# Fast Reachability As a Building Block for Verified Autonomy

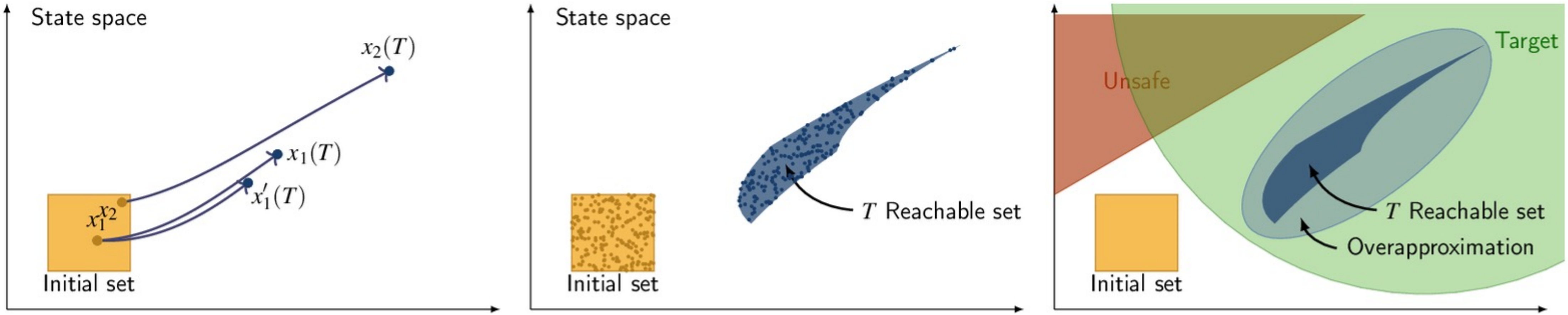
**Sam Coogan**
Associate professor
Electrical and Computer Engineering
Civil and Environmental Engineering

November 8, 2022

Georgia Tech

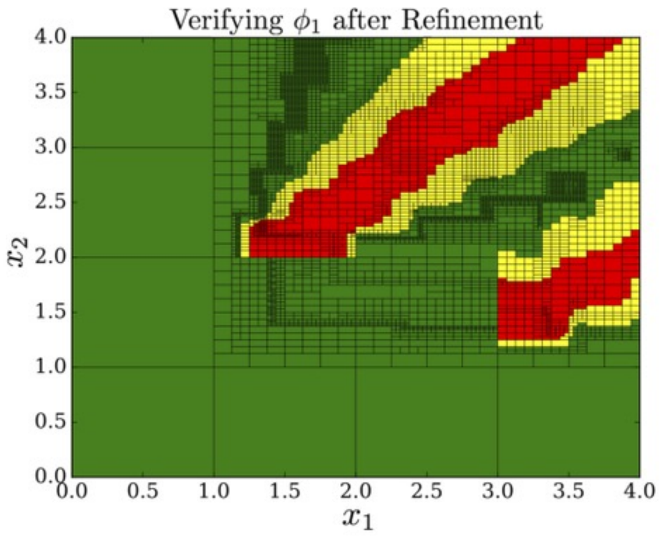
# Reachable sets of dynamical systems

**System:** $\dot{x} = f(x, w)$ **State:** $x \in \mathbb{R}^n$ **Disturbance:** $w \in \mathcal{W}$

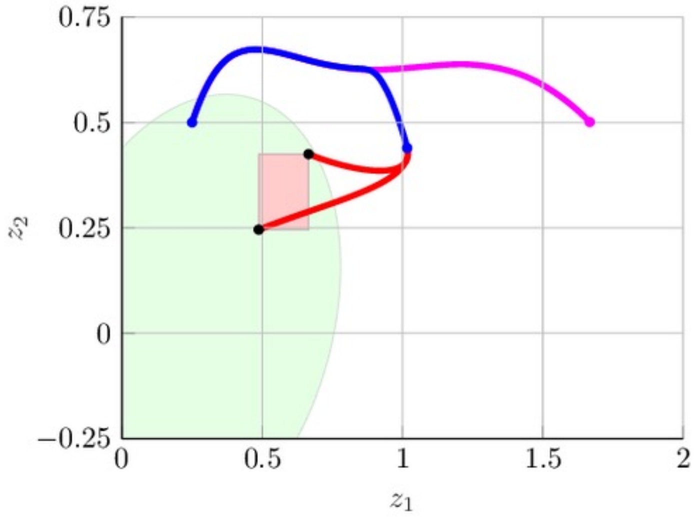



- Reachable sets characterize possible system evolution
- Overapproximations of reachable sets are appropriate for verification and safety

