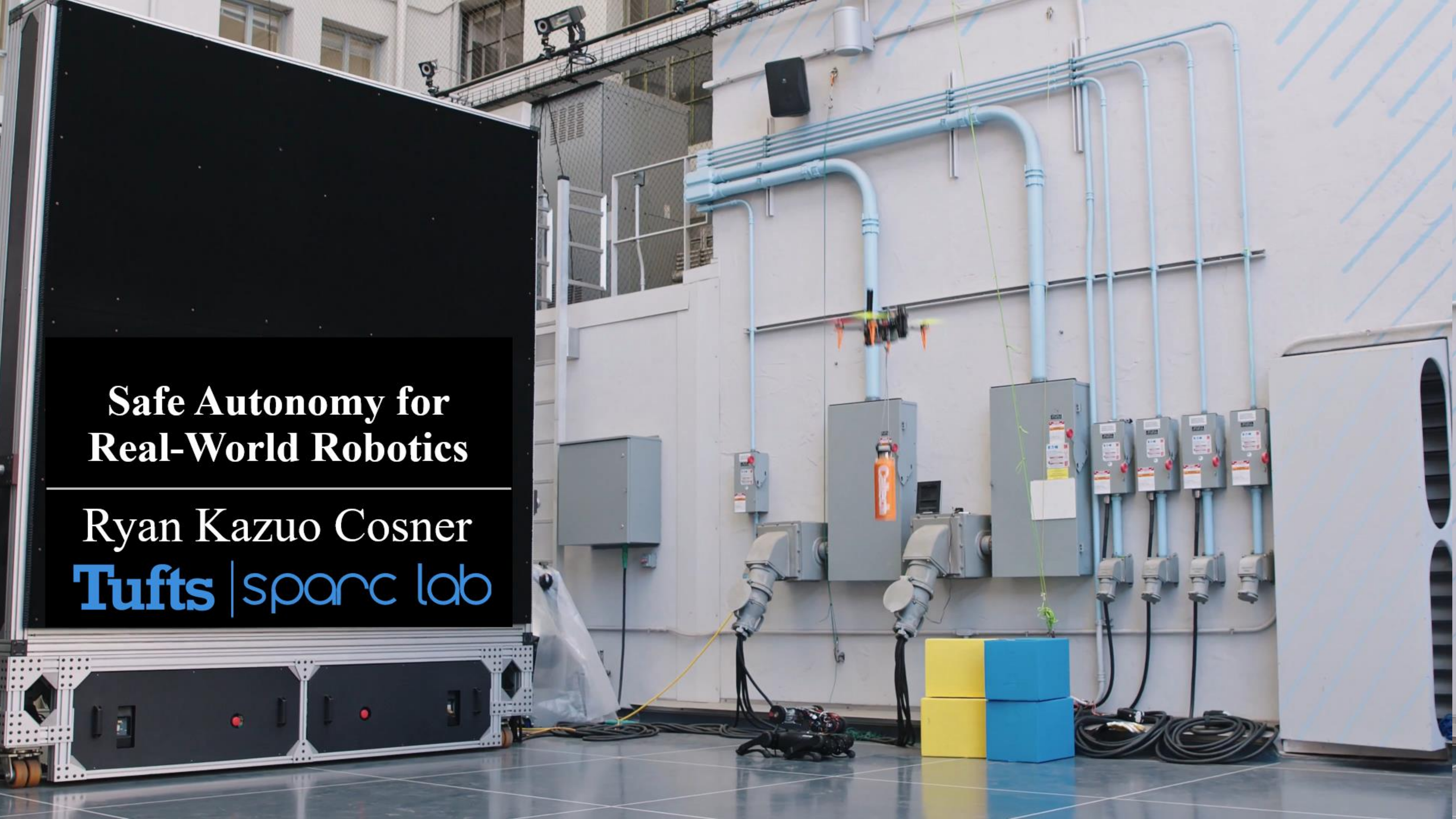


Safe Autonomy for Real-World Robotics

Ryan Kazuo Cosner

Tufts | sparc lab

Hi! I'm Ryan!

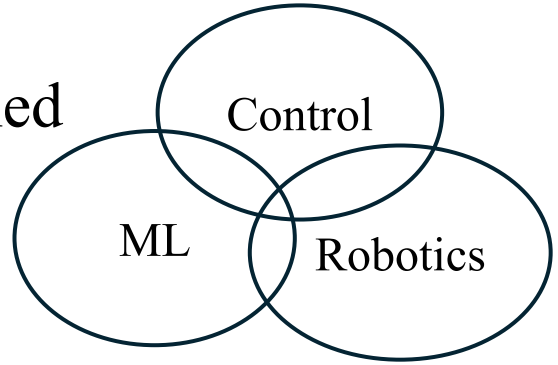
Ryan K. Cosner
Glenn R. Stevens Assistant Professor
Mechanical Engineering, Tufts University

Starting in January 2026:

Tufts University | sparc lab
The Safe + Performant Autonomous Robotics and Control Lab

Research Approach:

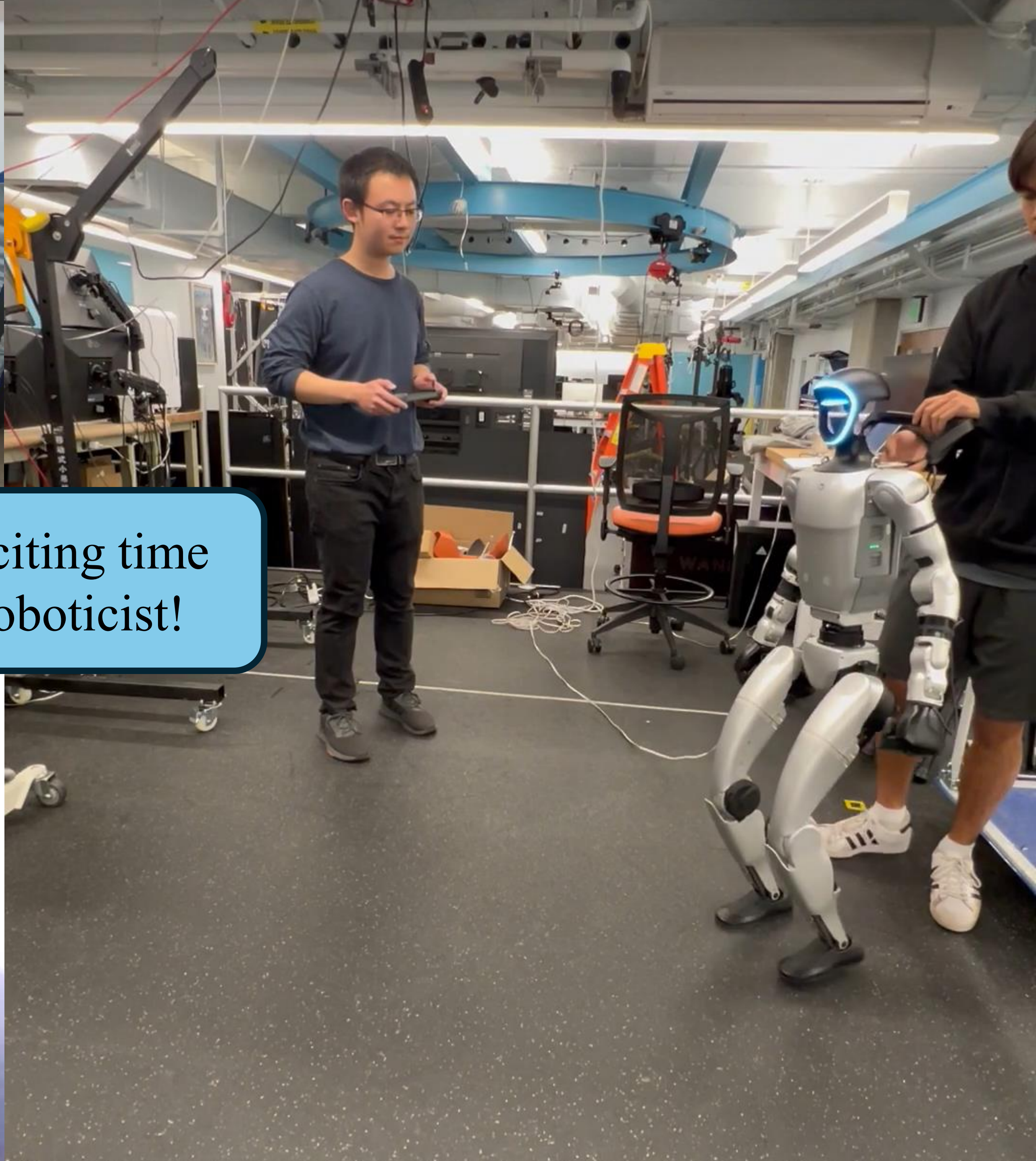
Control theory guarantees combined with ML improvements to create safe + performant robots.



Previously:

Berkeley University of California, Caltech, Ford, Squishy Robotics, NVIDIA

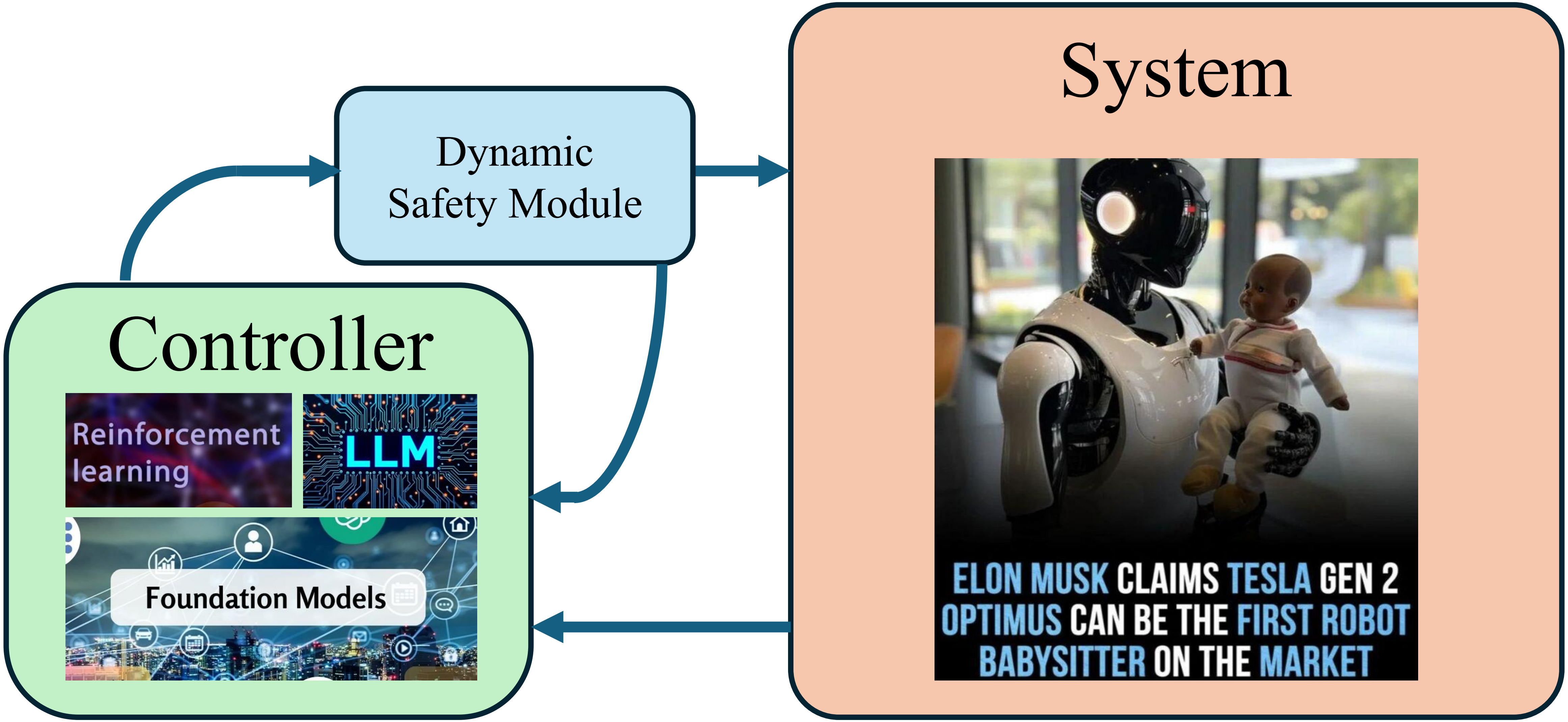
It’s an exciting time
to be a roboticist!

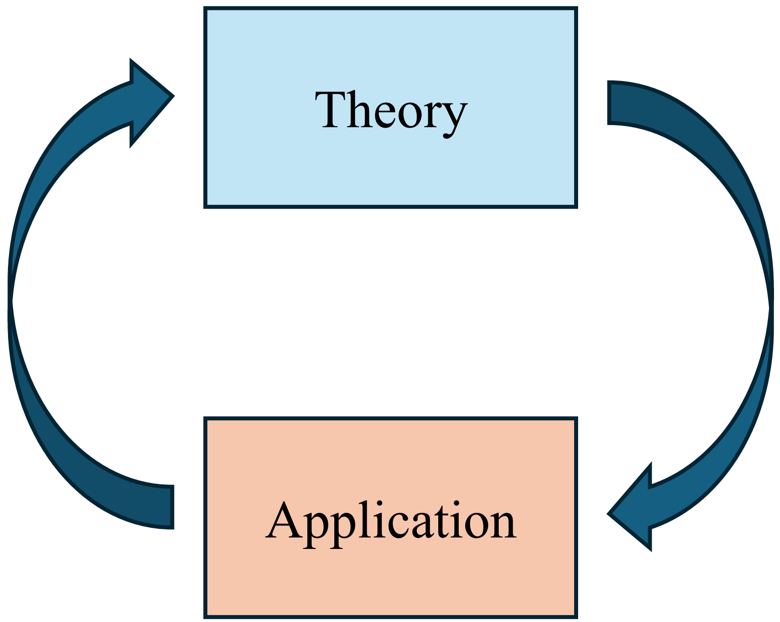


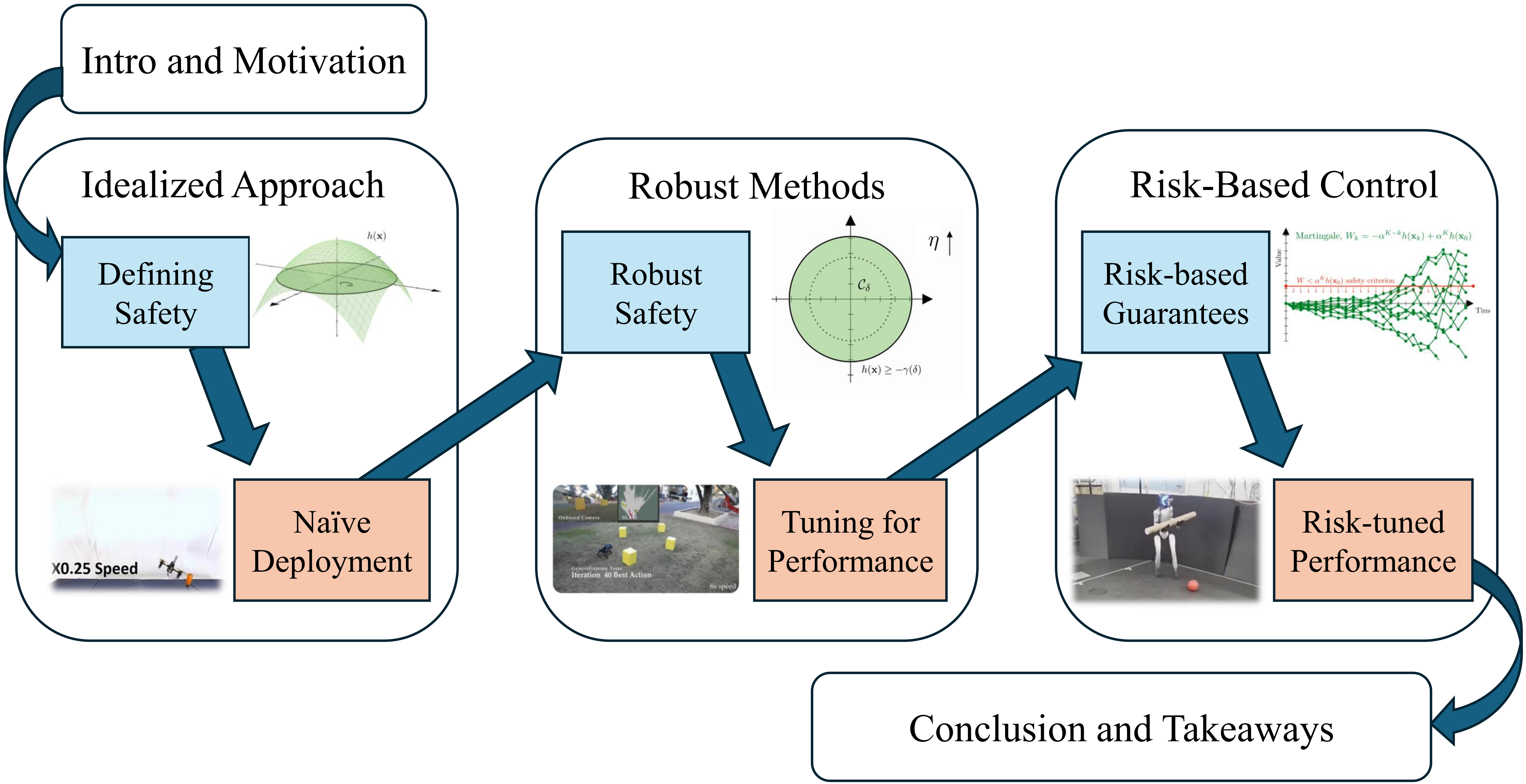
There's still lots to be done!

