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

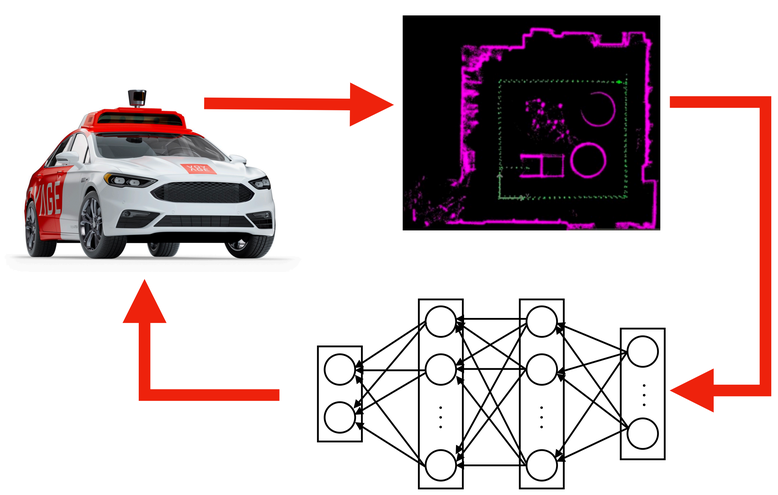
Yasser Shoukry, University of California, Irvine

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.

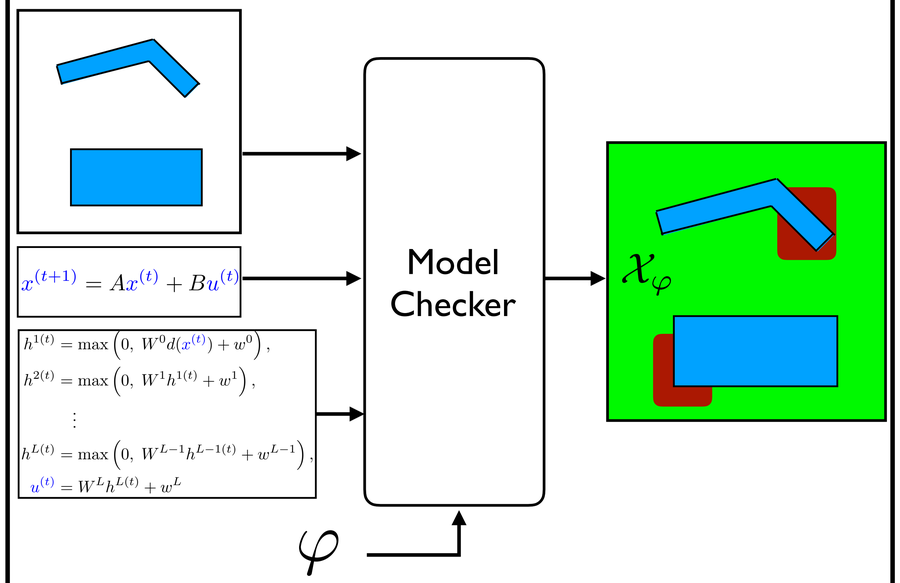
Objective

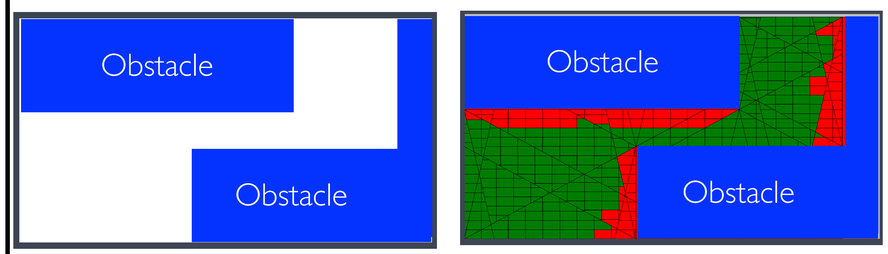
- Develop scalable formal methods to reason about the safety and reliability of AI-controlled CPS.
- Characterize the environments for which AI-controlled CPS are not safe to operate.
- Blame analysis in failed, yet formally verified AI-controlled CPS.

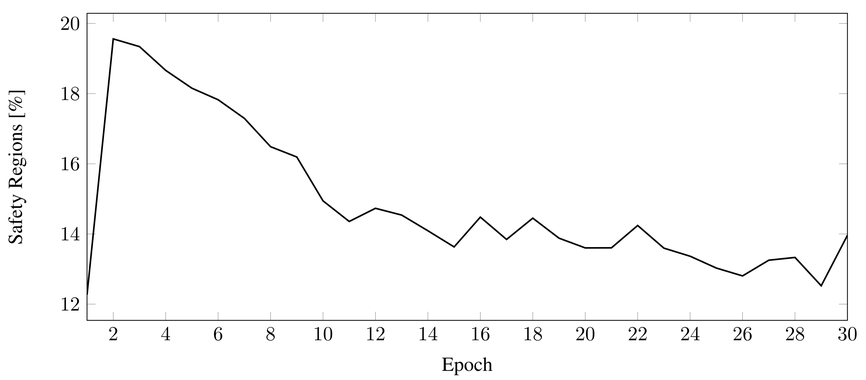


Model Checking of Deep RL

- **Given** a trained neural network and a safety specification.
- **Compute** the set of initial conditions, such that all trajectories starting from this set are guaranteed to satisfy the specification.

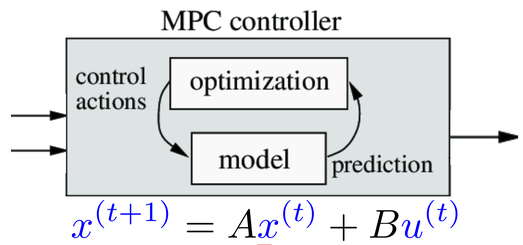


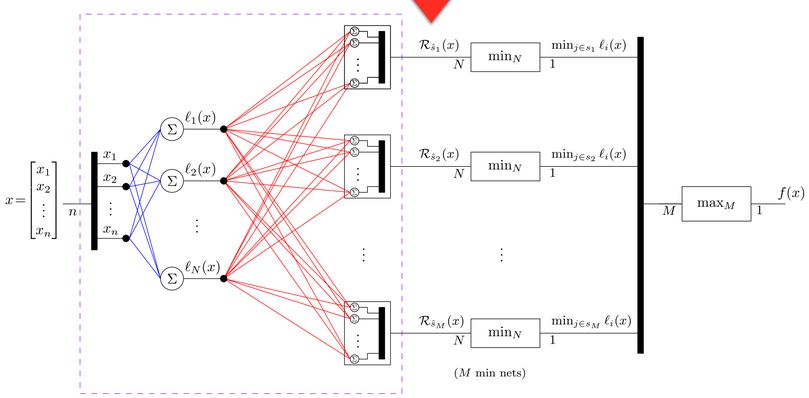




Certifiable DNN Architectures

- **Given** system dynamics, state constraints, input constraints, and a quadratic objective function.
- **Compute** a deep neural network architecture (number of layers and number of neurons per layer), such that there exists an assignment for the weights that render the network equivalent to a Model Predictive Controller.





Publications

- J. Ferlez and Y. Shoukry, "AReN: Assured ReLU NN Architecture for Model Predictive Control of LTI Systems," ArXiv 2019.
- X. Sun, H. Khder, and Y. Shoukry, "Formal Verification of Neural Network Controlled Autonomous Systems," HSCC 2018.