Robust Guarantees from Data: Bandits and Quantiles

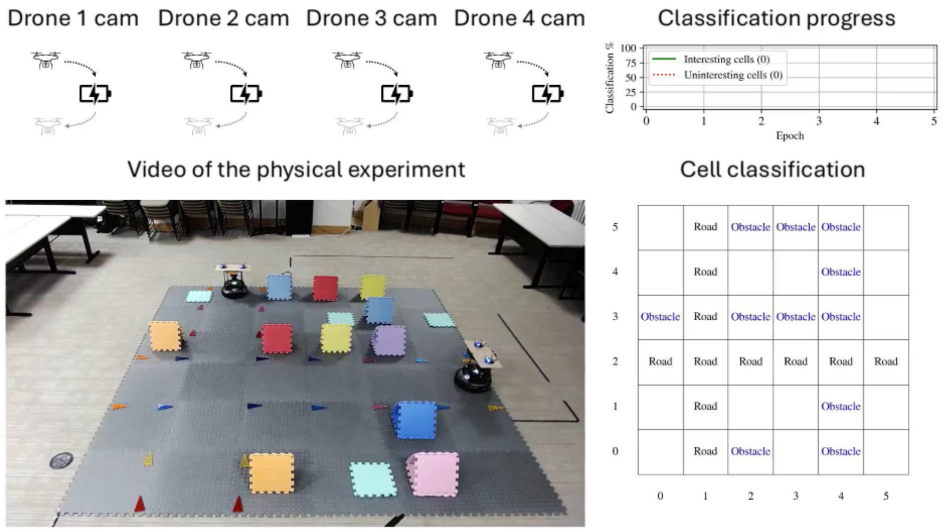
Abraham P. Vinod
Principal Research Scientist, Control for Autonomy Group

Trustworthy Data-Driven Control Workshop
August 23rd, 2026

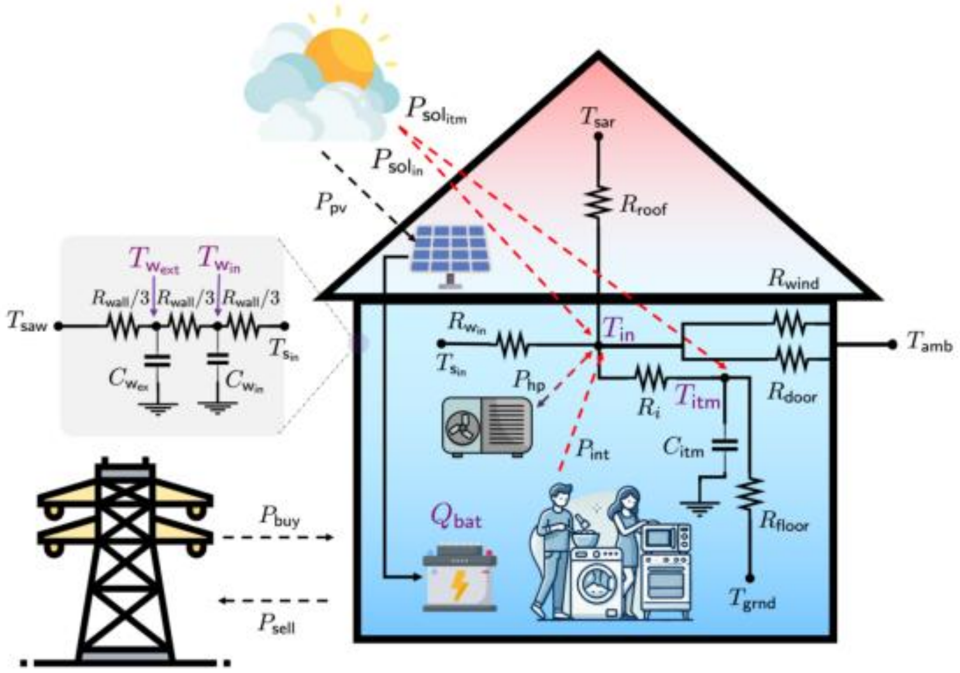
IFAC 2026 BUSAN

What will I be talking about?

- How to coordinate robot teams to quickly identify interesting regions in a search environment under constraints?
  - Secret sauce: Multi-arm bandits, temporal estimators, mixed-integer programming
  - Application: Disaster management, search-and-rescue, precision agriculture



- **(Bonus)** How can we use building disturbance forecast data to control grid-interactive buildings effectively?
  - Secret sauce: Sample quantiles, optimization, and time series analysis



Thank you to my collaborators (and amazing interns I have had in the past)

People @ MERL


S. Di Cairano


A. Chakrabarty


C. Laughman


A. Raghunathan

People outside MERL (Greiff and Germain were at MERL during this work)


M. Greiff (TRI)

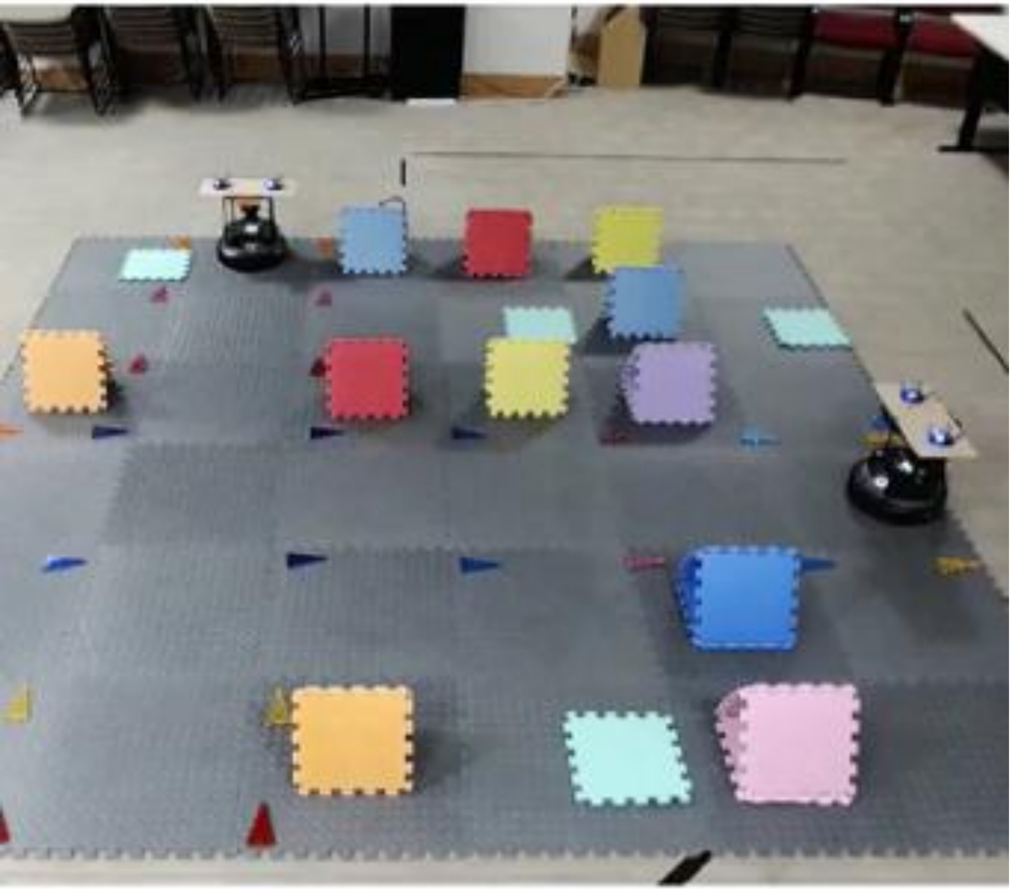

F. Germain (Shure)


G. Dasarathy (ASU)

Interns (during 2022-2025)

Top row, left to right
Xiaoshan Lin (UMN → Facebook)
Siddharth Nayak (MIT → Waymo)
Parth Thaker (ASU → Intuitive Surgicals)
Bottom row, left to right
Ali Buyukkocak (UMN → Mytra)
Khalid Jaffar (UMD)
Spencer Hutchinson (UCSB)

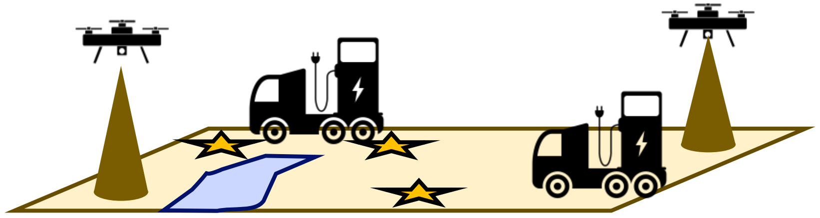
Internship opportunities



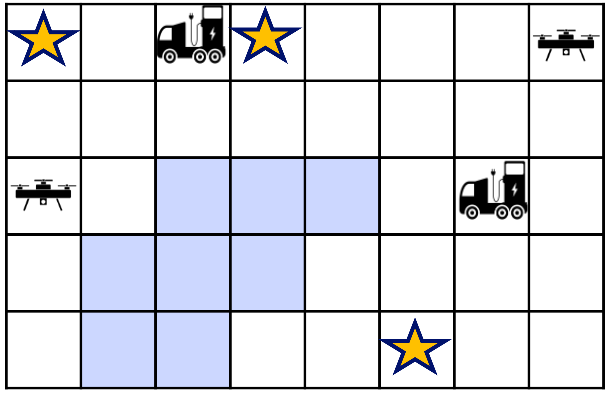
Spatio-temporal monitoring using energy-constrained heterogeneous robots

IFAC WC '23, ACC' 24, T-CST '26,
CDC '26 (accepted)

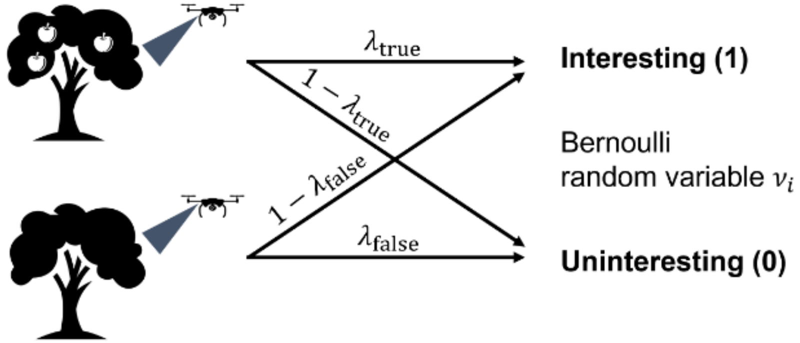
Problem overview: Static monitoring using heterogeneous robots



Sensor-equipped, energy-constrained drones identify interesting regions



Coordinate search over a discrete world



Perception introduces uncertainty (hard to model)

