A quick look at road safety and risk assessment for autonomous vehicles: The importance or virtualization.

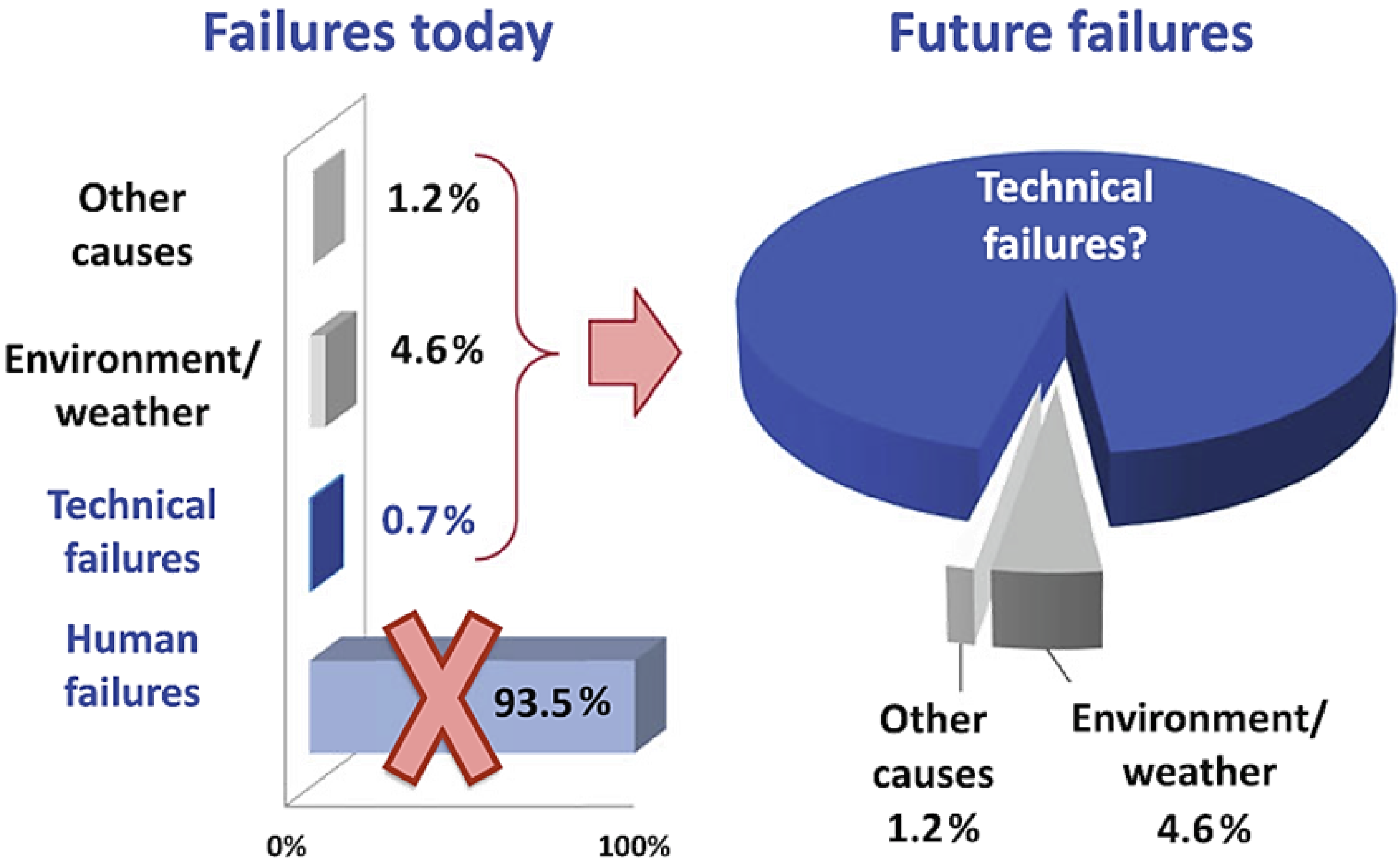
Professor Denis Gingras, Dr Ing.
Laboratory on Intelligent vehicles
Université de Sherbrooke
Quebec, Canada

OPAL-RT
TECHNOLOGIES

TABLE OF CONTENT

- □ Context
- □ Risks and safety for AVs
- □ Reliability and robustness
- □ Statistics based on VMTs
- □ System complexity and unconstrained ODD
- □ Usefulness of computational tools
- □ Conclusion

CAUSES OF ACCIDENTS SHIFTING



Source: T Winkle, GIDAS

Human errors are much more easily socially accepted than technological failures!

VEHICULAR ASSESSMENT ISSUES

- Vehicles are **no longer self-contained**; they observe and interact with their surroundings.

- **The control of the vehicle is no longer fully assumed by the driver**, if at all.

- On-board decision-making processes are complex and **based on multiple heterogeneous non-ergodic sources of information** to highly dimensional data spaces.

- **Open ODD.**

- **Vehicles are heterogeneous and evolving** with time (learning, updates).

CURRENT TYPES OF SAFETY METRICS

Based on leading measures (pre-collision):

- Computed on-board, integrated dynamic measure of driving abilities
- Safety zone determination and detection of violations (ego vehicle and surrounding objects dynamics)
- Stopping distance predictions
- Safety related infractions

Based on lagging measures (post-collision):

- Outcome measures, including crashes, injuries, etc. Data collected by Transportation agencies.
- Statistics based on VMTs (vehicle miles traveled)
- Statistics based on Hours Driven
- Statistics based on Disengagement Frequency

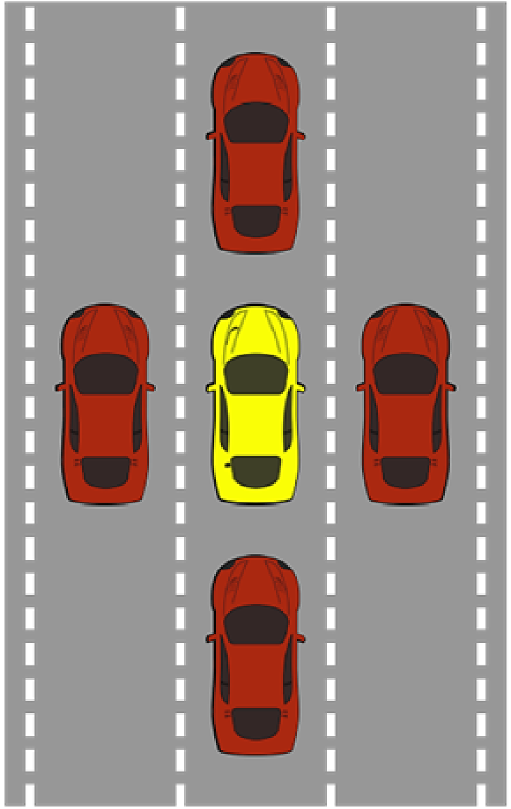
Turing tests approaches

- Measuring similarity/difference between human versus artificial driving

"Best practice" design approaches: lead to "over" complexity with excess of redundant sensors, software code etc. Expensive. Not transparent (Likely, society will not be satisfied just with the "best in class" argument when dealing with driving automation.

