Caltech

Can We Really Use Machine Learning in (Large-Scale) Safety-Critical Systems?

Richard M. Murray

California Institute of Technology

IPAM Workshop on Intersections between
Control, Learning and Optimization

27 February 2020

Where did this talk come from:

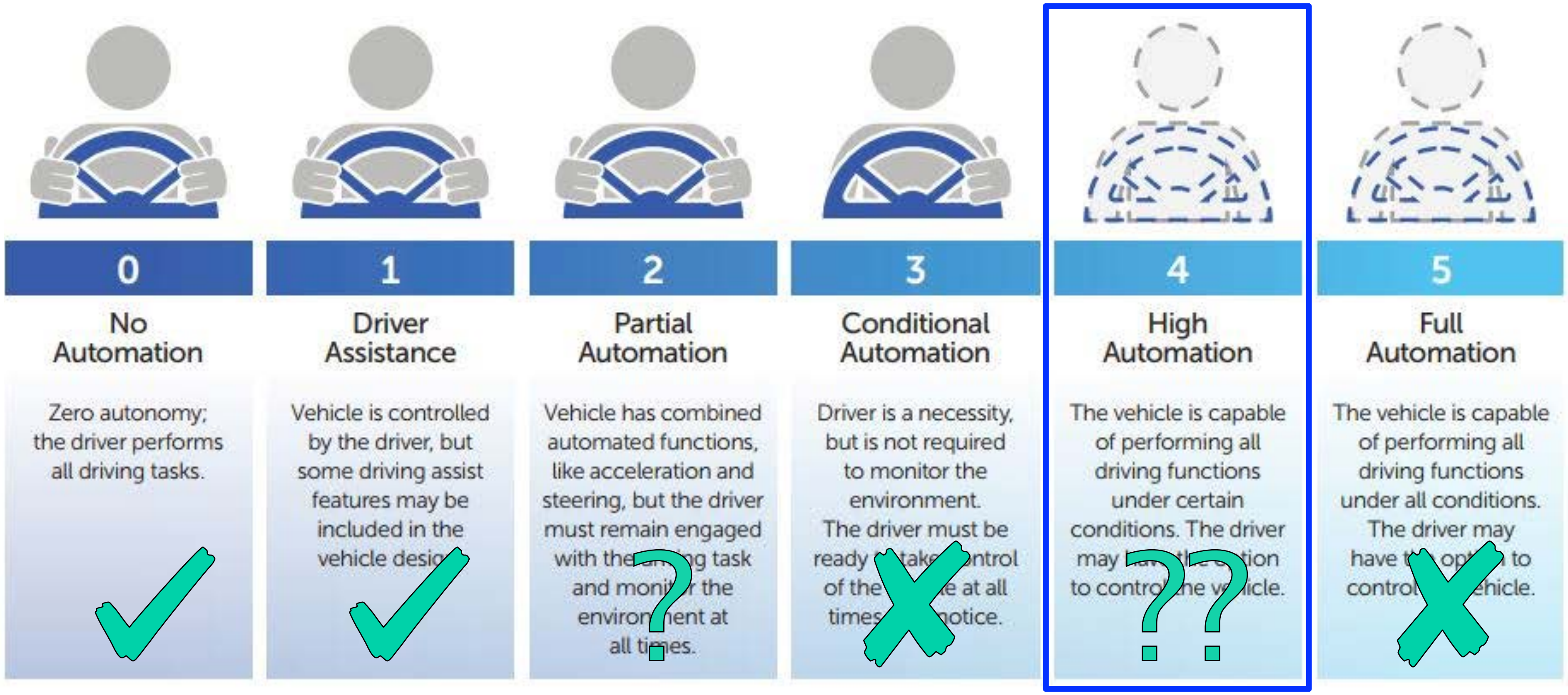
- ML is revolutionizing our approach to many problems (images/speech recognition, etc)
- ML is being applied to complex decision-making tasks in safety-critical systems
- I am worried: do we really understand how deploy ML in the physical world, at scale?

What this talk is about:

- How (large-scale) safety-critical systems are designed today (aerospace focus)
- Challenges of adopting those techniques to ML-based components
- Problems I would like to see more people working on (but not really what my group is doing)

Current Landscape: Self-Driving Cars

SAE Levels of Automation:



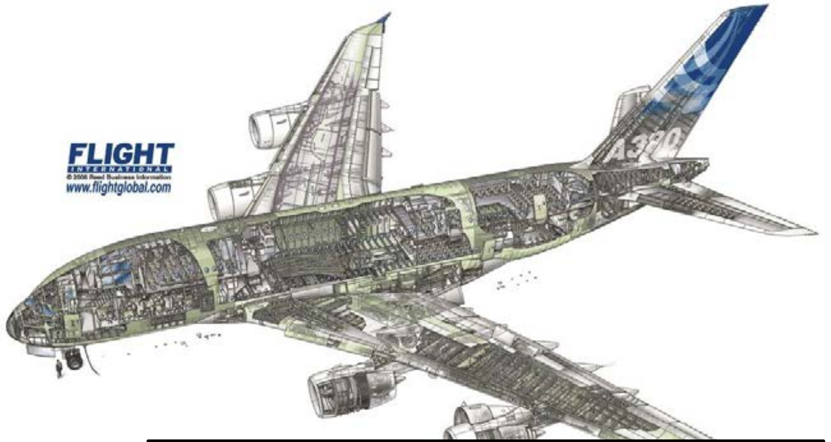
0	1	2	3	4	5
No Automation	Driver Assistance	Partial Automation	Conditional Automation	High Automation	Full Automation
Zero autonomy; the driver performs all driving tasks.	Vehicle is controlled by the driver, but some driving assist features may be included in the vehicle design.	Vehicle has combined automated functions, like acceleration and steering, but the driver must remain engaged with the driving task and monitor the environment at all times.	Driver is a necessity, but is not required to monitor the environment. The driver must be ready to take control of the vehicle at all times with notice.	The vehicle is capable of performing all driving functions under certain conditions. The driver may have the option to control the vehicle.	The vehicle is capable of performing all driving functions under all conditions. The driver may have the option to control the vehicle.
✓	✓	?	✗	??	✗

Safety Critical Autonomous Systems

Question: How safe do autonomous vehicles need to be?

