Session 1: ML/AI Verification in CPS Report

Session Chair: Marco Vieira (UNC Charlotte)

Scribe: Lishan Yang (George Mason University)

Motivation

- Trustworthy; assured; safe
- Autonomous CPS

Video released of Uber self-driving crash that killed woman in Arizona

New footage of the crash that killed Elaine Herzberg raises fresh questions about why the self-driving car did not stop

▲ Uber dashcam footage shows lead up to fatal self-driving crash – video

The Guardian, Mar 22 2018

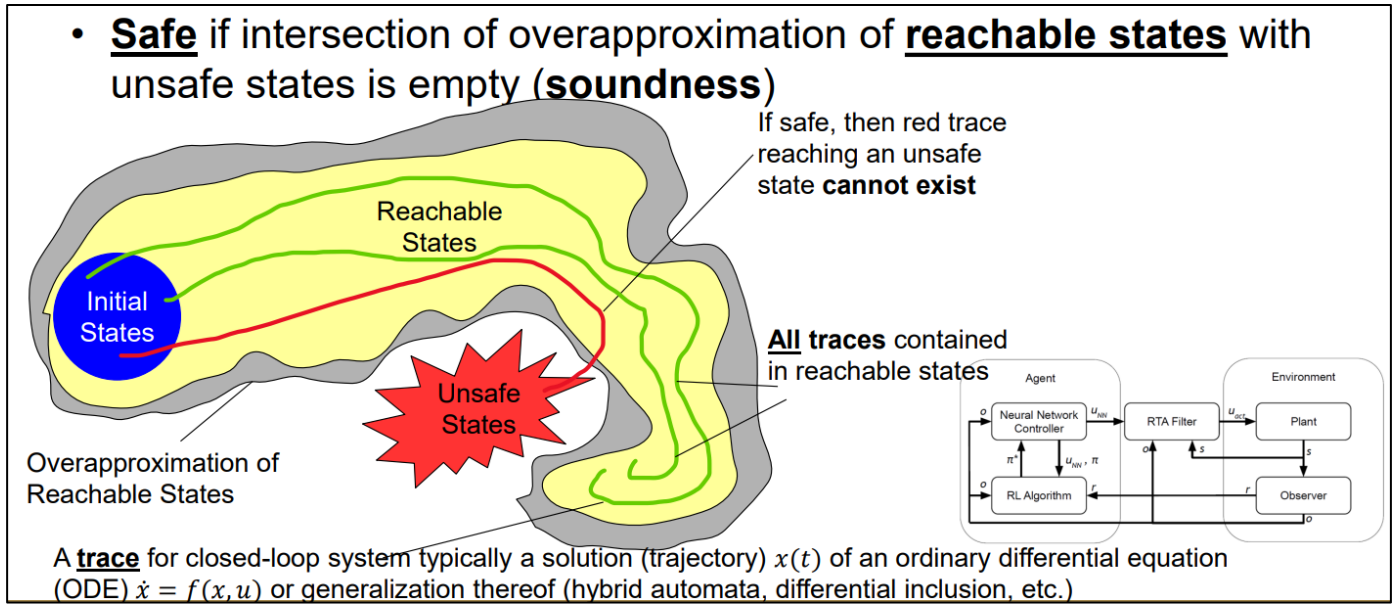
Assured Perception and Control of Autonomous Systems Using Formal Verification of Neural Networks, Yasser Shoukry, 85th IPIF WG 10.4 Meeting, 2024

Formal Verification Challenge

- High engineering cost
- State-space explosion
- Robustness of NNs

Formal Verification in ML & CPS

- Closed-Loop Verification with NNV
  - Safety properties
- Safety Verification of Closed-Loop Autonomous Systems with Reachability
  - Monitoring: Runtime Verification



Neural Network Verification in Autonomous Cyber-Physical Systems (CPS), Taylor Johnson, 85th IPIF WG 10.4 Meeting, 2024

Formal Verification in ML

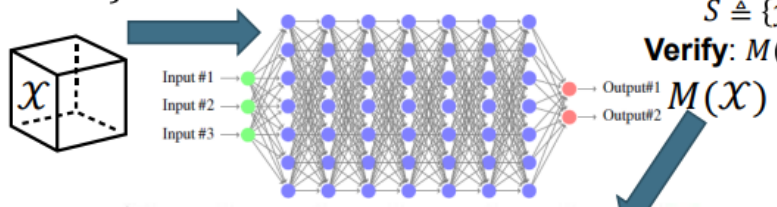
Given a NN $M: \mathbb{R}^n \mapsto \mathbb{R}^m$ & an input set $\mathcal{X} \subseteq \mathbb{R}^n$, the **output reachable set** of M is $Y = \{y \mid y = M(x), \forall x \in \mathcal{X}\} \subseteq \mathbb{R}^m$