Goal: Identify cells (set of all cells is $\mathcal{G}$) with high probability μ_i (unknown mean of ν_i),

$$\mathcal{S}_\theta = \{i \in \mathcal{G} \mid \mu_i \geq \theta\}$$

Additional desirables

✓ Cooperative search based on data collected online

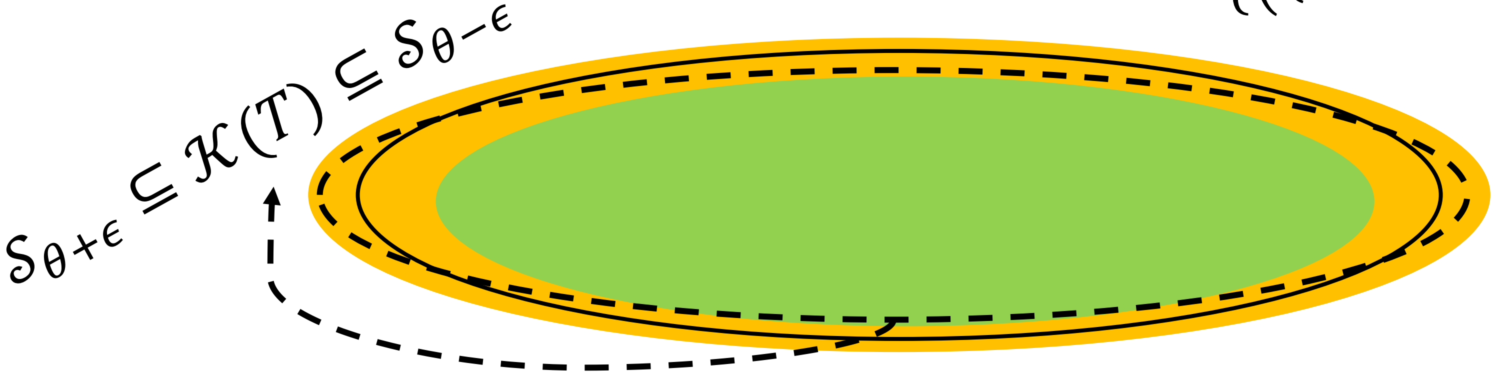
✓ <u>Prioritize finding interesting cells</u> over uninteresting cells

✓ Account for <u>physical constraints</u> --- Motion and energy constraints on drones

What constitutes a successful search?

- A successful search terminates at $T \in \mathbb{N}$ to returns $\{\mathcal{K}(p)\}_{p=0}^{T}$ must satisfy:

1. **Low labelling error at T:** Ensure $\mathbb{P}\left\{\left(\left(\mathcal{S}_{\theta+\epsilon} \setminus \mathcal{K}(T)\right) \cup \left(\mathcal{K}(T) \setminus \mathcal{S}_{\theta-\epsilon}\right)\right) \neq \emptyset\right\} \leq \delta$



$\mathcal{S}_{\theta-\epsilon}$

$\mathcal{A} \nsubseteq \mathcal{B} \Leftrightarrow \mathcal{A} \setminus \mathcal{B} \neq \emptyset$

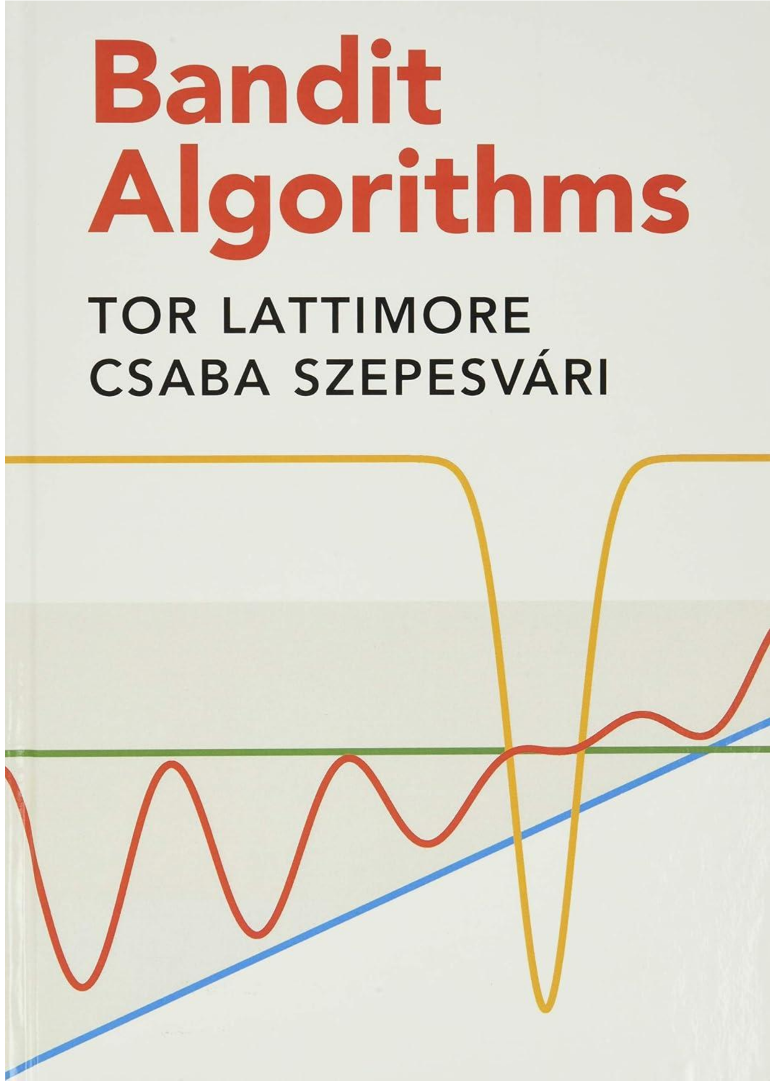
$\mathcal{S}_{\theta} = \{i \in \mathcal{G} \mid \mu_i \geq \theta\}$

$\mathcal{S}_{\theta+\epsilon}$

2. **Finite iterations to meet the above:** Environment-dependent bounds
3. **Monotone property:** Never flip a decision
4. **Anytime property:** Give useful decisions throughout

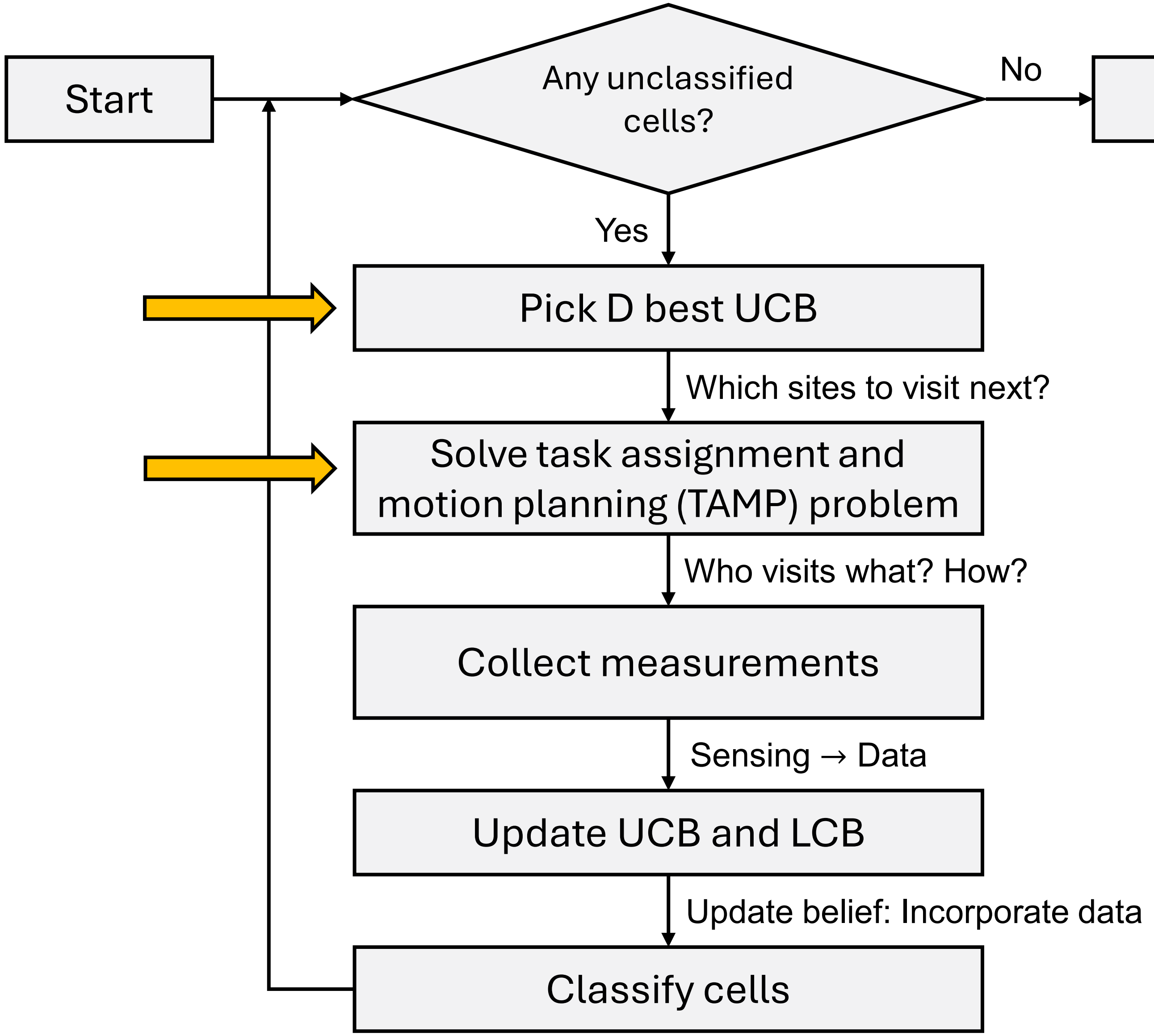
What are multi-arm bandits? Why is it relevant here?