- As safe as human-driven cars (7 deaths every 10^9 miles)
- As safe as buses and trains (0.1-0.4 deaths every 10^9 miles)
- As safe as airplanes (0.07 deaths every 10^9 miles)

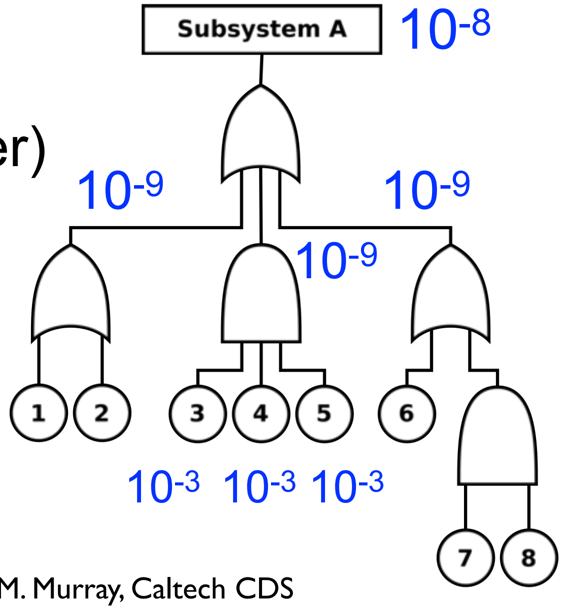
I. Savage, “Comparing the fatality risks in United States transportation across modes and over time”, *Research in Transportation Economics*, 43:9-22, 2013.

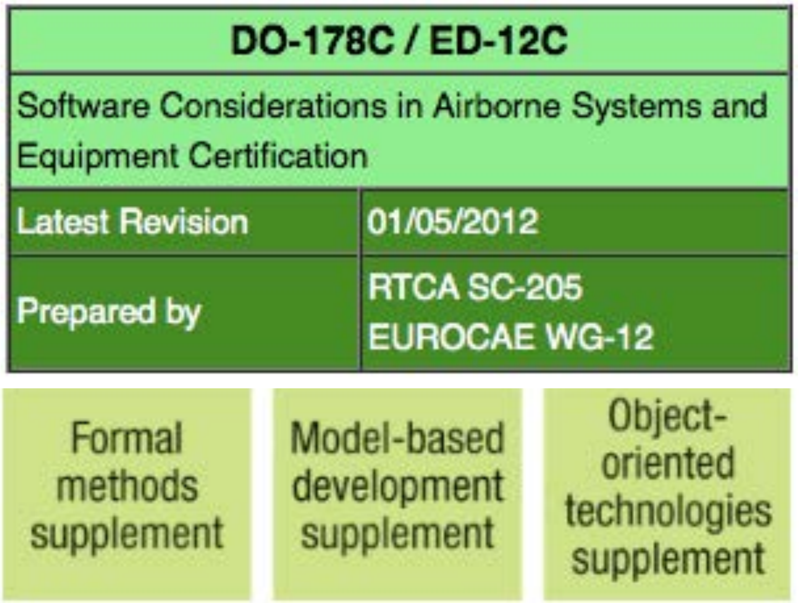


Hazard Class	SW Level	Failure/ Flight Hr
Catasophic	A	10^{-9}

How this is done in the aerospace industry?

- Strong certification requirements/process (DO-178C)
  - Fault tree analysis (1e-9 failure rates)
  - Model-based design + SIL, HIL testing
  - Fleet-wide analysis (⇒ rare cases matter)
- Very structured operating environments
- Well-trained personnel (pilots, FAs)
- Expensive vehicles (~$1M/passenger)