Real-world safety is hard:

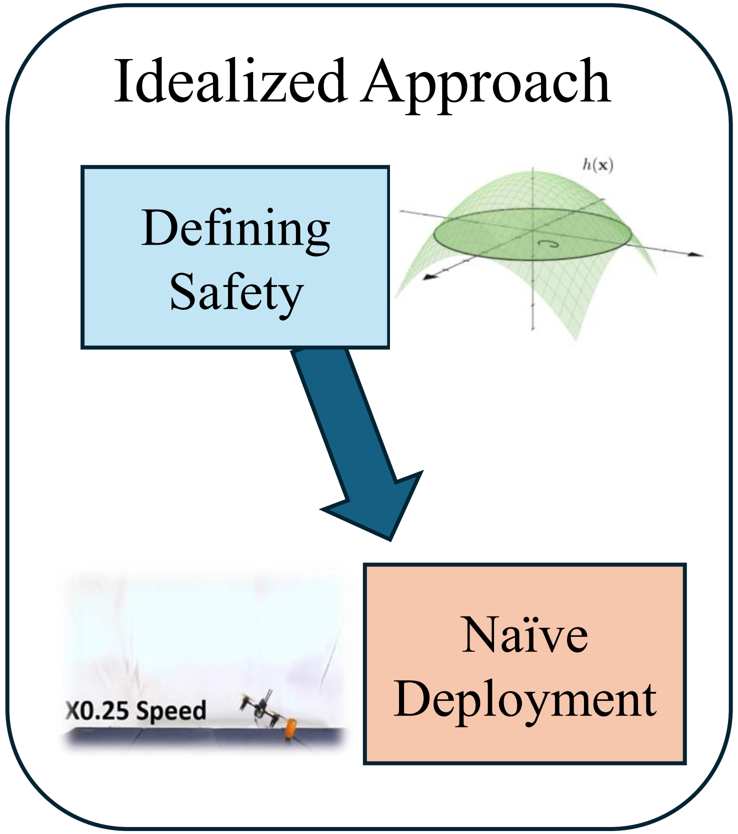
- Learned controllers
- Complex environments
- Noisy sensors
- Multiagent scenarios
- Sim-to-real gaps
- The list is endless …


System
Dynamic
Safety Module
Controller
Reinforcement
learning
LLM
Foundation Models
ELON MUSK CLAIMS TESLA GEN 2
OPTIMUS CAN BE THE FIRST ROBOT
BABYSITTER ON THE MARKET


Theory
Application



Intro and Motivation
Idealized Approach
Defining Safety
Naïve Deployment
X0.25 Speed
Robust Methods
Robust Safety
Tuning for Performance
Risk-Based Control
Risk-based Guarantees
Risk-tuned Performance
Conclusion and Takeaways


Idealized Approach
Defining Safety
$h(\mathbf{x})$
X0.25 Speed
Naïve Deployment

Several methods have emerged

System Model:

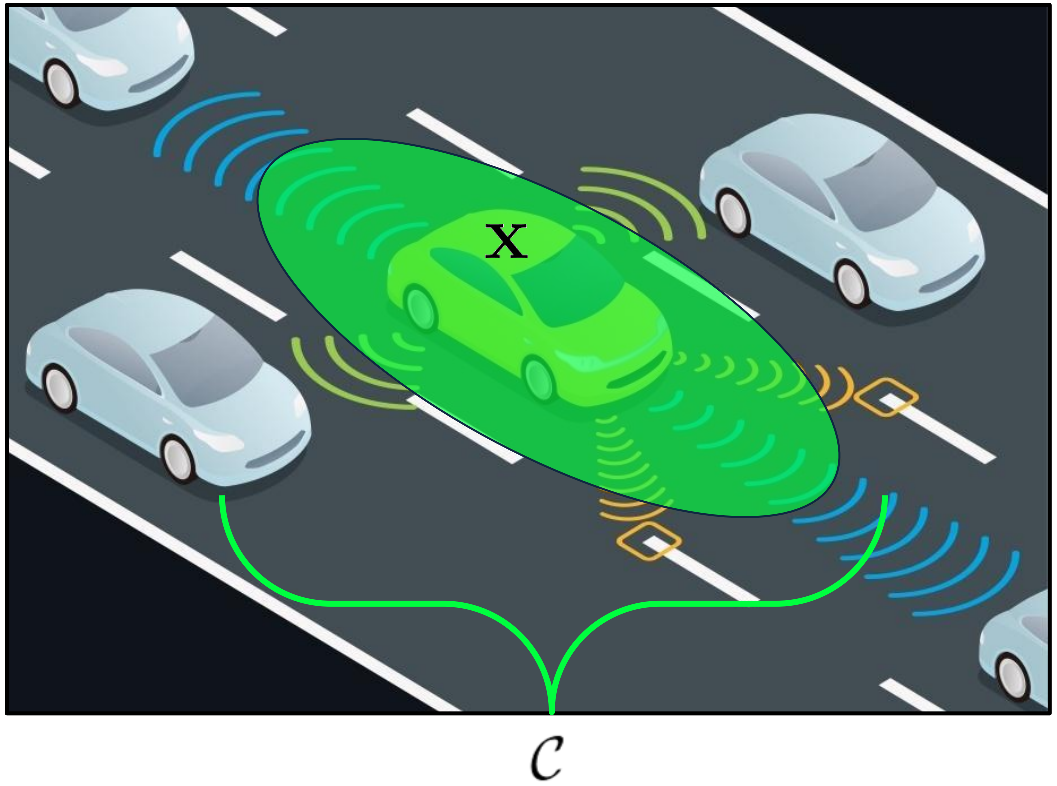
$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0$$

Safety as forward-invariance

Definition: Set Invariance and Safety

If $\mathbf{x}(t_0) \in \mathcal{C} \implies \mathbf{x}(t) \in \mathcal{C},\ \forall t \geq 0,$
then $\mathcal{C}$ is a forward-invariant set and *safe*.



Common methods:

- Hamilton Jacobi methods[1]
- State constraints in model predictive control (MPC)[2]
- Control Barrier Functions[3]

[1] Bansal, et al. *Hamilton-Jacobi reachability: A brief overview and recent advances.* CDC, 2017.
[2] Borrelli, et al. *Predictive control for linear and hybrid systems.* Cambridge University Press, 2017.
[3] Ames, et al. *Control barrier function based QPs for safety critical systems.* TAC, 2017.

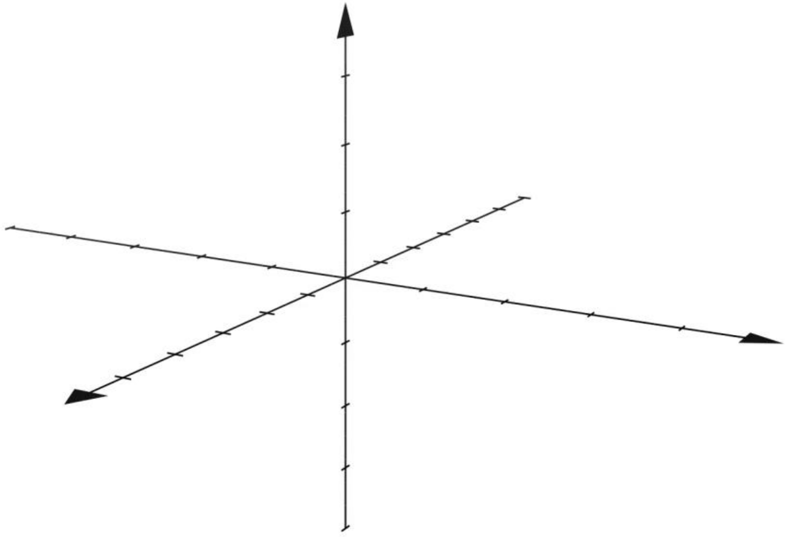
Defining Safety: Control Barrier Functions

User-Defined Safe Set: $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid h(\mathbf{x}) \geq 0\}$

Theorem: CBF Safety [3]

For $\alpha > 0$,

$$\frac{dh}{dt}(\mathbf{x}, \mathbf{u}) \geq -\alpha h(\mathbf{x}) \implies \text{safety.}$$



Growing popularity:

- 4710 publications since introduced in 2014
- Several conference sessions and workshops

[3] Ames, et al. *Control barrier function based quadratic programs for safety critical systems.* TAC, 2017.

Defining Safety: CBF Safety Filter

CBFs are often used in safety filters:

Safety definition: $h(\mathbf{x})$

User command: $\mathbf{k}_{\text{nom}}(t)$

System model: $\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$

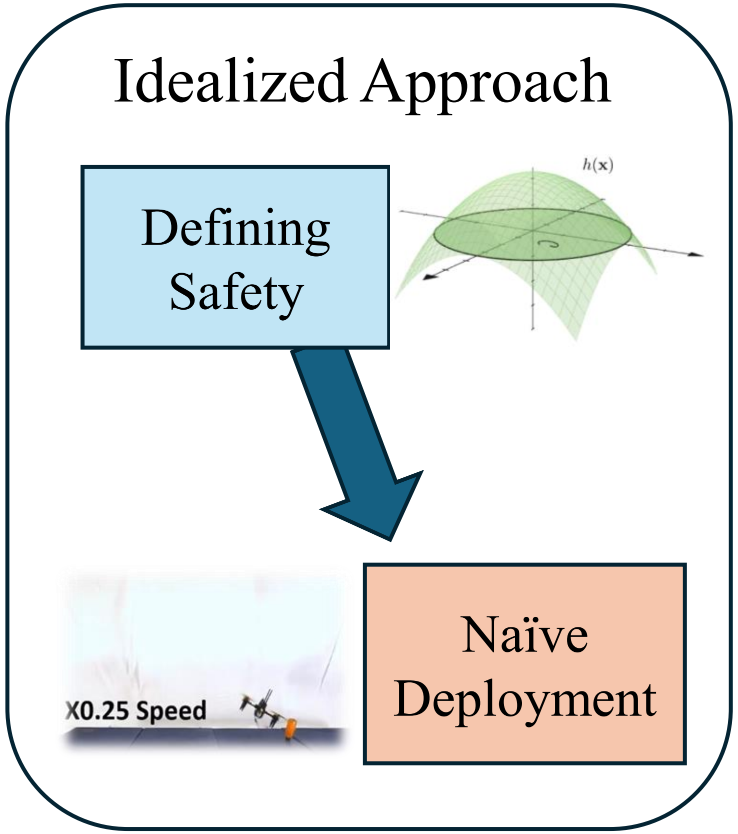
System state: $\mathbf{x}$

Safety Filter:

$$\begin{cases} \mathbf{k}_{\text{nom}}(t), & \text{if } \mathbf{k}_{\text{nom}}(t) \text{ is safe} \\ \mathbf{k}_{\text{nom}}(t) + \frac{L_{\mathbf{g}}h(\mathbf{x})^\top(-L_{\mathbf{f}}h(\mathbf{x}) - \alpha h(\mathbf{x}) - \mathbf{k}_{\text{nom}}(t))}{\|L_{\mathbf{g}}h(\mathbf{x})\|^2}, & \text{else.} \end{cases}$$

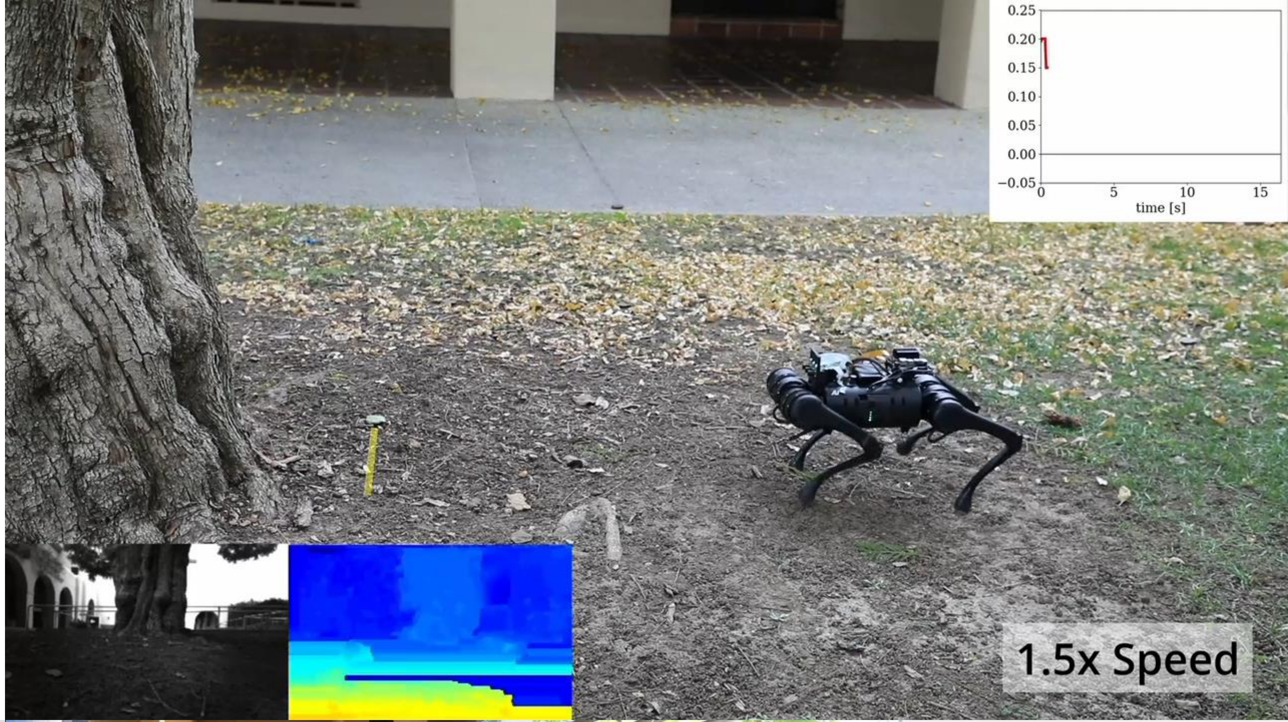
$\mathbf{k}_{\text{safe}}(\mathbf{x}, t)$

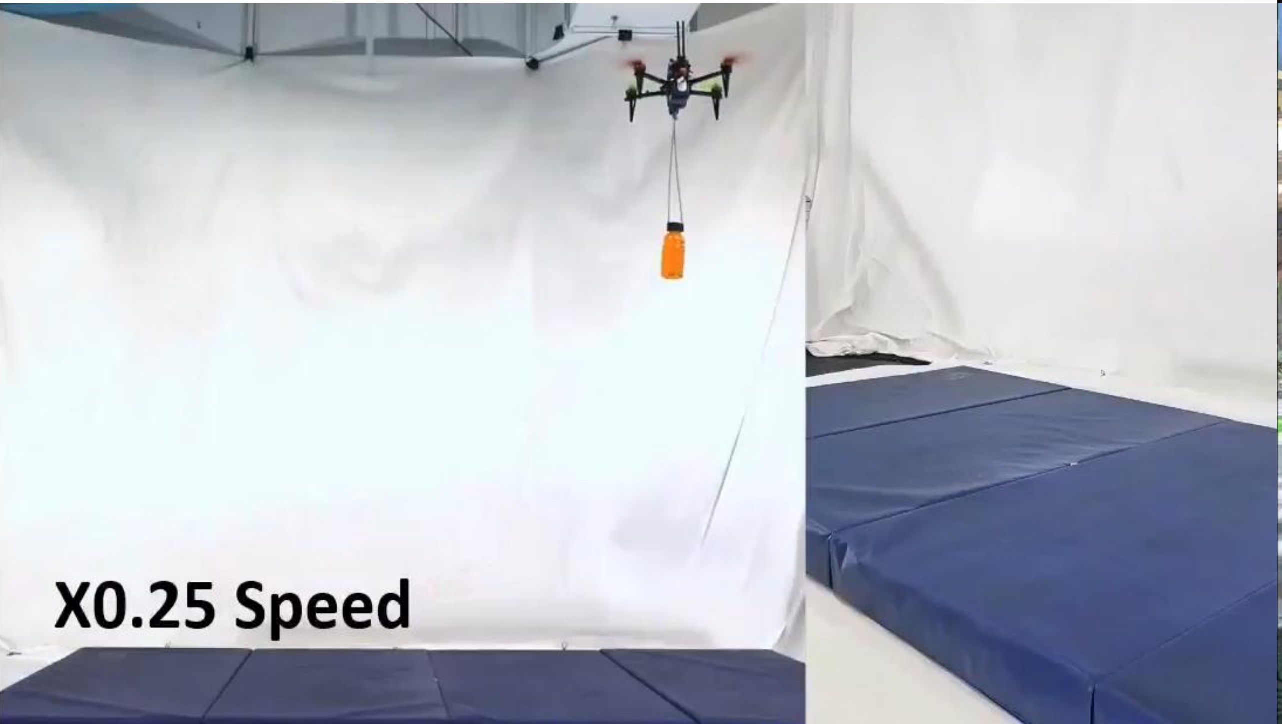
[4] Gurriet, et al. *Towards a Framework for Realizable Safety Critical Control through Active Set Invariance.* ICCPS, 2018.

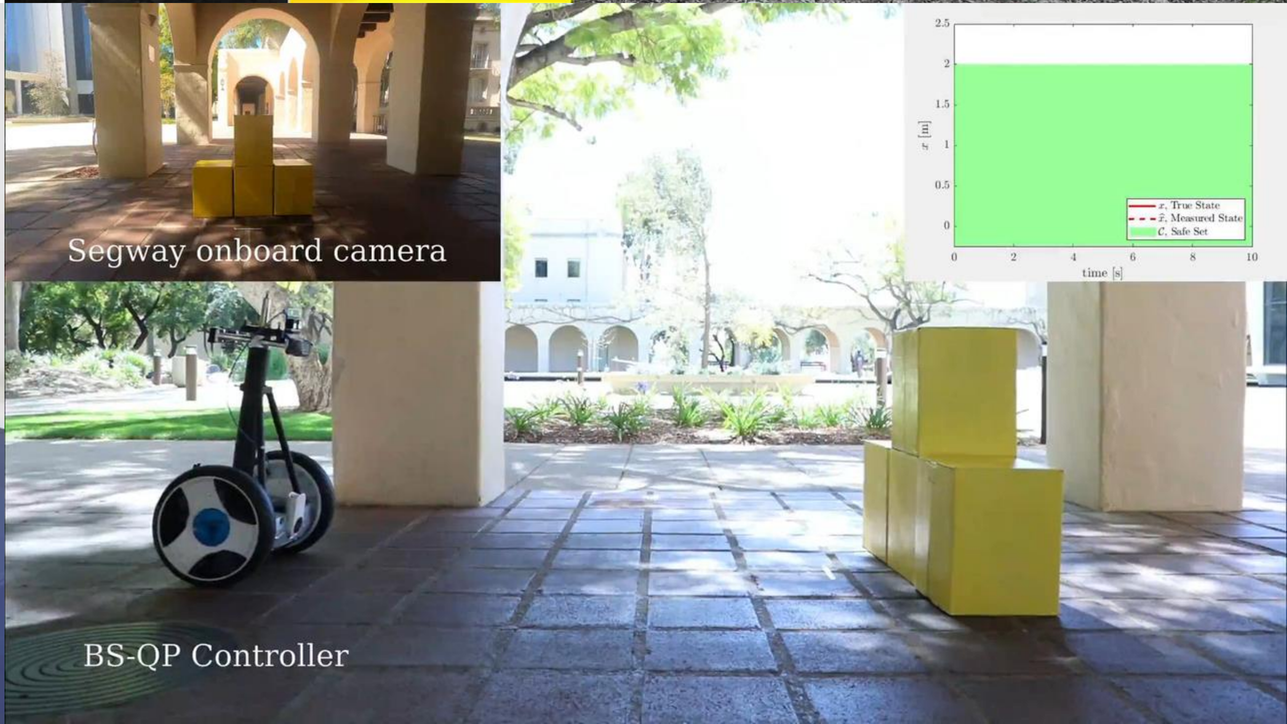

Idealized Approach
Defining Safety
h(x)
X0.25 Speed
Naïve Deployment

Naïve Application

Theorem: CBF-based Safety
$$\frac{dh}{dt}(\mathbf{x}, \mathbf{k}(\mathbf{x})) \geq -\alpha(h(\mathbf{x})) \implies \text{safety}$$