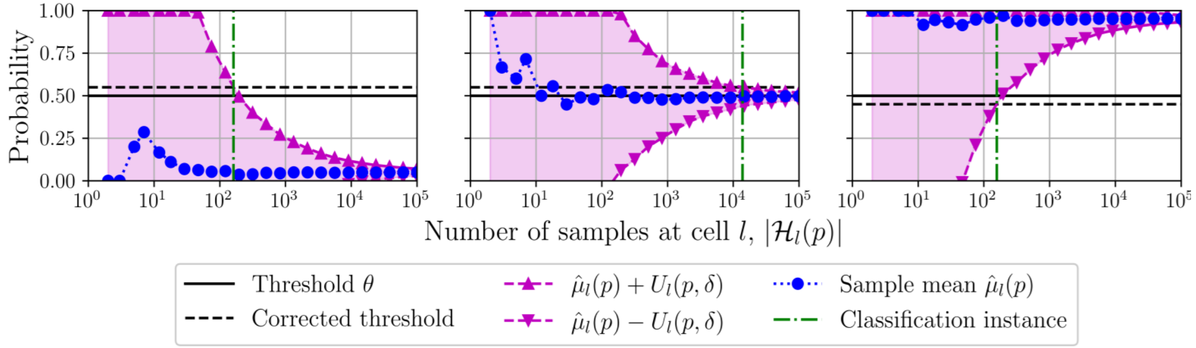
- Bandits are a special class of stochastic optimal control/reinforcement learning
  - Current actions do not affect the future
- Why are multi-arm bandits relevant here?
  - Treat each cell $i \in \mathcal{G}$ as an arm of the bandit
  - Each arm pull returns a Bernoulli reward with unknown mean μ_i
  - **Search problem:** Sequentially pull arms to terminate with $\mathcal{S}_{\theta+\epsilon} \subseteq \mathcal{K}(T) \subseteq \mathcal{S}_{\theta-\epsilon}$ with high prob. ⇒ Thresholding bandit.
- How are thresholding bandit problems solved?
  - Sequentially minimize a surrogate function
  - Return a set based on the surrogate function after an appropriate number of iterations
  - See Locatelli et al., 2016; Mason et al., 2020 for more details
- **Challenge:** *Robot constraints are not unaccounted.*



Solution framework

UCB: Upper confidence bound LCB: Lower confidence bound



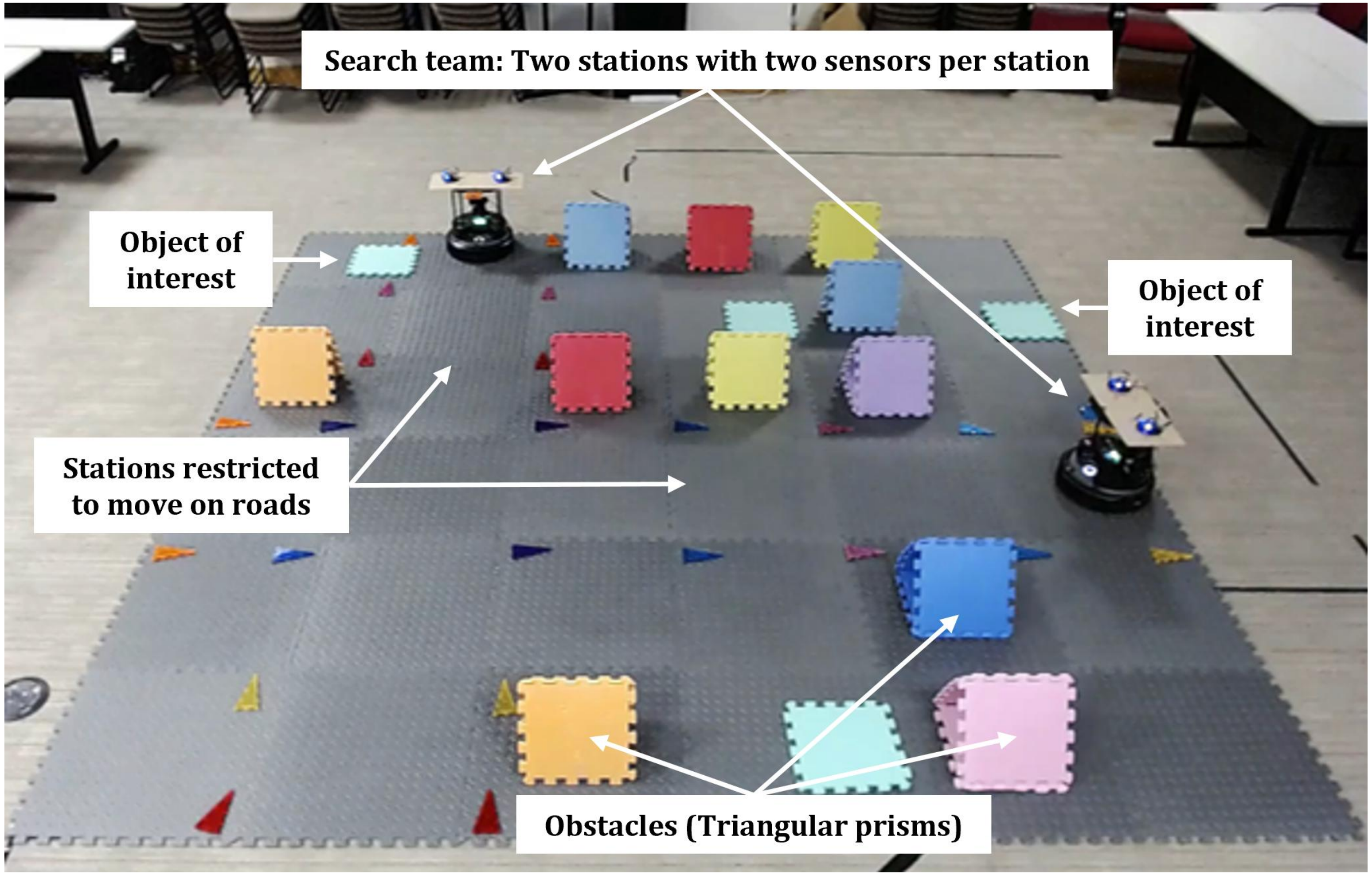


Classify cells using LCB and UCB

Use integer linear program or heuristics to solve the TAMP problem.

min.	Number of flight cycles,
s. t.	Search team respects motion constraints,
	Search team visits all cells in $\mathcal{E}_p$ but none in $\mathcal{O}$,
	Mobile sensors never run out of battery.

Demonstration on physical hardware: Set up



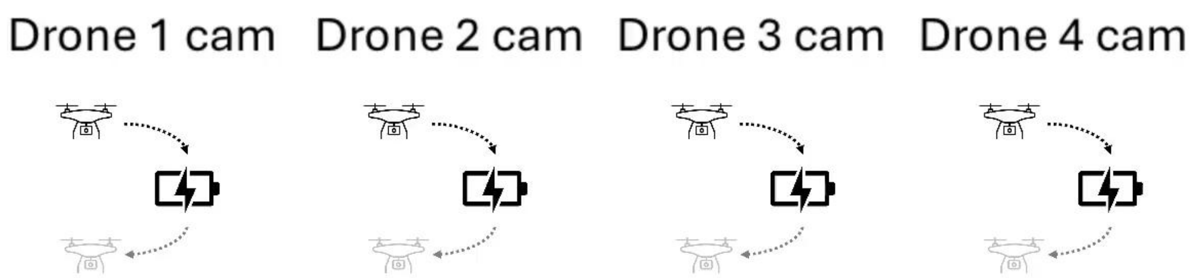
Team: 4 Crazyflie drones (sensors) + 2 Turtlebot robots (stations).

Obstacles: Sensors can not visit cells with triangular prisms.

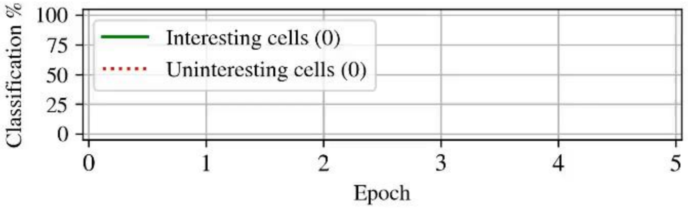
Road: Stations must stay on the 2^{nd} column and 3^{rd} row.

Spatial distribution of cells (apriori unknown): 4 green tiles indicate interesting cells, and other cells are uninteresting.

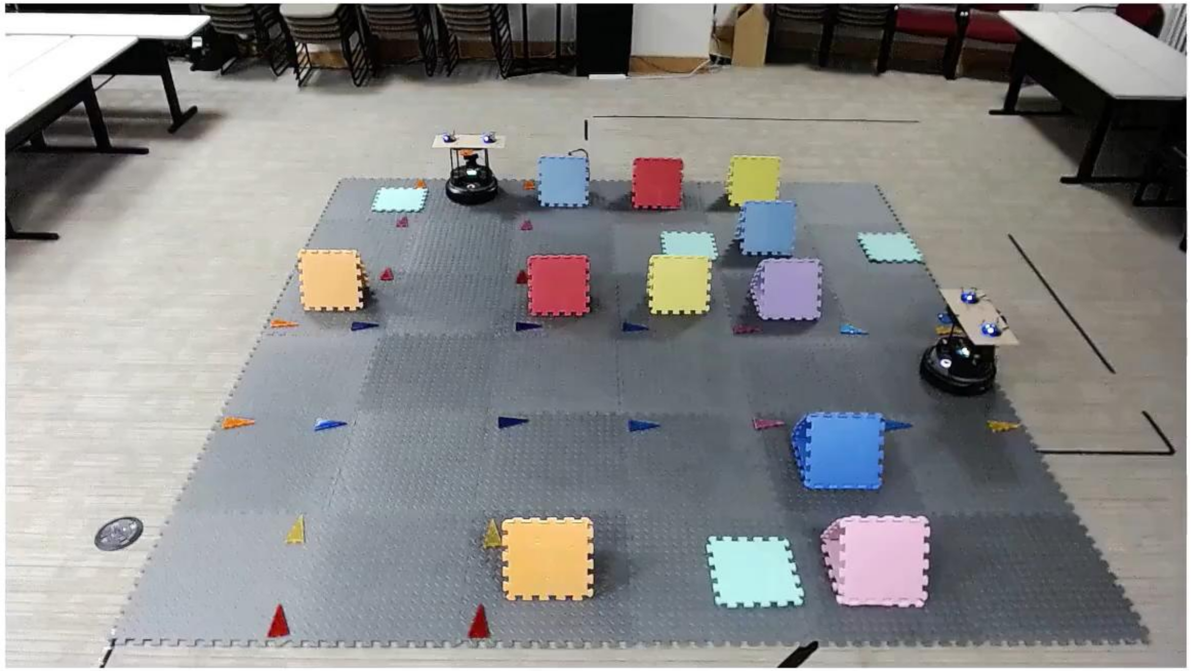
Demonstration on physical hardware: Video