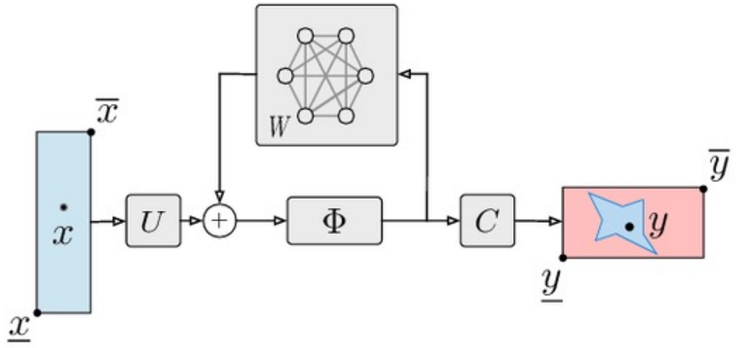
# Need for fast reachability methods




For **formal methods:**
Reachability from each region of a finite abstraction




For **safe autonomy:**
Reachability in the control loop for runtime assurances



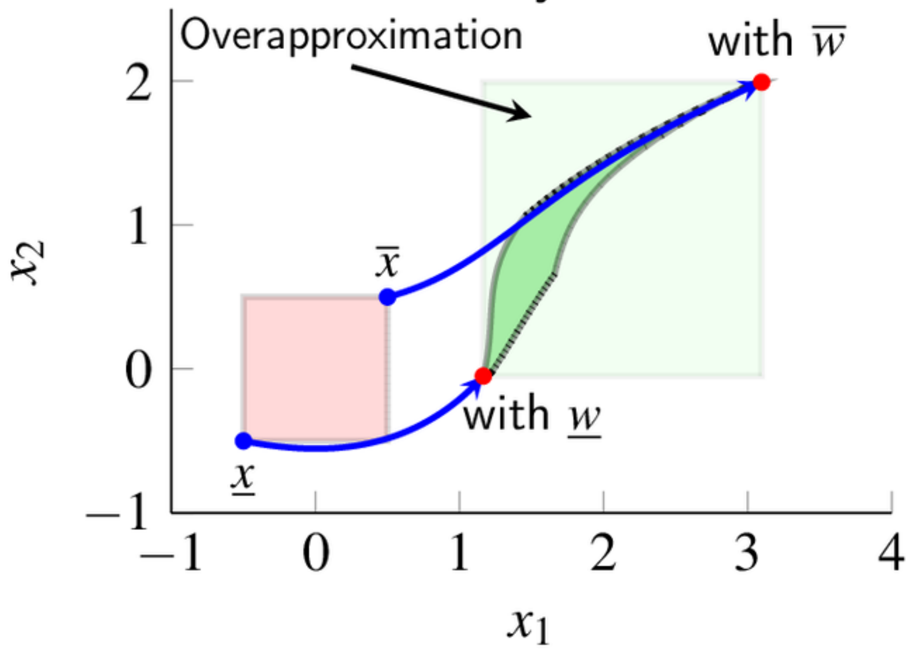

For **NN verification:**
High dimensional reachability

# Mixed monotonicity for interval reachability estimates

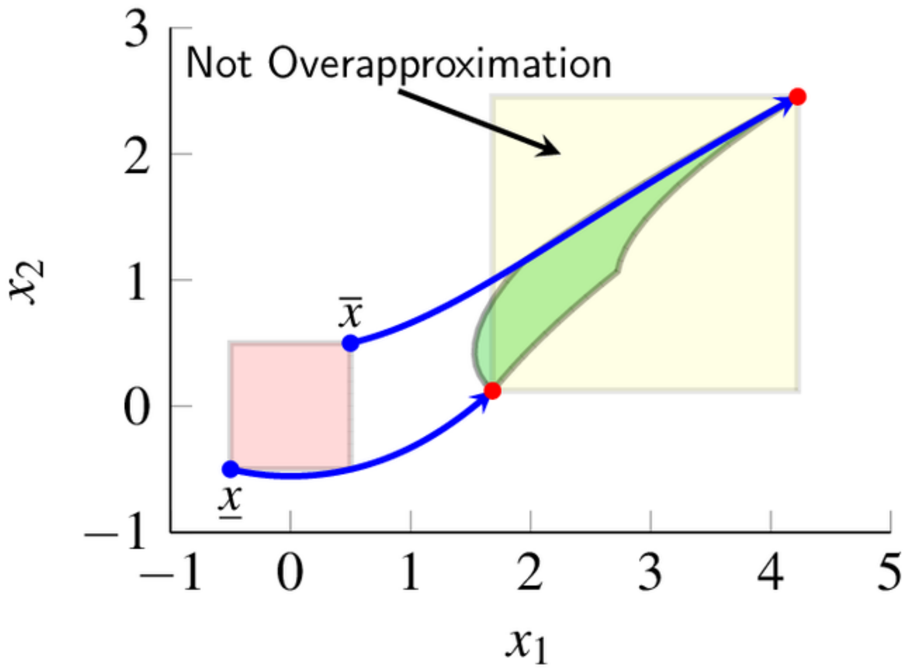
*Reachability analysis for monotone systems.* For a monotone system,

$$\text{Reachable set} \subseteq [\text{lower trajectory}, \text{upper trajectory}].$$