100% SAFE: REALLY ?

- □ Defining a safe action-taking by a car, it is impossible to achieve absolute safety, in all circumstances.
- □ A simple example: From the central yellow car’s perspective, no action can ensure that none of the surrounding cars will crash into it, and no
- □ action can help it escape this potentially dangerous situation.



Source:Shai Shalev-Shwartz, Shaked Shammah, Amnon Shashua, Mobile Eye, 2017

ROAD SAFETY DEPENDENCY ON SYSTEM PERFORMANCE

Demand for safety naturally increases with increasing automated driving tasks, since passengers must fully rely on flawless systems operation.

- ❑ **Safety** of passengers is directly dependent on the performance of vehicular embedded systems and sensors. “Safety” is a quality with a sense of duration in time and is based on two systems properties:
- ❑ **Reliability**: statistically guarantee minimum period of correct (safe) operation with a given level of confidence (time to failure). Affected by system complexity.
- ❑ **Robustness**: able to operate correctly under a wide range of conditions and able to sustain (insensitive) unexpected perturbations. Affected by ODD dimensions.

SAFETY AND VEHICLE COMPLEXITY

- ❑ Robustness means generalization.
- ❑ Generalization brings complexity.
- ❑ System complexity means lack of control.
- ❑ Lack of control brings unpredictability.
- ❑ Unpredictability means increased risk.

Driverless Car Mishap #13

SAFETY AND VEHICLE COMPLEXITY

- System complexity dramatically increases the number system behavioral responses (combinatorial NP hard problems) with
  - The number of embedded interconnected sensors, processors and controllers
  - Exploding software functionalities
  - Outsourcing and sharing information
- Safety and reliability margin will decrease
- The total cost of failure will increase dramatically
- User tolerance to failure will decrease
- Systems will need to be designed with increased modularity and testability.

RELIABILITY

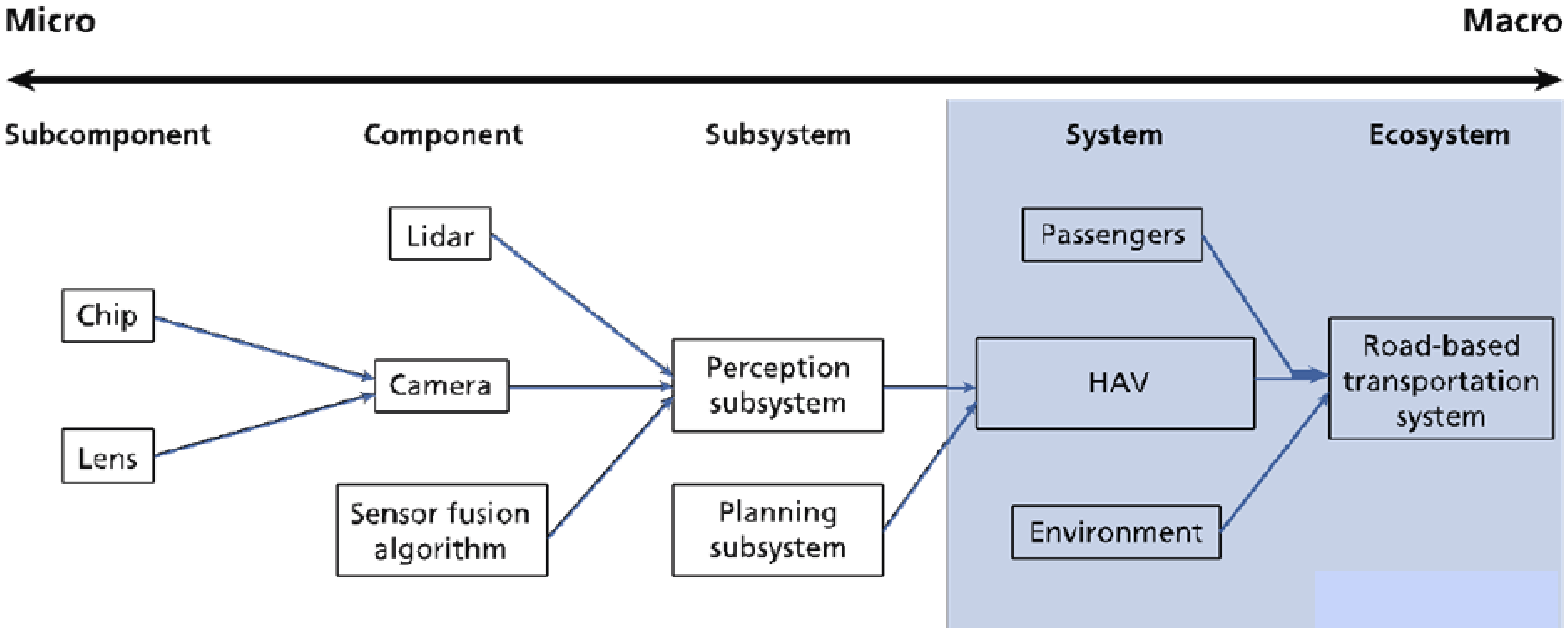
❑ For life-critical applications such as with driving, the validation process must establish that system reliability is extremely high.

❑ Historically, this very high reliability requirement has been translated into a probability of failure on the order of 10^{-7} to 10^{-9} for a 1- to 10-hour ride.