Classification progress



Video of the physical experiment



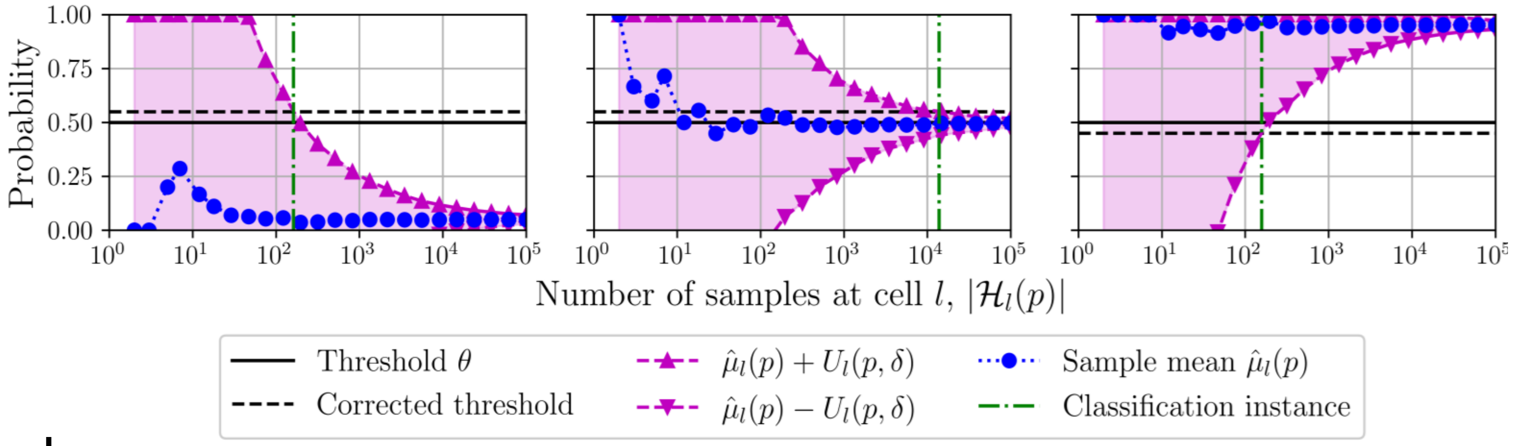
Cell classification

5		Road	Obstacle	Obstacle	Obstacle	
4		Road			Obstacle	
3	Obstacle	Road	Obstacle	Obstacle	Obstacle	
2	Road	Road	Road	Road	Road	Road
1		Road			Obstacle	
0		Road	Obstacle		Obstacle	
	0	1	2	3	4	5

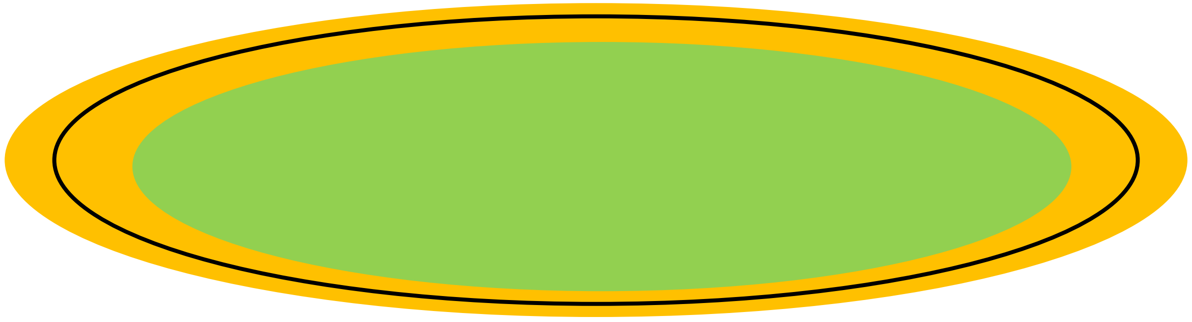
https://www.youtube.com/watch?v=gzulpOcVYzg

We meet all the criteria of a successful search!

- Set constructions implies **monotone decisions**



- **Anytime method:**
  With probability $1-\delta$, for all epoch p,
  $\mathcal{K}(p) \subseteq \mathcal{S}_{\theta-\epsilon}$, and $\mathcal{R}(p) \cap \mathcal{S}_{\theta+\epsilon} = \emptyset$.



$\mathcal{S}_{\theta-\epsilon}$

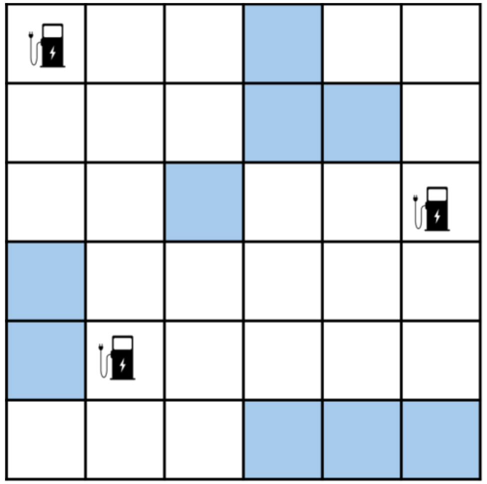
$\mathcal{S}_\theta = \{i \in \mathcal{G} \,|\, \mu_i \geq \theta\}$

$\mathcal{S}_{\theta+\epsilon}$

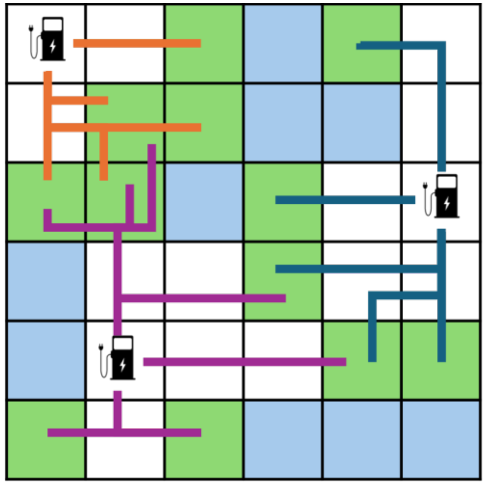
- **Finite iterations to correctly labelled termination in** $\mathcal{O}\left(\frac{\log(\delta \log(\delta))}{(|\mu_i - \theta| + \epsilon)^2}\right)$.

Extension 1/3: If I assume static stations, can I get more?

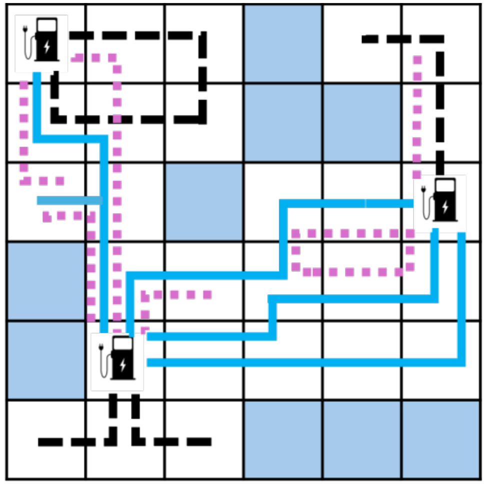
- Yes! Arms of the bandit are now paths. First, how do we design them?



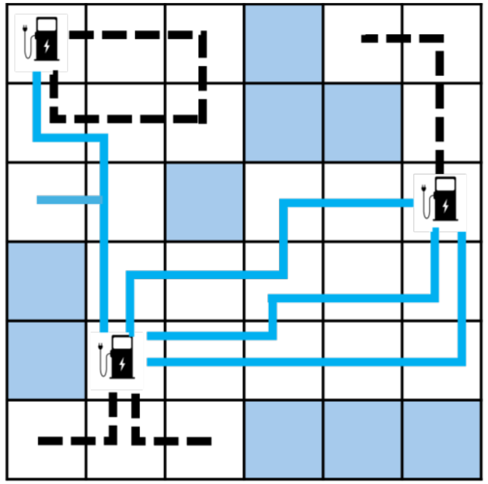
Environment



Shortest-path forest with 1/2 energy limit

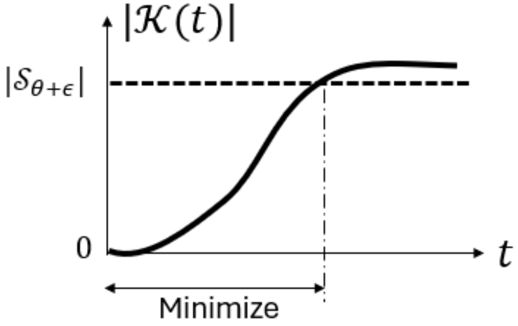


Join them to get self-loops and station-to-station path transfers

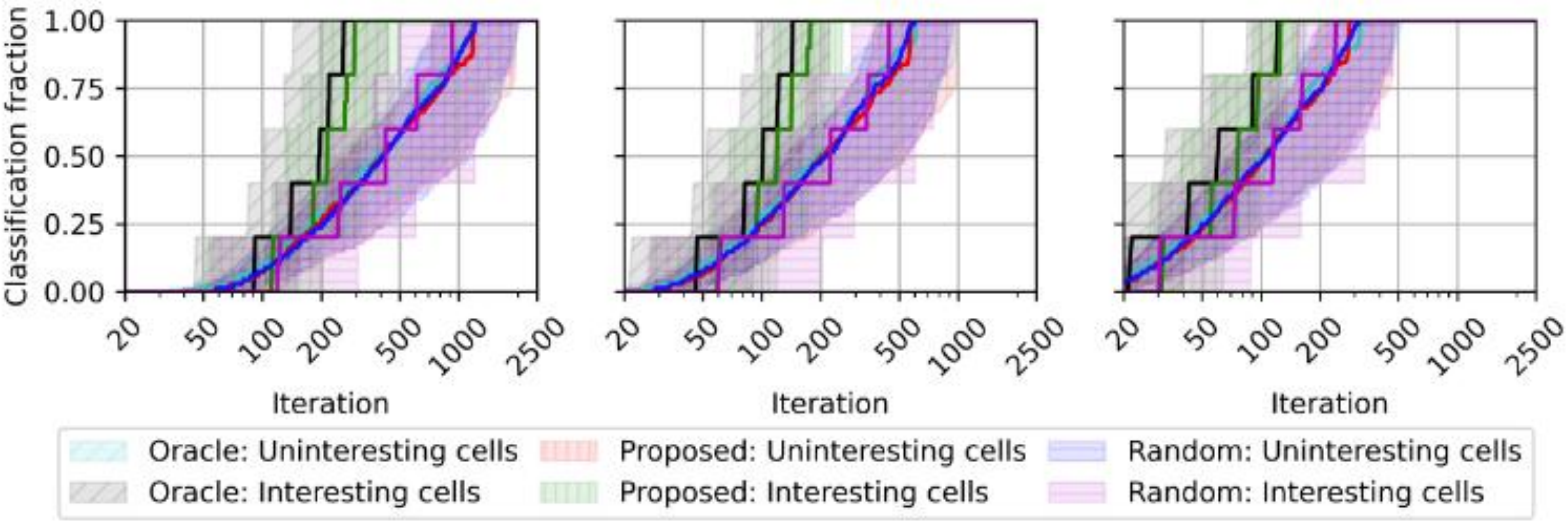


Set cover problem to find a minimum cardinality set

- **Guarantees:** What can we additionally say?
  - Sufficient conditions for priority labeling (Discover $\mathcal{S}_{\theta+\epsilon}$ ASAP!)
  - Refined bounds on the priority labeling and termination time
  - Minimax-optimality rates on the above bounds for a broad class of search problems