Monotone system




Nonmonotone system




Goal of **mixed monotonicity**: Embed nonmonotone system in a monotone system

# Reachability from embedding system

System: $\dot{x} = f(x,w)$, disturbance input $w \in \mathcal{W} = [\underline{w}, \overline{w}] = \{w : \underline{w} \le w \le \overline{w}\}$

$$\Downarrow$$

2$n$ dimensional embedding system: $\begin{bmatrix} \dot{x} \\ \dot{\widehat{x}} \end{bmatrix} = e(x,\widehat{x}) := \begin{bmatrix} d(x,\underline{w},\widehat{x},\overline{w}) \\ d(\widehat{x},\overline{w},x,\underline{w}) \end{bmatrix}$

- $d$ is a *decomposition function* constructed from the dynamics $f$.

- **MM is fast**: A single trajectory of the deterministic embedding bounds reachable sets of the original mixed monotone system
- **MM is scalable**: If $(\underline{x}_{\mathrm{eq}}, \overline{x}_{\mathrm{eq}})$ is an equilibrium for embedding system, then hyperrectangle $[\underline{x}_{\mathrm{eq}}, \overline{x}_{\mathrm{eq}}]$ is robustly forward invariant

# A simple example

**Mixed Monotone System:**

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}x_2^2+2\\ x_1\end{bmatrix}$$

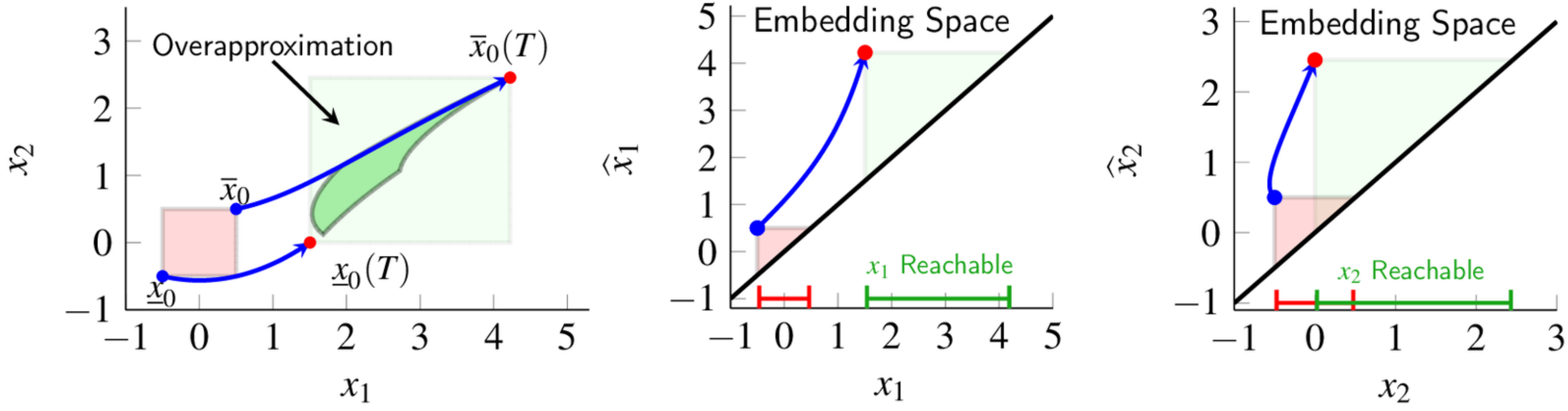
$$[\underline{x},\overline{x}] = [(-0.5,-0.5),(0.5,0.5)]$$