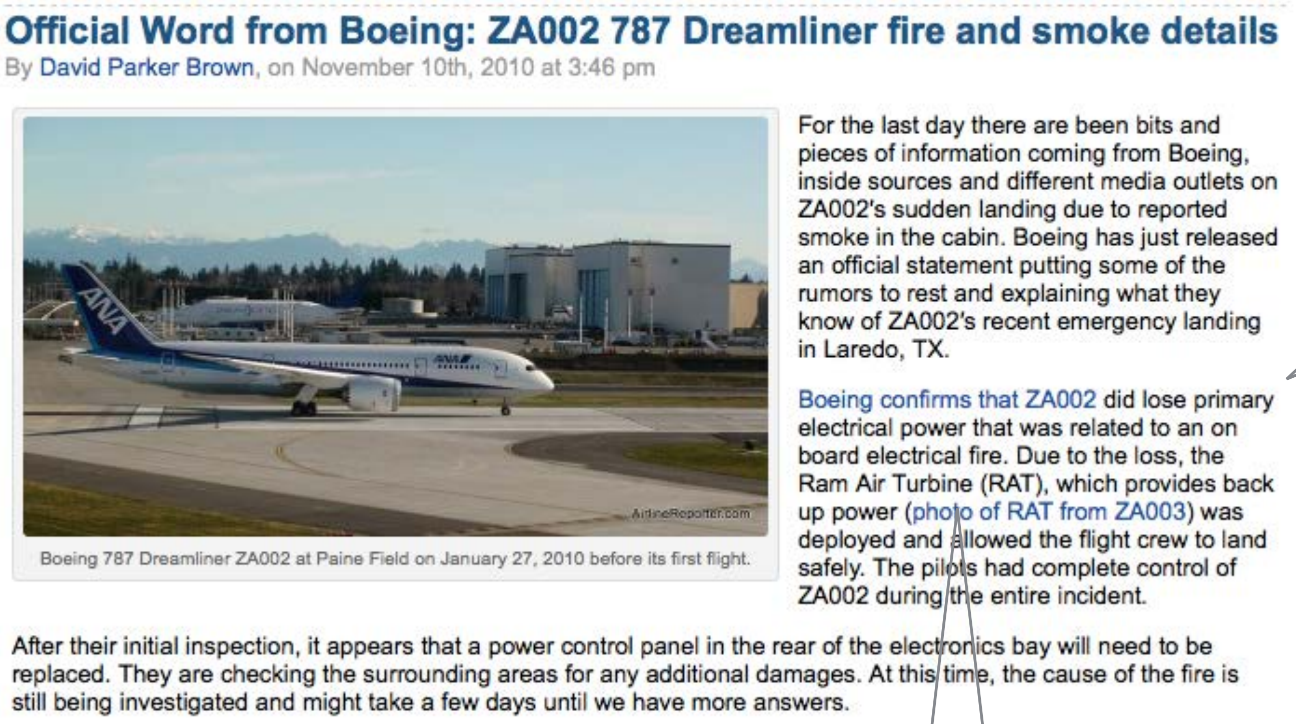




DO-178C / ED-12C	
Software Considerations in Airborne Systems and Equipment Certification	
Latest Revision	01/05/2012
Prepared by	RTCA SC-205 EUROCAE WG-12

Formal methods supplement	Model-based development supplement	Object-oriented technologies supplement

What Goes Wrong: ZA002, Nov 2010



Official Word from Boeing: ZA002 787 Dreamliner fire and smoke details

By David Parker Brown, on November 10th, 2010 at 3:46 pm

Boeing 787 Dreamliner ZA002 at Paine Field on January 27, 2010 before its first flight.

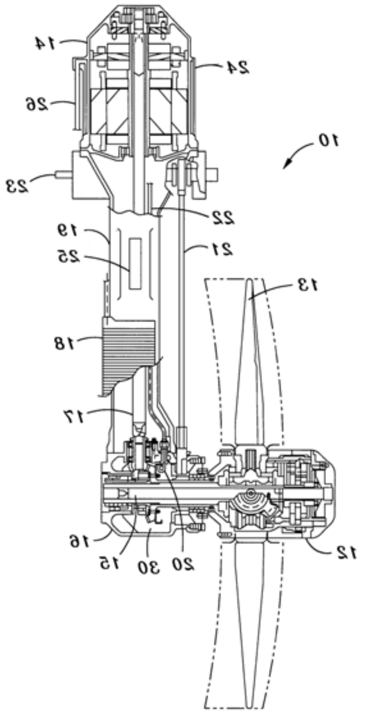
For the last day there are been bits and pieces of information coming from Boeing, inside sources and different media outlets on ZA002's sudden landing due to reported smoke in the cabin. Boeing has just released an official statement putting some of the rumors to rest and explaining what they know of ZA002's recent emergency landing in Laredo, TX.

Boeing confirms that ZA002 did lose primary electrical power that was related to an on board electrical fire. Due to the loss, the Ram Air Turbine (RAT), which provides back up power (photo of RAT from ZA003) was deployed and allowed the flight crew to land safely. The pilots had complete control of ZA002 during the entire incident.

After their initial inspection, it appears that a power control panel in the rear of the electronics bay will need to be replaced. They are checking the surrounding areas for any additional damages. At this time, the cause of the fire is still being investigated and might take a few days until we have more answers.

Loss of primary electrical power => cockpit goes “dark”