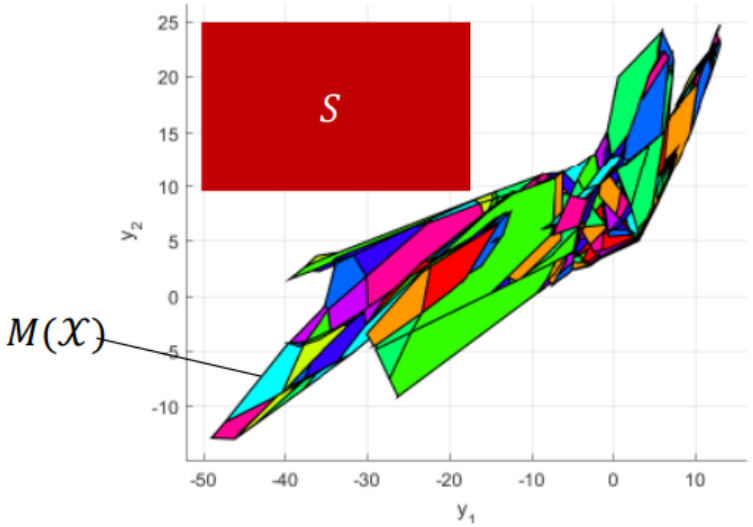
Input set: $\mathcal{X} \triangleq \{x \in \mathbb{R}^3 \mid \|x\|_\infty \leq 1\}$

Specification:
$S \triangleq \{y \in \mathbb{R}^2 \mid -50 \leq y_1 \leq -20 \wedge 10 \leq y_2 \leq 25\}$

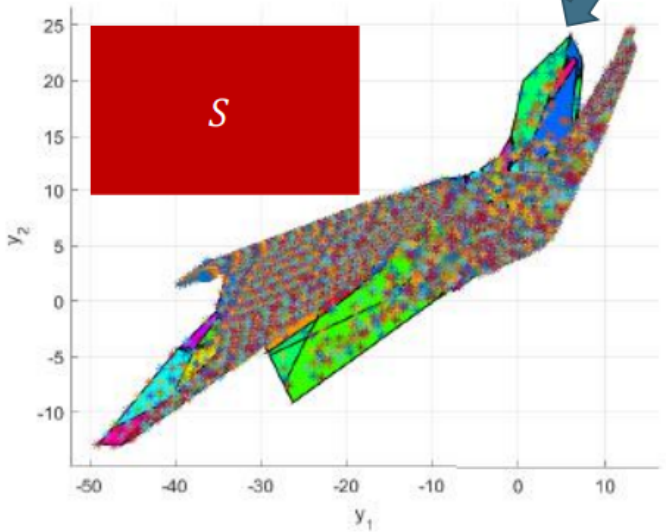
Verify: $M(\mathcal{X}) \cap S = \emptyset$?

M: simple feedforward NN with 3 inputs, 2 outputs, 7 hidden layers of 7 neurons each, ReLU activations; $M: \mathbb{R}^3 \to \mathbb{R}^2$





Output reachable set $Y = M(\mathcal{X})$: union of 1250 polytopes, shown in different colors



8000 randomly generated outputs (evaluating M on points, e.g., $M(x)$ for 8000 points $x \in \mathcal{X}$)