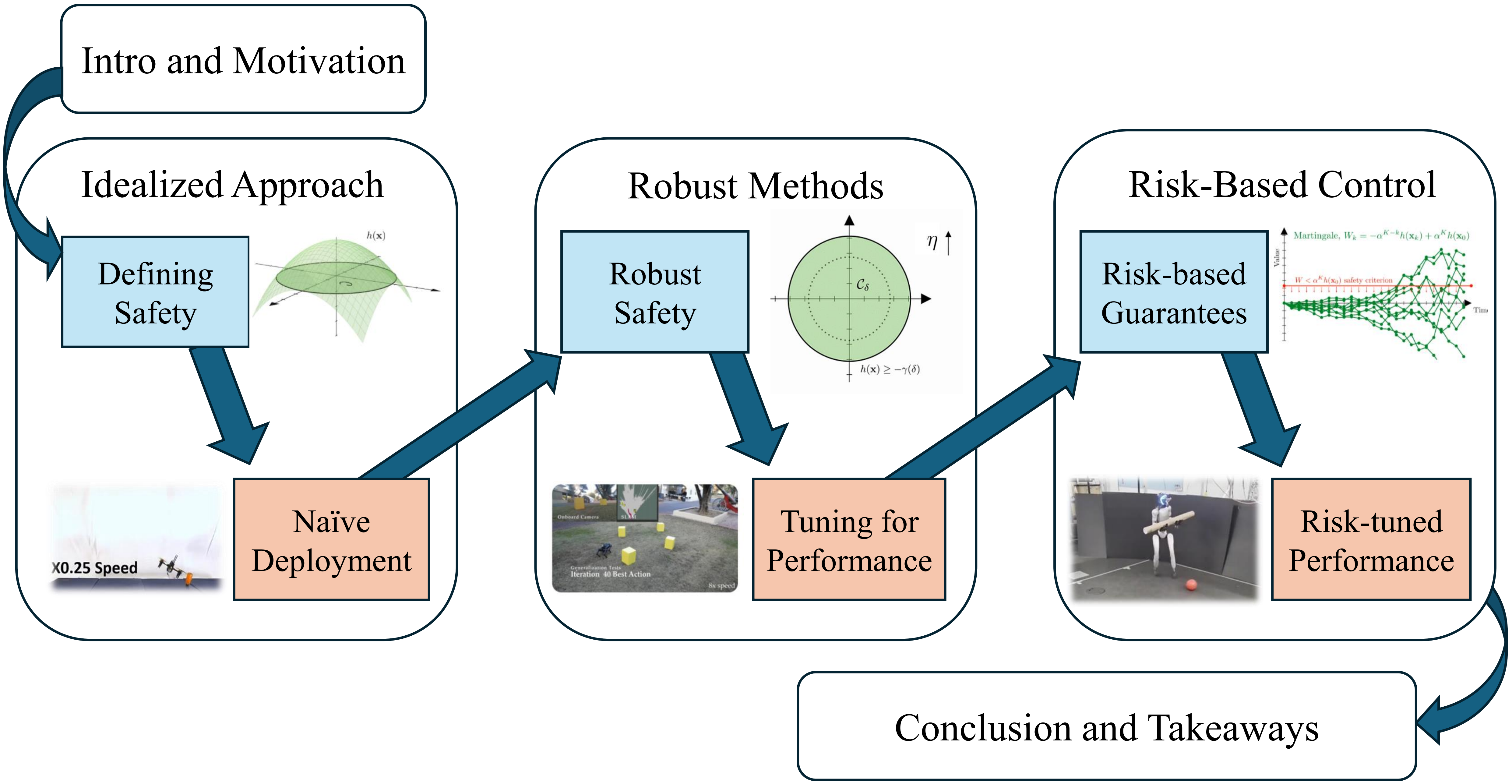


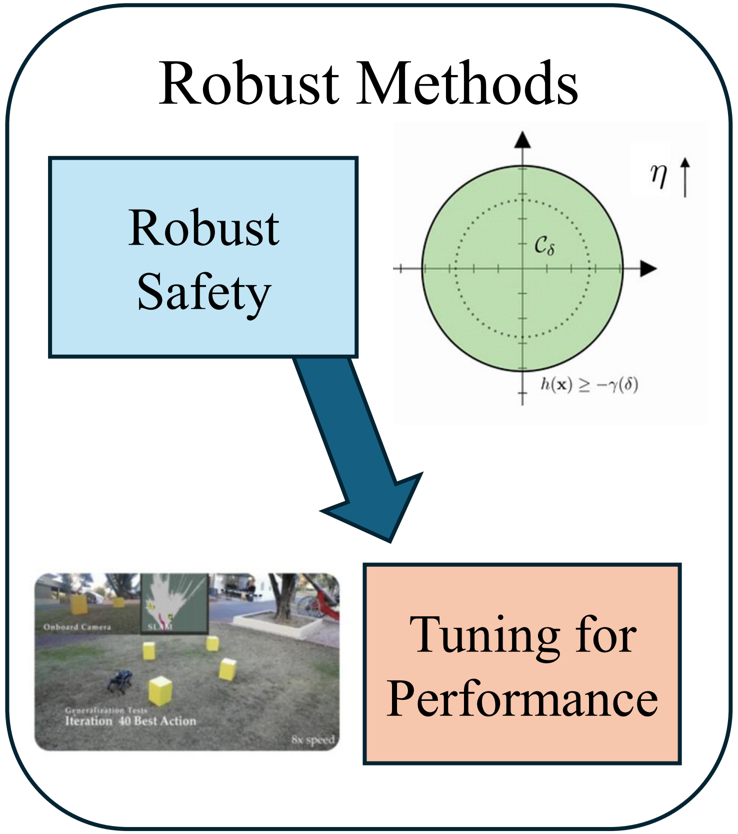
Should we throw away our theory?

No! But we should reexamine our assumptions:

Model Error	**Measurement Error**
The true dynamics are known: $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$	The true state is known: $\widehat{\mathbf{x}} = \mathbf{x}$

Learning Error	**Infeasible Safety**
Perfect controller imitation $\mathbf{k}(\mathbf{x}) = \mathbf{k}_\theta(I(\mathbf{x}))$	The CBF inequality is feasible: $\dot{h}(\mathbf{x}, \mathbf{u}) \geq -\alpha(h(\mathbf{x}))$


Intro and Motivation
Idealized Approach
Defining Safety
$h(\mathbf{x})$
X0.25 Speed
Naïve Deployment
Robust Methods
Robust Safety
$\eta \uparrow$
$\mathcal{C}_\delta$
$h(\mathbf{x}) \geq -\gamma(\delta)$
Iteration 40 Best Action
8x speed
Tuning for Performance
Risk-Based Control
Risk-based Guarantees
Martingale, $W_k = -\alpha^{K-k}h(\mathbf{x}_k) + \alpha^K h(\mathbf{x}_0)$
Risk-tuned Performance
Conclusion and Takeaways


Robust Methods
Robust Safety
η
$\mathcal{C}_\delta$
$h(\mathbf{x}) \geq -\gamma(\delta)$
Iteration 40 Best Action
8x speed
Tuning for Performance

Real-World Safety

Key assumptions, one at a time:

Model Error

The true dynamics are known:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$$

Measurement Error

The true state is known:

$$\widehat{\mathbf{x}} = \mathbf{x}$$

Learning Error

Perfect controller imitation

$$\mathbf{k}(\mathbf{x}) = \mathbf{k}_\theta(I(\mathbf{x}))$$

Infeasible Safety

The CBF inequality is feasible:

$$\dot{h}(\mathbf{x}, \mathbf{u}) \geq -\alpha(h(\mathbf{x}))$$

Safety with Bounded Dynamics Uncertainty

Adding bounded disturbances, $\|\mathbf{d}(t)\| \leq \delta$

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})(\mathbf{u} \boxed{+ \mathbf{d}(t)})$$

Expanded Worst-Case Safe Set:

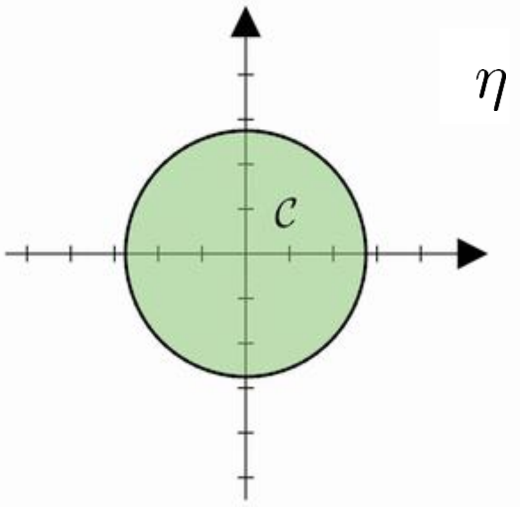
$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid h(\mathbf{x}) \geq 0\}$$

$$\mathcal{C}_\delta = \{\mathbf{x} \in \mathbb{R}^n \mid h(\mathbf{x}) \geq -\gamma(\delta, \eta)\}$$

Theorem: Input-to-State Safe CBF [9]

$$\frac{\partial h}{\partial \mathbf{x}}(\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}) \boxed{- \frac{1}{\eta}\left\|L_{\mathbf{g}} h(\mathbf{x})\right\|^2} \geq -\alpha(h(\mathbf{x}))$$

renders $\mathcal{C}_\delta$ forward invariant for $\gamma(\delta, \eta) = \dfrac{\delta^2 \eta}{4}$.



[9] Kolathaya, et al. *Input-to-State Safety with Control Barrier Functions.* CSL, 2018.

Learning Dynamics Online

Learn dynamics residuals $\mathbf{f}(\mathbf{x}) - \widehat{\mathbf{f}}(\mathbf{x}),\ \mathbf{g}(\mathbf{x}) - \widehat{\mathbf{g}}(\mathbf{x})$ via online sampling

Keep robust constraint feasible:

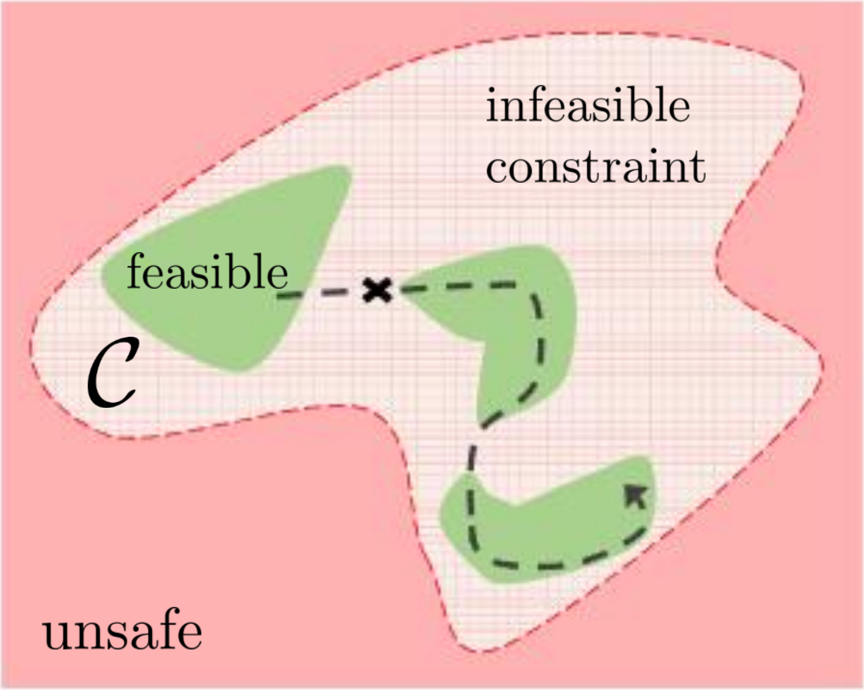
$$\text{Lower Bound}_{(1-\delta)}\left(\frac{\partial h}{\partial \mathbf{x}}(\widehat{\mathbf{f}}(\mathbf{x}) + \widehat{\mathbf{g}}(\mathbf{x})\mathbf{u})\right) \geq -\alpha(h(\mathbf{x})) + \frac{\epsilon}{2}$$

Theorem: Recovering Safety Feasibility [10]

If everything is Lipschitz, h is a ϵ-robust CBF, the dynamics residuals belong to an RKHS, and data sampling is at least this fast:

$$\Delta_t \leq \frac{\epsilon}{2\mathcal{L}_\alpha \mathcal{L}_h \mathcal{L}_{\dot{x}} N_{\max}(\delta)}$$

Then the system is safe with a probability $1-\delta$.



[10] Capone, Cosner, et al. *Learning safe control via on-the-fly bandit exploration.* ICML, 2025.

Real-World Safety

Key assumptions, one at a time:

Model Error	Measurement Error
The true dynamics are known: $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$	The true state is known: $\widehat{\mathbf{x}} = \mathbf{x}$

Learning Error	Infeasible Safety
Perfect controller imitation on data $\mathbf{k}(\mathbf{x}) = \mathbf{k}_\theta(I(\mathbf{x}))$	The CBF inequality is feasible: $\dot{h}(\mathbf{x}, \mathbf{u}) \geq -\alpha(h(\mathbf{x}))$

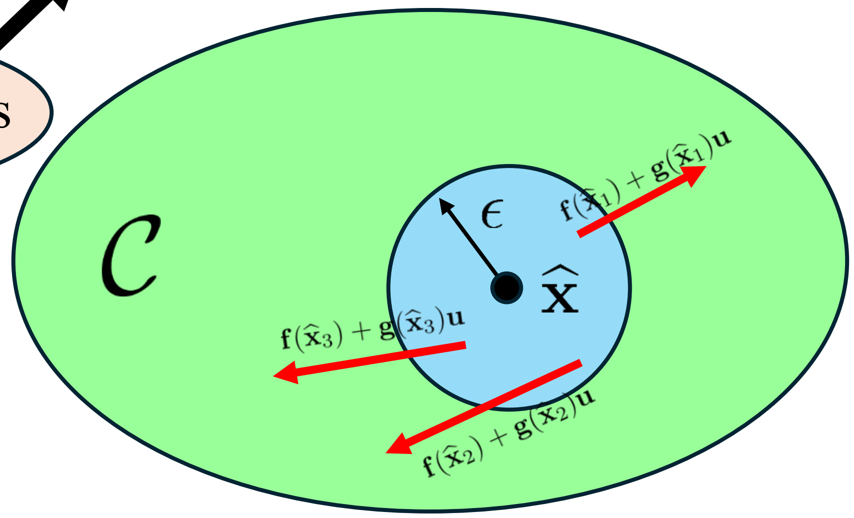
Safety with Bounded Measurement Uncertainty

Adding measurement uncertainty $\|\mathbf{x} - \widehat{\mathbf{x}}\| \leq \epsilon$ ← Worst-case bound

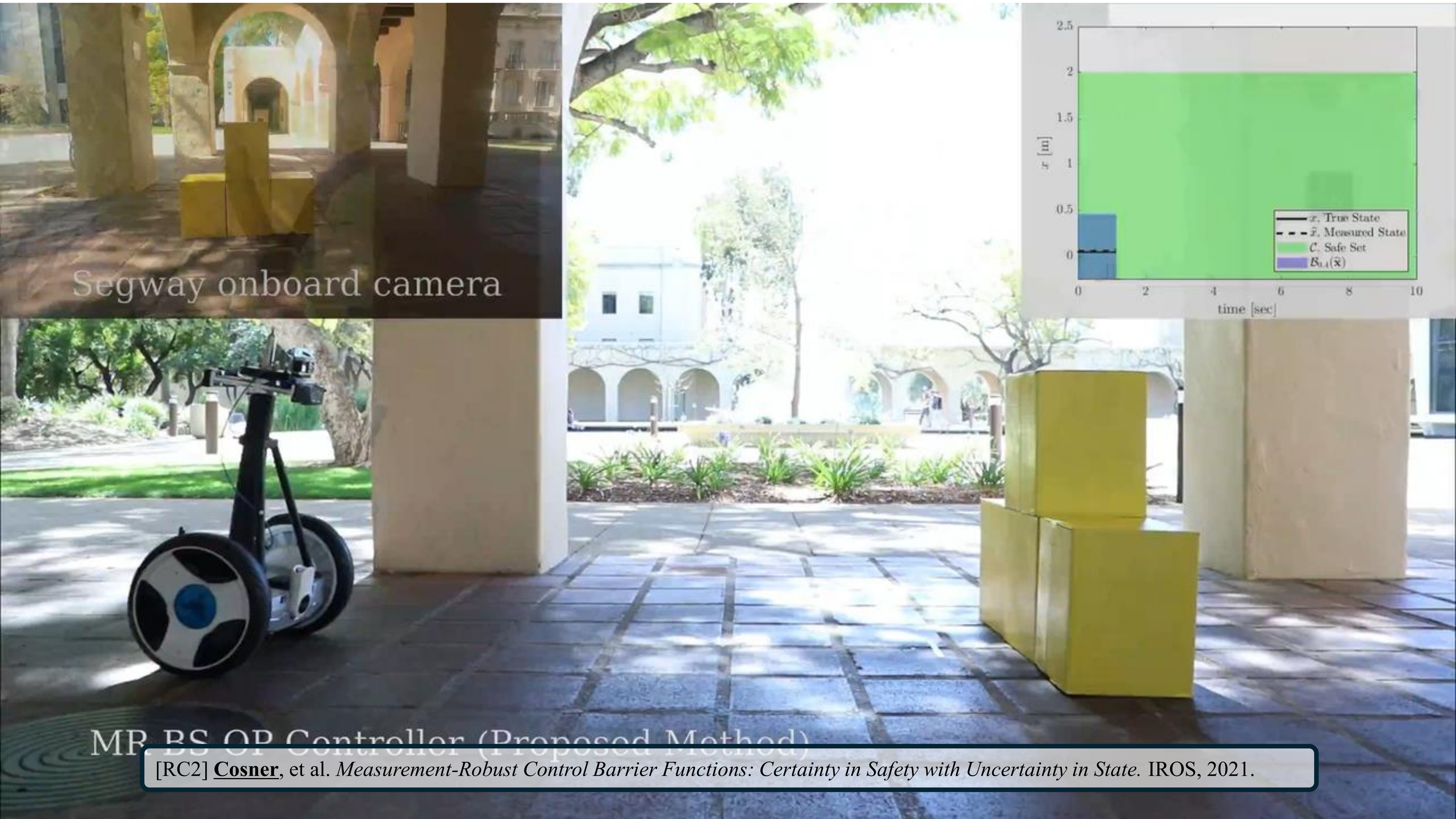
Theorem: Measurement Robust CBFs [11]

$$\dot{h}(\widehat{\mathbf{x}}, \mathbf{u}) \boxed{- \varphi(\widehat{\mathbf{x}}, \mathbf{u})} \geq -\alpha h(\widehat{\mathbf{x}}) \implies \text{safety}$$

Lipschitz Bounds



[11] Dean, Taylor, Cosner, et. al. *Guaranteeing safety of learned perception modules via measurement-robust control barrier functions*. CoRL, 2020.



[RC2] **Cosner**, et al. *Measurement-Robust Control Barrier Functions: Certainty in Safety with Uncertainty in State.* IROS, 2021.