Ram Air Turbine (RAT) deployed and allows safe landing

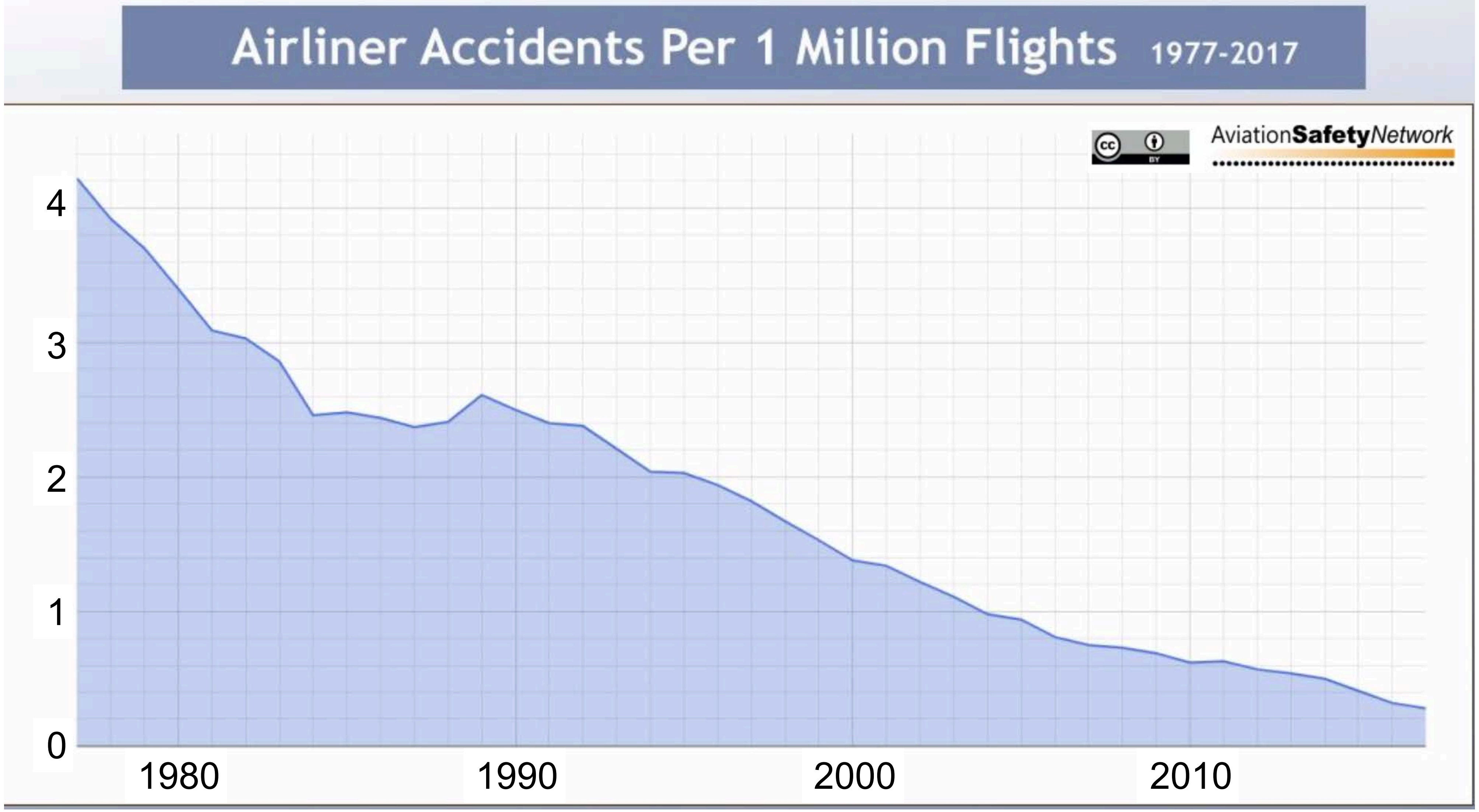


RAT stats

- ~100K flights/ day globally => 35M flights/year
- ~6 documented RAT deployments in the last 20 years
- Assume 10X that amount => 3 per year => 1 in every 10M flights (!)