$$T = 1$$

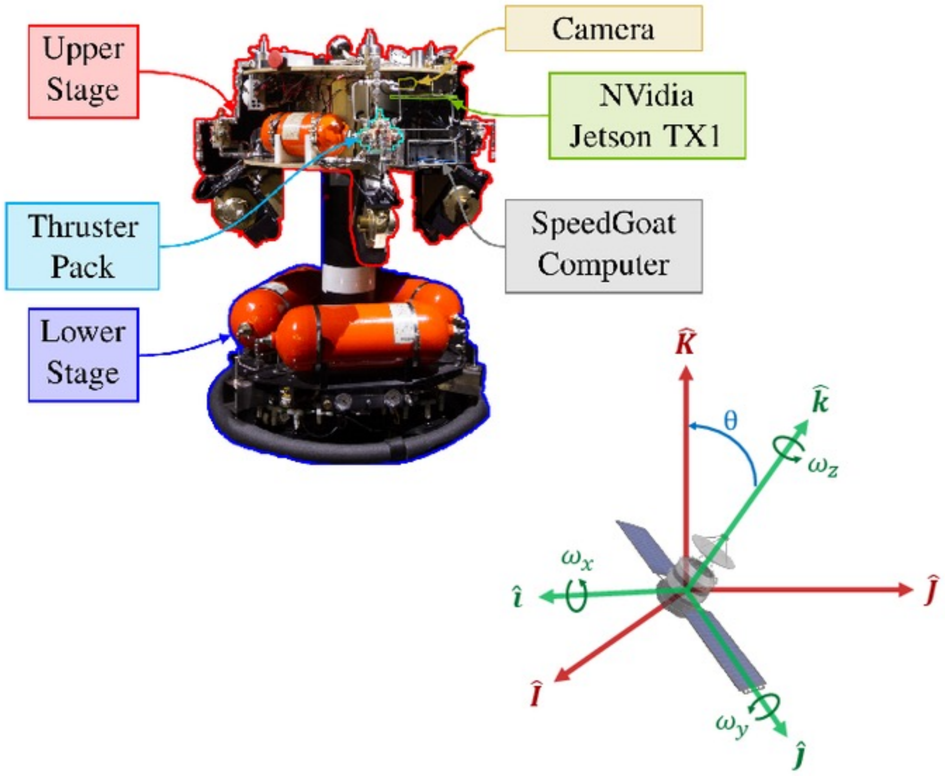
**Decomposition Function:**

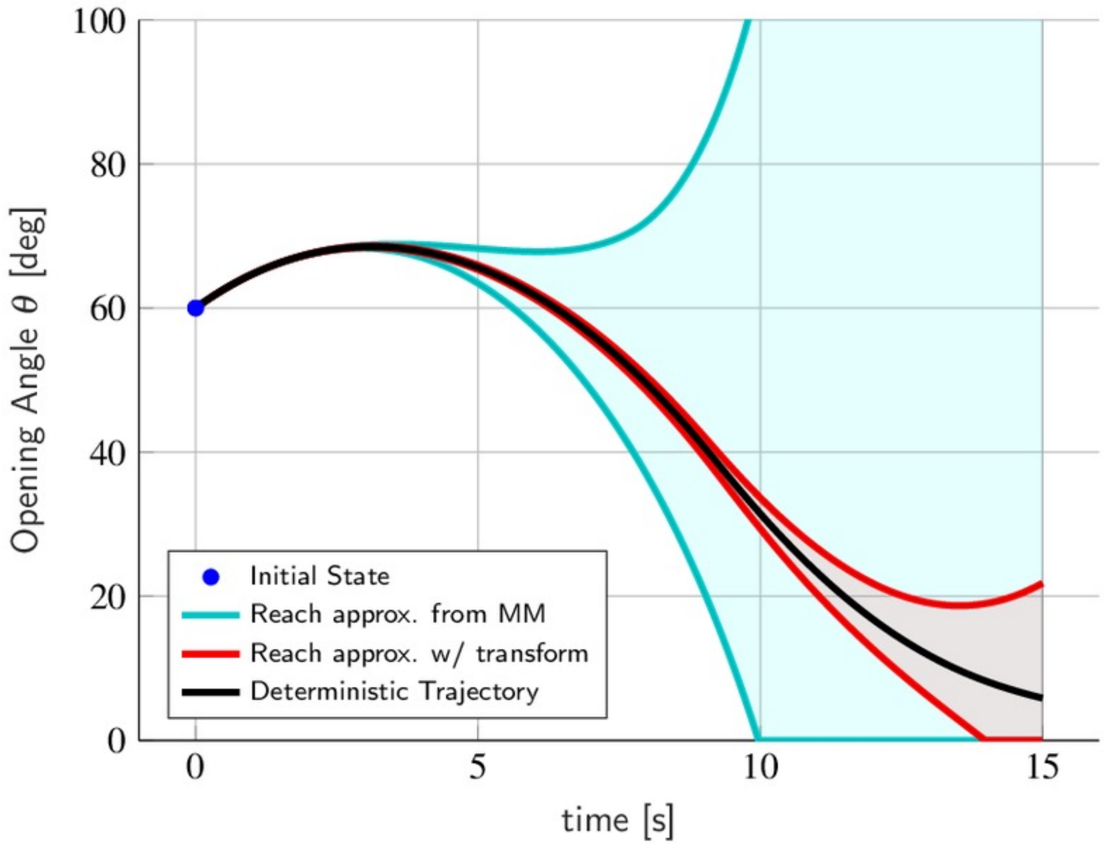
$$d_1(x,\widehat{x}) = \begin{cases} x_2^2+2 & \text{if } x_2 \geq 0 \text{ and } x_2 \geq -\widehat{x}_2, \\ \widehat{x}_2^2+2 & \text{if } \widehat{x}_2 \leq 0 \text{ and } x_2 < -\widehat{x}_2, \\ 2 & \text{if } x_2 < 0 \text{ and } \widehat{x}_2 > 0. \end{cases}$$

$$d_2(x,\widehat{x}) = x_1$$



- Torque-controlled nonlinear spacecraft system in free rotational motion
- Safety objective: Line-of-sight angle constraint, $\theta \leq \theta_{max}$