Real-World Safety

Key assumptions, one at a time:

Model Error	Measurement Error
The true dynamics are known: $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$	The true state is known: $\widehat{\mathbf{x}} = \mathbf{x}$
Learning Error	**Infeasible Safety**
Perfect controller imitation on data $\mathbf{k}(\mathbf{x}) = \mathbf{k}_\theta(I(\mathbf{x}))$	The CBF inequality is feasible: $\dot{h}(\mathbf{x}, \mathbf{u}) \geq -\alpha(h(\mathbf{x}))$

Error in Imitation Learning

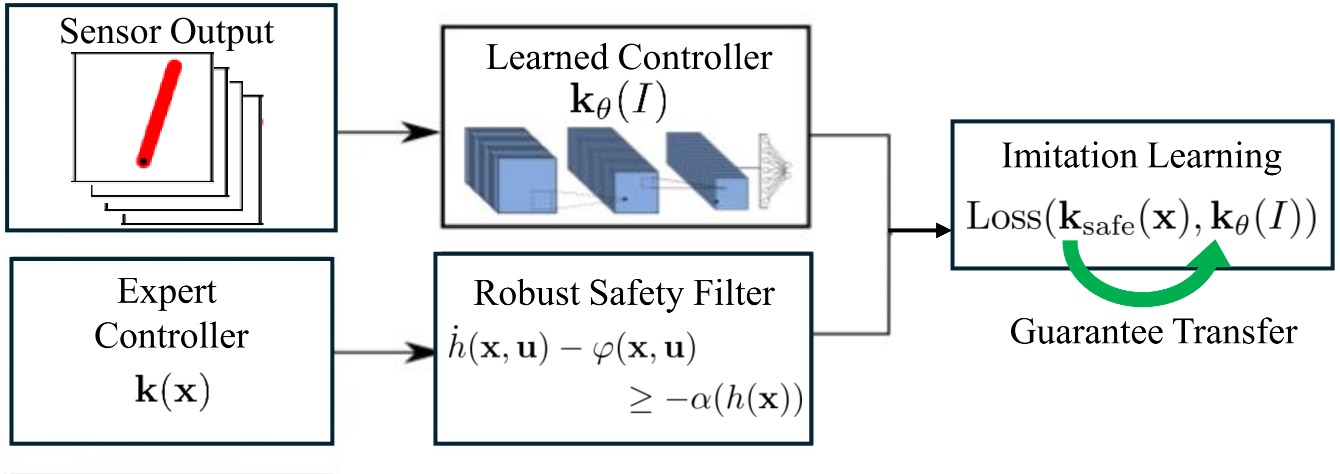
Increasingly popular in robotics

Errors in imitation/generalization can cause safety failures

[12]

Goal:

transfer safety from the robustified expert to learned end-to-end controller



[12] Fu, et al. *Mobile ALOHA: Learning bimanual mobile manipulation with low-cost whole-body teleoperation.* CoRL, 2024.

Error in Imitation Learning

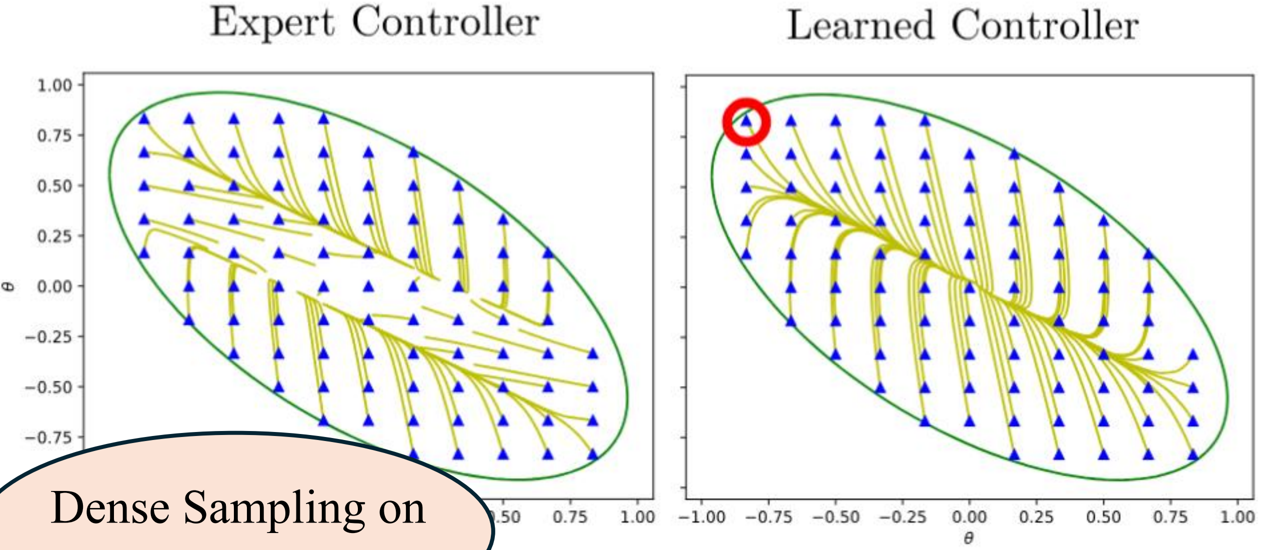
Lipschitz constants
& worst-case bounds

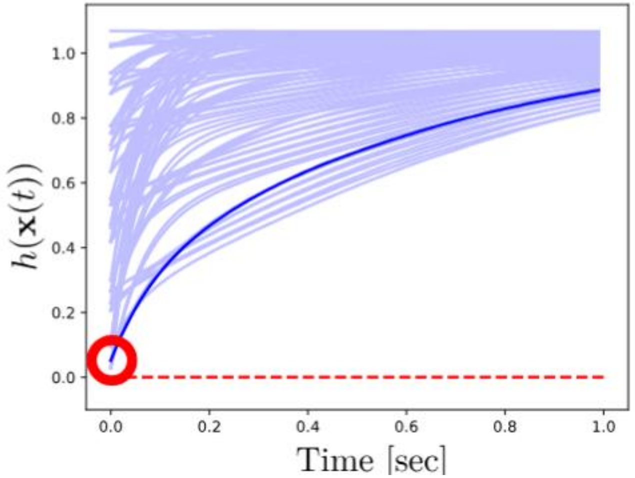
Thm: Transferring Safety Guarantees [13]

If the expert controller is robustly safe, the learned controller is Lipschitz, and data is sufficiently dense on the safe set boundary, $\partial\mathcal{C}$,

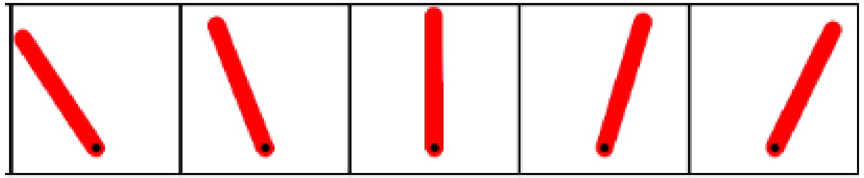
then the closed-loop system with the learned controller is safe.

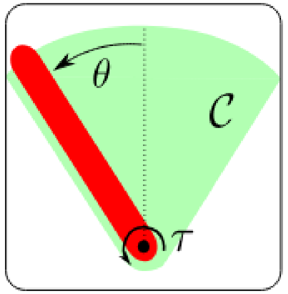
Dense Sampling on
$\partial\mathcal{C}$





Learned controller: images to torques





[13] Cosner, et al. *End-to-end imitaiton learing with sfaety guarantees using CBFs.* CDC, 2022.

Real-World Safety

Key assumptions, one at a time:

Model Error

The true dynamics are known:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$$

Measurement Error

The true state is known:

$$\widehat{\mathbf{x}} = \mathbf{x}$$

Learning Error

Perfect controller imitation on data

$$\mathbf{k}(\mathbf{x}) = \mathbf{k}_\theta(I(\mathbf{x}))$$

Infeasible Safety

The CBF inequality is feasible:

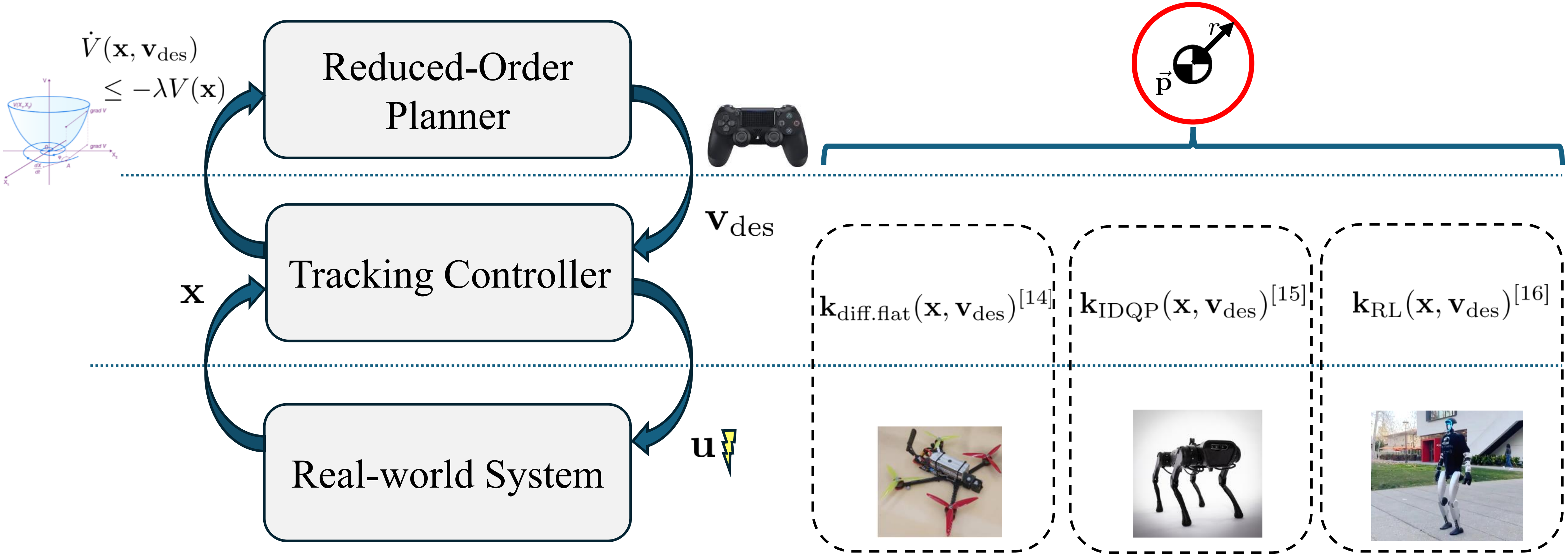
$$\dot{h}(\mathbf{x}, \mathbf{u}) \geq -\alpha(h(\mathbf{x}))$$

Synthesizing CBFs

Real-world System

Synthesizing CBFs

Leverage the hierarchical structure of robotic systems:



[14] Lee, et al. *Geometric tracking control of a quadrotor UAV on SE(3).* CDC, 2010.
[15] Buchli, et al. *Compliant quadruped locomotion over rough terrain.* IROS, 2009.
[16] Radosavovic, et al. *Real-world humanoid locomotion with reinforcement learning.* Science Robotics, 2024.

Synthesizing CBFs

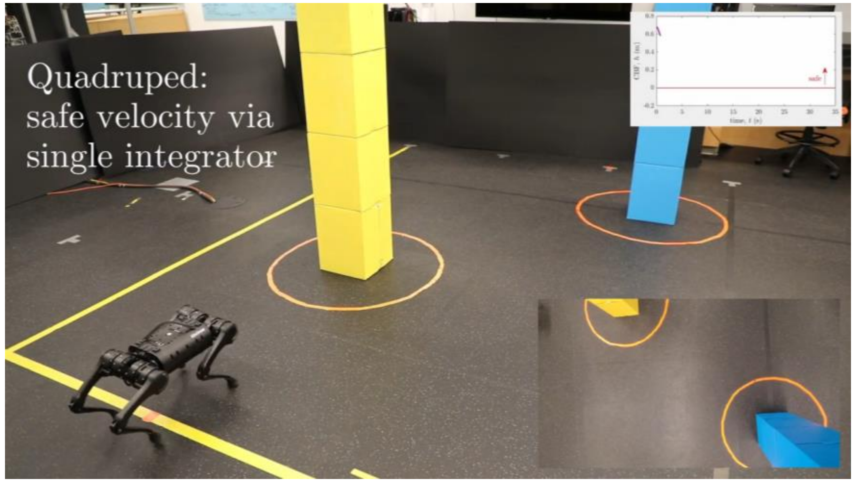
Theorem: Model-Free CBF[17, 18, 19]

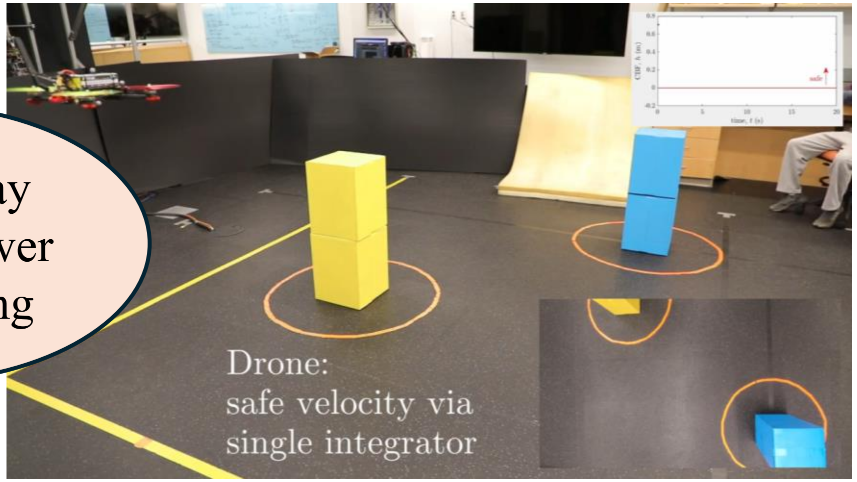
The simple model can be used to guarantee safety, if the true system tracks it fast enough.

- "Fast enough" for $\dot{h}_0(\mathbf{x}) \geq -\boxed{\alpha} h_0(\mathbf{x})$

Safety decay must be slower than tracking

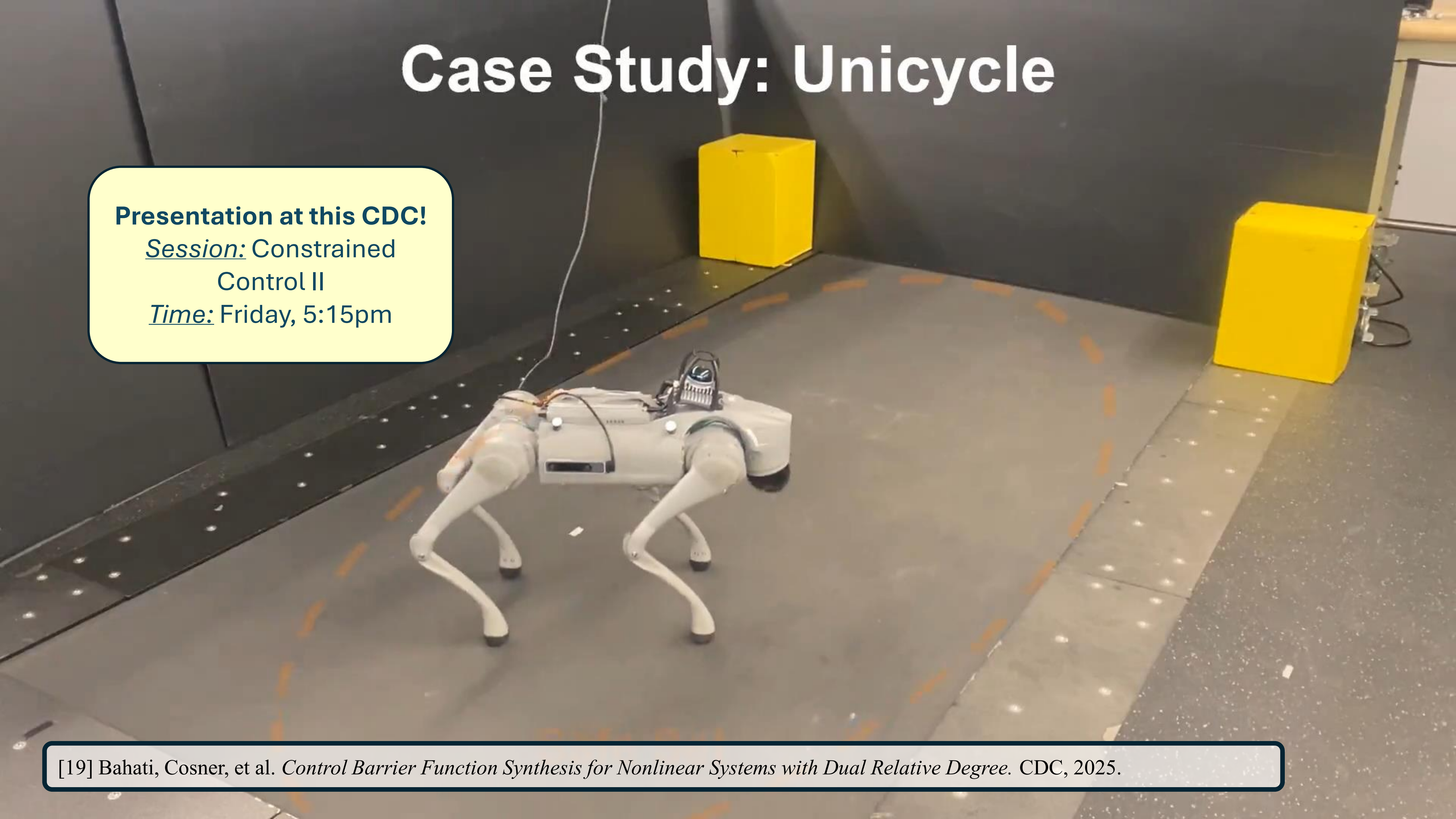
- Key idea: $h(\mathbf{x}) = \underbrace{h_0(\mathbf{x})}_{\text{safety requirement}} - \underbrace{V(\mathbf{x})}_{\text{tracking metric}}$





[17] Molnar, Cosner, et al. *Model-Free Safety Critical Control for Robotic Systems.* RAL, 2021.
[18] Cohen, Cosner, et al. *Constructive Safety-Critical Control: Synthesizing CBFs for Partially Feedback Linearizable Systems.* CSL, 2024.
[19] Bahati, Cosner, et al. *Control Barrier Function Synthesis for Nonlinear Systems with Dual Relative Degree.* CDC, 2025.

Case Study: Unicycle



Presentation at this CDC!
Session: Constrained Control II
Time: Friday, 5:15pm

[19] Bahati, Cosner, et al. *Control Barrier Function Synthesis for Nonlinear Systems with Dual Relative Degree.* CDC, 2025.