Key point: aerospace engineers worry about the worst case

Continuous Improvement Over Time



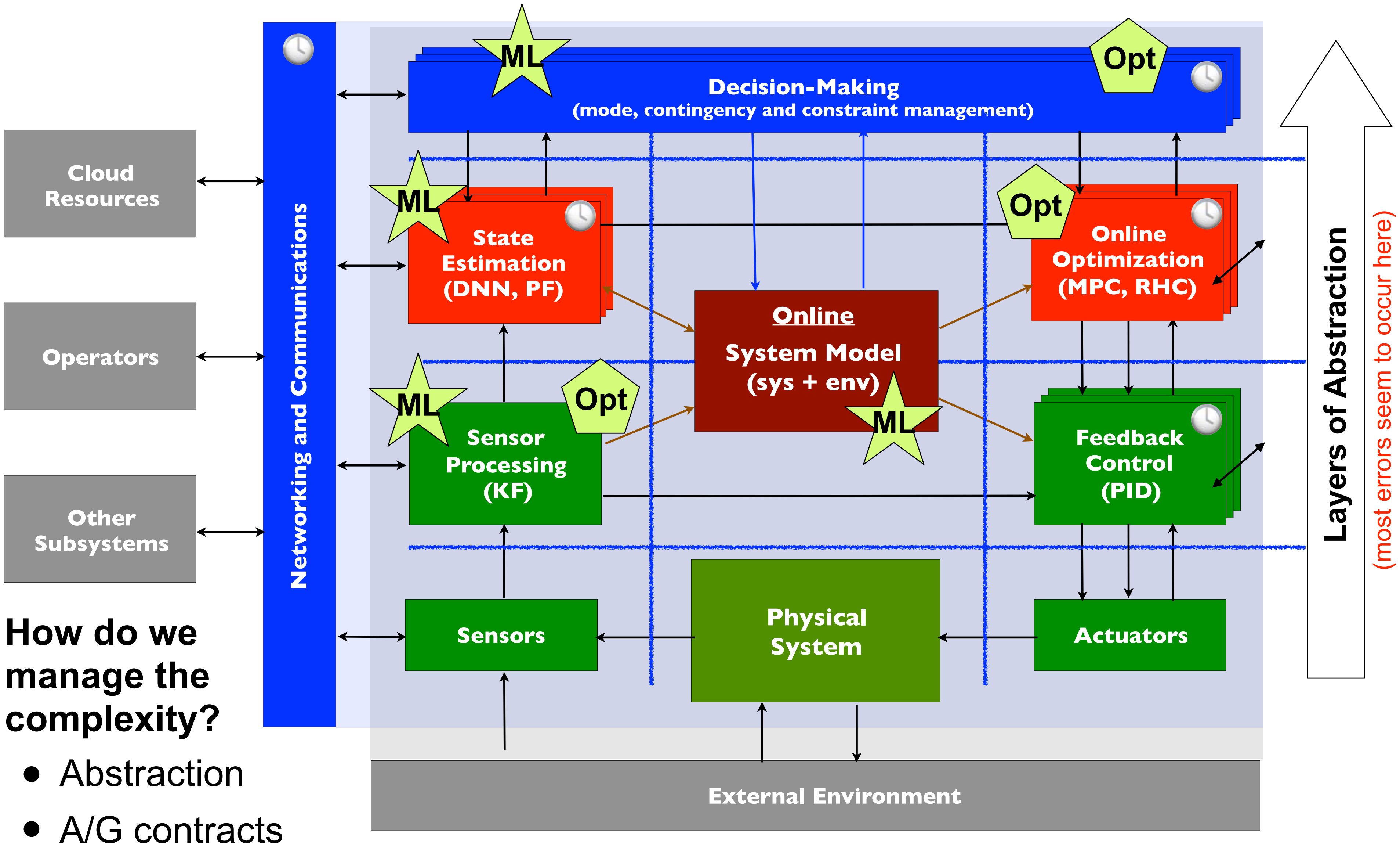
Early history

- Failures led to government regulation
- Industry groups developed standards

Challenges for self-driving cars

- Already starting with a pretty low accident rate
- 10X improvement could take 40+ years (!)
- Economics are very different…

Design of Modern (Networked) Control Systems



Examples

- Aerospace systems
- Autonomous vehicles
- Factory automation/ process control
- Smart buildings, grid, transportation

Challenges

- How do we define the layers/interfaces (vertical contracts)
- How do we scale to *many* devices (horizontal contracts)
- Safety, robustness, security, privacy

How do we manage the complexity?

- Abstraction
- A/G contracts
- Formal methods for verification/synthesis + model- & data-driven sims/testing

Thoughts on ML and Control ("Easy" Problems)

ML Challenges

Failure rates are too high, w/ poor metrics

- 1 hour = 10K frames => 1B hours = …
- Classification error is not that useful

Data requirements are unknown (but large)

- Size of error vs amount of training data?
- How do we catch corner cases?

Focus on ML output vs system behavior

- Classification error is not what we actually care about; do we hit anything?

Early adoption in safety-critical settings

- Use of ML for decision-making is not ready
- Advice: ML for *performance*, optimization and control for safety and robustness

Controls Perspective

Stability margins with uncertainty balls

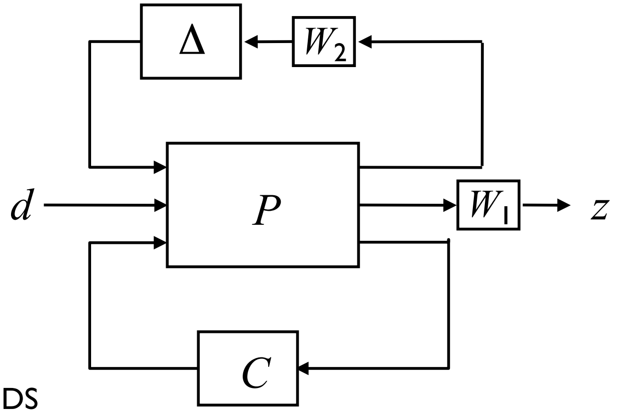
- Bounds on disturbances, uncertainty
- Model/analyze temporal response

Model-based, parametric representations

- Constrain model class (TFs, ARMAX, etc)
- Reason over worst case behavior

Input/output focus

- Focus on outputs that matter for the task and impact of uncertainty on those outputs



$$\|z\|_2 \leq \gamma \, \|d\|_2$$

for all

$$\|\Delta\| \leq 1$$

Thoughts on ML and Control (Hard Problems)

Autonomous Vehicles for Urban Mobility
Emilio Frazzoli, ETH Zurich & Aptiv

… [As] we move past the peak of the hype cycle, the industry is bracing for a development timeline that is much longer than many early predictions.

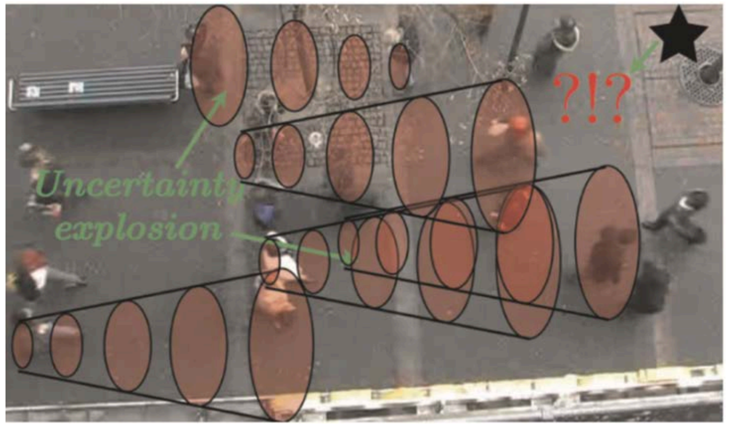
… fundamental issues that remain essentially unresolved, and will require a concerted effort by industry, academia, and regulatory bodies to address.