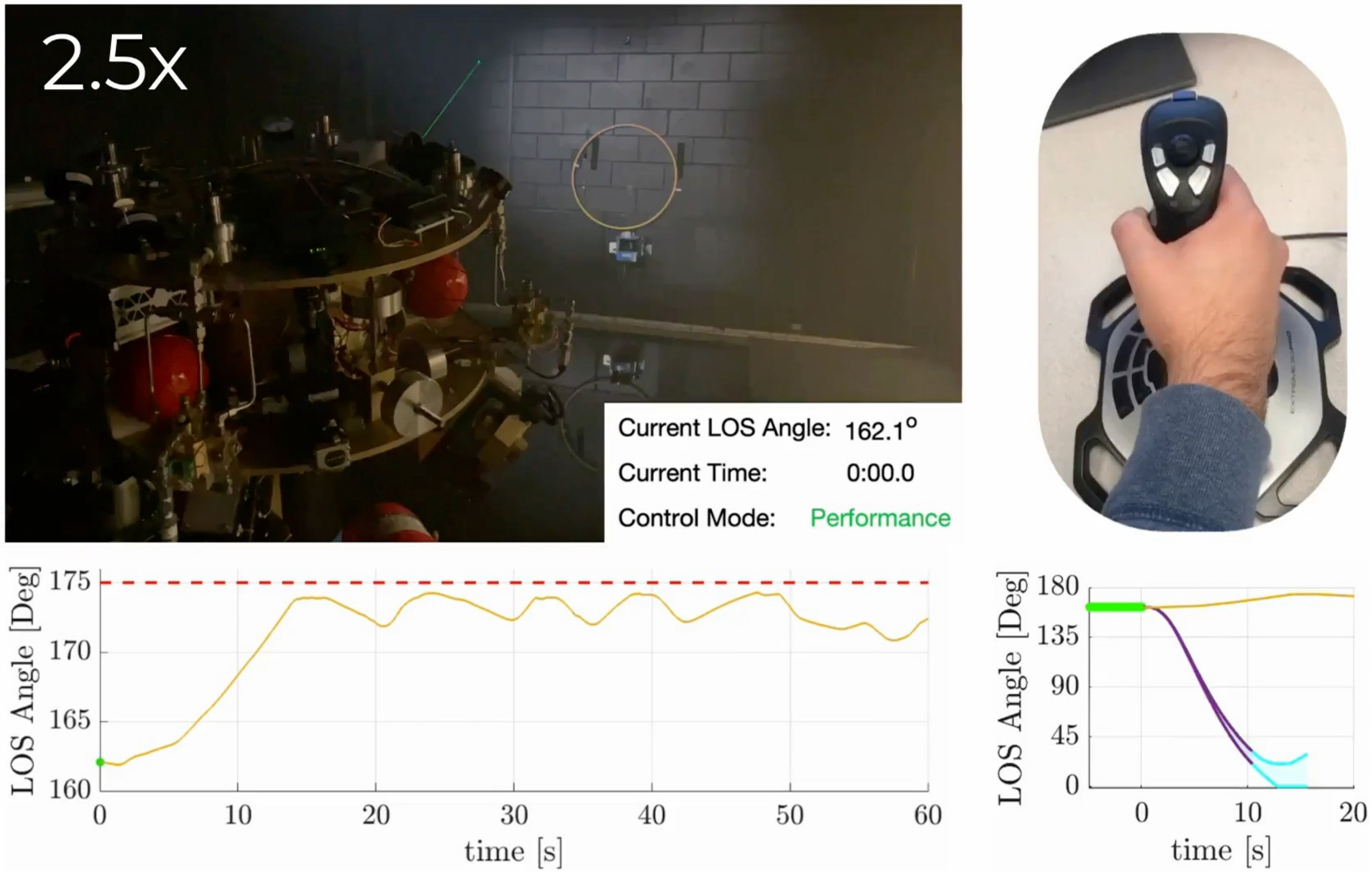



M. Abate, M. Mote, M. Dor, C. Klett, S. Phillips, K. Lang, P. Tsiotras, E. Feron, S. Coogan, in submission.

# Fast reachability for runtime assurance (RTA) from mixed monotonicity

- RTA mechanism overrides human input when reachable set could become unsafe




M. Abate, M. Mote, M. Dor, C. Klett, S. Phillips, K. Lang, P. Tsiotras, E. Feron, S. Coogan, in submission.

# Fast reachability for formal methods verification

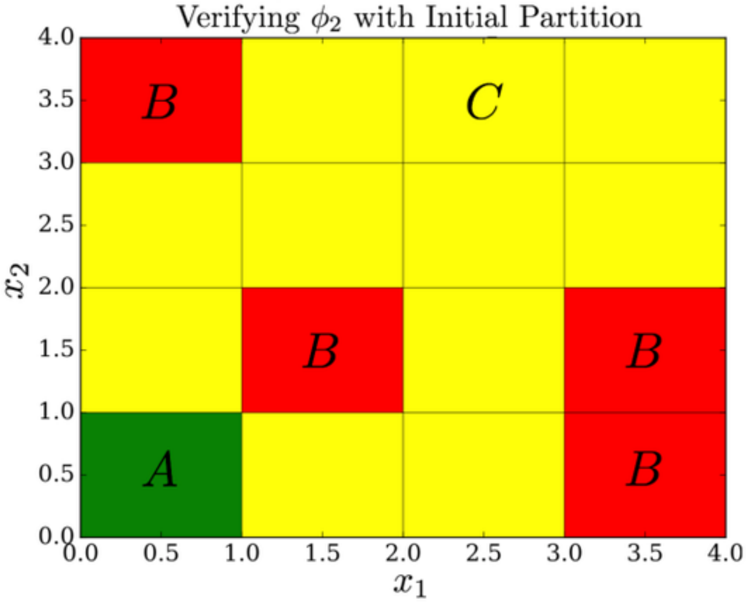