Real-World Safety

Key assumptions:

Model Error	Measurement Error
The true dynamics are known: $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$	The true state is known: $\widehat{\mathbf{x}} = \mathbf{x}$

Learning Error	Infeasible Safety
Perfect controller imitation on data $\mathbf{k}(\mathbf{x}) = \mathbf{k}_\theta(I(\mathbf{x}))$	The CBF inequality is feasible: $\dot{h}(\mathbf{x}, \mathbf{u}) \geq -\alpha(h(\mathbf{x}))$

Real-World Safety

Key assumptions:

Model Error

The true dynamics are known:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$$

Measurement Error

The true state is known:

$$\widehat{\mathbf{x}} = \mathbf{x}$$

Learning Error

Perfect controller imitation on data

$$\mathbf{k}(\mathbf{x}) = \mathbf{k}_\theta(I(\mathbf{x}))$$

Infeasible Safety

The CBF inequality is feasible:

$$\dot{h}(\mathbf{x}, \mathbf{u}) \geq -\alpha(h(\mathbf{x}))$$

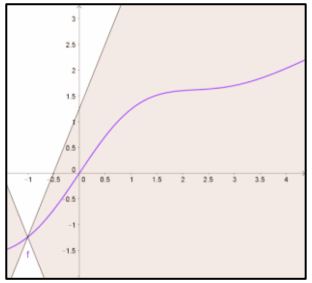
Robustification Techniques

- Worst-case over-approximations

$$\|\mathbf{d}(t)\|_\infty \leq \delta$$

- Lipschitz bounds

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \leq \mathfrak{L}_{\mathbf{f}}\|\mathbf{x} - \mathbf{y}\|$$



- Dense sampling near boundary

$$\forall \mathbf{x} \in \partial\mathcal{C}\ \exists \mathbf{x}_{\mathbf{D}} \in \mathfrak{D} \quad \text{s.t.} \quad \|\mathbf{x} - \mathbf{x}_D\| \leq \epsilon$$

- Stability rate

$$\dot{V}(\mathbf{x}) \leq -\lambda V(\mathbf{x})$$

Real-World Safety

Theorem: Tunable Robust CBF [20]

$$\frac{\partial h}{\partial \mathbf{x}}\left(\widehat{\mathbf{f}}(\widehat{\mathbf{x}}) + \widehat{\mathbf{g}}(\widehat{\mathbf{x}})\mathbf{u}\right) - \mathrm{robustification}(a, b, c, \widehat{\mathbf{x}}, \mathbf{u}) \geq -\alpha(h(\widehat{\mathbf{x}}))$$

achieves safety even when these assumptions are violated:

Perfect Model

The true dynamics are known:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$$

Perfect Measurement

The true state is known:

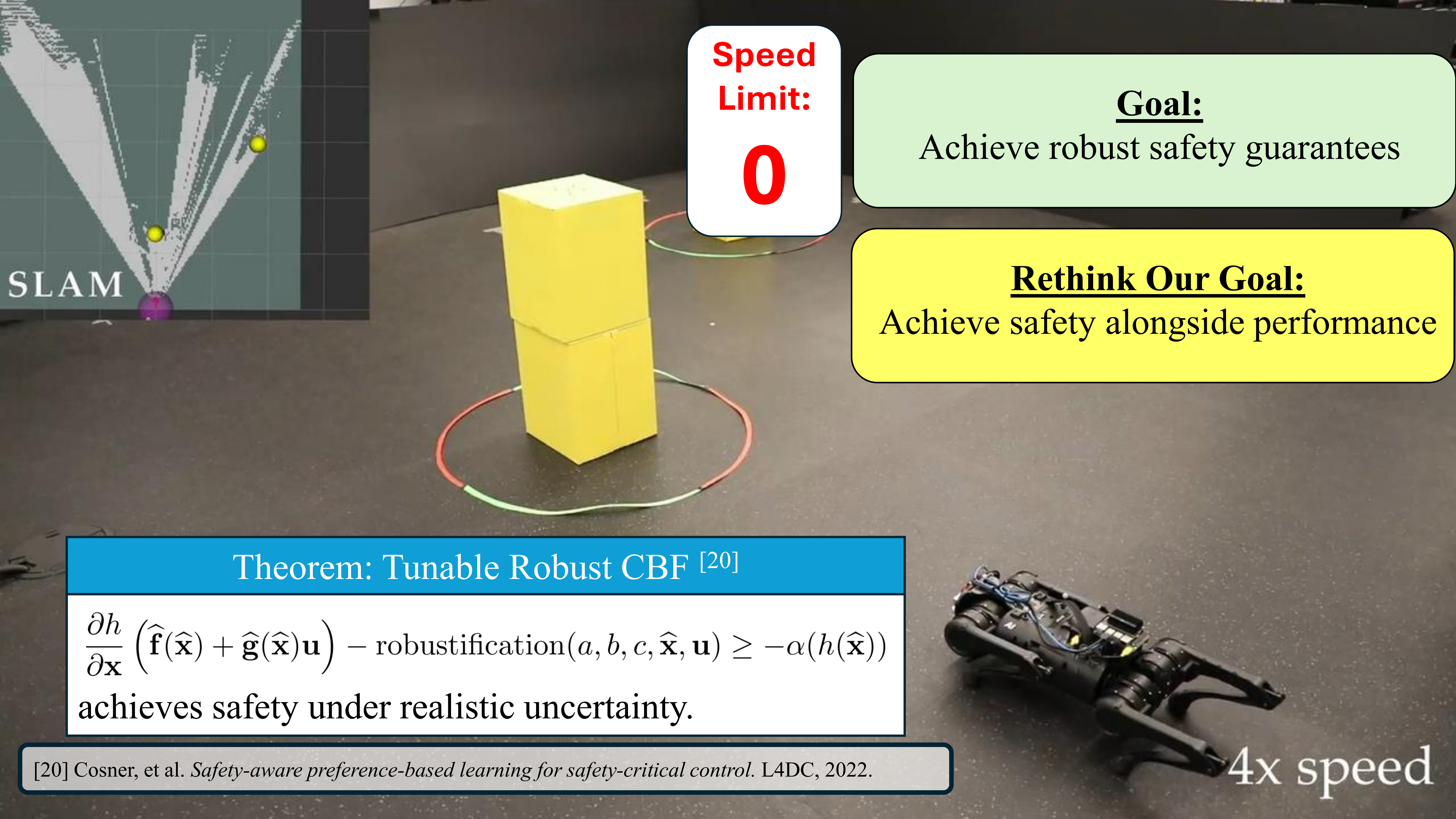
$$\widehat{\mathbf{x}} = \mathbf{x}$$

Well-Defined Safety

The CBF inequality is feasible:

$$\dot{h}(\mathbf{x}, \mathbf{u}) \geq -\alpha(h(\mathbf{x}))$$

[20] Cosner, et al. *Safety-aware preference-based learning for safety-critical control.* L4DC, 2022.



Speed Limit:

0

Goal:

Achieve robust safety guarantees

Rethink Our Goal:

Achieve safety alongside performance

Theorem: Tunable Robust CBF [20]

$$\frac{\partial h}{\partial \mathbf{x}}\left(\widehat{\mathbf{f}}(\widehat{\mathbf{x}}) + \widehat{\mathbf{g}}(\widehat{\mathbf{x}})\mathbf{u}\right) - \mathrm{robustification}(a, b, c, \widehat{\mathbf{x}}, \mathbf{u}) \geq -\alpha(h(\widehat{\mathbf{x}}))$$

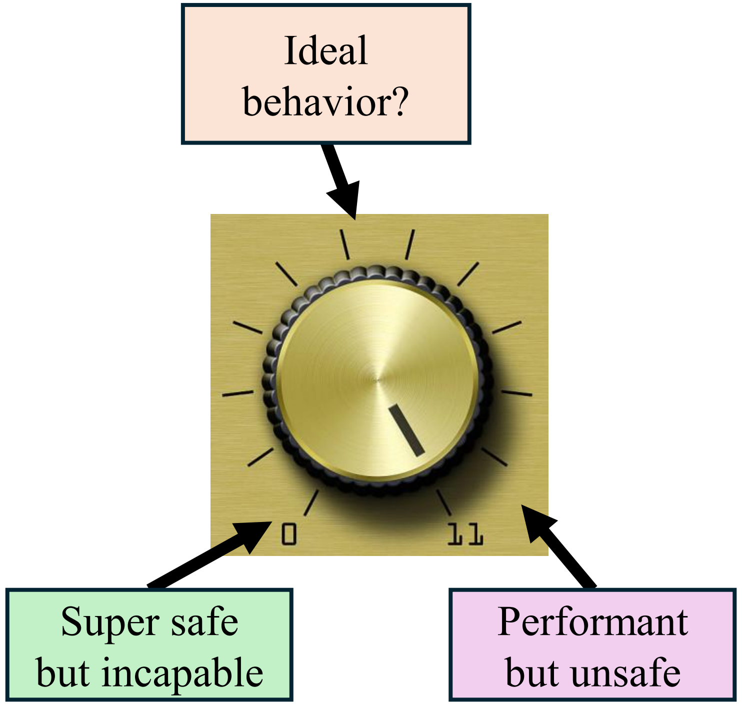
achieves safety under realistic uncertainty.

[20] Cosner, et al. *Safety-aware preference-based learning for safety-critical control.* L4DC, 2022.

Should we throw away our theory?

No! Use theory to guide learning-based performance.

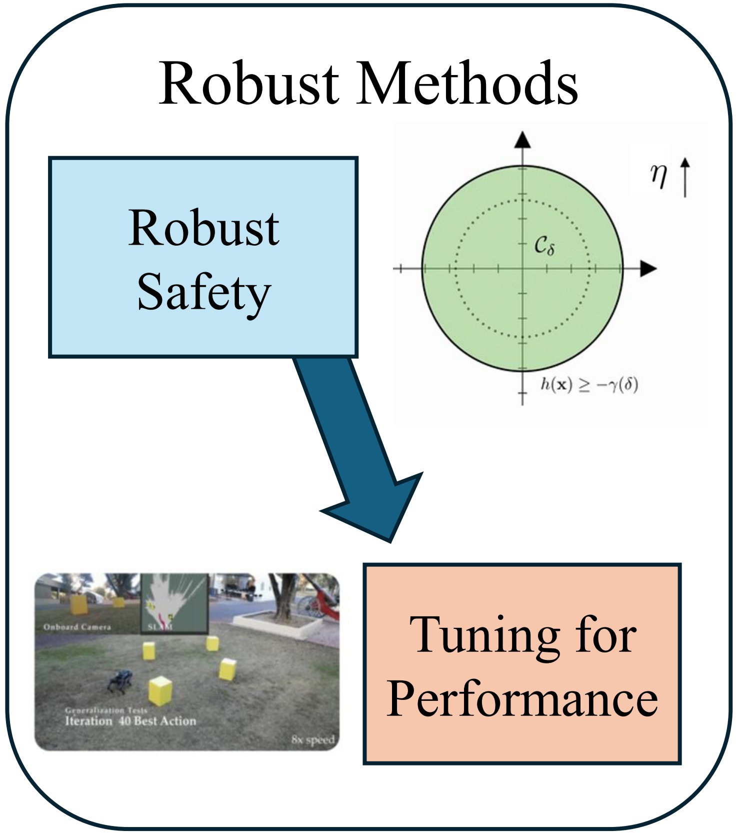
Theory reveals relevant tuning dials



Theory reveals important system characteristics

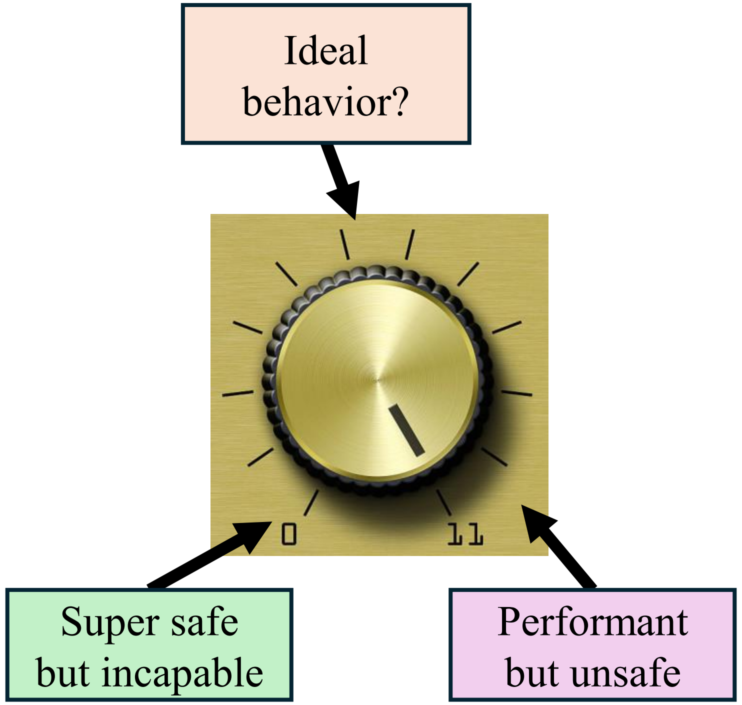
Robustification Techniques

- Worst-case bounds: $\|\mathbf{d}(t)\|_\infty \leq \delta$
- Lipschitz constants
  $$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \leq \mathfrak{L}_{\mathbf{f}}\|\mathbf{x} - \mathbf{y}\|$$
- Dense sampling
  $$\forall \mathbf{x} \in \partial\mathcal{C}\ \exists \mathbf{x}_{\mathbf{D}} \in \mathfrak{D} \text{ s.t. } \|\mathbf{x} - \mathbf{x}_D\| \leq \epsilon$$
- Stability rate $\dot{V}(\mathbf{x}) \leq -\lambda V(\mathbf{x})$


Robust Methods
Robust Safety
η
$\mathcal{C}_\delta$
$h(\mathbf{x}) \geq -\gamma(\delta)$
Onboard Camera
SLAM
Generalization Tests
Iteration 40 Best Action
8x speed
Tuning for Performance

Theory-driven, learning-based performance

Theory reveals relevant tuning dials



Theory reveals important system characteristics

Robustification Techniques

- Worst-case bounds: $\|\mathbf{d}(t)\|_\infty \leq \delta$
- Lipschitz constants
  $$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \leq \mathfrak{L}_{\mathbf{f}}\|\mathbf{x} - \mathbf{y}\|$$
- Dense sampling
  $$\forall \mathbf{x} \in \partial\mathcal{C}\ \exists \mathbf{x}_{\mathbf{D}} \in \mathfrak{D} \text{ s.t. } \|\mathbf{x} - \mathbf{x}_D\| \leq \epsilon$$
- Stability rate $\dot{V}(\mathbf{x}) \leq -\lambda V(\mathbf{x})$

Theory Reveals Tuning Knobs

Choose parameters for preferred behavior instead of assumed bounds

$$\frac{\partial h}{\partial \mathbf{x}}\left(\widehat{\mathbf{f}}(\widehat{\mathbf{x}}) + \widehat{\mathbf{g}}(\widehat{\mathbf{x}})\mathbf{u}\right) - \text{robustification}(\boxed{a, b, c,} \widehat{\mathbf{x}}, \mathbf{u}) \geq -\alpha(h(\widehat{\mathbf{x}}))$$

Method:

- Preference-Based Learning
- Safety-Aware Region of Interest Sampling

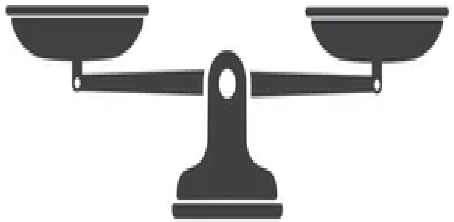
- Learn from sparse, noisy user feedback

Assumption Bounds

Perfect Model	Perfect Measurement	Well-Defined Safety
The true dynamics are known: $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$	The true state is known: $\widehat{\mathbf{x}} = \mathbf{x}$	The CBF inequality is feasible: $\dot{h}(\mathbf{x}, \mathbf{u}) \geq -\alpha(h(\mathbf{x}))$

Preferred Behavior

Robustness | Performance

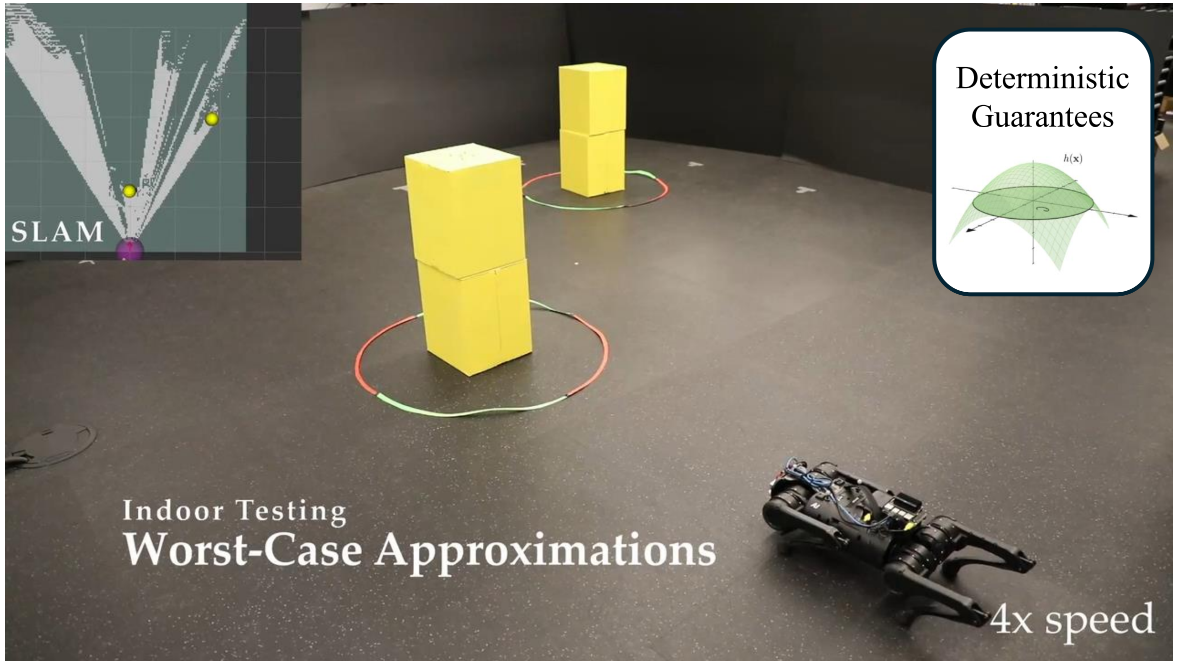


[20] Cosner, et al. *Safety-aware preference-based learning for safety-critical control.* L4DC, 2022.

Better Tuning: Preference-Based Learning

Subject: 

37 Iterations of Preference-Based Learning



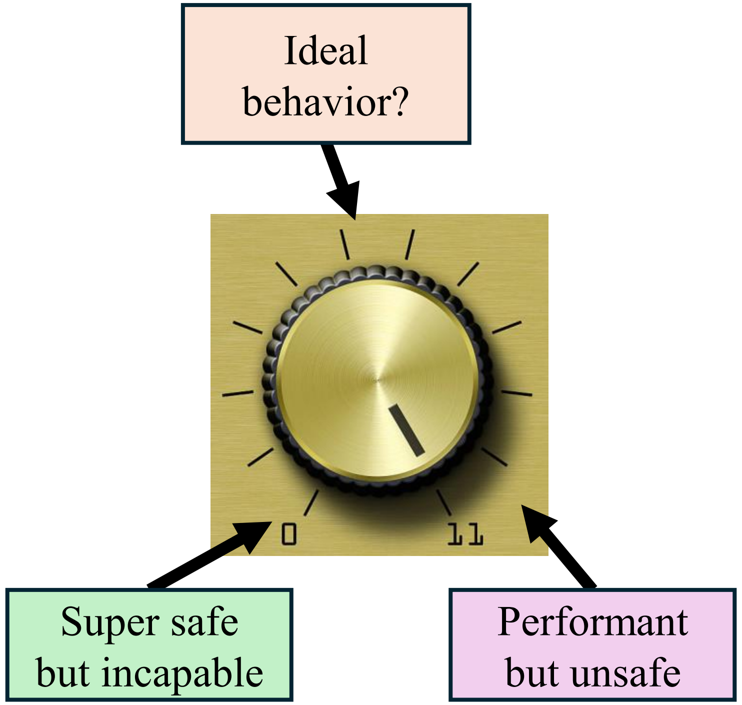
[20] Cosner, et al. *Safety-aware preference-based learning for safety-critical control.* L4DC, 2022.

Online CBF synthesis from vision data

[20] Cosner, et al. *Safety-aware preference-based learning for safety-critical control.* L4DC, 2022.