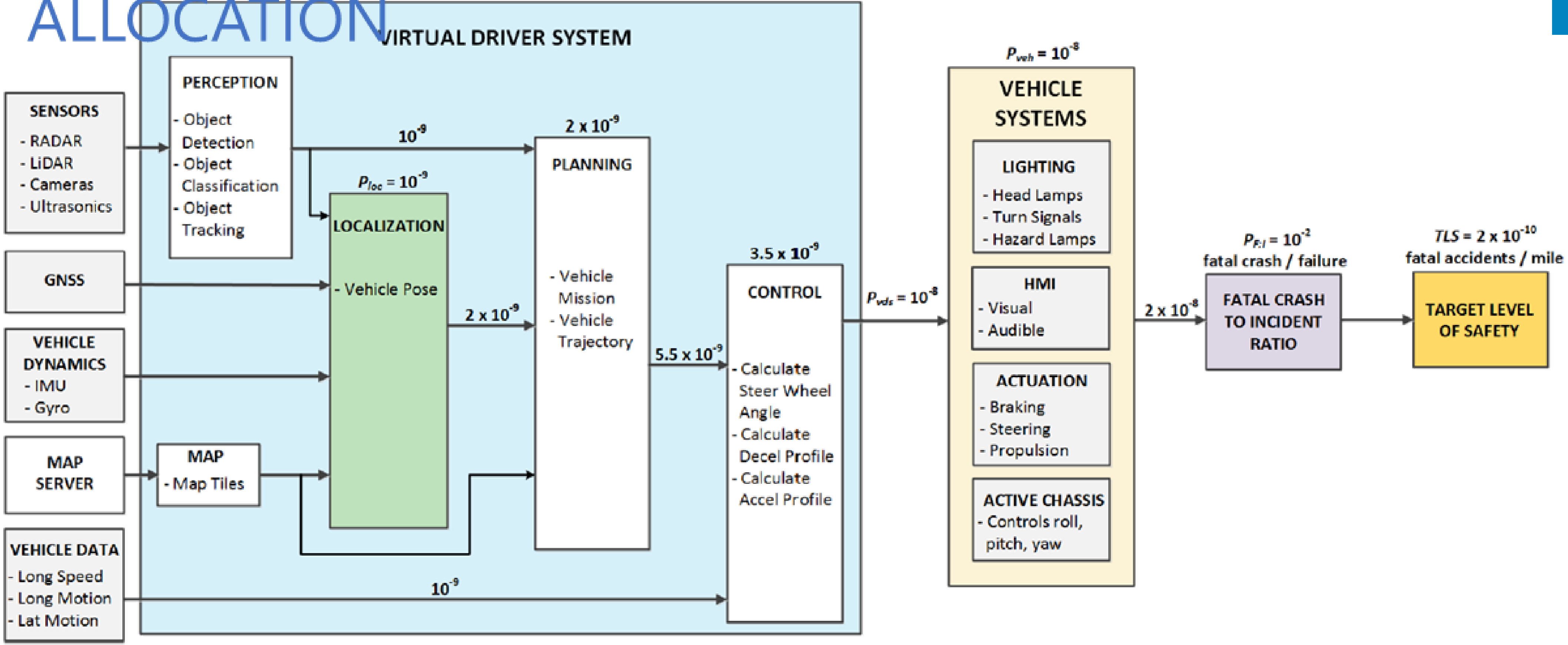
Level of reliability	Failure rate
High	$< 10^{-7}$
Moderate	10^{-3} to 10^{-7}
Low	$> 10^{-3}$

TESTING FOR SAFETY WITH COMPLEX SYSTEMS

Testing and validation protocols will differ greatly from one level of complexity to another. Granularity must be taken into account in the assessment process.



VIRTUAL DRIVER SYSTEM INTEGRITY RISK ALLOCATION

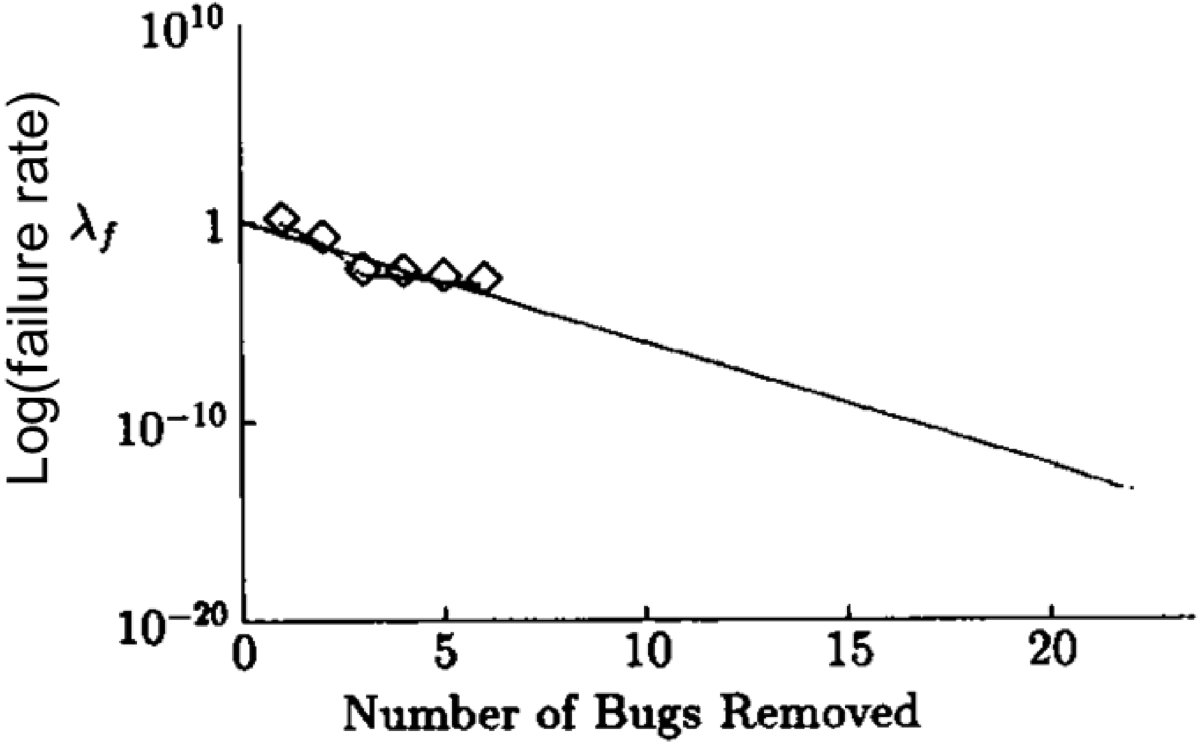


Values are given as failures per mile. Failures in localization output are assumed to lead directly to failures in planning.

RELIABILITY GROWTH MODELS

Typical sequence of events

- ❑ An intelligent system is subjected to inputs until it fails.
- ❑ The cause of failure is determined.
- ❑ Then the system is fixed and is subjected to new sequences of inputs.