Scalability Challenge: for ReLU activations, this problem is NP-complete

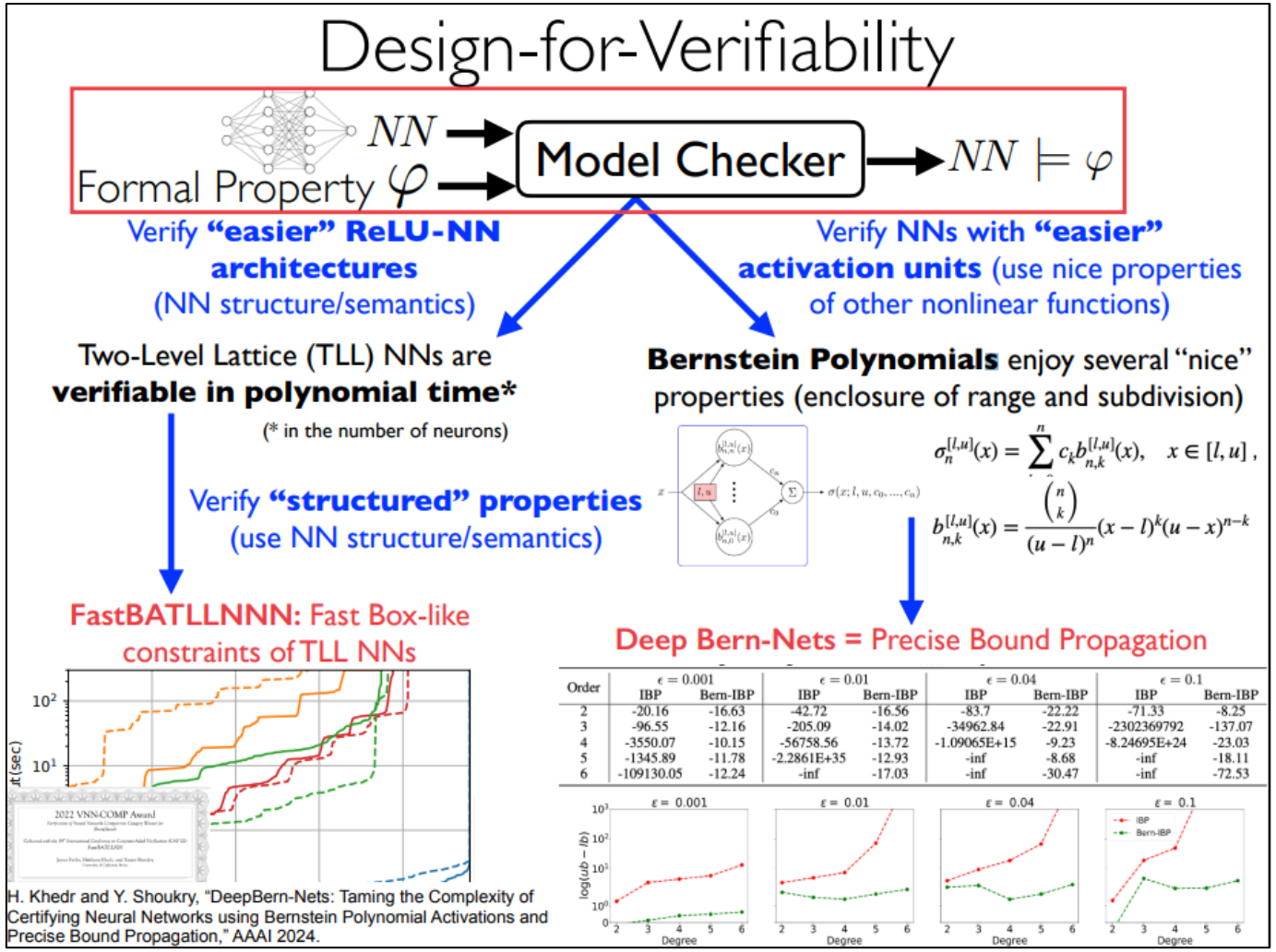
Intuition: number of polytopes may grow exponentially in number of ReLUs due to case splitting

- Can verify MNIST, VGG19, …
- Still long way to go

Neural Network Verification in Autonomous Cyber-Physical Systems (CPS), Taylor Johnson, 85th IPIF WG 10.4 Meeting, 2024

Assured NN-Based Perception & Control

- Design-for-Verifiability
  - Verify “easier” ReLU-NN architectures
  - Verify NNs with “easier” activation units