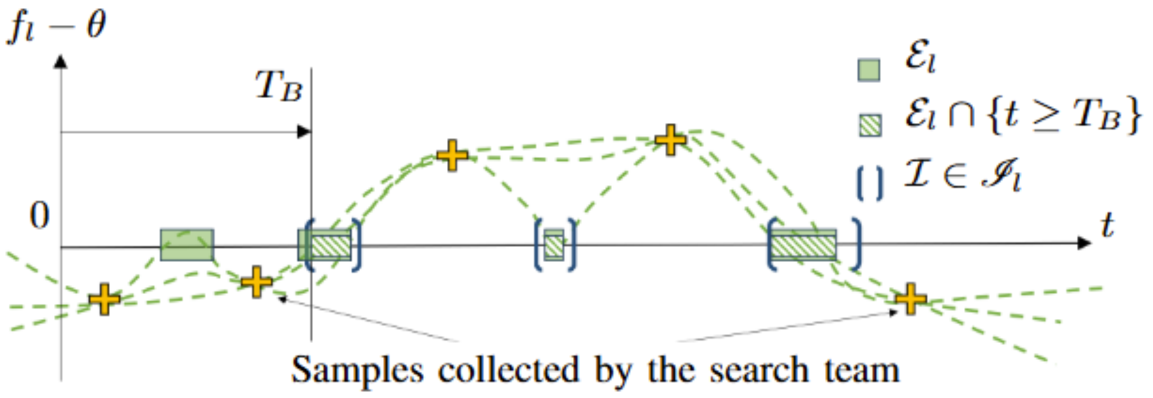
Extension 1/3: If I assume static stations, can I get more?

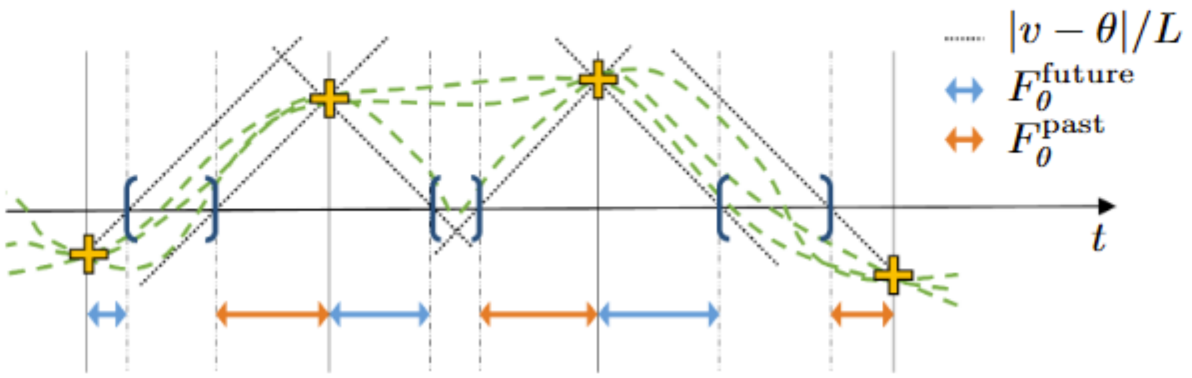


Oracle here knows which paths contain at least one interesting cell.
Our approach without that knowledge approaches that performance!

Extension 2/3: What if the environment is not static?

- **Problem:** Identify zero crossings of an unknown function using data collected by a team of robots
- **Model:** Each cell has an unknown function $f_l \in \mathfrak{F}$, $\mathfrak{F}$ is known (e.g., Lipschitz continuous, Lipschitz gradients)

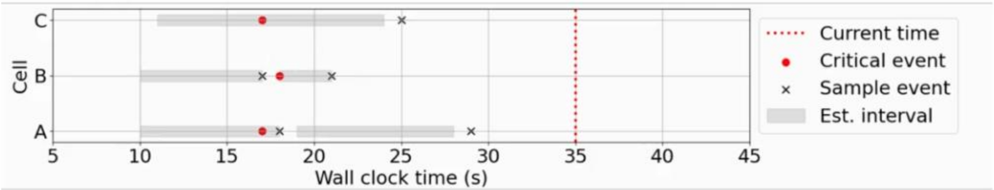




- **Solution:** Replace UCB with temporal estimators that identify earliest occurrence of critical events, and then do receding horizon-style to deploy the team with least delay ⇒ Persistent monitoring
- **Guarantees:** Sufficient conditions (static charging + sufficient battery) for
  - Recursive feasibility,
  - Generated interval set will not miss any critical event,
  - Sufficiently separated critical events → Every interval has at most one critical event.

Extension 2/3: What if the environment is not static?

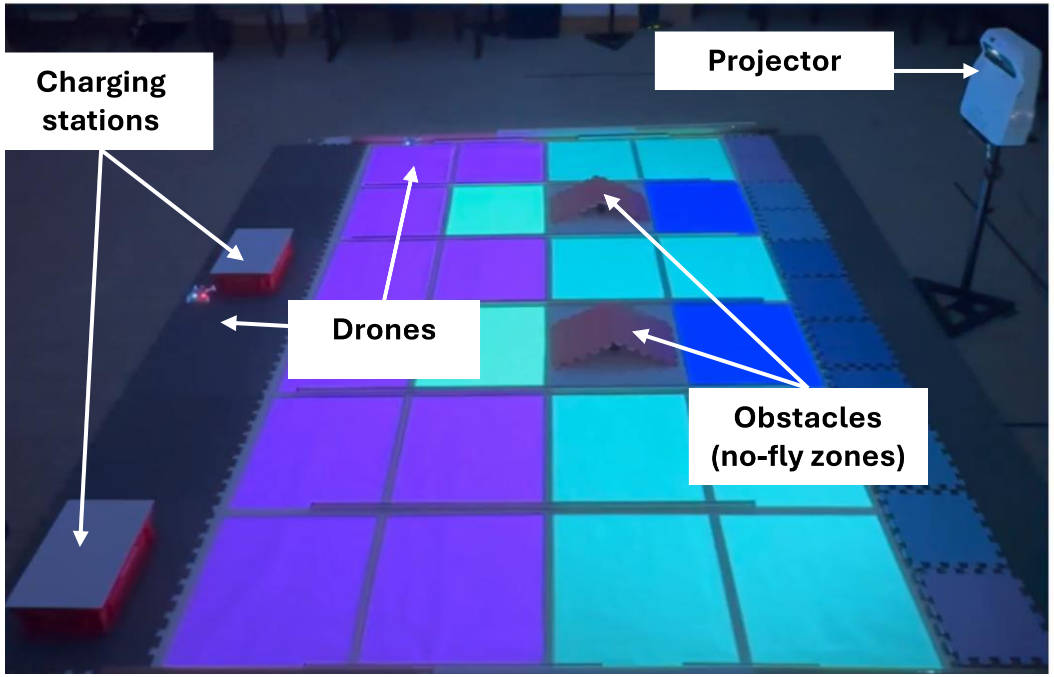
Resulting intervals



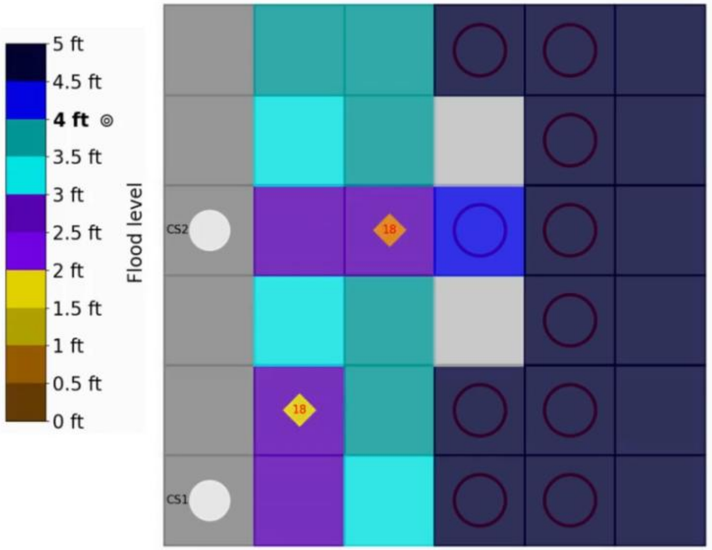
Drone 1 cam

Drone 2 cam

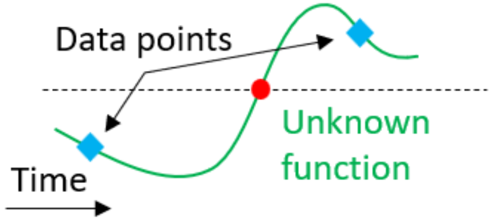
Video of physical experiment



Last sample info.