UNCONSTRAINED ODD: CALLING FOR ROBUSTNESS

A product is always designed to operate correctly only under nominal specifications. Proper operation is not guaranteed if the system operates "outside" nominal specifications.

Above: Glowing from the sun and high dynamic range of illumination;

Below: Typical snow storms in northern countries.

OPERATIONAL DESIGN DOMAINS (ODD)

❑ ODD: conditions in which the CAV is intended to operate, "where" and "when.

❑ Important for levels 4-5 AVs without the aid of a human driver

❑ ODD's definition must be in terms that are identifiable or inferable by the CAVs.

❑ But precision in ODD definition is very complicated. The definition should involve inclusions and exclusions.

❑ Guaranteeing coverage in areas included or not excluded might be unachievable because there is an infinity of possible driving scenarios.

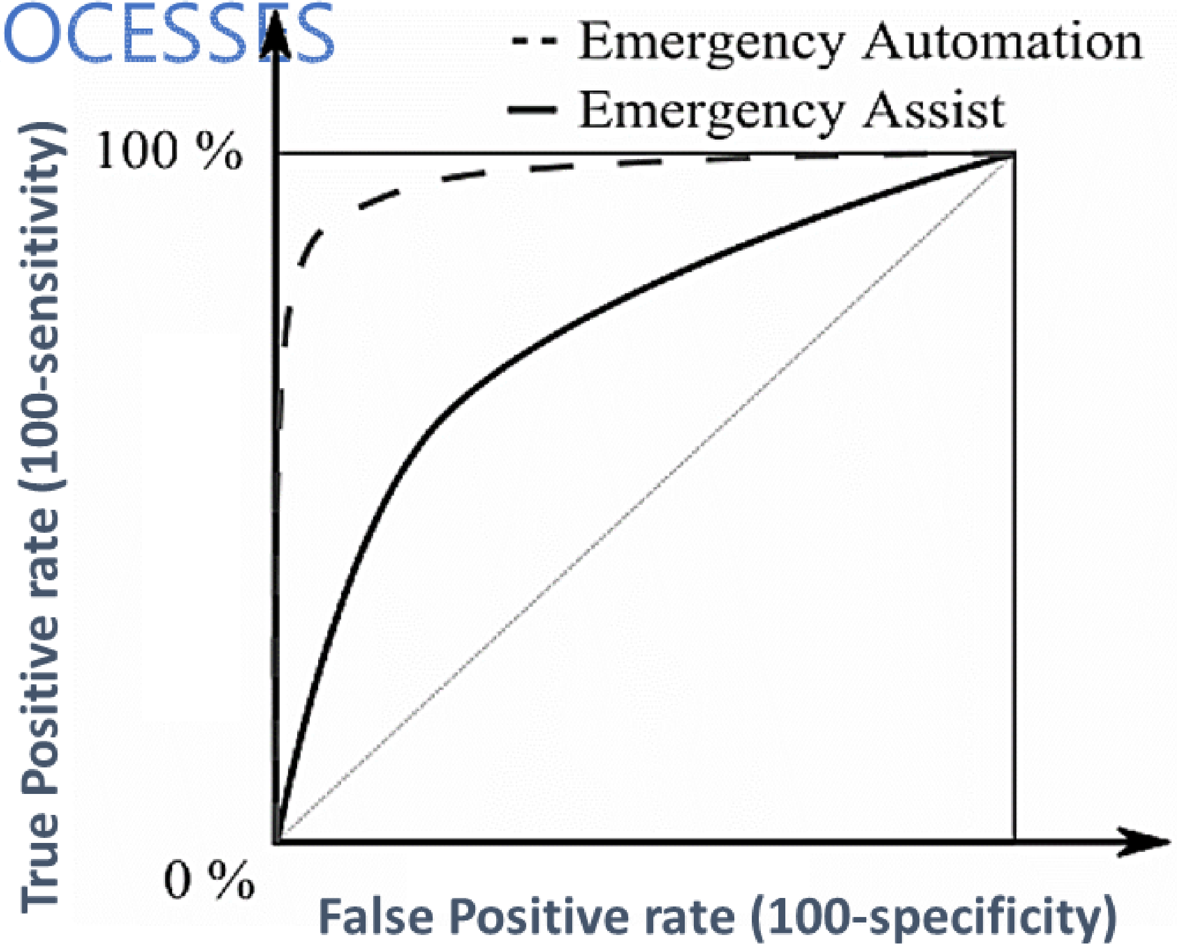
❑ CAVs operate in open-world environment; The ODD is open.

❑ ODD's definition requires common standard terminology: ontology

❑ Comparison of various CAVs safety performance should be based on same ODD!

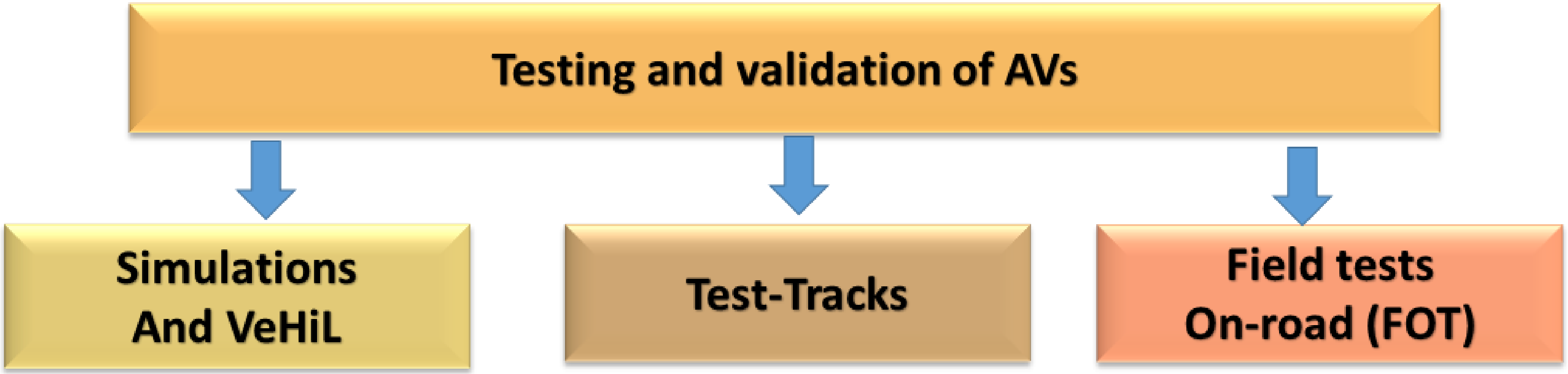
OPTIMIZING THE ROC CURVE IN DECISION MAKING PROCESSES

A big challenge to achieve robustness and reliability is to have an extremely high sensitivity (true positive) as well as a close to 100% specificity (true negative).