Theory-driven, learning-based performance

Theory reveals relevant tuning dials



Theory reveals important system characteristics

Robustification Techniques

- Worst-case bounds: $\|\mathbf{d}(t)\|_\infty \leq \delta$
- Lipschitz constants
  $$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \leq \mathfrak{L}_{\mathbf{f}} \|\mathbf{x} - \mathbf{y}\|$$
- Dense sampling
  $$\forall \mathbf{x} \in \partial \mathcal{C} \ \exists \mathbf{x}_{\mathbf{D}} \in \mathfrak{D} \ \text{s.t.} \ \|\mathbf{x} - \mathbf{x}_D\| \leq \epsilon$$
- Stability rate $\dot{V}(\mathbf{x}) \leq -\lambda V(\mathbf{x})$

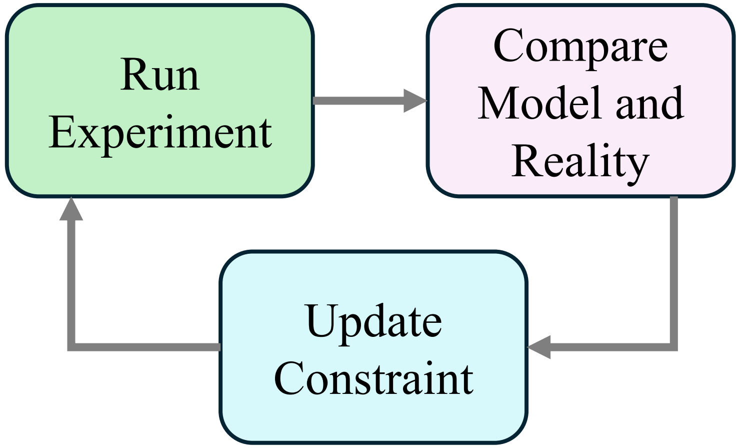
Following theory's intuition

Follow intuition from robust theory:

- Learn residuals and regularize weights to reduce Lipschitz constants

$$\begin{aligned} \mathbf{e}_{\mathbf{f},\theta}(\mathbf{x}) &= \mathbf{f}(\mathbf{x}) - \widehat{\mathbf{f}}(\mathbf{x}) \\ \mathbf{e}_{\mathbf{g},\phi}(\mathbf{x}) &= \mathbf{g}(\mathbf{x}) - \widehat{\mathbf{g}}(\mathbf{x}) \end{aligned}, \quad \text{Loss} \mathrel{+}= \|\theta\| + \|\phi\|$$

- Run iteratively to collect data near boundary



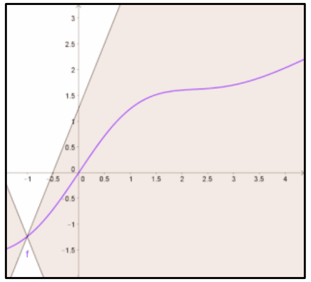
Robustification Techniques

- Worst-case over-approximations

$$\|\mathbf{d}(t)\|_\infty \le \delta$$

- Lipschitz bounds

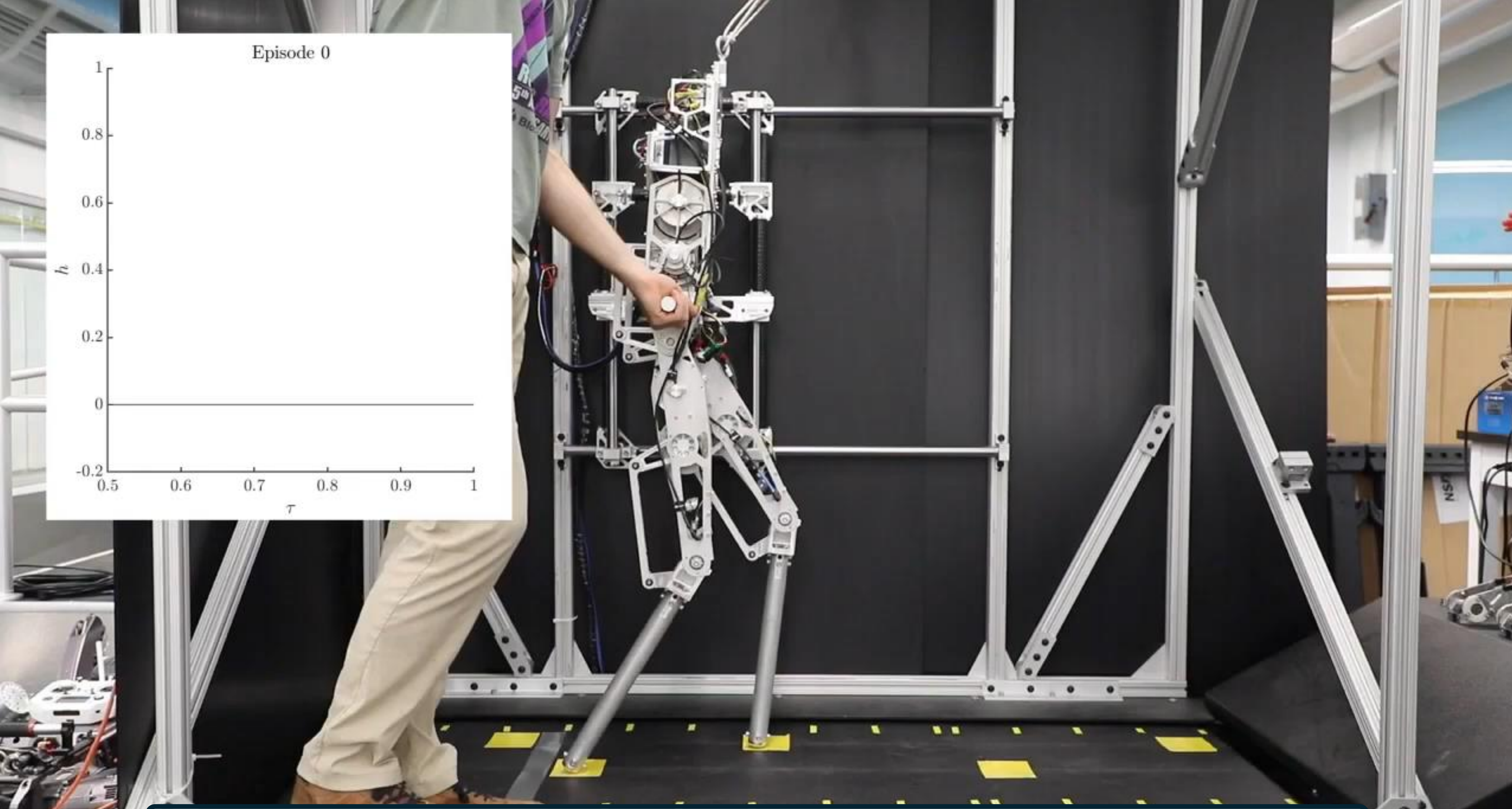
$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \le \mathfrak{L}_{\mathbf{f}}\|\mathbf{x} - \mathbf{y}\|$$



- Dense sampling near boundary

$$\forall \mathbf{x} \in \partial\mathcal{C}\ \exists \mathbf{x}_{\mathbf{D}} \in \mathfrak{D} \quad \text{s.t.} \ \|\mathbf{x} - \mathbf{x}_D\| \le \epsilon$$

[21] Csomay-Shanklin*, Cosner*, Dai*, et al. *Episodic learning for safe bipedal locomotion with CBFs and PSSf.* L4DC, 2021.



[21] Csomay-Shanklin*, Cosner*, Dai*, et al. *Episodic learning for safe bipedal locomotion with CBFs and PSSf.* L4DC, 2021.

Following theory's intuition

New Paradigm!
Stochastic uncertainty

Problem setting: residual learning on a system with more complicated uncertainty

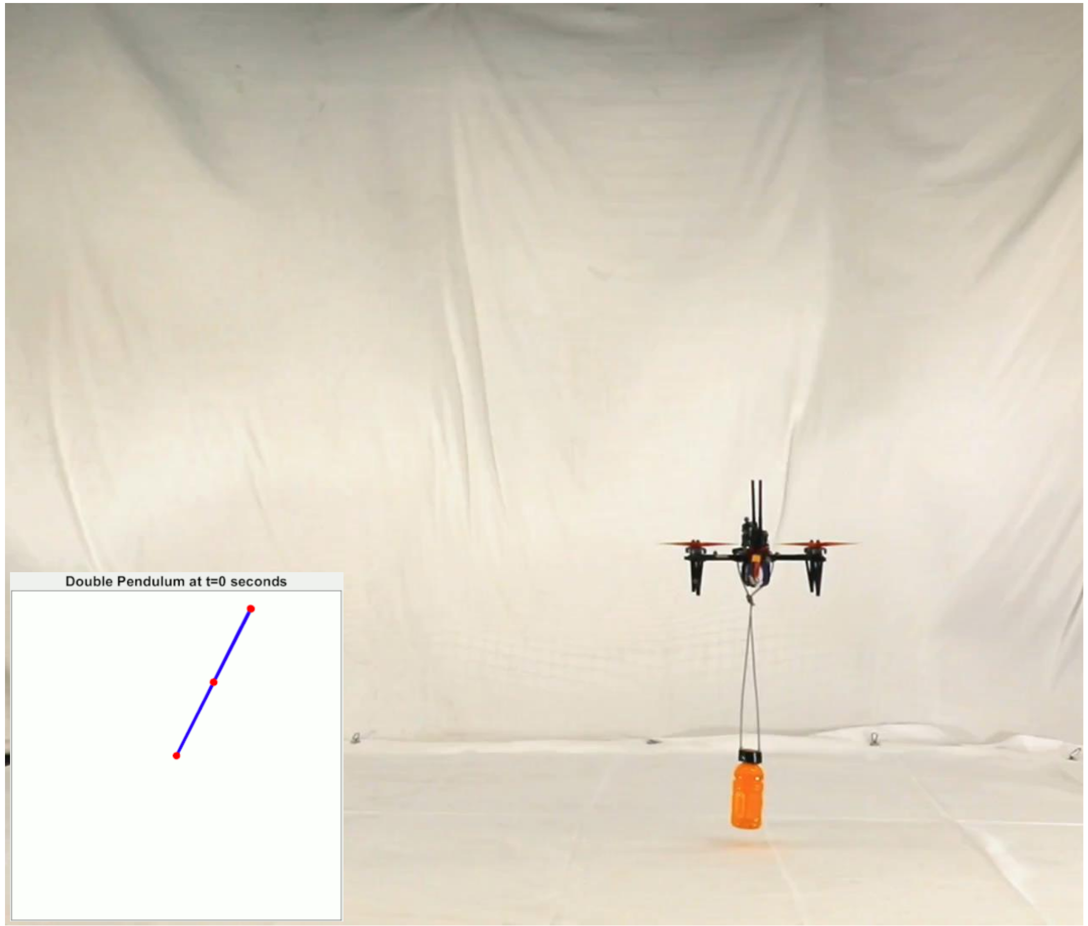
- Unknowns: bottle state and mass
- Safety: don't touch the ground
- Learn: drone state → disturbance

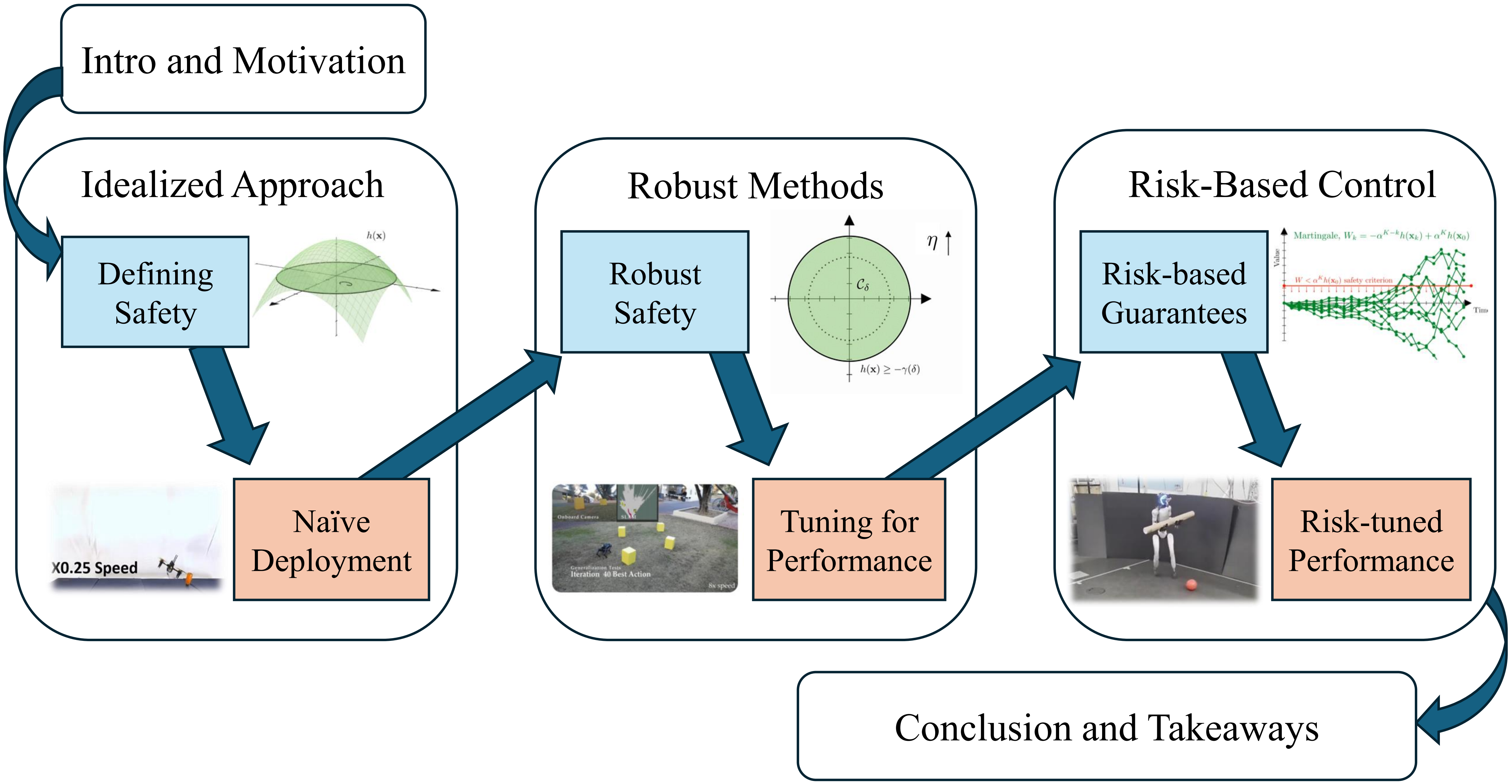
$$\mathbf{d}_{\text{bottle}}(\mathbf{x}_{\text{drone}})$$

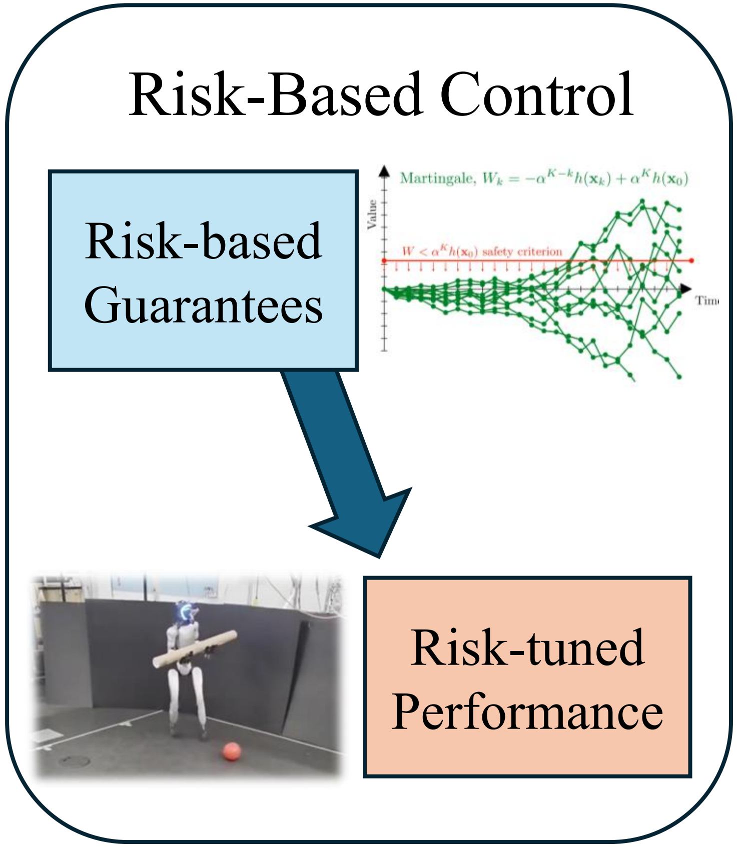
This learning problem isn't well posed!

- System is chaotic, infinite DoF
- No one-to-one mapping

$$\mathbf{d}_{\text{bottle}}(\mathbf{x}_{\text{drone}}, \mathbf{x}_{\text{bottle}}, \mathbf{x}_{\text{environment}}, \dots)$$




Intro and Motivation
Idealized Approach
Defining Safety
$h(\mathbf{x})$
X0.25 Speed
Naïve Deployment
Robust Methods
Robust Safety
η
$\mathcal{C}_\delta$
$h(\mathbf{x}) \geq -\gamma(\delta)$
Tuning for Performance
Risk-Based Control
Risk-based Guarantees
Martingale, $W_k = -\alpha^{K-k}h(\mathbf{x}_k) + \alpha^K h(\mathbf{x}_0)$
Risk-tuned Performance
Conclusion and Takeaways


Risk-Based Control
Risk-based Guarantees
Martingale, $W_k = -\alpha^{K-k} h(\mathbf{x}_k) + \alpha^K h(\mathbf{x}_0)$
Value
$W < \alpha^K h(\mathbf{x}_0)$ safety criterion
Risk-tuned Performance

Switching to Risk-based guarantees

Switch discrete time system with stochastic uncertainty:

Continuous Time CBF

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u} + \mathbf{d}$$

$$\frac{dh}{dt}(\mathbf{x}, \mathbf{u}) \geq -\alpha h(\mathbf{x})$$

Safety goal:
Forward Invariance

Discrete Time CBF

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{d}_k$$

$$\mathbb{E}[h(\mathbf{F}(\mathbf{x}, \mathbf{u})) + \mathbf{d} \mid \mathbf{x}_k] \geq \rho h(\mathbf{x})$$

Safety goal:
Bound Risk of Failure
Over a Finite Horizon

DCBF Guarantees

DCBF:
$$\mathbb{E}[\ h(\mathbf{F}(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{d}) \mid \mathbf{x}_k\ ] \geq \rho h(\mathbf{x}_k)$$

Assume: $h(\mathbf{x})$ is upper-bounded by $M > 0$

Lemma: Ville's Inequality [22]

Theorem: Ville's DCBFs [24,25]

$$\mathbb{E}[\ h(\mathbf{F}(\mathbf{x}, \mathbf{k}(\mathbf{x})) + \mathbf{d}) \mid \mathbf{x}\ ] \geq \rho h(\mathbf{x})$$
$$\implies \mathbb{P}_{\text{unsafe}}(K, \mathbf{x}_0) \leq 1 - \rho^K \frac{h(\mathbf{x}_0)}{M}.$$

Assume:

Lower-bounded uncertainty:

$\ldots k] - \delta$

Super Martingale:
$$\mathbb{E}[\ W_{k+1} \mid \mathscr{F}_k\ ] \leq W_k$$

Bounded variance:
$$\mathrm{Var}(h(\mathbf{x_{k+1}}) | \mathbf{x}_k) \leq \sigma^2$$

Lemma: Freedman's Inequality [23]

Theorem: Freedman's DCBFs [26]

$$\mathbb{E}[\ h(\mathbf{F}(\mathbf{x}, \mathbf{k}(\mathbf{x})) + \mathbf{d}) \mid \mathbf{x}\ ] \geq \rho h(\mathbf{x})$$
$$\implies \mathbb{P}_{\text{unsafe}}(K, \mathbf{x}_0) \leq H\left(\frac{\rho^K h(\mathbf{x}_0)}{\delta}, \frac{\sigma\sqrt{K}}{\delta}\right).$$

[22] Ville. *Etude critique de al notion de collectif*. 1939.
[23] Freedman. *On tail probabilities for martingales*. 1975.
[24] Cosner, et al. *Robust safety under stochastic uncertainty with discrete-time control barrier functions*. RSS, 2023.
[25] Kushner. *Stochastic stability and control*.1967
[26] Cosner, et al. *Bounding Stochastic Safety: Leveraging Freedman's Inequality with Discrete-Time Control Barrier Functions*. LCSS, 2024.