Search team uses online data to identify **when and where** the water level reaches 4ft --- a *critical event.*



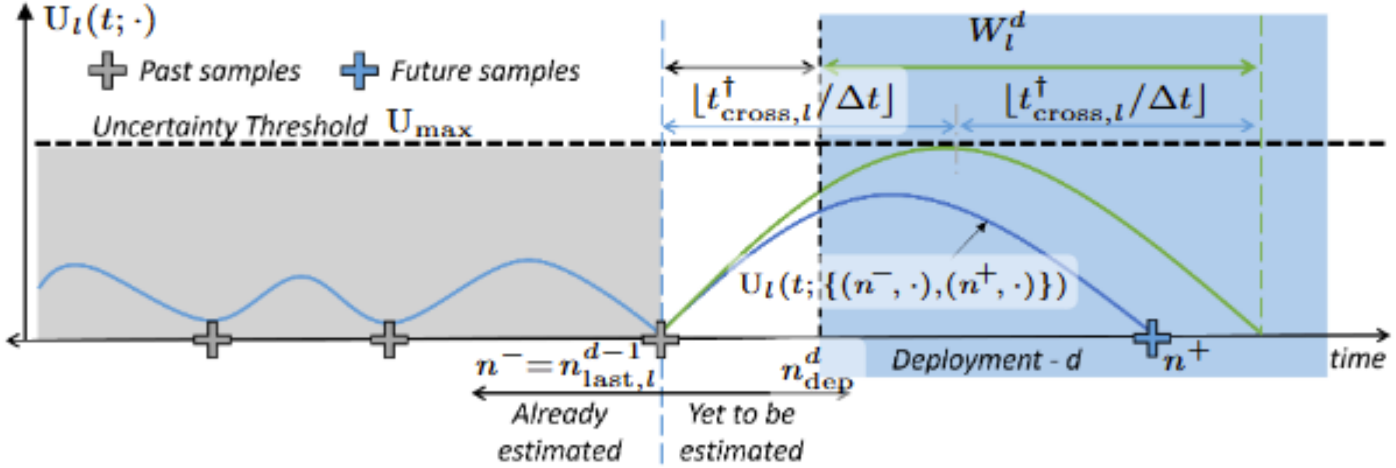
Compute intervals guaranteed to cover critical events using:

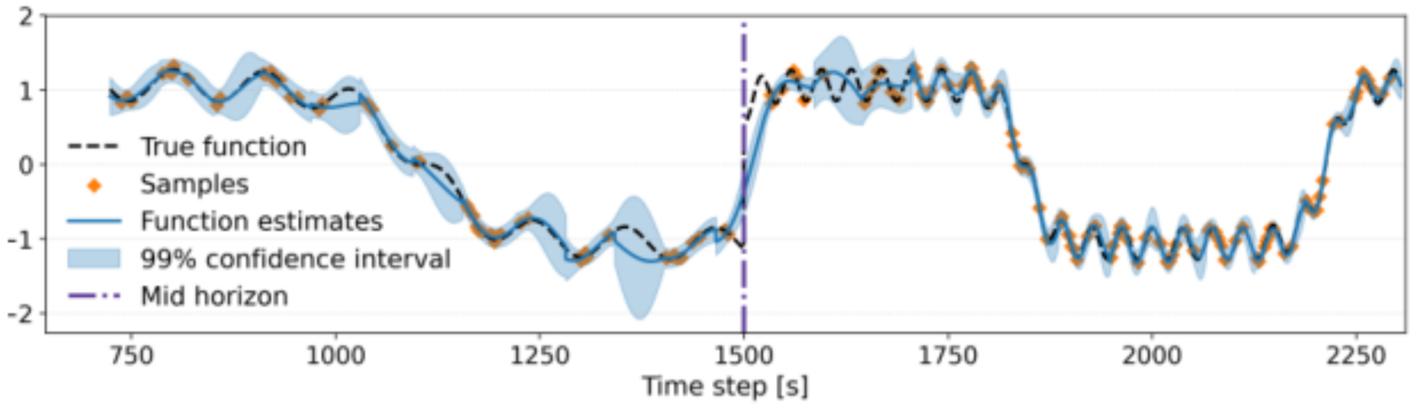
1. noisy data collected online,
2. limited prior information.

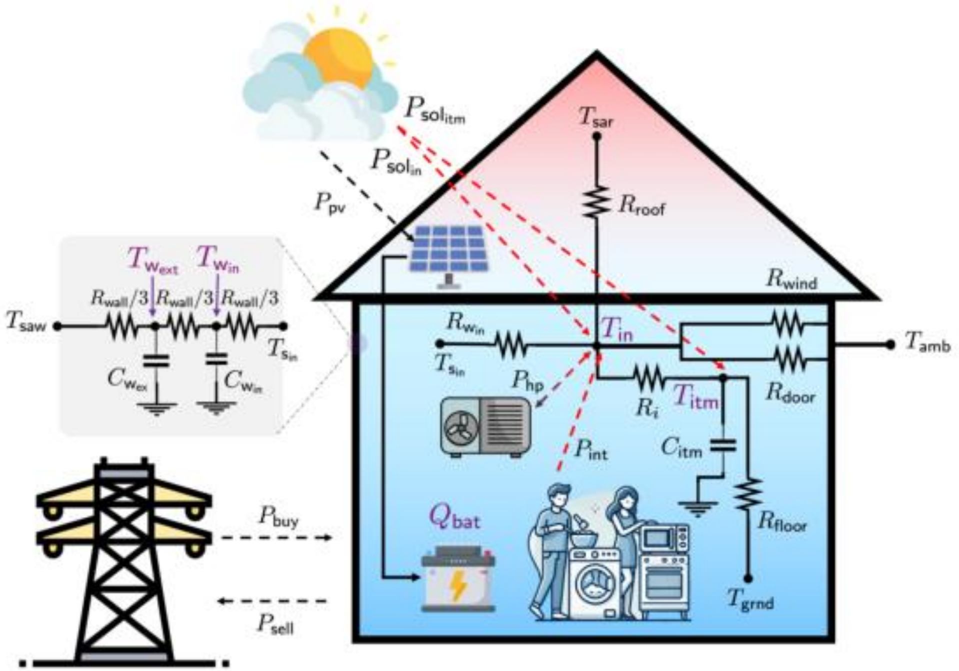
UAV respect physical constraints, and land on stations for recharging.

Extension 3/3: What if the family class $\mathfrak{F}$ is changing over time?

- **Problem:** Estimate f_l with tolerable uncertainty using data collected by a team of robots
- **Model:** Gaussian process at each cell with hyperparameters that are to be learnt online
- **Solution:** Generate sampling frequency for each cell based on Gaussian process regression, and then do receding horizon-style to deploy the team to meet the sampling frequency.
- **Guarantee:** If sampling requirements are met and hyperparameter update is ok, then we have guarantee of tolerance satisfaction.







Sample-quantile based stochastic control

ACC ’23, UrbanAI ’25 @ NeurIPS

Problem setting

- **Problem:** Solve a (possibly non-convex) separable, chance constraint program

$$\begin{array}{rl} \text{minimize} & f_0(z) \\ \text{subject to} & z \in \mathcal{Z}, \\ \forall i \in \mathbb{N}_{[1,N]}, & \mathbb{P}_{\boldsymbol{w}}\{f_i(z) + g_i(\boldsymbol{w}) \leq 0\} \geq \delta_i \end{array}$$

with decision variable $z \in \mathbb{R}^{n_z}$, random vector $\boldsymbol{w} \in \mathbb{R}^{n_w}$, known functions f_i, g_i and a set of deterministic constraints $\mathcal{Z}$ and known risk thresholds δ_i.

The probability measure $\mathbb{P}_w$ is unknown, but we have access to a sample generator for w

- **Main idea:** Use sample-based quantile functions to obtain high-quality approximations

- **Key features of the approach:** Apart from being able to handle non-convexity,
  - Approximation preserves any structural properties
  - Constraints in the approximation do not grow with number of samples
  - Less susceptible to outliers in the generated samples

Existing approaches for chance constrained programming

- Sample-based approaches
  - Assumes no prior knowledge of the stochastic uncertainty and less structural assumptions
  - Two groups based on how constraints are enforced in the approximation
    - All-sample approximations: Convex and non-convex programs
      - **Advantages:** Preserves structural properties leading to computational tractability
      - **Limitations:** Susceptible to outliers and typically more conservative
      - Campi and Garatti (2011); Vidyasagar (2015); Grammatico, Zhang, Margellos, Goulart, and Lygeros (2015); Campi, Garatti, and Ramponi (2018)
    - Fraction-sample approximations: Typically convex programs
      - **Advantages:** Less susceptible to outliers and typically less conservative
      - **Limitations:** Does not preserve structural properties resulting in prohibitively expensive computation
      - Blackmore, Ono, Bektassov, and Williams (2010); Sivaramakrishnan and Oishi (2020); Pena-Ordieres, Luedtke, and Wachter (2020)
- Sample-free approaches
  - Assumes knowledge of the stochasticity of the uncertainty and makes structural assumptions
  - Examples include quantile reformulation, moment-based, and Bernstein approximations
    - Nemirovski and Shapiro (2007); Ono, Blackmore, and Williams (2008); Prekopa (2013); Oldewurtel, Jones, Parisio, and Morari (2013); Paulson, Buehler, Braatz, and Mesbah (2020); Priore, Petersen, and Oishi (2022);

We retain the advantages of both sample-based approaches for non-convex, chance constrained programs with separable constraints.

Functions associated with random variables and DKW inequality

- Given a random variable $\boldsymbol{x}$
- *Cumulative distribution function* $\Phi_{\boldsymbol{x}}$, *Quantile function* $Q_{\boldsymbol{x}}$

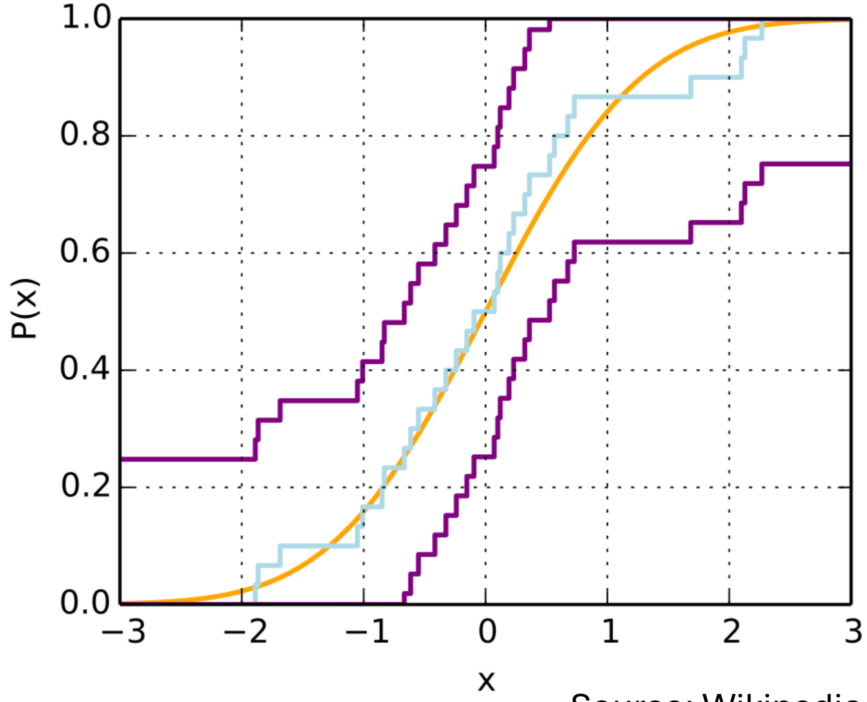
$$\Phi_{\boldsymbol{x}}(x) = \mathbb{P}_{\boldsymbol{x}}\{\boldsymbol{x} \leq x\},$$
$$Q_{\boldsymbol{x}}(p) = \inf\{x \in \mathbb{R} : \Phi_{\boldsymbol{x}}(x) \geq p\}$$

- Given M samples $\left\{x_{\mathrm{sample}}^{(j)}\right\}_{j=1}^{M}$, their <u>*empirical*</u> counterparts

$$\hat{\Phi}_{\boldsymbol{x}}^{M}(x) \triangleq \frac{1}{M}\sum_{j=1}^{M} \mathbb{I}\{x_{\mathrm{sample}}^{(j)} \leq x\},$$
$$\hat{Q}_{\boldsymbol{x}}^{M}(p) \triangleq \inf\{x \in \mathbb{R} : \hat{\Phi}_{\boldsymbol{x}}^{M}(x) \geq p\}.$$

Lemma 2 (DVORETZKY–KIEFER–WOLFOWITZ [16, SEC. 2.3.2]). *Let $\boldsymbol{x}$ have a continuous distribution $\Phi_{\boldsymbol{x}}$. Then, for any $\varepsilon > 0$ and $\hat{\Phi}_{\boldsymbol{x}}^{M}$ defined using M i.i.d. samples,* $\mathbb{P}\left\{\sup_{x\in\mathbb{R}} |\Phi_{\boldsymbol{x}}(x) - \hat{\Phi}_{\boldsymbol{x}}^{M}(x)| > \varepsilon\right\} \leq 2\exp(-2M\varepsilon^2)$.