Less than 90% chance of: reaching a $B$ state if it eventually always remains in $A$ state, and always stays outside of $B$ state if it reaches a $C$ state

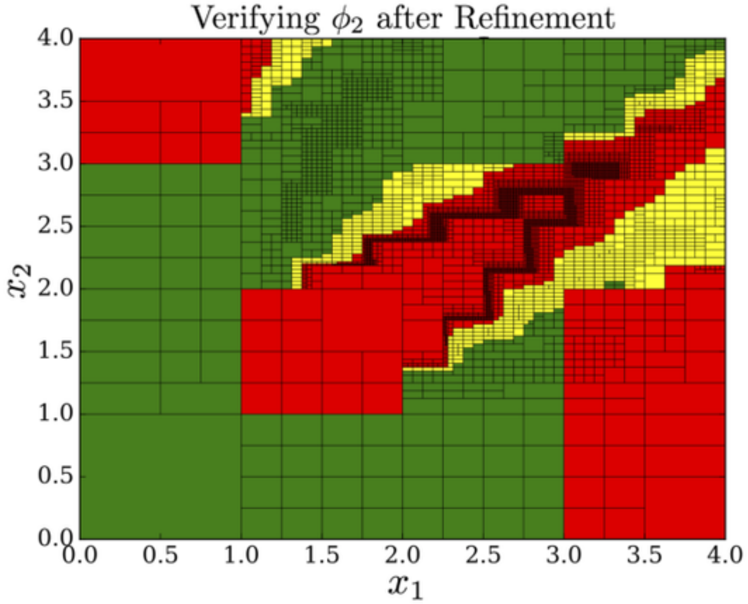
$$\mathcal{P}_{\leq 0.90}[(\Diamond\Box A \rightarrow \Diamond B) \wedge (\Diamond C \rightarrow \Box\neg B)]$$

$$x_1^+ = x_1 + (-ax_1 + x_2)\cdot\Delta T + w_1$$

$$x_2^+ = x_2 + \left(\frac{(x_1)^2}{(x_1)^2+1} - bx_2\right)\cdot\Delta T + w_2$$