Stochastic DCBFs Guarantees

Connections to stochastic process theory

Failure probability is governed by:

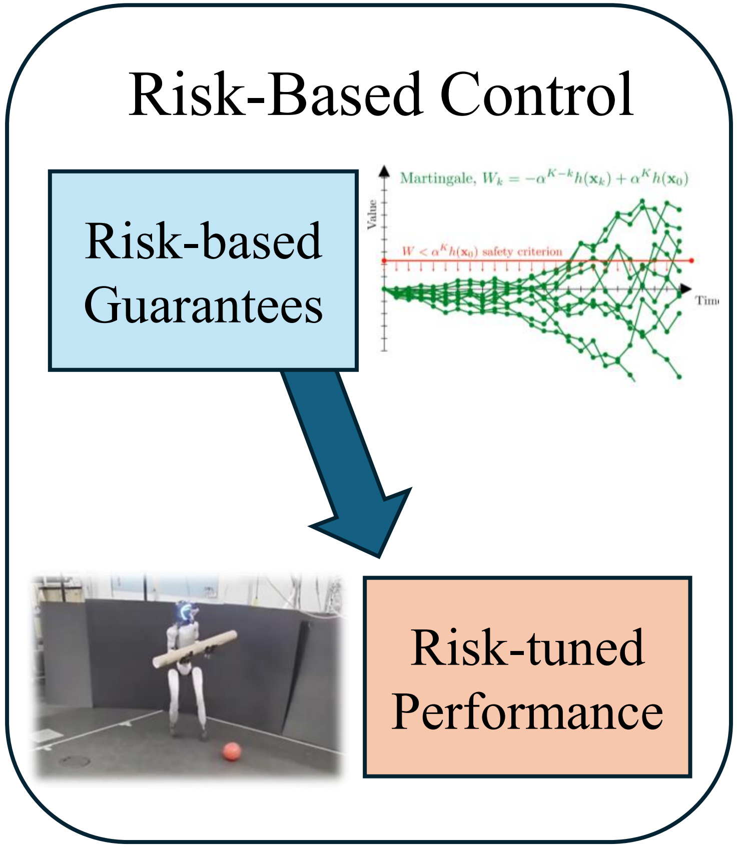
- the initial condition: $\mathbf{x}_0$
- the horizon length: K
- distribution information: $\sigma, \mathfrak{D}, p(\mathbf{d})$
- the safety decay rate: $\rho \in (0, 1)$

Theorem: Stochastic Safety [24, 26]

Stochastic DCBFs

$\implies$ bounded risk of failure.

[24] Cosner, et al. *Robust safety under stochastic uncertainty with discrete-time control barrier functions.* RSS, 2023.
[26] Cosner, et al. *Bounding Stochastic Safety: Leveraging Freedman's Inequality with Discrete-Time Control Barrier Functions.* LCSS, 2024.


Risk-Based Control
Risk-based Guarantees
Martingale, $W_k = -\alpha^{K-k}h(\mathbf{x}_k) + \alpha^K h(\mathbf{x}_0)$
Value
$W < \alpha^K h(\mathbf{x}_0)$ safety criterion
Risk-tuned Performance

Enforcing Stochastic Safety

How do we enforce this constraint in practice?

Theoretical

Stochastic DCBF Constraint

$$\boxed{\mathbb{E}[}\ h(\mathbf{F}(\mathbf{x},\mathbf{u})+\mathbf{d}) \mid \mathbf{x}\,] \geq \rho h(\mathbf{x})$$

Jensen's Inequality

$$h(\mathbf{F}(\mathbf{x},\mathbf{u})+\mathbb{E}[\mathbf{d}|\mathbf{x}]) + \frac{\max\{\lambda_{\max},0\}}{2}\mathrm{tr}(\mathrm{cov}(\mathbf{d}|\mathbf{x})) \geq \rho h(\mathbf{x})$$

Learning (Generative Modeling)

$$h(\boxed{\mathbf{F}(\mathbf{x},\mathbf{u})+}\,\mu_\theta(\mathbf{d}|\mathbf{x})) + \frac{\max\{\lambda_{\max},0\}}{2}\mathrm{tr}(\Sigma_\theta(\mathbf{d}|\mathbf{x})) \geq \rho h(\mathbf{x})$$

Sampled-Data Approximations

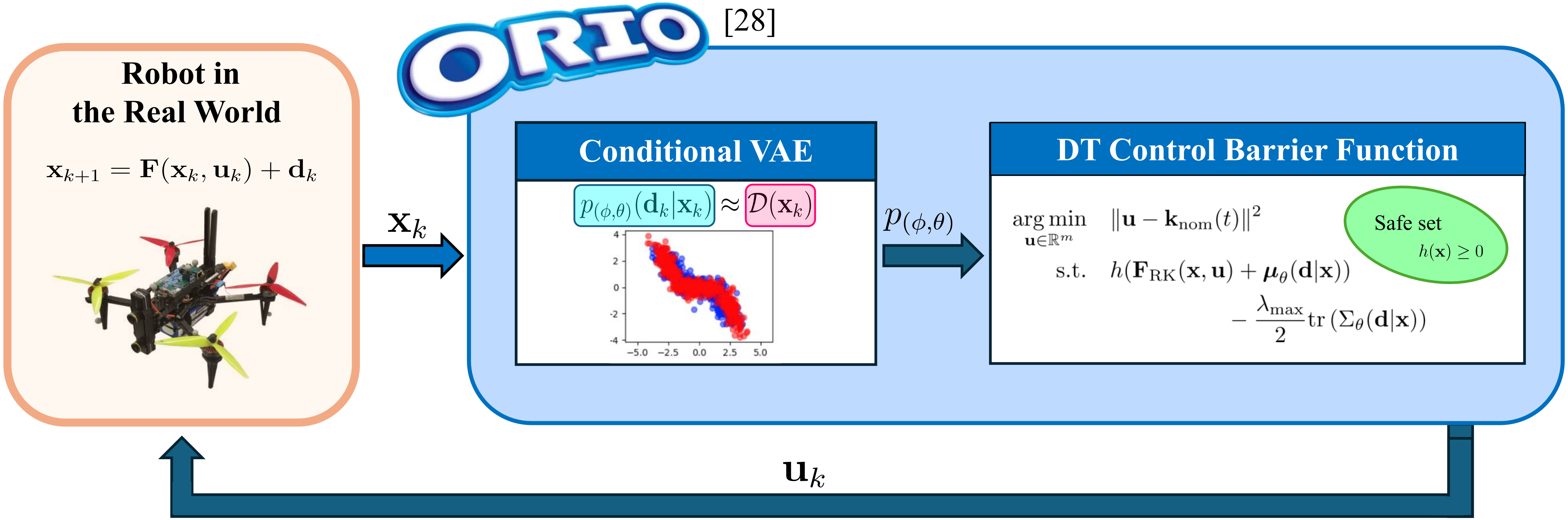
$$h(\mathbf{F}_{\mathrm{RK}}(\mathbf{x},\mathbf{u})+\mu_\theta(\mathbf{d}|\mathbf{x})) + \frac{\max\{\lambda_{\max},0\}}{2}\mathrm{tr}(\Sigma_\theta(\mathbf{d}|\mathbf{x})) \geq \rho h(\mathbf{x})$$ [27]

Practical

[27] Taylor, Dorobantu, Cosner, et al. *Safety of Sampled-Data Systems with CBFs via Approximate Discrete Time Models.* CDC, 2022.

Enforcing Stochastic Safety

Online Risk-Informed Optimization (ORIO) Controller:



[28] Cosner, et al. *Generative Modeling of Residuals for Real-Time Risk-Sensitive Safety with Discrete-Time Control Barrier Functions.* ICRA 2024.

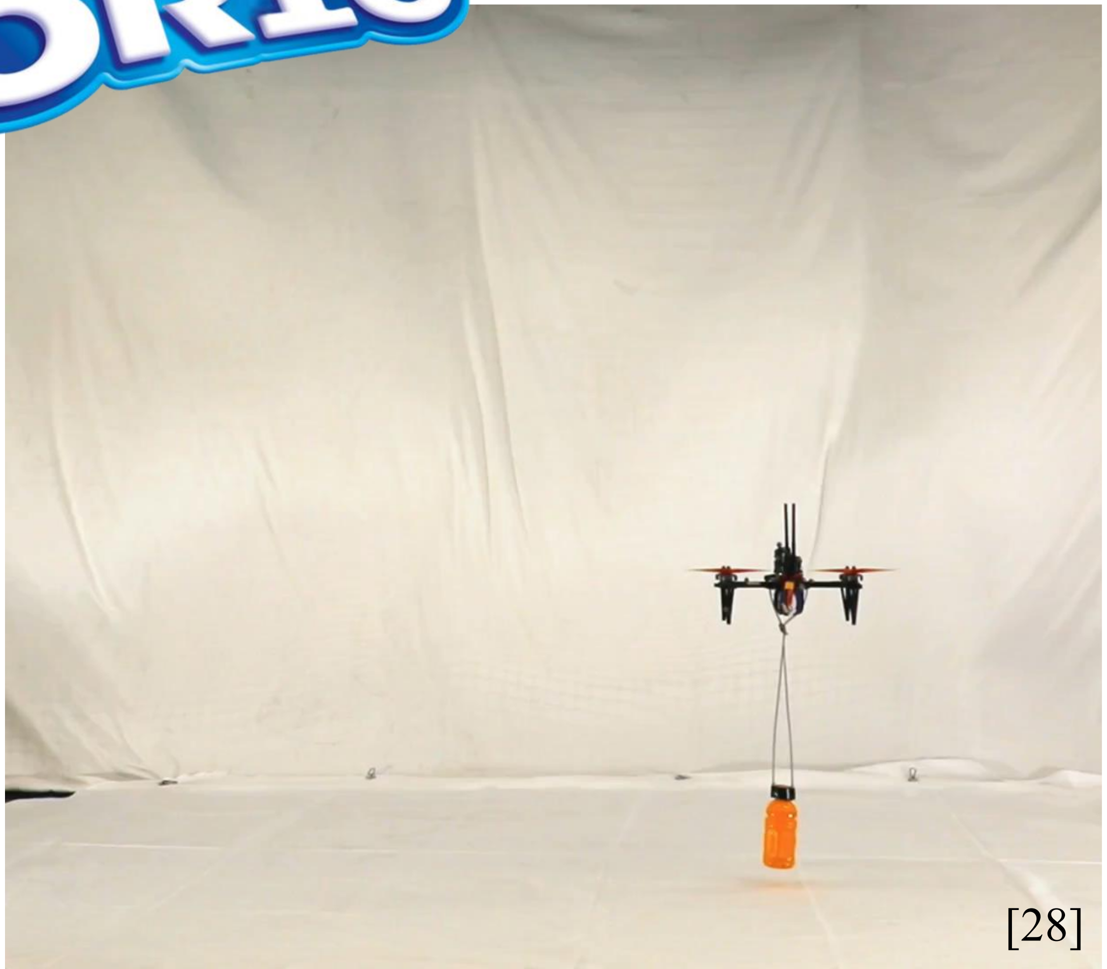
Application

- Learned distribution with generative modeling
- Enforce stochastic safety

Theorem: Stochastic Safety [24,26]

Stochastic DCBFs $\implies$ bounded risk of failure.

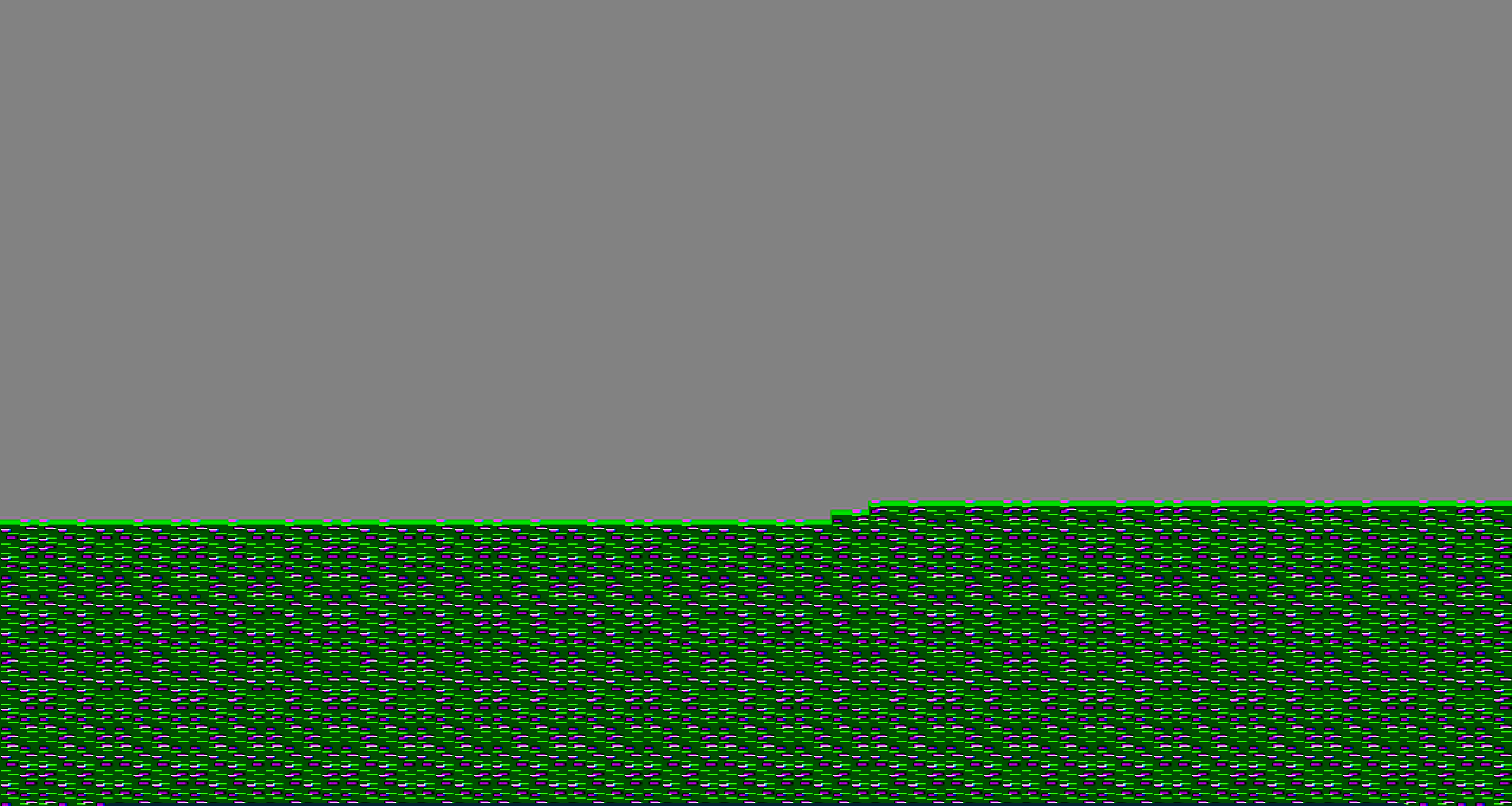
ORIO



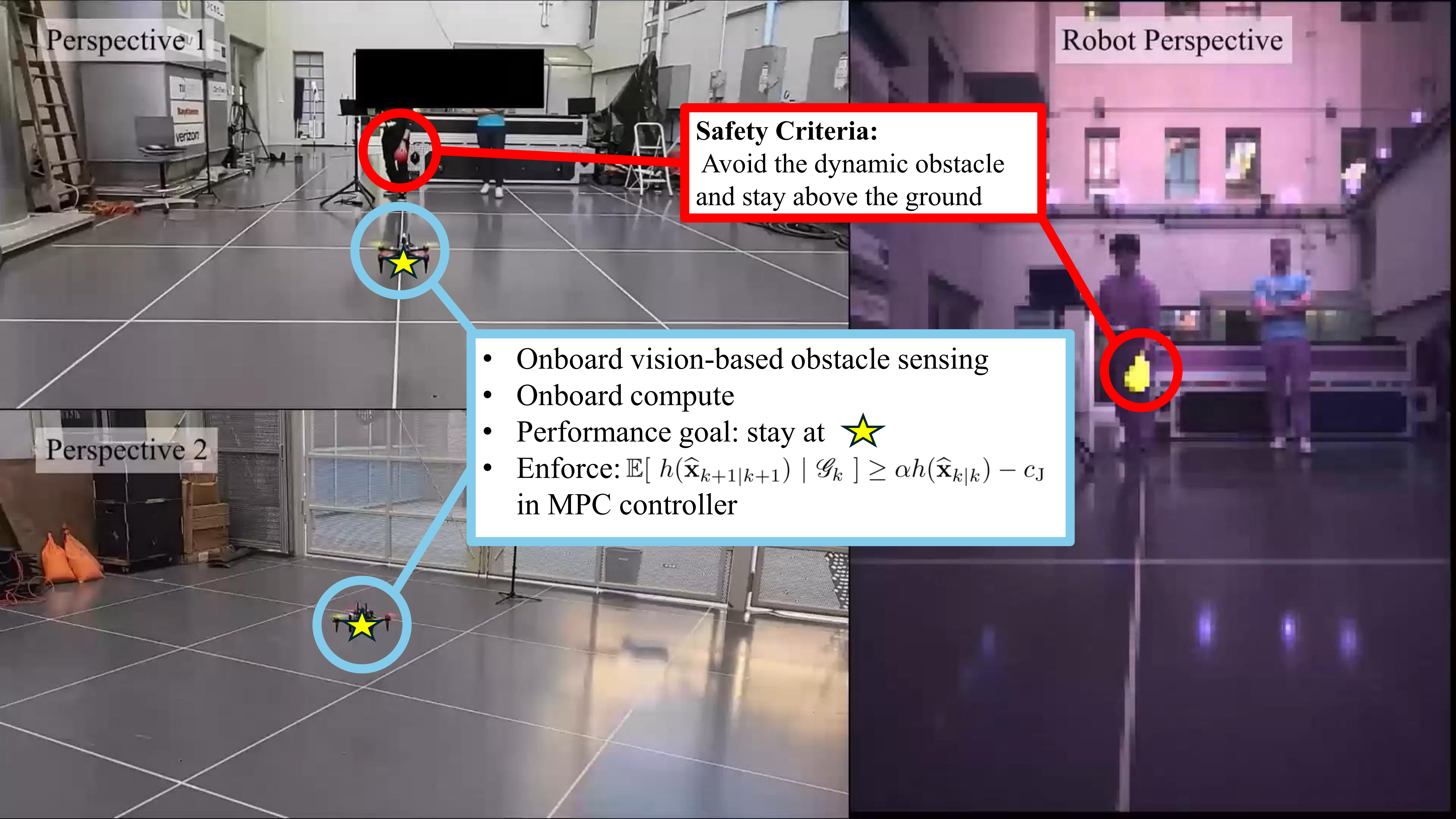
[28]