CURRENT APRROACHES FOR ASSESSING AVs

<u>There are currently three main approaches to assess AVs technologies</u>



Testing and validation of AVs

Simulations And VeHiL

Test-Tracks

Field tests On-road (FOT)

- □ Fast and reproducible
- □ Relatively inexpensive
- □ Large sampling of ODD
- □ Controlled scenarios
- □ Can generate ergodic data
- □ Can focus on high-risk situations
- □ Both leading and lagging safety measures

- □ Controlled environment;
- □ Very limited sampling of ODD;
- □ Very expensive;
- □ Leading safety measures only;
- □ Large infrastructure
- □ No statistics, "snapshot"

- □ Testing in real-world conditions
- □ Take very long time
- □ Lagging safety measures
- □ Very expensive (fleet)
- □ Not reproducible
- □ Not ergodic data

OPAL-RT TECHNOLOGIES

THE NEED FOR RANDOMNESS

❑ AVs safety assessment is a highly dimensional NP problems; Can be intractable (ex. infinite number of trajectories on a continuous space)

❑ Physical testing on test-tracks generate only very **sparse data** in this highly dimensional space; Sparsity lead to over approximated solution.

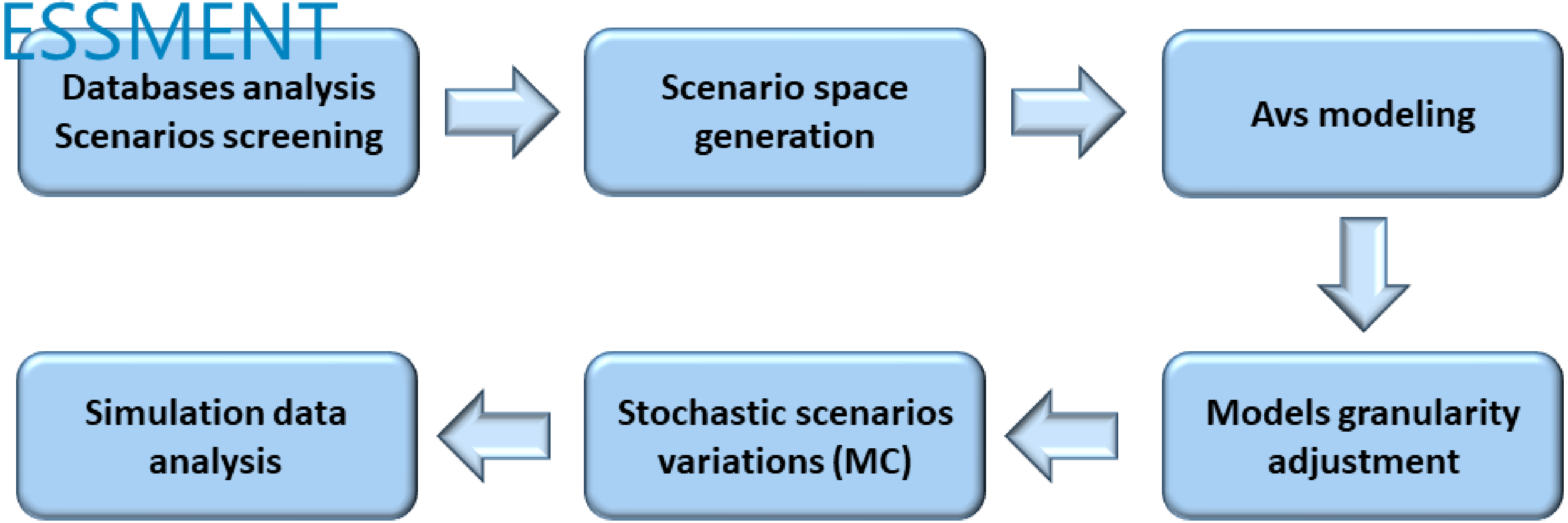
❑ Randomized/stochastic algorithms, such as Monte Carlo Las Vegas, Importance Sampling, etc., as well as kernel methods and bootstrapping techniques are required to adequately model uncertainties and behavioral scenarios of the traffic participants in an open ODD space;

❑ Worst-case analysis may be used, but leads to conservative processing and control system design, which may limit the functional performance

❑ The guaranteed quality of simulation outcomes at a certain level of confidence will depend on sample size, i.e. the number of simulations performed.

❑Sample size is bounded if the desired level of accuracy and reliability is finite; These bounds are rather conservative.

PROCESS OF VIRTUAL ASSESSMENT



Databases analysis
Scenarios screening

Scenario space generation

Avs modeling

Models granularity adjustment

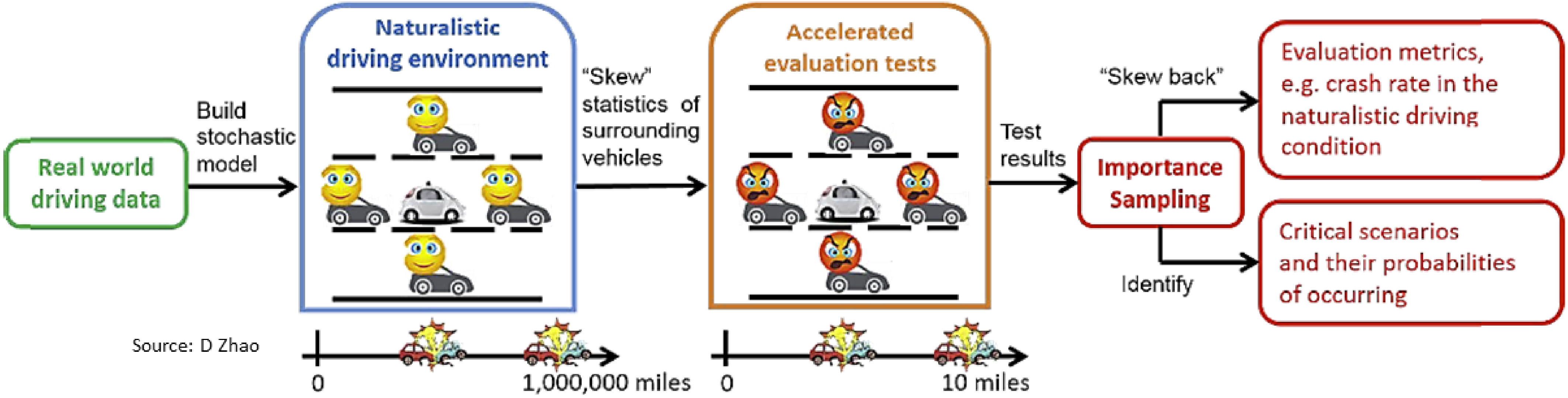
Stochastic scenarios variations (MC)