$w_1, w_2$ truncated Gaussian

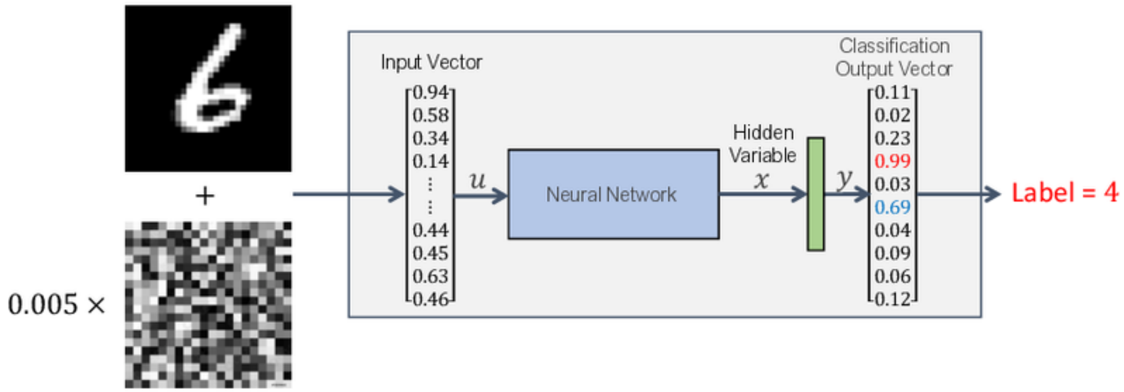



M. Dutreix, S. Coogan, *IEEE TAC*, 2021.

# Scalable mixed monotonicity for robustness analysis/training in NNs

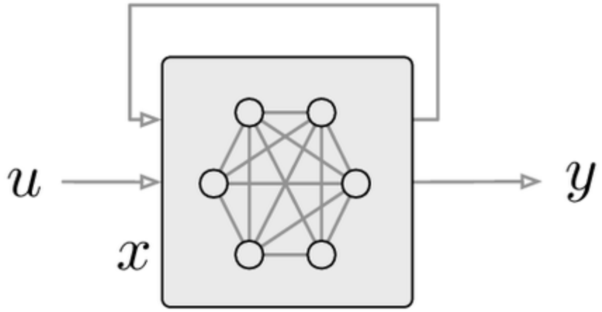
A well known problem: Neural networks are brittle




Increased interest in *Implicit* Neural Networks defined by fixed point equation:

