2$
- $a^{lat}_{min,brake} = 3\ m/s^2$
- $a^{lat}_{max,brake} = 3\ m/s^2$
- $\rho = 0.5$
- $\mu = 0.4\ m$

Safety Charts

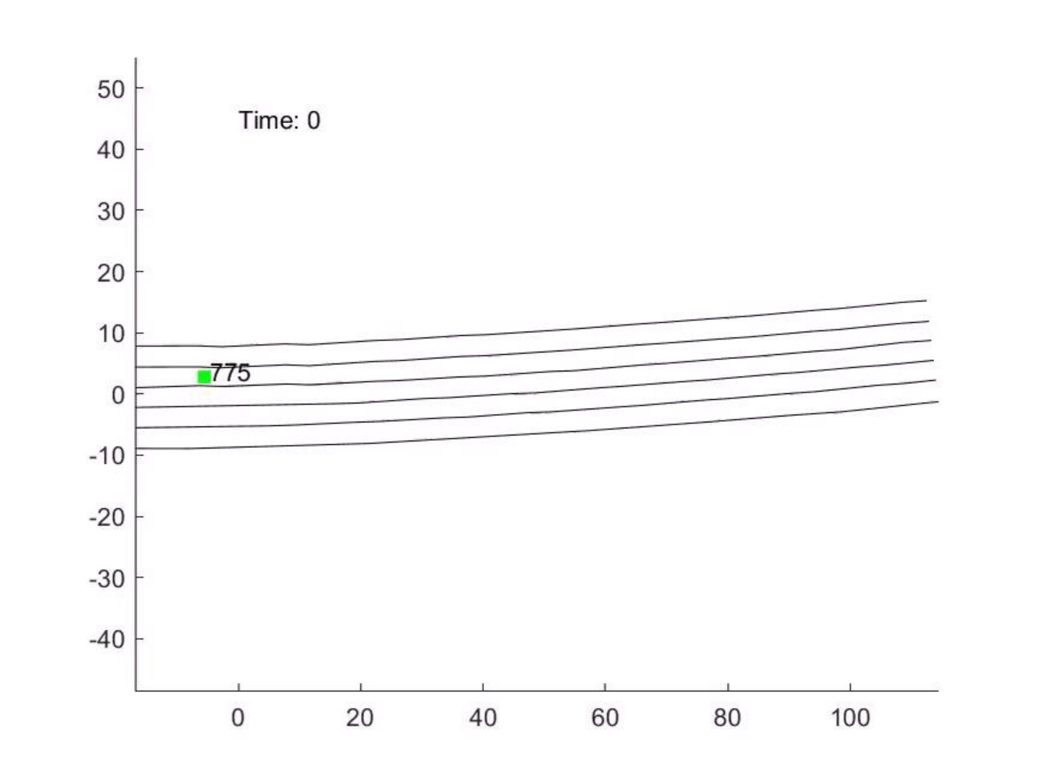
$$\varphi^{lon} \equiv \Box\left(\left(\neg S_{l,r}^{lat} \wedge S_{b,f}^{lon} \wedge \bigcirc\left(\neg S_{l,r}^{lat} \wedge \neg S_{b,f}^{lon}\right)\right) \to \bigcirc P^{lon}\right)$$

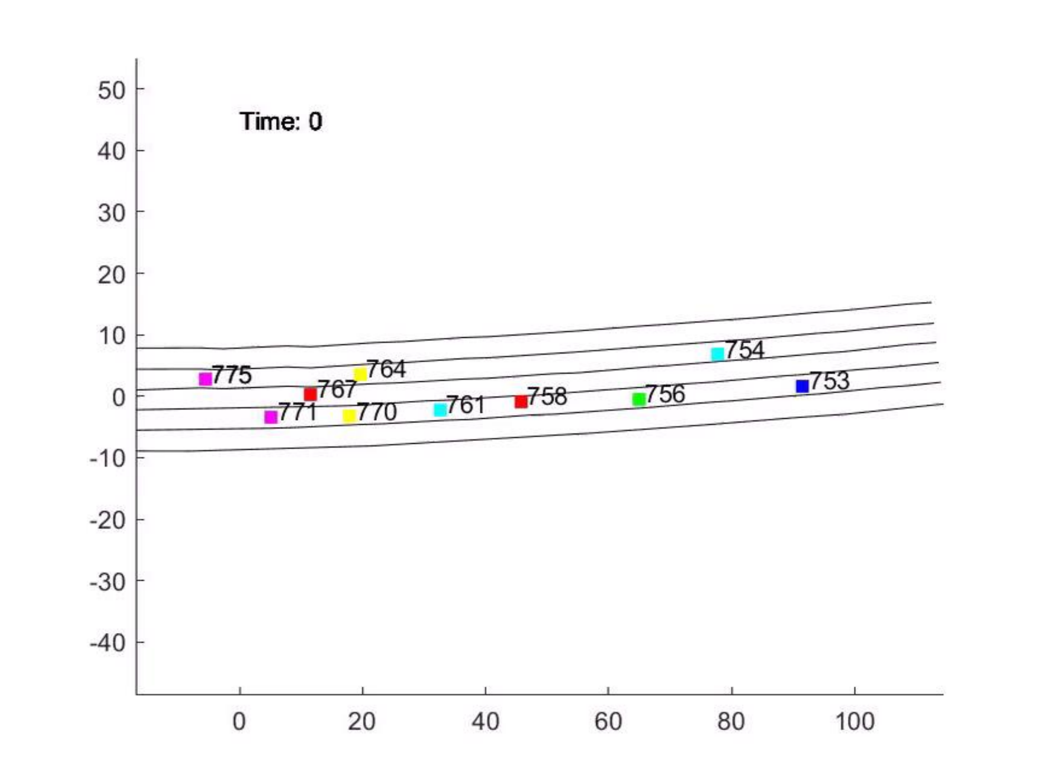
$$P^{lon} \equiv \left(S_{b,f}^{lon} \bar{\mathcal{R}}_{[0,\rho)}\left(A_{b,maxAcc}^{lon} \wedge A_{f,maxBr}^{lon}\right)\right) \wedge \left(S_{b,f}^{lon} \bar{\mathcal{R}}_{[\rho,+\infty)}\left(A_{b,minBr}^{lon} \wedge A_{f,maxBr}^{lon}\right)\right)$$