Simulation data analysis

- ❑ Virtual assessment has the ability to perform controlled experiments in high-risk traffic environments, without the “real” danger.
- ❑ Another key advantage of virtualization is the generation of adequate sample sizes:
- ❑ Stochastic model validity is a central requirement for proper use of virtualization;
- ❑ Fast and easily adaptable; virtual samples can be produced fast and with minimal cost;
- ❑ Virtual testing can capture any data desired for evaluation;
- ❑ Several safety metrics can be applied for performance comparison.

SIMULATORS FOR SPEEDING UP ASSESSMENT

To avoid years of test drive…

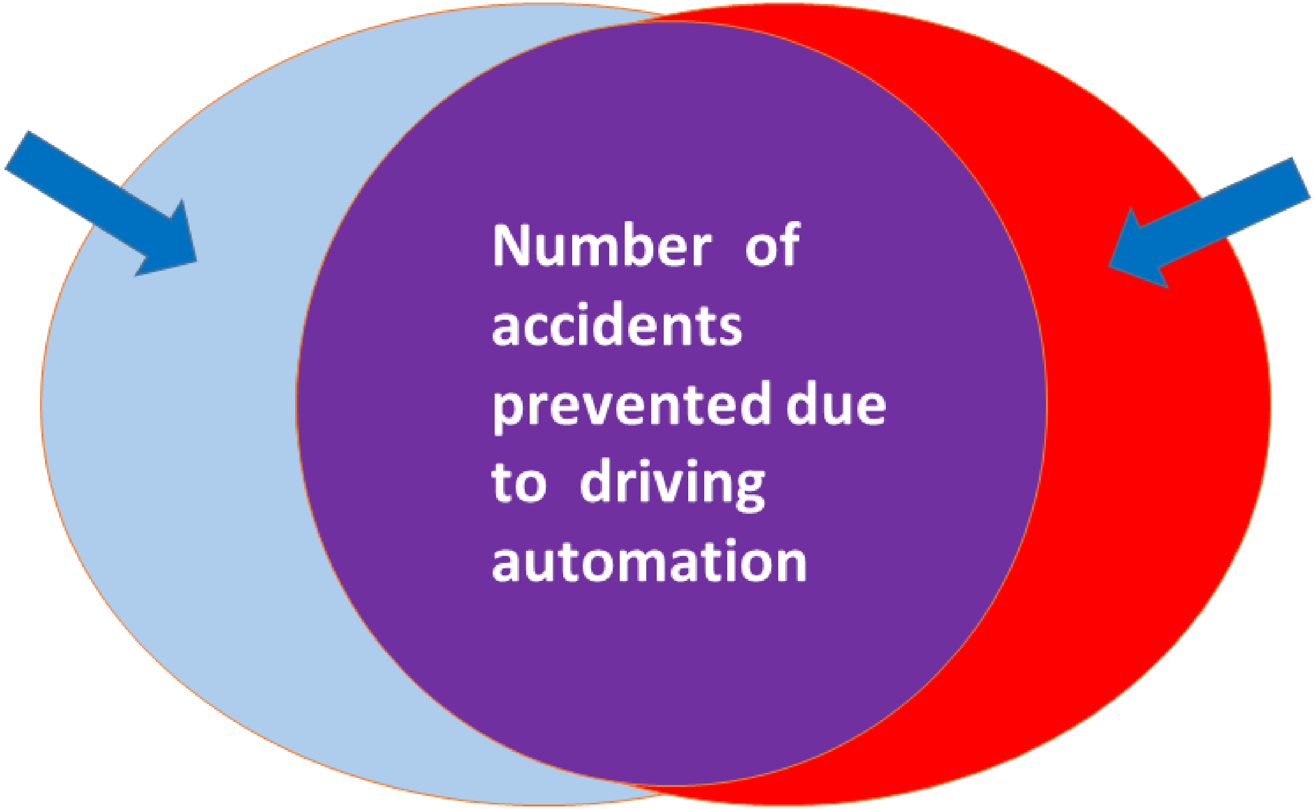

Real world driving data
Build stochastic model
Naturalistic driving environment
"Skew" statistics of surrounding vehicles
Accelerated evaluation tests
Test results
Importance Sampling
"Skew back"
Identify
Evaluation metrics, e.g. crash rate in the naturalistic driving condition
Critical scenarios and their probabilities of occurring
0
1,000,000 miles
0
10 miles

Source: D Zhao

- ❑ Because major accidents are very rare events, their statistics (distributions) need to be skewed to emphasize the safety-critical scenarios in daily driving.
- ❑ The "accelerated tests" with modified statistics are conducted.
- ❑ Stochastic methods are then used to "skew back" the results, thus predicting in a much shorter time the safety benefits of a system.

SAFETY METRIC USING VMT: REDUCING ~~FATALI~~TIES

Number of accidents due to human driving



Number of accidents prevented due to driving automation

Number of accidents due to risks in driving automation

Source: T Winkle, GIDAS

SAFETY METRIC USING VMT: RISK DEFINITION

Safety is generally described as the absence of “unreasonable” risks. This risk is usually defined as a product of the probability of an accident and the severity of that accident.