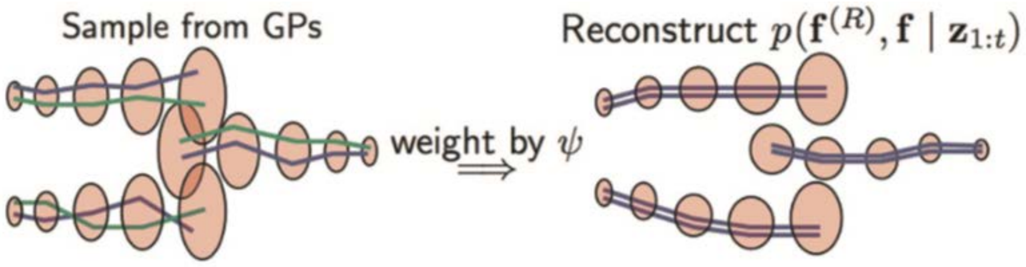
These issues essentially go beyond the (very hard, but in a sense "standard" and well studied) problems of control, perception, etc. and revolve around making sound decisions on precisely how we want these vehicles to behave, both at the individual, single-car level, and at the fleet level. In other words, how we want these vehicles to behave when interacting with pedestrians, cyclists, or other cars, and what effect we want them to have on urban mobility, including, e.g., their impact on the urban environment, public transit, and society.

http://www.exempelbanken.se/examples/347

Trautman, Ma, M and Krause
IJRR, 2014

Some Prior Work: Navigation in Crowds




Sample from GPs
Reconstruct $p(\mathbf{f}^{(R)}, \mathbf{f} \mid \mathbf{z}_{1:t})$
weight by ψ

$$p(\mathbf{f}^{(R)}, \mathbf{f}|\mathbf{z}_{1:t}) = \frac{1}{Z}\psi(\mathbf{f}^{(R)}, \mathbf{f}) \prod_{i=R}^{n} p(\mathbf{f}^{(i)}|\mathbf{z}_{1:t}^{(i)})$$

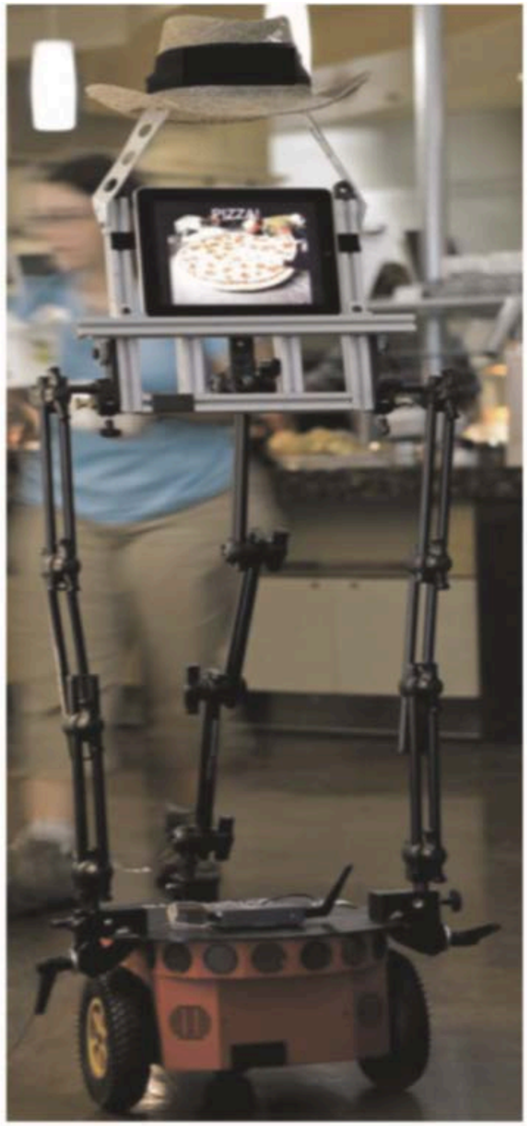
$$\psi(\mathbf{f}^{(R)}, \mathbf{f}) = \prod_{i=R}^{n}\prod_{j=i+1}^{n}\prod_{\tau=t}^{T}(1-\alpha \exp(-\frac{1}{2h^2}|\mathbf{f}^{(i)}(\tau) - \mathbf{f}^{(j)}(\tau)|))$$

