CAREER: Decision Procedures for High-Assurance AI-Controlled CPS

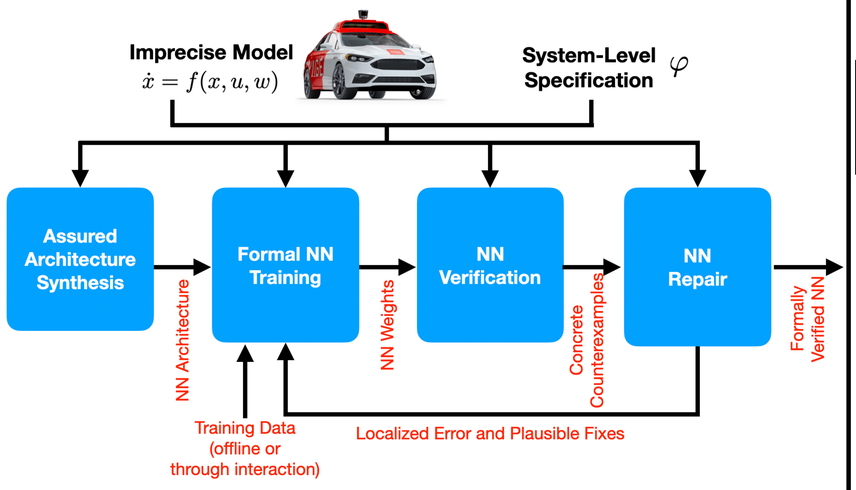
Yasser Shoukry, University of California, Irvine

Motivations

- Safety and reliability of AI-Controlled CPS are understudied problems.
- Lack of widely-accepted, precise, mathematical specifications capturing the correct behavior of AI-agents.
- Even a formally verified system may still fail in real scenarios due to the discrepancy between models used for verification and the real system.

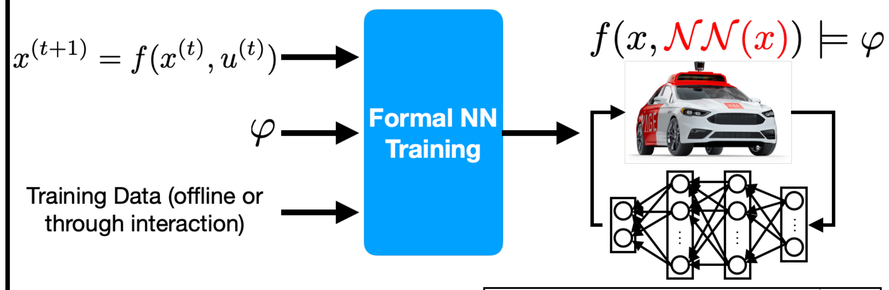
Objective

- Develop scalable formal methods to reason about the safety and reliability of AI-controlled CPS.
- Characterize the environments for which AI-controlled CPS are not safe to operate.
- Blame analysis in failed, yet formally verified AI-controlled CPS.



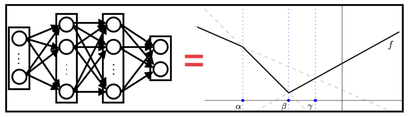
Provably Correct Training of Neural Network Controllers

- **Given** a model of the physical system and a formal specification
- **Train** a neural network controller that renders the closed loop system provably correct.



Core idea:

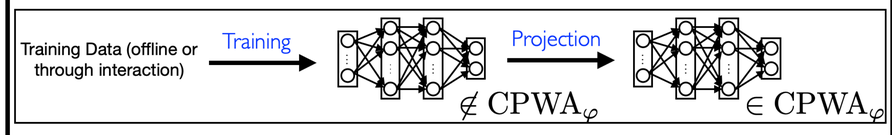
- Regression ReLU NN are Continuous Piece-Wise Affine (CPWA) functions