Monitoring Demo





Experimental results

Longitudinal Predicates	# of Violations φ^{lon}	# of Violations φ^{lon}_{lat}
safe_long	2	2
safe_lat	1	0
a_ego_lt_max_acc	18	18
a_ego_gt_min_brake	**190**	**184**
a_front_max_brake	9	9
Lateral Predicates	**# of Violations φ^{lon}**	**# of Violations φ^{lon}_{lat}**
safe_long	0	0
safe_lat	9	8
a_ego_lat_lt_max_acc	**188**	**186**
a_ego_lat_lt_min_brake	0	0
a_right_lat_max_acc	**256**	**256**
a_right_lat_min_brake	0	0
stopped_ego_lat	39	36
stopped_right_lat	0	0
ego_lat_velocity_neg	0	0
right_lat_velocity_pos	0	0

Lateral & Longitudinal Predicates	# of Violation $\bar{\varphi}^{lat,lon}$	# of Violation $\varphi^{lat,lon}$
safe_long	0	0
safe_lat	0	0
a_ego_lat_lt_max_acc	0	0
a_ego_lat_lt_min_brake	0	0
a_right_lat_max_acc	5	3
a_right_lat_min_brake	0	0
stopped_ego_lat	0	0
stopped_right_lat	0	0
ego_lat_velocity_neg	0	0
right_lat_velocity_pos	0	0
a_ego_lt_max_acc	0	0
a_ego_gt_min_brake	4	0
a_front_max_brake	1	1

Execution Statics		
Total violation	**722**	**703**
Violation percentage	**5.9%**	**5.74%**

Experimental results (cont’)

Lateral & Longitudinal Predicates	# of Violation $\overline{\varphi}^{\neg lat,\neg lon}$	# of Violation $\varphi^{\neg lat,\neg lon}$
safe_long	0	0
safe_lat	0	0
a_ego_lat_lt_max_acc	**172**	**166**
a_ego_lat_lt_min_brake	0	0
a_right_lat_max_acc	**177**	**161**
a_right_lat_min_brake	0	0
stopped_ego_lat	**420**	**350**
stopped_right_lat	0	1
ego_lat_velocity_neg	0	0
right_lat_velocity_pos	0	0
a_ego_lt_max_acc	6	7
a_ego_gt_min_brake	5	3
a_front_max_brake	0	1

Execution Statics		
Total violation	**780**	**689**
Violation percentage	**6.37%**	**5.63%**

item	
Average runtime per monitor execution (*ms*)	21
Average number of cars in each scenario	48
Average number of surrounding cars to be monitored	8.8
Average length of trajectories per car (*s*)	6.8

Sensitivity Analysis

parameter	values								
$a^{lon}_{max,acc}$	2.75			5.5			8.25		
$a^{lat}_{max,acc}$	1.5			3			4.5		
$a^{lon}_{max,brake}$	5			10			15		
$a^{lon}_{min,brake}$	6			4			2		
$a^{lat}_{min,brake}$	4.5			3			1.5		
ρ	0.3	0.5	2	0.3	0.5	2	0.3	0.5	2
Violations %	**0.5%**	**0.8%**	**11%**	**2.3%**	**5.2%**	**15.5%**	**6.7%**	**15%**	**23.1%**

Conclusions

- Translation of the Responsibility-Sensitive Safety (RSS) rules into Signal Temporal Logic (STL)
- The encoded formulas could be used for
  - ADS model verification
  - Automated test case generation for discovering control software bugs (our Sim-ATAV framework*)
  - Test the control and perception system stack against the RSS model
- We utilized the STL formulas to monitor off-line naturalistic driving data provided with CommonRoad.
- Computation is efficient
- The RSS rules are satisfied in the majority of the actual vehicle trajectories (assuming fast reaction times by the drivers).
- **Future works:**
  - We are completing all the RSS rules in our translation.
  - Formalize in STL the RSS rules concerning different road geometries.

* C. E. Tuncali, G. Fainekos, H. Ito, and J. Kapinski, "**Simulationbased adversarial test generation for autonomous vehicles with machine learning components**," in IEEE Intelligent Vehicles Symposium (IV), 2018.

Thank You!

Acknowledgement:
This work was partially supported by NSF 1350420 and by a gift from Intel.