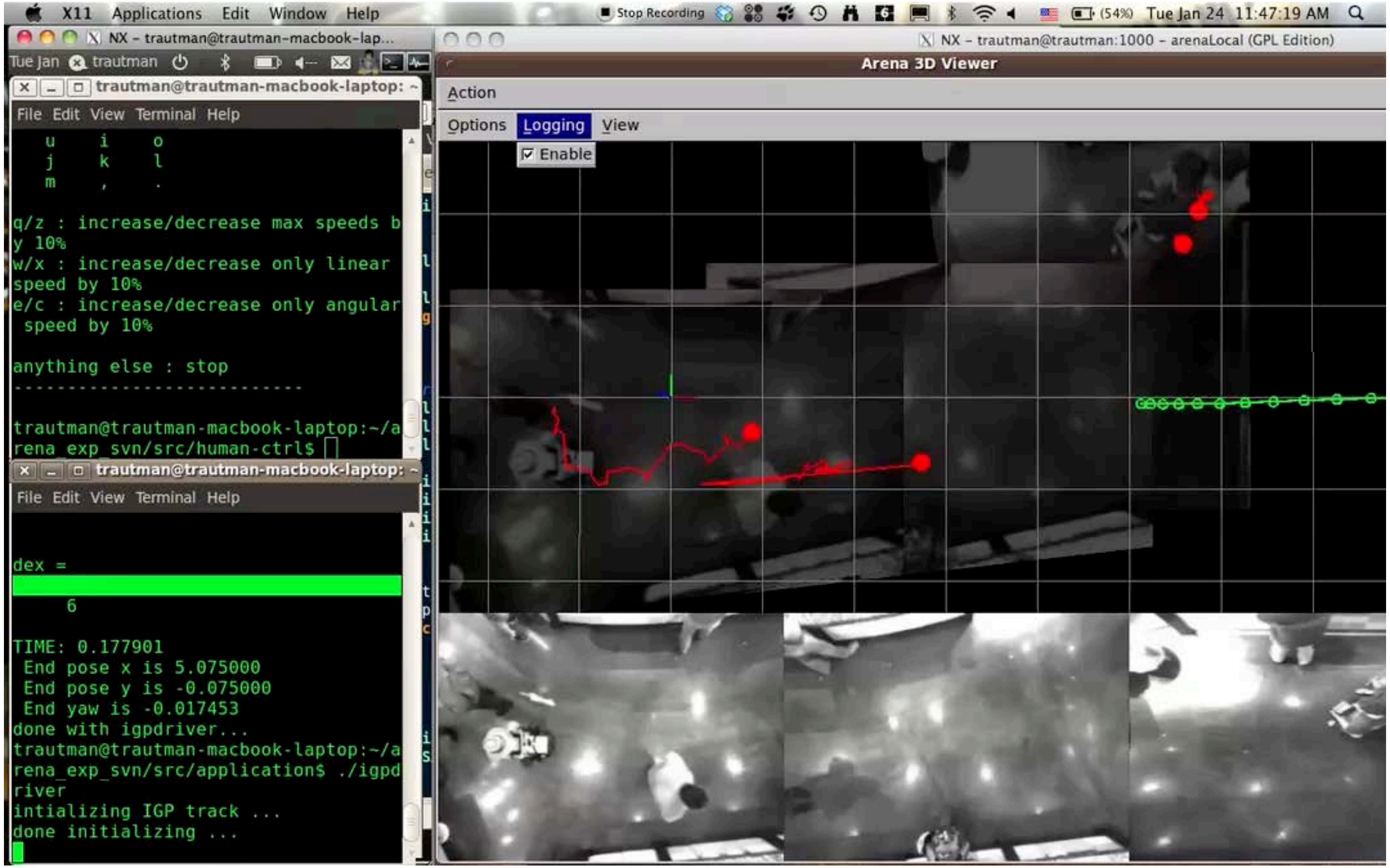
Key results

- Address “freezing robot problem”: planner decides that all forward paths are unsafe and freezes in place
- Approach: *interacting Gaussian processes*
  - captures cooperative collision avoidance
  - allows goal-driven nature of human decision making
- Validation in Caltech staff cafeteria
  - Performs comparably with human teleoperators
  - non-cooperative planner exhibits unsafe behavior
  - reactive planner fails for crowd densities > 0.55 ppl/m^2