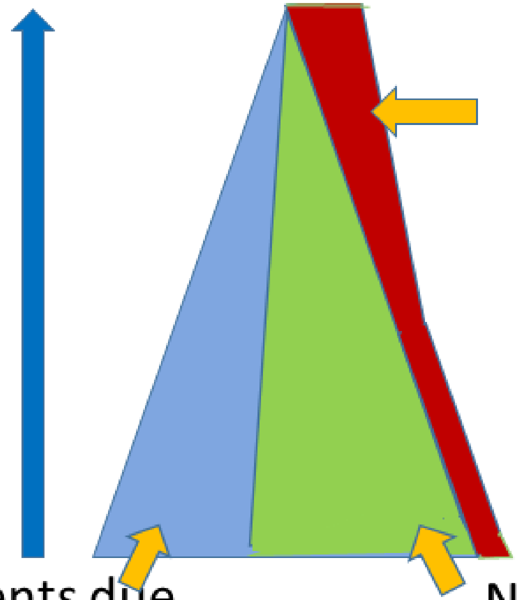
Severity

Fatality

Serious injury

Light injury

Property damages



Number of accidents due to risk in driving automation (accidents may be less numerous on average, but more severe)

Number of accidents due to human driving

Number of accidents prevented due to driving automation

Source: T Winkle, GIDAS

SAFETY METRIC USING VMT: RISK DEFINITION

For acceptance purpose, we would like ideally the following ratio to be as small as possible.

$$ExtraRisk_{automation} \square \sum_{\substack{Accidents\ due \\ to\ automation}} S_A \cdot P(A)$$

$$Risk_{avoided} \square \sum_{\substack{Accidents\ avoided \\ by\ automation}} S_A \cdot P(A)$$

Ideally,

$$Safety\ gain = SG \square \frac{1 + Risk_{avoided}}{1 + ExtraRisk_{automation}} \square\ 1$$

STATISTICS BASED ON VMT

□ Number of years per person: 55 years

□ Average number of km/year/car driven in the US: 21,000 km

□ Total km/person in a lifetime in the US: 1.15 x 10^6 km

□ Lifetime total number of hours driven (average speed of 50 km/hr): 23,000 hours

□ VMT per severe accident: approximately 1 x 10^8 km

□ Number of lifetimes per severe (lethal or with major injuries) accident: 34

□ Probability of having a severe accident in a lifetime/person: 3%

□ Although human drivers make millions of mistakes, very few lethal or severe accidents occur as a consequence. These accidents are very rare events. Assuming they are independent, we can use a simple Poisson model.

□ What are the necessary traveled distance between severe accidents for the approval of autonomous vehicles?

STATISTICS BASED ON VMT

Assuming we wish the same performance for autonomous vehicles as for human driving:

The safety gain SG = 1. The Poisson model:

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

The rate λ: ratio of the total distance performed for the test (same with and without automation), which is the average distance between fatal accident for human driving times a distance factor, D, divided by the desired (relative) safety performance for autonomous vehicles, that is the safety gain times the average distance between fatal accident for human driving.

$$Safety\ gain = SG \square \frac{1 + Risk_{avoided}}{1 + ExtraRisk_{automation}}$$

$$\lambda \square \frac{d_{test}}{d_{AVperf}} = \frac{D \cdot \bar{d}_{Human}}{S_G \cdot \bar{d}_{Human}} = \frac{D}{S_G}$$

STATISTICS BASED ON VMT

For S_G = 1 and D = 1, $\lambda = 1$, the probability of having zero or one fatal accident P(0) = P(1) = 0.37. This is the probability in human driving situations, that is every 100 million km (D = 1).

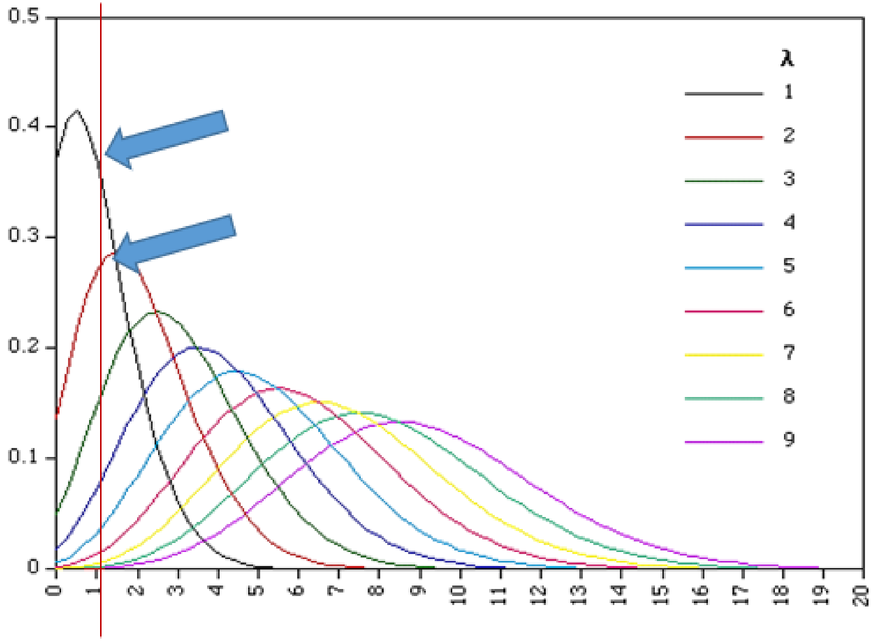
$$\lambda \;\square\; \frac{d_{test}}{d_{AVperf}} = \frac{D \cdot \bar{d}_{Human}}{S_G \cdot \bar{d}_{Human}} = \frac{D}{S_G}$$

For S_G= 2 and D = 2, $\lambda = 1$, for CAVs, the automated driving still gives P(1) = 0.37.

With human driving, S_G = 1 and D = 2, $\lambda = 2$, P(1) = 0.27.

Here, even when automated driving is assumed twice as secure as human driving, with D = 2, its probability of having 1 fatal accident is higher than for human driving!

If D is too small, it may also lead to false conclusions!



STATISTICS BASED ON VMT

❑ To reach D = 1 at a rate of 25,000 km driven/day (Waymo CAVs fleet, 2018) it would take almost 14 years!

❑ For λ much smaller than 1, (assuming SG = 1 but D very small) the probability of not observing a fatal accident is very high (for D = 0.1, P(0) = 0.9). No conclusions, then, can be made.

❑ D must be much larger than d_{human} in order to be able to draw a conclusion with a sufficiently high significance about the performance of autonomous driving.

❑ D = 3 to reach a probability fewer than 5% that a worse vehicle than the comparison group is not involved in a fatal accident. At Waymo's rate, this would take 42 years of driving with their actual fleet and VMT!

STATISTICS BASED ON VMT

□ But with D = 3, the confidence probability, P, to achieve SG = 1 is also 5%.

□ To determine if S_G = 1 is really achieved, we need a much greater probability and a greater distance factor.

□ With P = 90% and SG = 1, we need D to be about 10,7. This give us a minimum of 1.3 billion km to test drive!

□ At Waymo’s rate, this would require 150 years to cumulate proper statistics ! Note that for such a driven distance, there is also 90% probability of having 15 fatal accidents or less (either human or AV with S_G = 1).