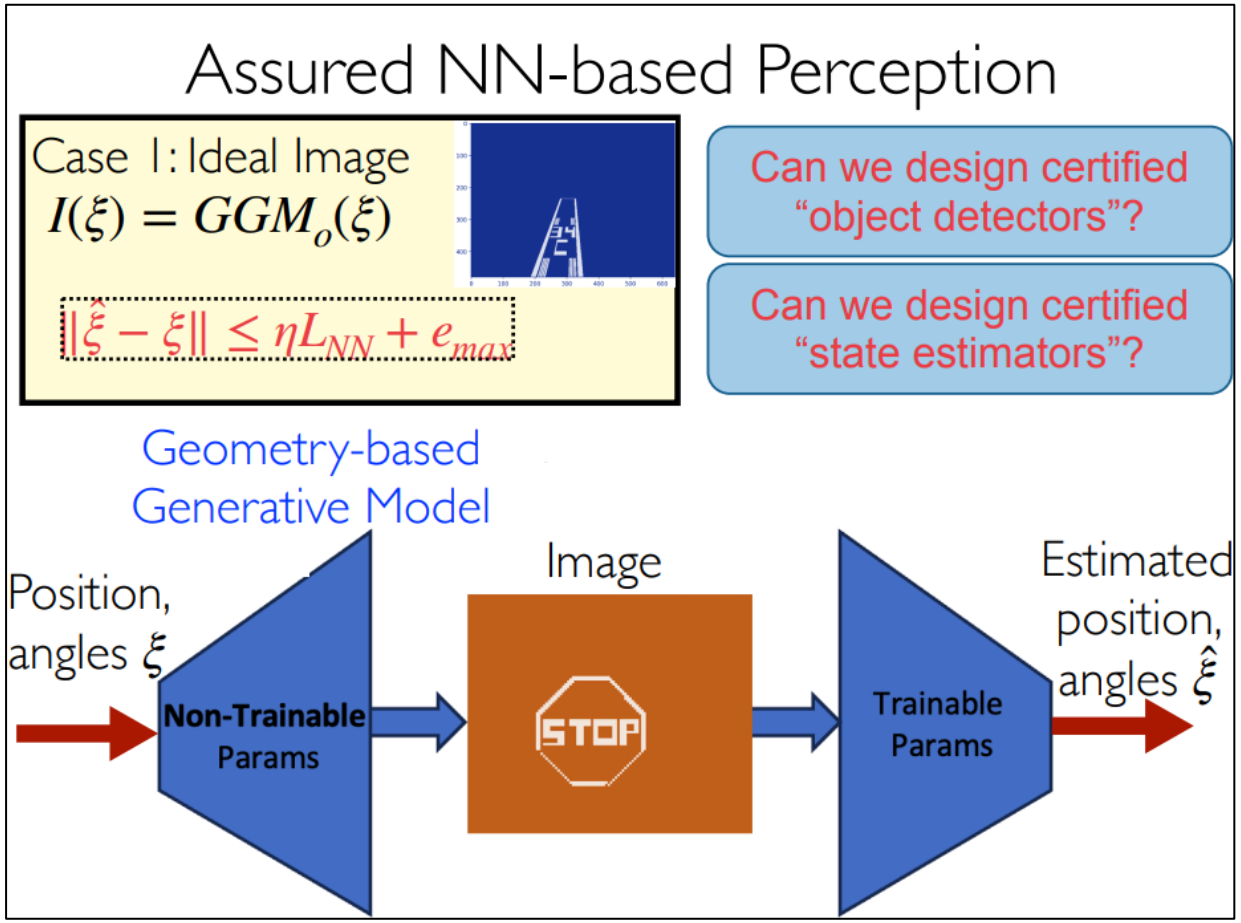


Order	ε = 0.001 IBP	ε = 0.001 Bern-IBP	ε = 0.01 IBP	ε = 0.01 Bern-IBP	ε = 0.04 IBP	ε = 0.04 Bern-IBP	ε = 0.1 IBP	ε = 0.1 Bern-IBP
2	-20.16	-16.63	-42.72	-16.56	-83.7	-22.22	-71.33	-8.25
3	-96.55	-12.16	-205.09	-14.02	-34962.84	-22.91	-2302369792	-137.07
4	-3550.07	-10.15	-56758.56	-13.72	-1.09065E+15	-9.23	-8.24695E+24	-23.03
5	-1345.89	-11.78	-2.2861E+35	-12.93	-inf	-8.68	-inf	-18.11
6	-109130.05	-12.24	-inf	-17.03	-inf	-30.47	-inf	-72.53

Assured Perception and Control of Autonomous Systems Using Formal Verification of Neural Networks, Yasser Shoukry, 85th IPIF WG 10.4 Meeting, 2024