DKW inequality provides finite-sample confidence bounds on the CDF



Source: Wikipedia

Orange: True cumulative distribution
Light blue: Empirical cumulative distribution
Purple: Confidence bounds

Key insight with a bit more details

$$\begin{aligned} \text{minimize} \quad & f_0(z) \\ \text{subject to} \quad & z \in \mathcal{Z}, \\ \forall i \in \mathbb{N}_{[1,N]}, \quad & \mathbb{P}_{\boldsymbol{w}}\{f_i(z) + g_i(\boldsymbol{w}) \leq 0\} \geq \delta_i \end{aligned}$$

M samples $\Longrightarrow$

$$(\text{Safe-Approx}) : \begin{array}{ll} \min. & f_0(z) \\ \text{s. t.} & z \in \mathcal{Z}, \\ \forall i \in \mathbb{N}_{[1,N]}, & f_i(z) + \hat{Q}^M_{\boldsymbol{g}_i}(\Delta_i^{\text{inner}}) \leq 0 \end{array}$$

$$(\text{Unsafe-Approx}) : \begin{array}{ll} \min. & f_0(z) \\ \text{s. t.} & z \in \mathcal{Z}, \\ \forall i \in \mathbb{N}_{[1,N]}, & f_i(z) + \hat{Q}^M_{\boldsymbol{g}_i}(\Delta_i^{\text{outer}}) \leq 0 \end{array}$$

- **Motivation:** Use (exact) quantile-based reformulation and replace δ_i with Δ_i

$$\begin{aligned} \text{minimize} \quad & f_0(z) \\ \text{subject to} \quad & z \in \mathcal{Z}, \\ \forall i \in \mathbb{N}_{[1,N]}, \quad & \mathbb{P}_{\boldsymbol{w}}\{f_i(z) + g_i(\boldsymbol{w}) \leq 0\} \geq \delta_i \end{aligned}$$

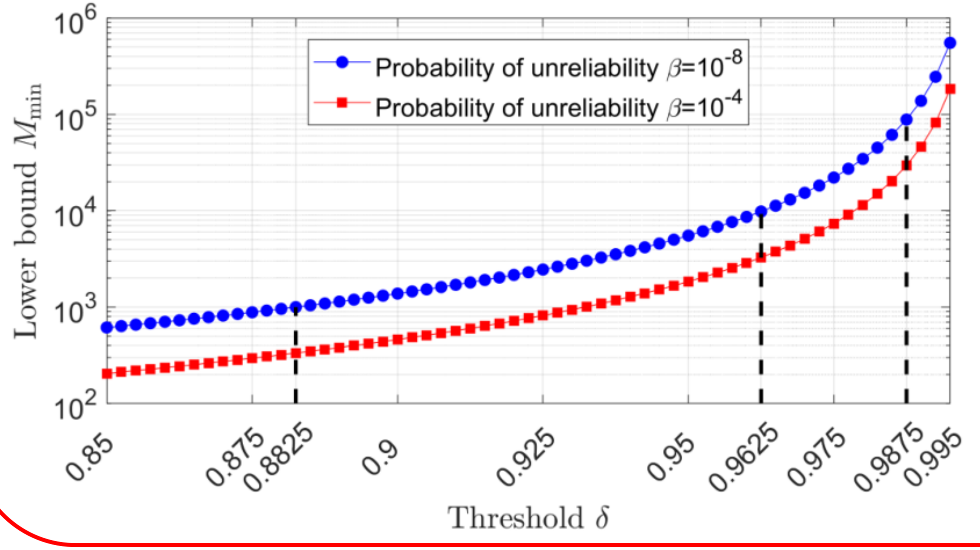
Quantile reformulation $\Longleftrightarrow$

$$\begin{aligned} \text{minimize} \quad & f_0(z) \\ \text{subject to} \quad & z \in \mathcal{Z}, \\ \forall i \in \mathbb{N}_{[1,N]}, \quad & f_i(z) + Q_{\boldsymbol{g}_i}(\delta_i) \leq 0 \end{aligned}$$

- **Key contribution:** Number of samples M and Δ_i^{inner} and Δ_i^{outer} as a function of number of constraints N, probability threshold δ, and probability of unreliability β

Some observations on the proposed sample-based approximation

Variation of $M_{\min}$ with δ for $N = 1$



Comparison with convexity case

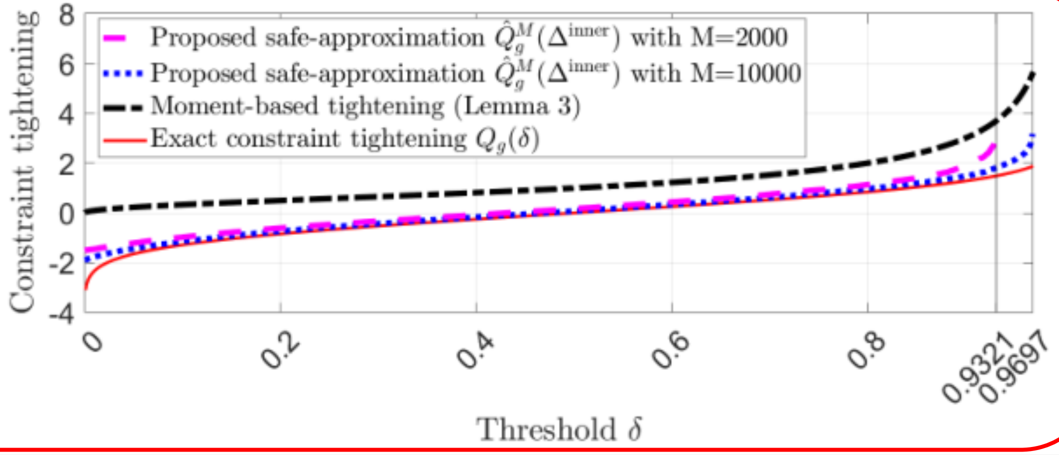
$$M_{\min}^{\text{convex}} = \frac{2}{1-\delta}(\ln(1/\beta) + n_z)$$

$$M_{\min} \triangleq \frac{\ln(N) + \ln(1/\beta)}{2(1-\delta)^2}$$

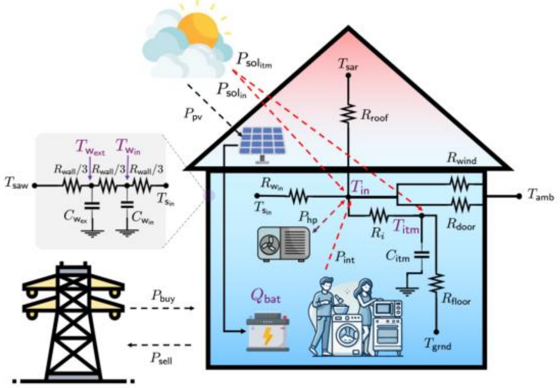
$M_{\min}^{\text{convex}}$ is from M. Campi and S. Garatti, 2011.
(setting the functions f_i and the set $\mathcal{Z}$ be convex).

Less conservative than Chebyshev-Cantelli

(Chebyshev-Cantelli bound) Let the random variable $g(\boldsymbol{w})$ have a finite mean $\mu_{\boldsymbol{g}}$ and standard deviation $\sigma_{\boldsymbol{g}}$. Then, for any $z \in \mathbb{R}^{n_z}$ and threshold $\delta \in (0,1)$, $f(z)+\mu_{\boldsymbol{g}}+\sigma_{\boldsymbol{g}}\sqrt{\delta/(1-\delta)} \leq 0 \Rightarrow \mathbb{P}_{\boldsymbol{w}}\{f(z)+g(\boldsymbol{w}) \leq 0\} \geq \delta$.

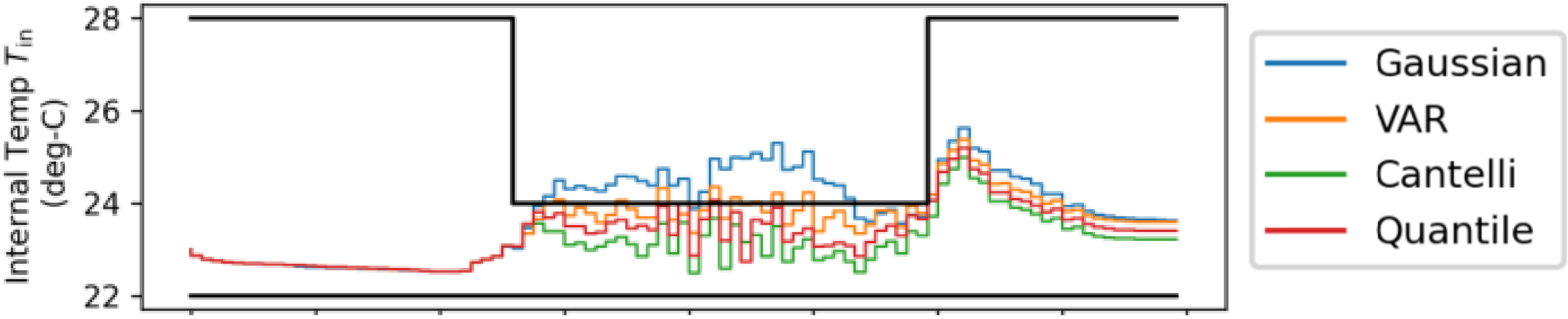


Results on a real disturbance dataset at the Mitsubishi Electric SUSTIE building



Algorithm	Winter / Jan 2023 TD	Winter / Jan 2023 Cost	Spring / April 2023 TD	Spring / April 2023 Cost	Summer / Jul 2023 TD	Summer / Jul 2023 Cost	Fall / Oct 2023 TD	Fall / Oct 2023 Cost
Gaussian	3.26	−16.01	9.66	6.02	21.14	19.48	16.86	14.60
VAR	1.04	−16.90	1.73	−2.39	2.03	10.15	2.90	6.02
Cantelli	**0.10**	−13.98	**0.02**	−1.97	**0.06**	9.78	**0.17**	5.16
Quantile	0.47	**−17.88**	0.43	**−3.33**	0.66	**8.97**	0.36	**3.49**