Trautman, Ma, M and Krause
IJRR, 2014

Some Prior Work: Navigation in Crowds

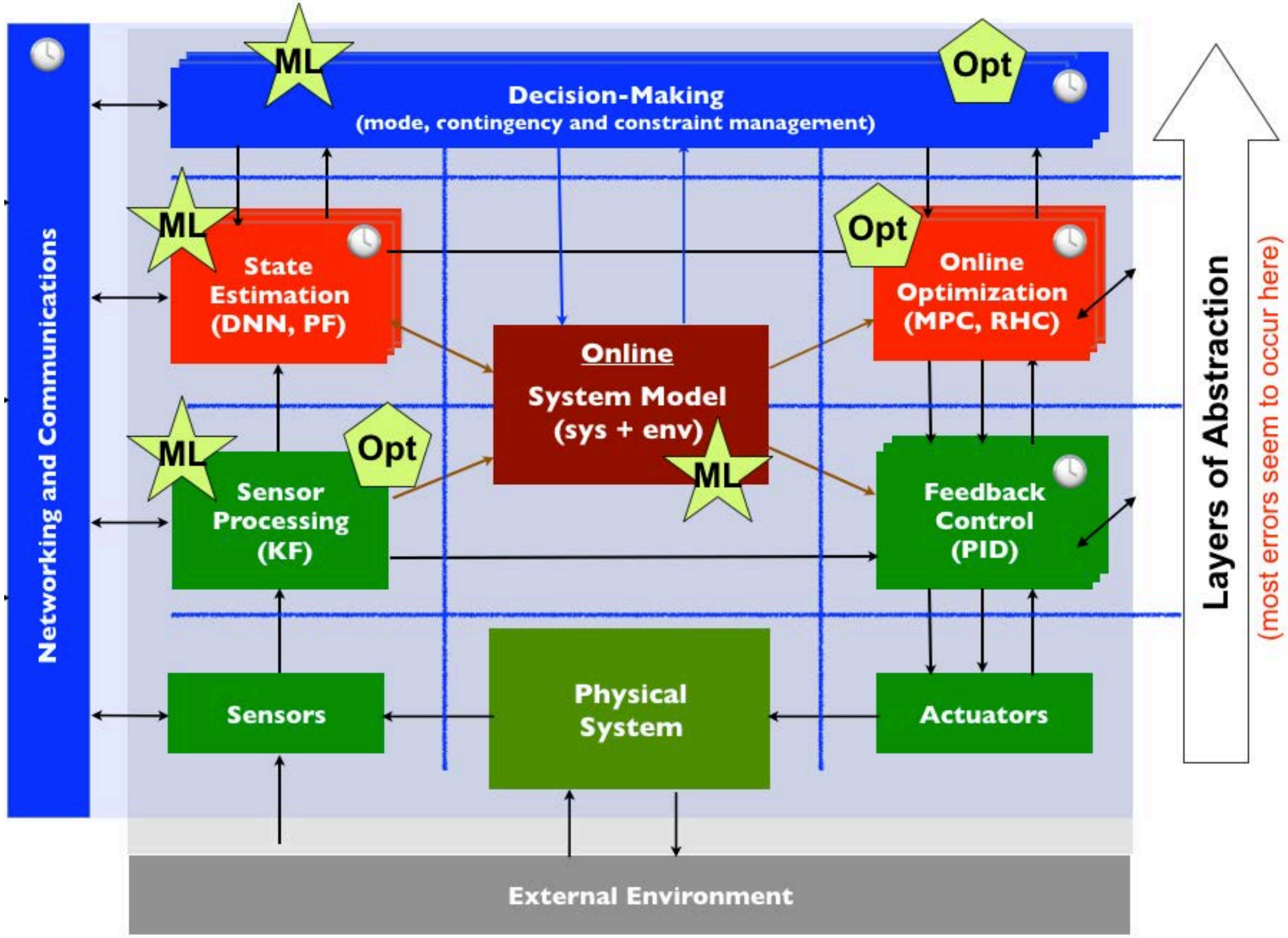



X11 Applications Edit Window Help
Stop Recording
(54%) Tue Jan 24 11:47:19 AM
NX - trautman@trautman-macbook-lap...
Tue Jan trautman
trautman@trautman-macbook-laptop: ~
File Edit View Terminal Help
NX - trautman@trautman:1000 - arenaLocal (GPL Edition)
Arena 3D Viewer
Action
Options Logging View
☑ Enable

```
   u    i    o
   j    k    l
   m    ,    .

q/z : increase/decrease max speeds b
y 10%
w/x : increase/decrease only linear
speed by 10%
e/c : increase/decrease only angular
 speed by 10%

anything else : stop
---------------------------

trautman@trautman-macbook-laptop:~/a
rena_exp_svn/src/human-ctrl$
```

```
dex =

     6

TIME: 0.177901
 End pose x is 5.075000
 End pose y is -0.075000
 End yaw is -0.017453
done with igpdriver...
trautman@trautman-macbook-laptop:~/a
rena_exp_svn/src/applications ./igpd
river
intializing IGP track ...
done initializing ...
```

Current Assessment: Wait for Others to Figure out ML…



Assume/guarantee contracts

- Assume: properties of other components in the system
- Guarantee: properties that will hold for my component

$$A_i \Rightarrow G_i$$

$$G_2 \wedge G_3 \Rightarrow A_1,\ G_1 \wedge G_3 \Rightarrow A_2,\ \ldots$$

- Contracts can be horizontal (within a layer) or vertical (between two layers)

Integrating ML (eventually)

- Wait for smart people to create ML w/ A/G contracts
- Think about how to best integrate these into the larger NCS architecture

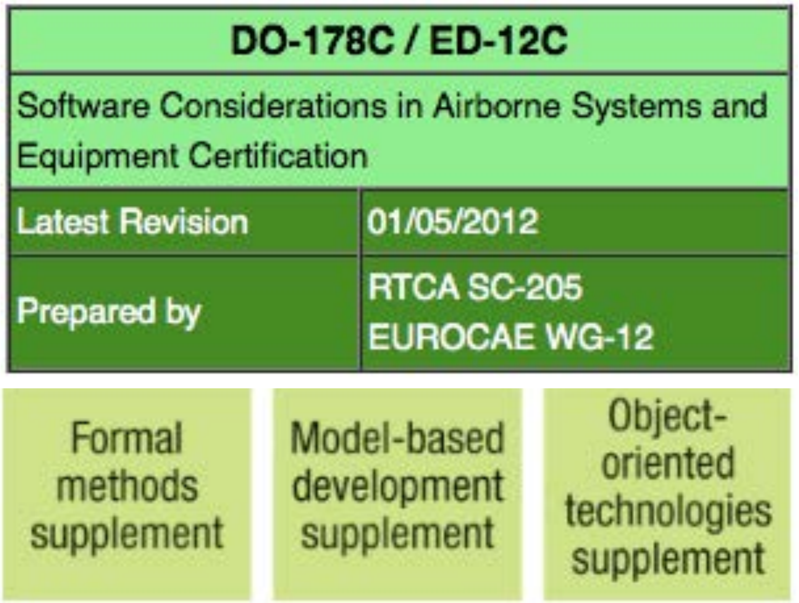
Machine Learning in Safety-Critical Systems

Hazard Class	SW Level	Failure/ Flight Hr
Catasophic	A	10^{-9}
Hazardous	B	10^{-7}
Major	C	10^{-5}
Minor	D	—
No Effect	E	—

0.07 deaths every 10^9 miles ← 7 deaths every 10^9 miles

??

35K/year (US)



DO-178C / ED-12C	
Software Considerations in Airborne Systems and Equipment Certification	
Latest Revision	01/05/2012
Prepared by	RTCA SC-205 EUROCAE WG-12

Formal methods supplement

Model-based development supplement

Object-oriented technologies supplement

Claim: ML can solve problems that we can't solve otherwise

Q: How do we move ML into safety-critical applications?

- Certification methodology for ML-based components
- Error rates (of decisions) measured in 1 per billions of hrs/miles
- Robust operation across wide range of conditions