$$x = \Phi(Ax + Bu)$$

$$xy = Cx$$



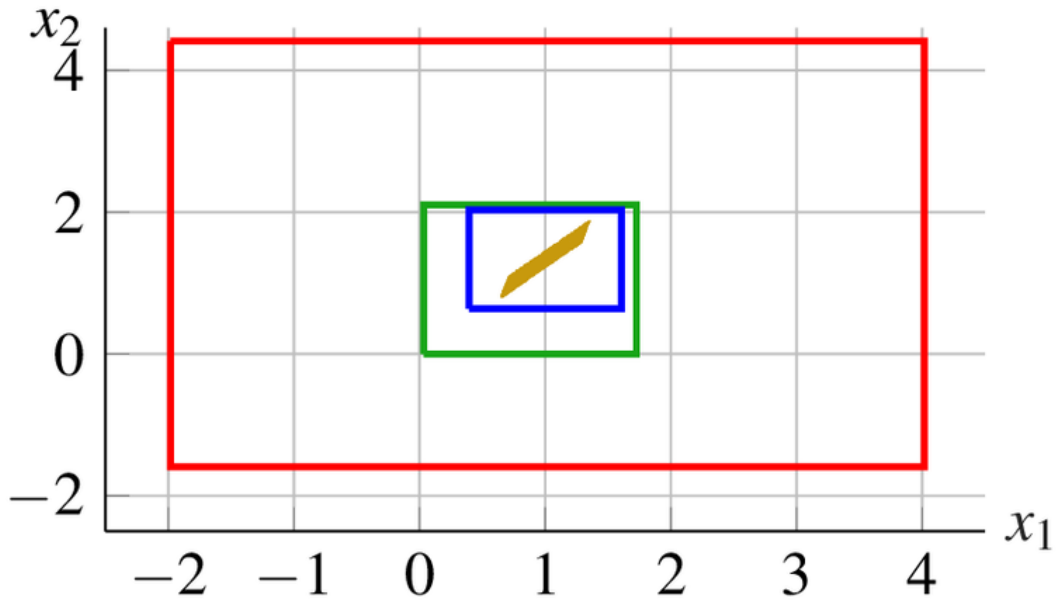

**Mixed monotonicity provides computationally and theoretically scalable local robustness certificates**

S. Jafarpour, M. Abate, A. Davydov, F. Bullo, S. Coogan, L4DC, 2022.

# A simple example

Consider $x = \Phi(Ax + Bu)$ with output $y = x$ and

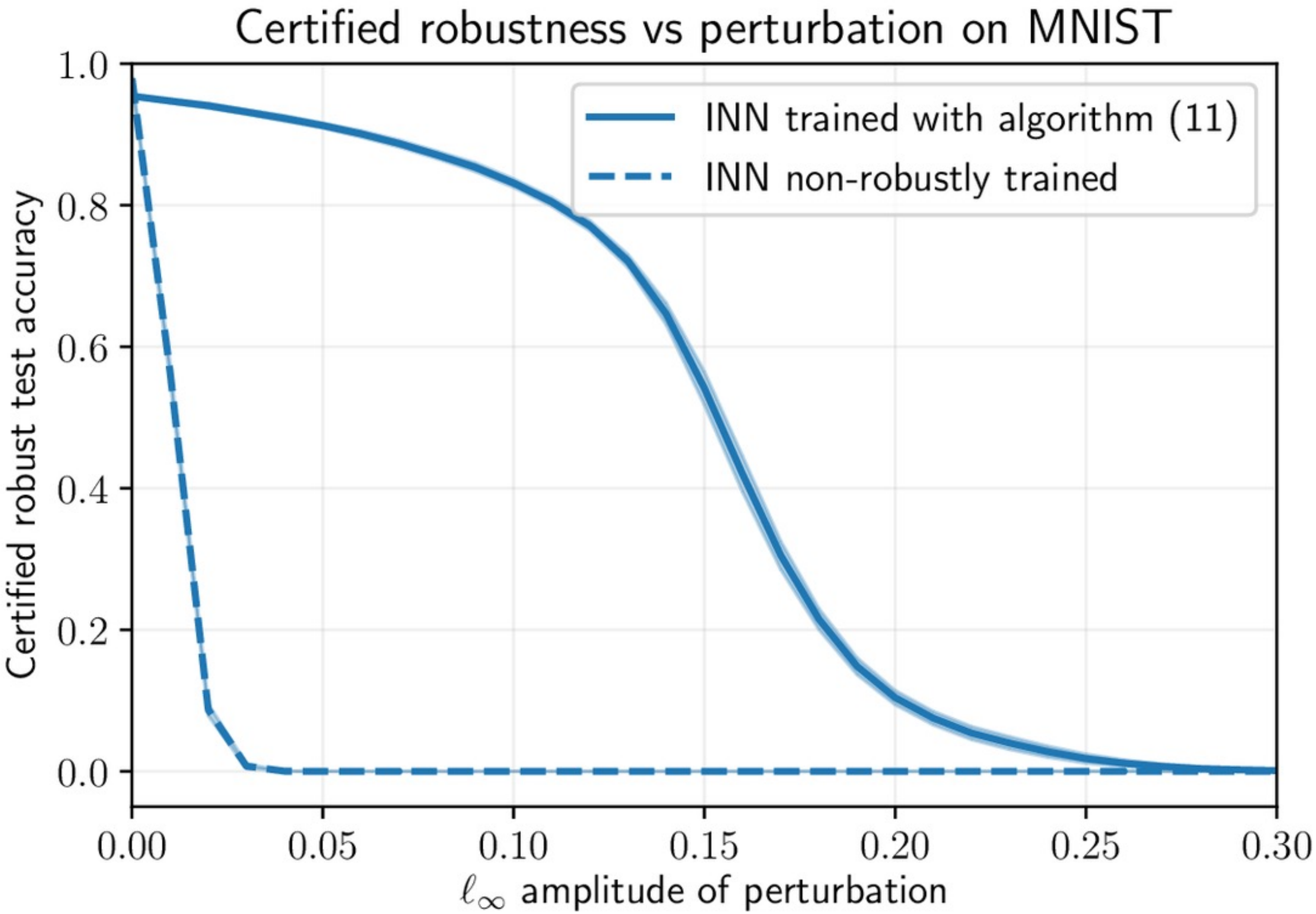
$$A = \begin{bmatrix} -0.25 & -0.25 \\ 0.75 & -0.25 \end{bmatrix},\ B = \begin{bmatrix} 0.5 & 1 \\ 1 & 0.5 \end{bmatrix},\ \Phi\left(\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}\right) = \begin{bmatrix} \mathrm{ReLU}(s_1) \\ \mathrm{ReLU}(s_2) \end{bmatrix},\ \underline{u} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \overline{u} = \begin{bmatrix} 1/3 \\ 2 \end{bmatrix}$$




- Red: Lipschitz bound
- Green: Interval bound propagation
- Blue: Embedding equilibrium

S. Jafarpour, M. Abate, A. Davydov, F. Bullo, S. Coogan, L4DC, 2022.

# Application to MNIST handwriting dataset




S. Jafarpour, M. Abate, A. Davydov, F. Bullo, S. Coogan, L4DC, 2022.

# Concluding thoughts

- Reachability analysis is a fundamental building block for verified autonomy
- Need for fast and scalable approximation methods
- Some challenges:
  1. NN in closed-loop with control systems
  2. Accommodating learning of dynamics in the reachable set computations
  3. Trade-off between offline computation of safe regions vs. online/runtime intervention
  4. Accommodating more complex safety specs than invariance (e.g., temporal logic)
  5. How to intervene to maintain safety (e.g., CBFs, switch to backup controller)
  6. Geometries besides rectangles

Papers available at

`coogan.ece.gatech.edu`