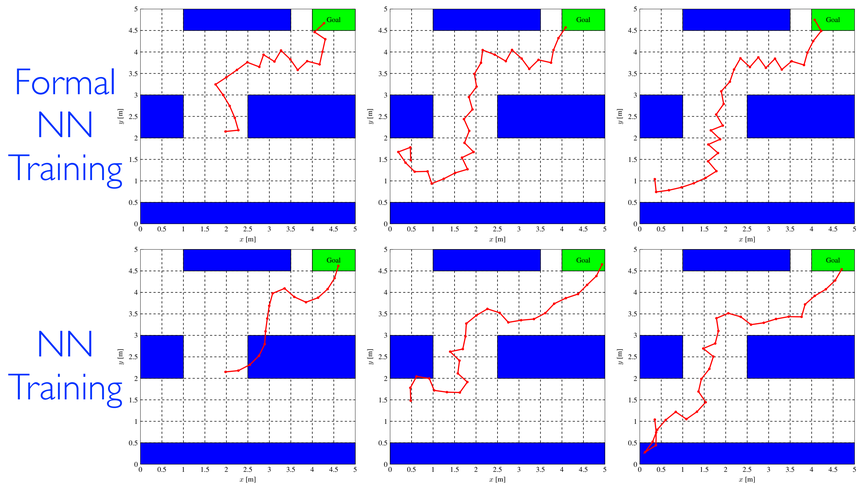


- Use reachability analysis to identify families of CPWA functions that satisfy the specifications.

$$f(x, K_{\text{CPWA}}(x)) \models \varphi \quad \forall\, K_{\text{CPWA}} \in \text{CPWA}_\varphi$$

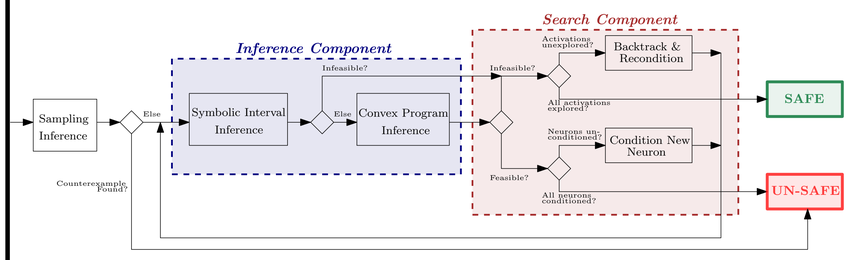
- Train the NN followed by "projecting" the NN weights to the identified family of CPWA functions.

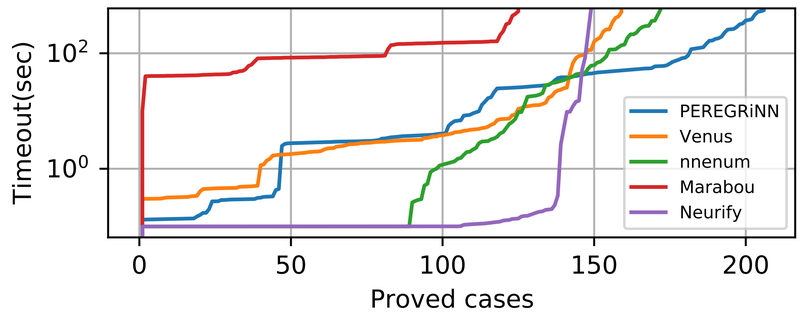




Scalable Formal Verification of Neural Networks

- **Given** a Neural Network and an input-output property.
- **Verify** that the NN follows the input-output property.
- **Solution:** PEREGRiNN: Penalized-Relaxation Greedy Neural Network Verifier.





Publications

- H. Khedr, J. Ferlez, and Y. Shoukry, "PEREGRiNN: Penalized-Relaxation Greedy Neural Network Verifier," CAV, 2021.
- X. Sun and Y. Shoukry, "Provably Correct Training of Neural Network Controllers Using Reachability Analysis," arXiv 2021.
- J. Ferlez, X. Sun, and Y. Shoukry, "Two-Level Lattice Neural Network Architectures for Control of Nonlinear Systems," CDC 2020.
- J. Ferlez and Y. Shoukry, "Bounding the Complexity of Formally Verifying Neural Networks: A Geometric Approach," arXiv 2020.

CPS Lab Resilient

NSF