STATISTICS BASED ON VMT

Drawbacks

❑ Bad, the better the AVs, the larger the number of test kilometers that must be driven to get reliable statistics; If S_G increases, so must D as well.

❑ Worse, if we change the CAV to be tested, the tests must be started all over again!

❑ Different OEMs produce different CAVs with different architectures and performances;

❑ Those systems are also evolving, being updated and learning continuously.

❑ Currently infeasible: To guarantee a p probability per VMT, one requires at least $1/p$ VMT must be driven.

❑ Not transparent: Unlikely that society will be satisfied with the statistical argument only.

PROBLEMS WITH EVOLVING AVs

CAVs on the road will evolve with time. They will require continuous monitoring of performance. Evolution may stem from:

- ❑ Learning capacity of the vehicle. It gains experience.
- ❑ Accumulation of data on its internal state
- ❑ Accumulation of environmental data, updating maps
- ❑ Updating software systems (such as in decision-making, risk calculation, data fusion, control, etc.).
- ❑ Updating hardware (e.g. sensors, processors)

This dynamic metamorphosis and steadily evolution of the AVs greatly complicate the anticipation of the vehicle’s behavior and its validation.

Worse: various manufacturers will put different CAV models on the roads.
AVs won’t have a homogeneous behavior. Therefore data are NOT ergodic!

PROBLEMS WITH EVOLVING AVs

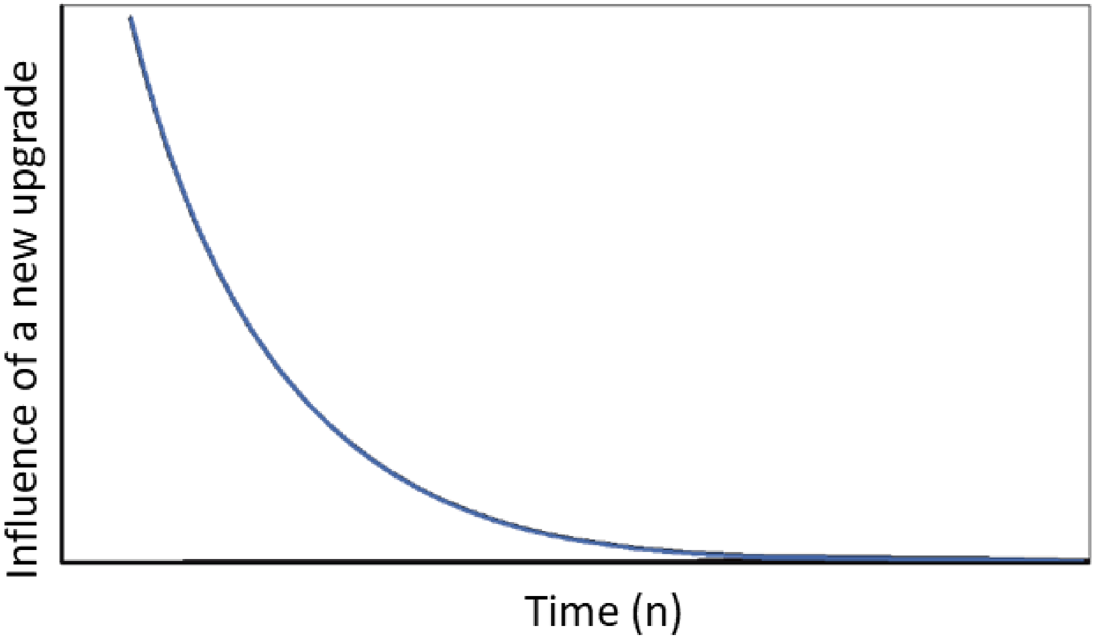
CAVs will evolve with time

- Short-term and long-term learning capabilities
- Software updates
- Hardware updates

Metrics of safety in an upgradable environment should forgive but not forget.

A simple weighted (W) first order AR to model a dynamic safety metric (SM), which has a Markovian property, can be useful:

Decay of Influence of a change over Time



$$SM(n) = W \cdot X(n) + (1-W)SM(n-1)$$

SOME PARTIAL REMEDIES

❑ Take a modular approach with already validated technologies and systems;

❑ Exploit virtualization and computing tools to identify most risky driving scenarios, speed-up assessment and reduce the cost of the validation process;

❑ Exploit virtualization and computing tools to optimize sensing suite configurations and combinations (fusion) in order to maximize reliability, robustness and safety;

❑ pin-point most risky driving scenarios, speed-up assessment and reduce the cost of the validation process;

❑ Use VeHil (Vehicle in the loop) approach to reduce modeling complexity issues.

❑ “Guardian angel” approach (driving automation runs in background of human driving) to cumulate data safely over long period of time (large VMT);

❑ Minimize the risk (short-term) by reducing speed, maximizing the use of a priori info and constraints on ODD (ex. using HD maps with low speed autonomous shuttle to solve the last km problem);

❑ Use other safety metrics, tolerate and accept some unknown risks.

CONCLUSIONS

- ❑ Safety has a cost;
- ❑ Safety can never reach 100%;
- ❑ AVs safety relies mainly on reliability and robustness;
- ❑ Extremely high system complexity and unconstrained ODD are two main challenges;
- ❑ Our ability to design complex systems currently exceeds our ability to test them;
- ❑ DMPs (decision-making processes) requires simultaneously extremely high sensitivity and selectivity. Hard to achieve;
- ❑ Real-time requirements, latency in DMPs and data registration are big issues;
- ❑ Dealing with extremely rare events and very high safety level requires the processing and analysis of huge amount of data;
- ❑ Real data is not ergodic nor stationary. Make usual statistical approaches difficult;
- ❑ Discretization in DMPs (discrete Markov chains, Bayesian networks) is an issue for adequate modeling.

CONCLUSIONS

□ Detailed physical modeling and microscopic simulation tools is a must for realistic behavior predictions;

□ Simulation and virtualization will be come the dominant approach for AVs validation similar to aerospace industry;

□ Fusion of FOT, tests-on-track and simulation data likely to be the way to go;

□ Test track data's role will shift from technology validation to ground truth data for calibrating/validating the simulation models;

□ FOT data yet too small for lagging safety analysis. But currently useful for technology debugging and for knowledge building on driving scenarios and vehicular behavior as well as extraction/screening of the most risky/relevant scenarios used in simulation;

□ Further problem: AVs are heterogeneous (different OEMs) and evolving (learning);

□ Urgent need for standards, standards, standards…