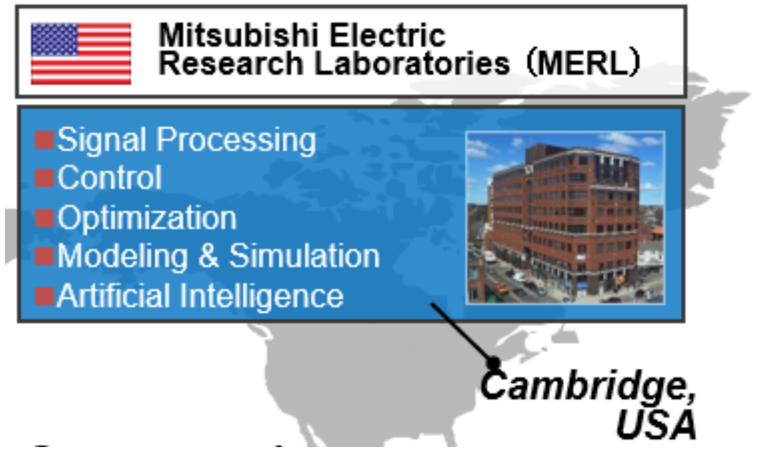
Table 1: Simulation results and baseline algorithms across seasonal test weeks. The best in each column is set in **bold** and the second best is underlined. Thermal discomfort (TD) is in units of deg-C·hr and cost is in scaled currency units. A negative cost implies no grid interaction; i.e. self-sustained operation.



VAR refers to Vector Autoregressive Model, and not Value-at-Risk

Background on MERL

We are a research arm of our parent company, Mitsubishi Electric Corporation (MELCO)



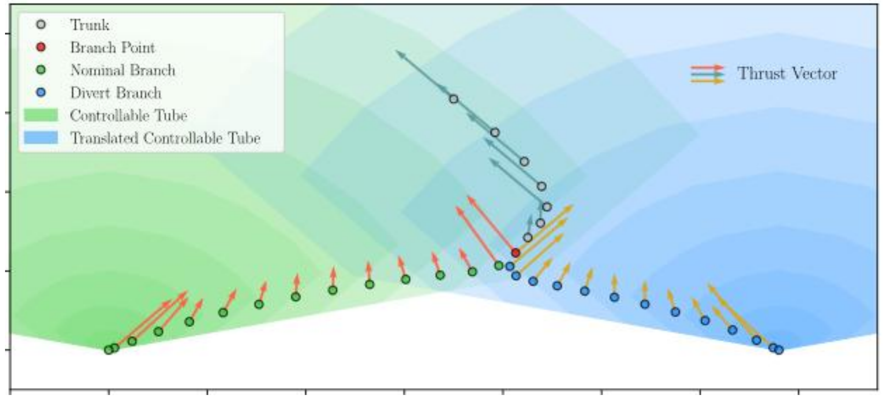
- MERL: An open industrial research lab with 30+ year history!
- Mid/long-term research in areas beneficial to MELCO
- 57 leading researchers including 5 Fellows (4 IEEE, 1 Optica)
- >150 pubs/year publications in leading conferences and journals

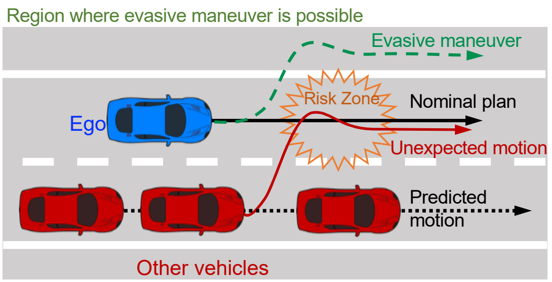
Internship opportunities

Research interests + other topics that I didn't talk about today

Intersection of *constrained control*, *multi-agent systems, learning*, and *perception* for robotics, space, and infrastructure

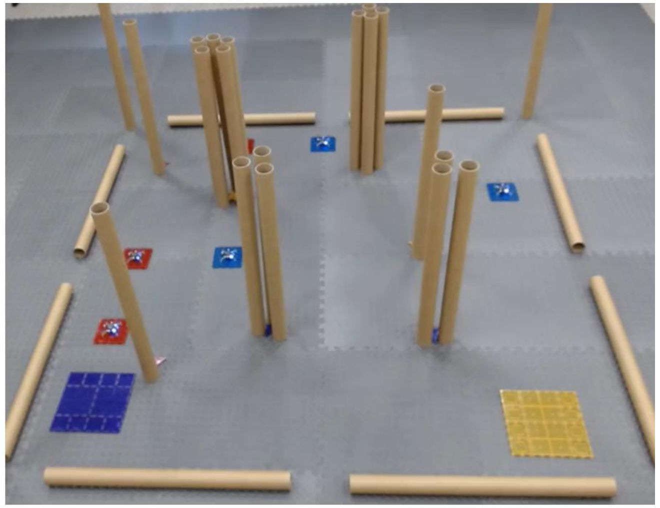
Set-based control and analysis



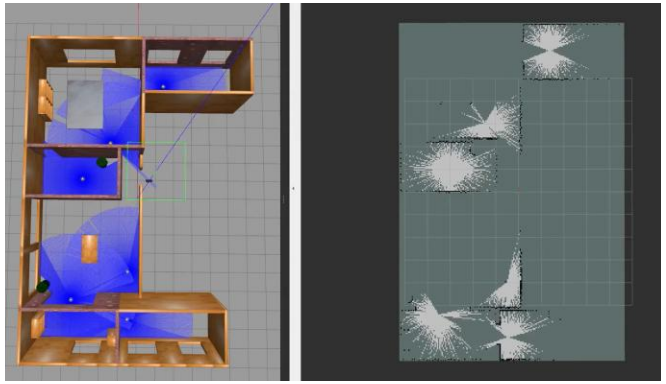


pycvxset (think MPT3+ET+CORA in Python)

Safe multi-agent trajectory design with filtered RL



Autonomous map building with energy constrained robots



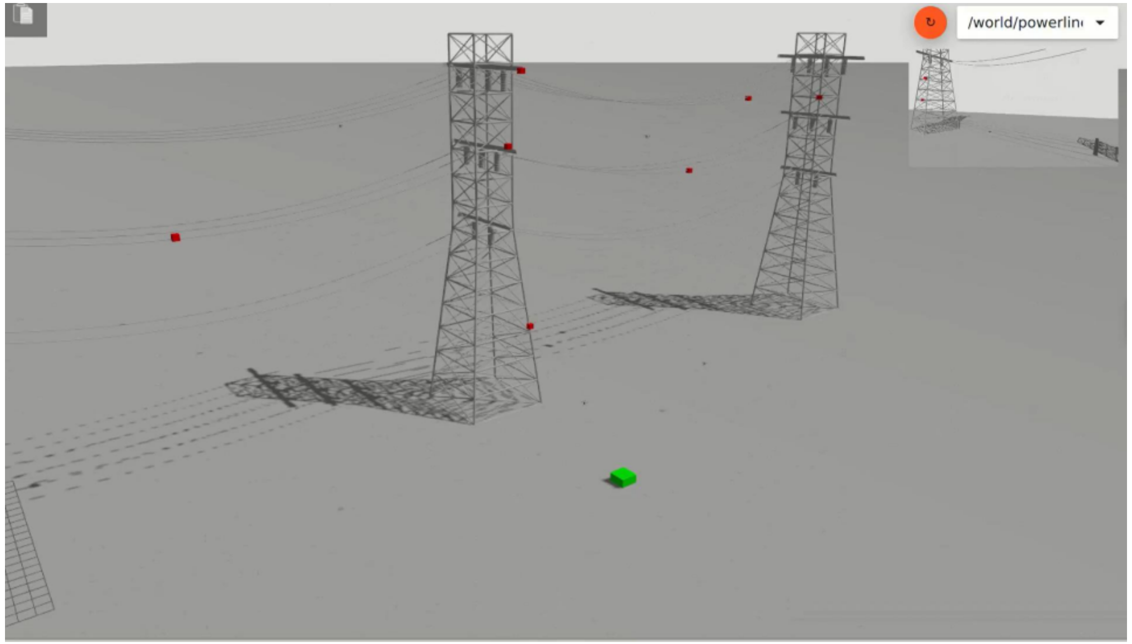
Email: *vinod@merl.com / abraham.p.vinod@ieee.org*

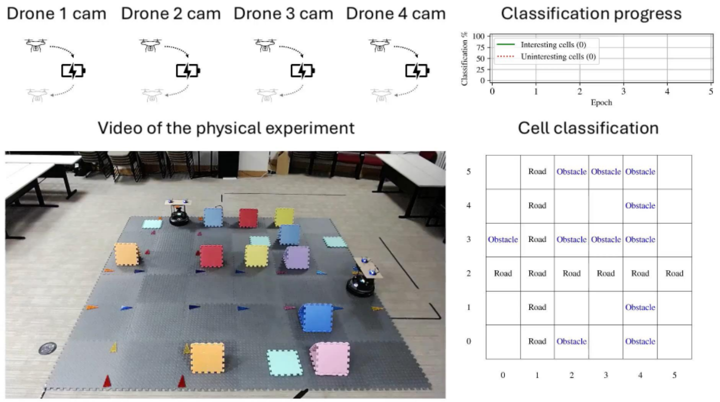
Key takeaway

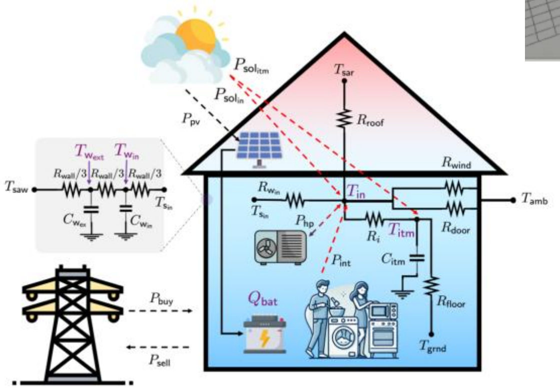
Reinforce that modular approach of control techniques + data-driven reasoning

→ Finite-sample performance guarantees and safety

→ Opens up several real-world applications







Come to my poster on Thursday B04.6!

Internship opportunities

Email: vinod@merl.com / abraham.p.vinod@ieee.org

MITSUBISHI
ELECTRIC

Changes for the Better

MITSUBISHI ELECTRIC
RESEARCH LABORATORIES