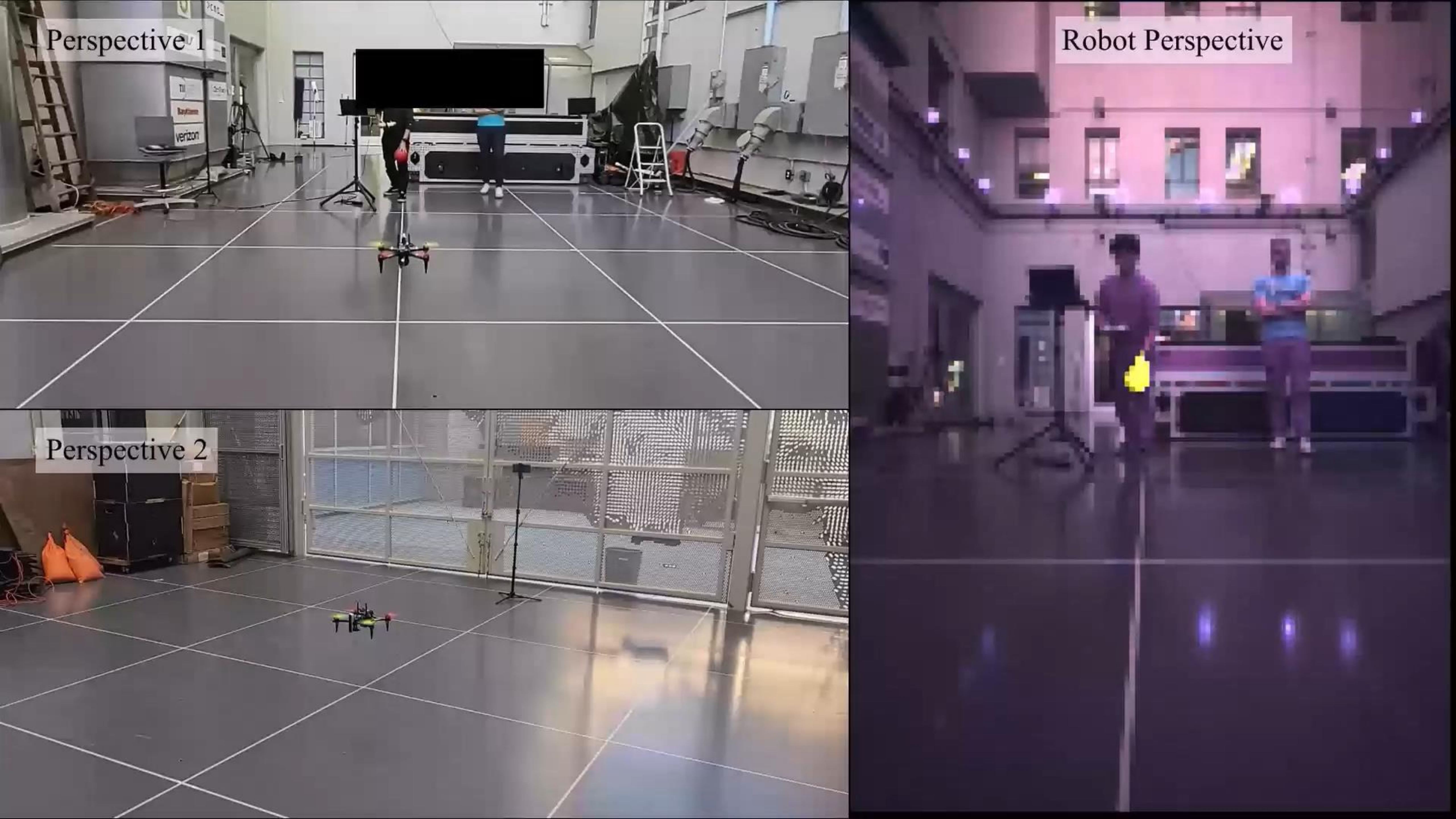
[24] Cosner, et al. *Robust safety under stochastic uncertainty with discrete-time control barrier functions.* RSS, 2023.
[26] Cosner, et al. *Bounding Stochastic Safety: Leveraging Freedman's Inequality with Discrete-Time Control Barrier Functions.* LCSS, 2024.
[28] Cosner, et al. *Generative Modeling of Residuals for Real-Time Risk-Sensitive Safety with Discrete-Time Control Barrier Functions.* ICRA 2024.

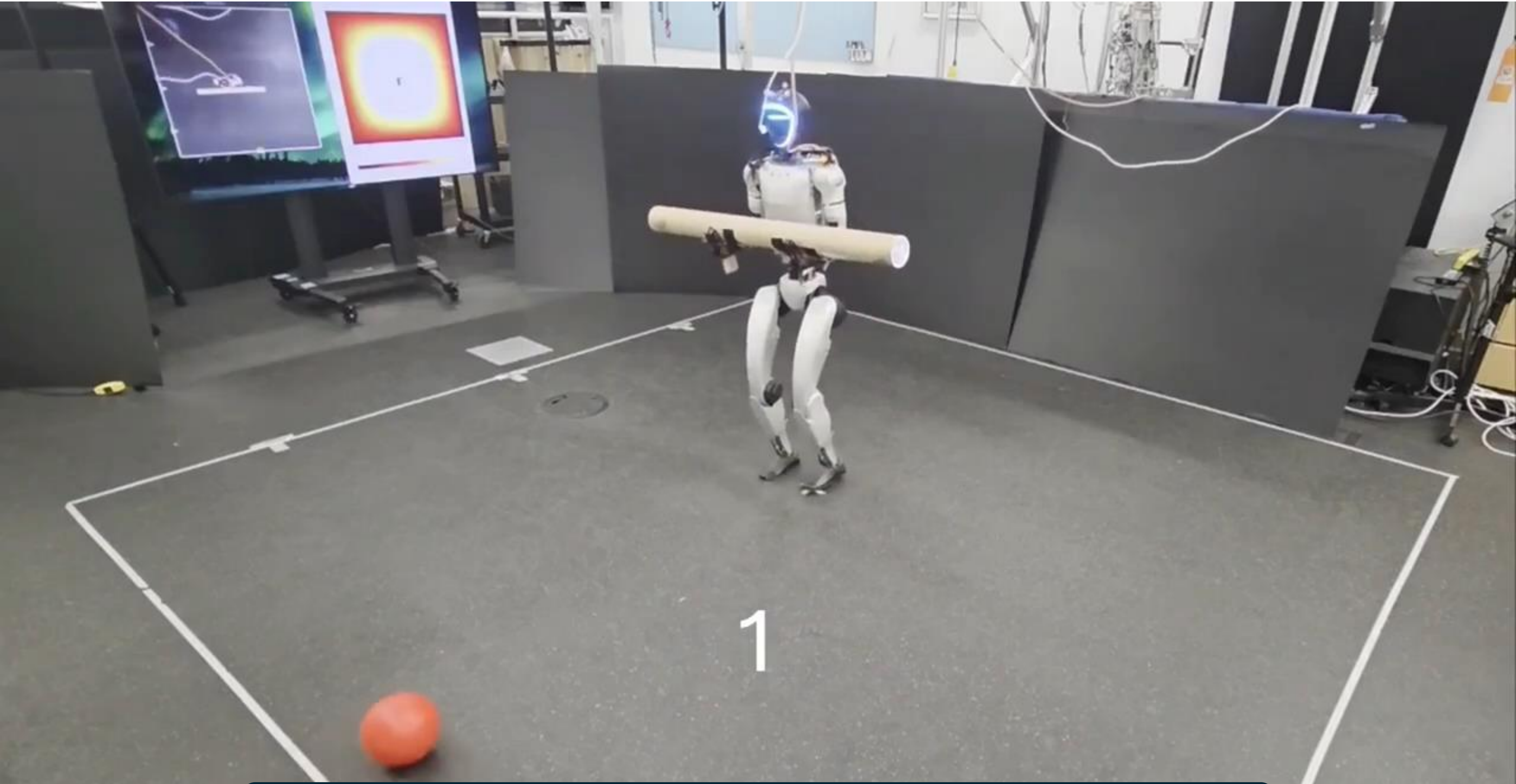

Caltech



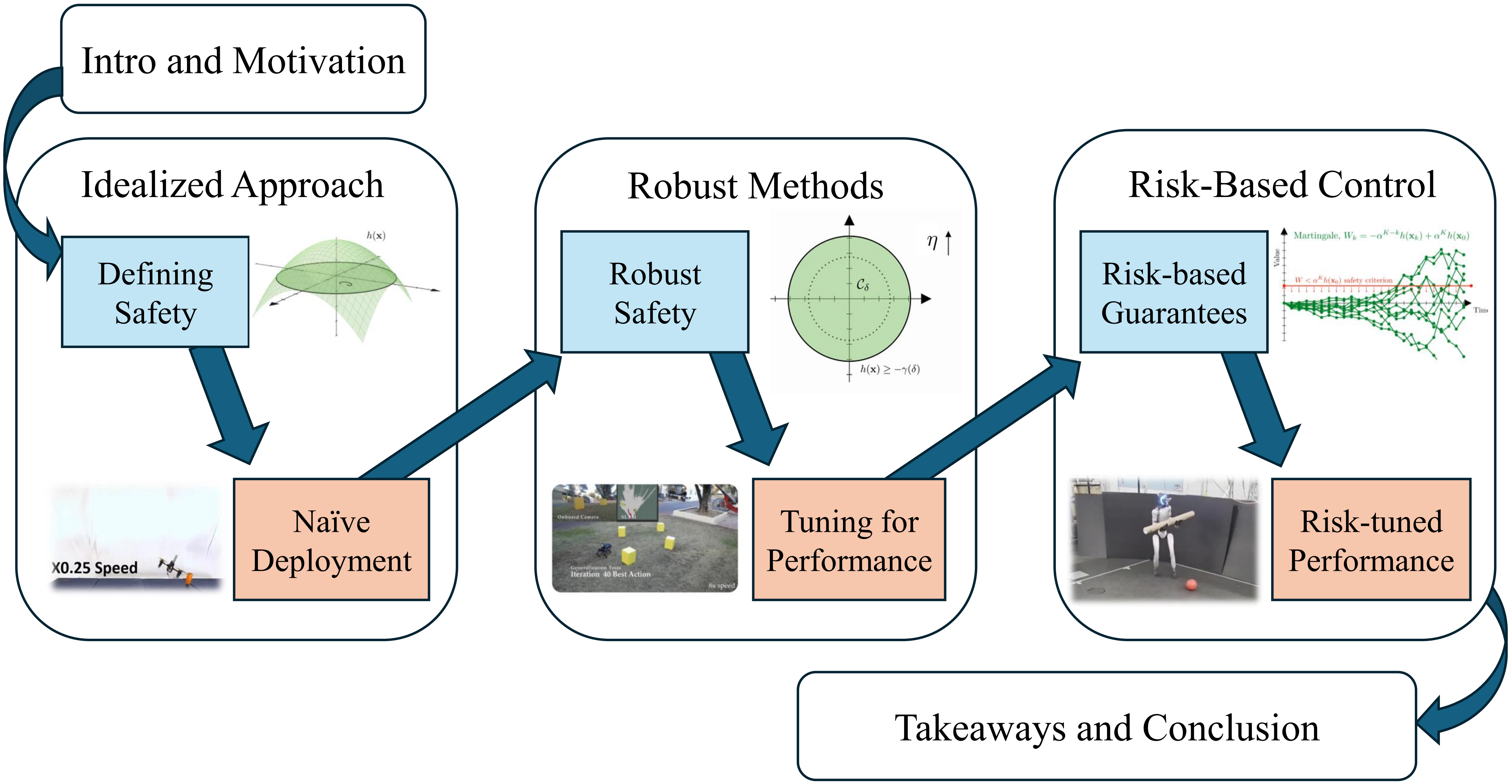
[29] Yang, Werner, Cosner, et al. *SHIELD: Safety on Humanoids via CBFs In Expectation on Learned Dynamics*. IROS, 2025.


Perspective 1
Perspective 2
Robot Perspective
Safety Criteria:
Avoid the dynamic obstacle and stay above the ground
• Onboard vision-based obstacle sensing
• Onboard compute
• Performance goal: stay at ☆
• Enforce: $\mathbb{E}[\ h(\widehat{\mathbf{x}}_{k+1|k+1}) \mid \mathscr{G}_k\ ] \geq \alpha h(\widehat{\mathbf{x}}_{k|k}) - c_{\mathrm{J}}$ in MPC controller


Perspective 1
verizon
Perspective 2
Robot Perspective



[30] Bena, Bahati, Werner, Cosner, et al. *Geometry-aware predictive safety filters on humanoids: From PSFs to CBF-constrained MPC*. Humanoids, 2025.



Intro and Motivation
Idealized Approach
Defining Safety
$h(\mathbf{x})$
X0.25 Speed
Naïve Deployment
Robust Methods
Robust Safety
η
$\mathcal{C}_\delta$
$h(\mathbf{x}) \geq -\gamma(\delta)$
Tuning for Performance
Risk-Based Control
Risk-based Guarantees
Risk-tuned Performance
Takeaways and Conclusion

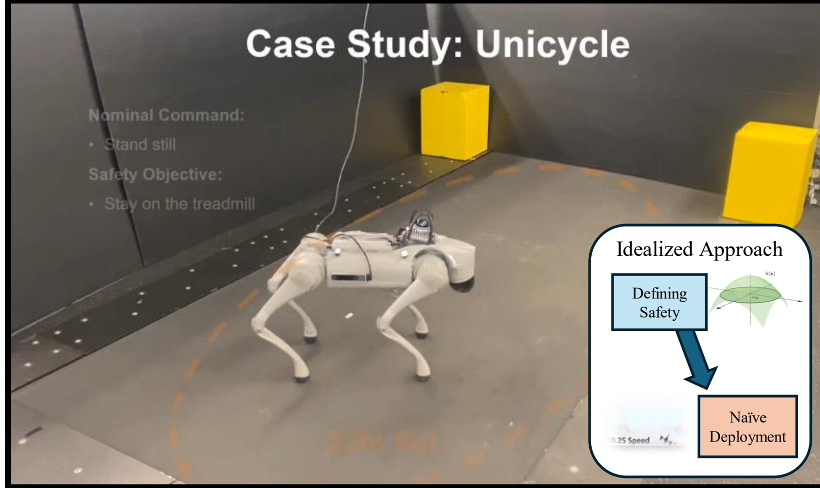
Conclusion

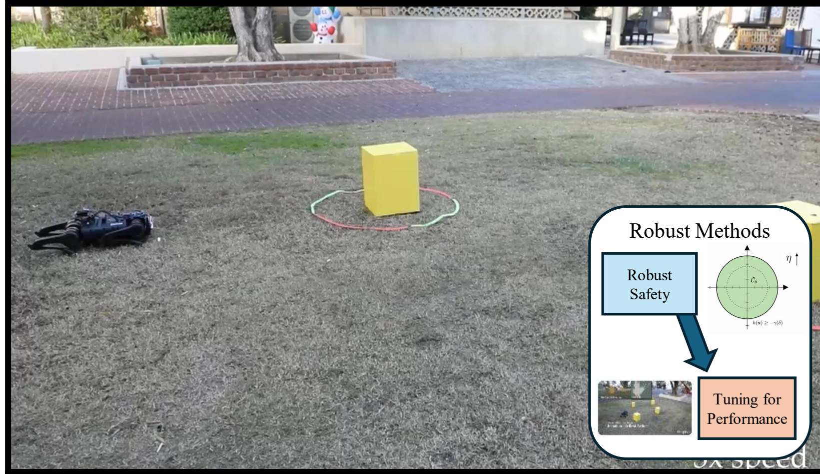
Contributions:

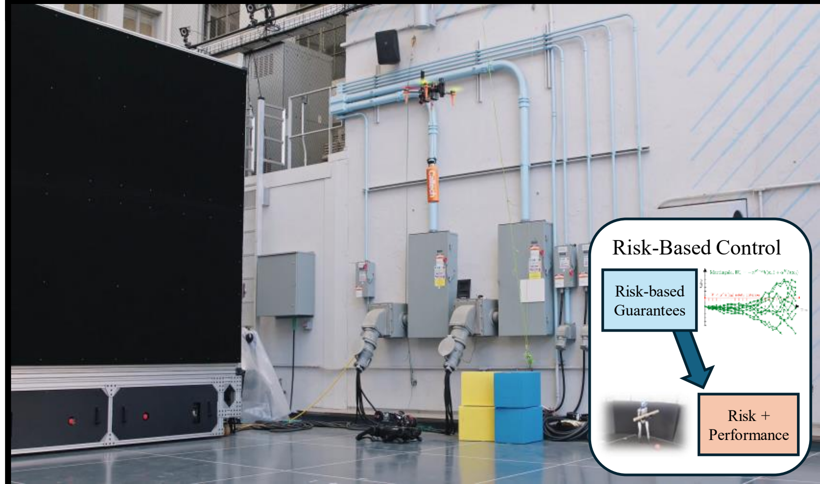
- Robust theoretical safety guarantees
- Theory-guided machine learning for safe performance
- Safety with tunable risk-based guarantees

Key Takeaways:

- Theoretical guarantees elucidate key characteristics
- Formal methods for safety + machine learning for performance



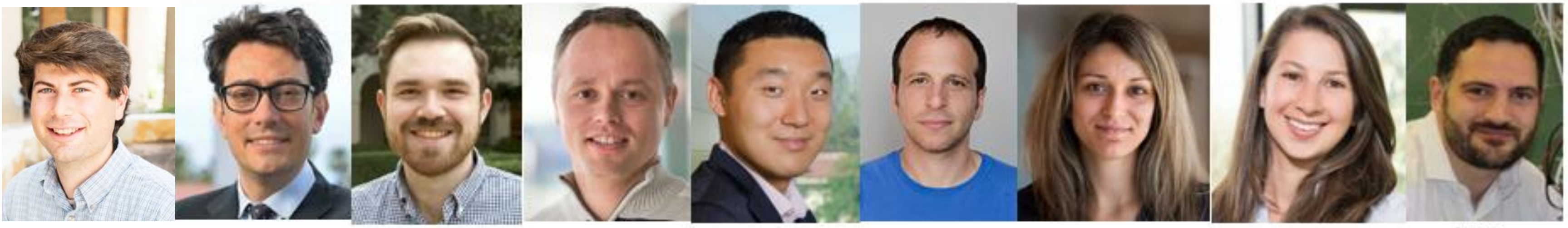




Acknowledgements: People

Lizhi Yang
Yuxiao Chen
Max H. Cohen
Ivan Jimenez-Rodriguez
Ugo Rosolia
Devnash Agrawal
Hardik Parwana
Andrew Singletary
Min Dai
Maegan Tucker
Victor Dorobantu
Gilbert Bahati

Blake Werner
Shushant Veer
Ryan Bena
Andrew J Taylor
Kejun Li
Tamas Molnar
Wyatt Ubellacker
Anil Alan
Kunal Garg
Sarah Dean
Noel Csomay-Shanklin
Karen Leung



David Fridovich-Keil
Marco Pavone
Preston Culbertson
Gabor Orosz
Yisong Yue
Ben Recht
Dimitra Panagou
Katherine Bouman
Aaron Ames

Acknowledgements: Support

Tufts UNIVERSITY

Caltech

Berkeley UNIVERSITY OF CALIFORNIA

UNIVERSITY OF MICHIGAN

NVIDIA

AeroVironment™

NSF

bp

Questions?

Lab website: sites.tufts.edu/sparc

sparc lab

The Safe + Performant Autonomous Robotics and Control Lab

Tufts UNIVERSITY

Ryan K. Cosner
Glenn R. Stevens Assistant Professor
Mechanical Engineering, Tufts University