Assured NN-Based Perception & Control

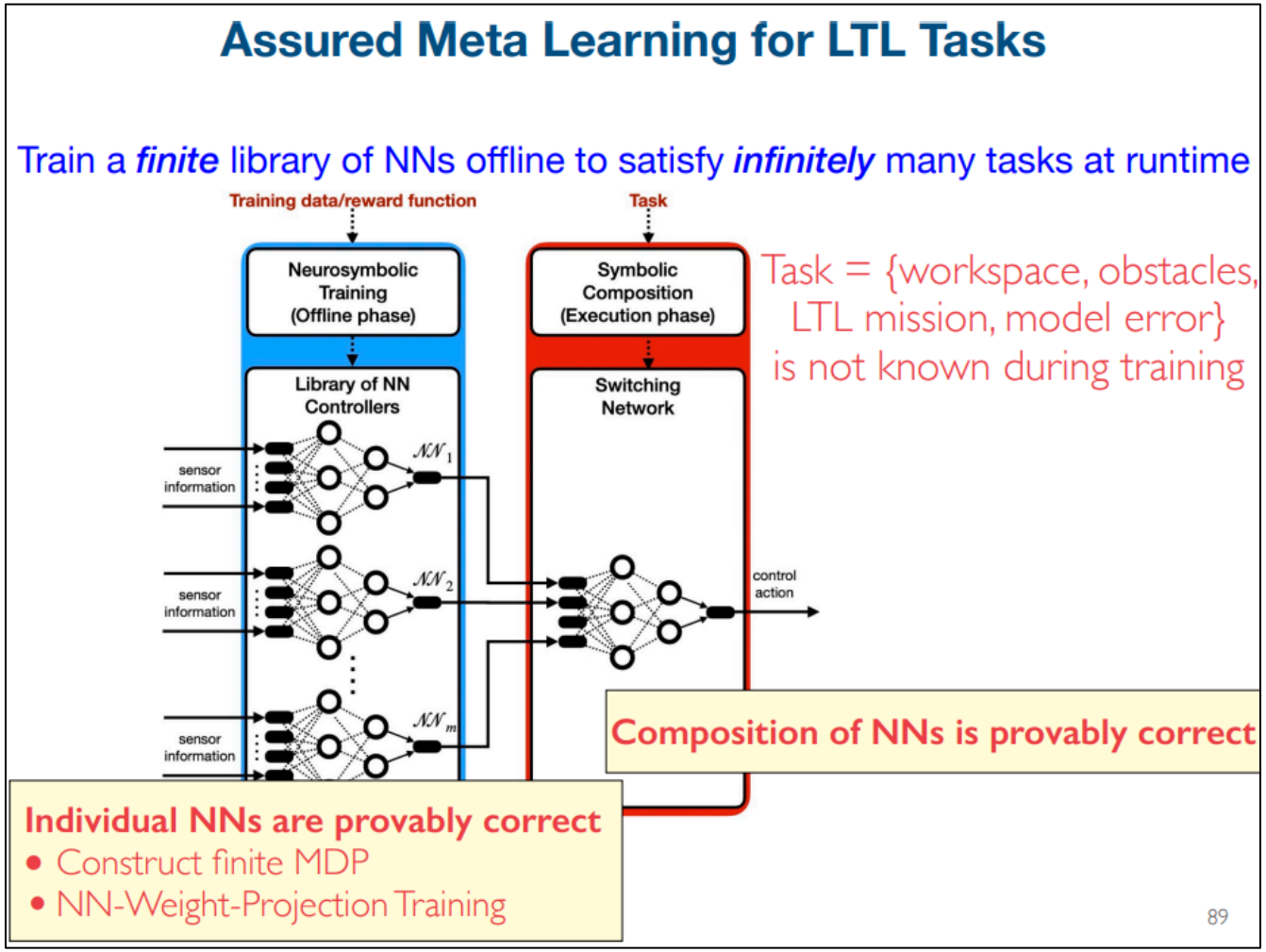
- Design-for-Verifiability
  - Verify “easier” ReLU-NN architectures
  - Verify NNs with “easier” activation units
- Assured NN-Based Perception
  - train NNs with provable guarantees



Assured Perception and Control of Autonomous Systems Using Formal Verification of Neural Networks, Yasser Shoukry, 85th IPIF WG 10.4 Meeting, 2024

Assured NN-Based Perception & Control

- Design-for-Verifiability
  - Verify “easier” ReLU-NN architectures
  - Verify NNs with “easier” activation units
- Assured NN-Based Perception
  - train NNs with provable guarantees
- Assured NN-Based Control
  - Assured meta learning



Assured Perception and Control of Autonomous Systems Using Formal Verification of Neural Networks, Yasser Shoukry, 85th IPIF WG 10.4 Meeting, 2024