

NEURAL INFORMATION PROCESSING SYSTEMS

40th Annual Conference on Neural Information Processing Systems

Act, Validate, Adapt: Closing the Causal Discovery-Control Loop

Taqiya Ehsan, Shuren Xia, Jorge Ortiz

Electrical & Computer Engineering, Rutgers University

Atlanta, GA

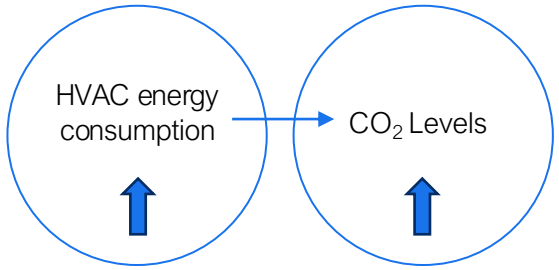
December 2026

RUTGERS UNIVERSITY

Correlation vs Causation

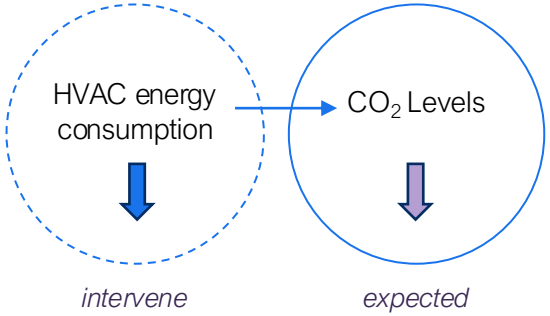
Building Management System (BMS)

1 WHAT IT SAW



always rise together

2 WHAT IT DID

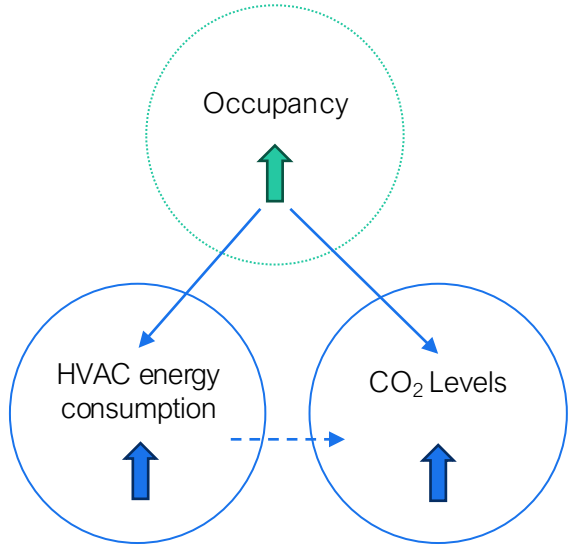


✓ HVAC reduced

✗ CO_2 is unchanged

✗ Occupants suffer

3 WHAT WAS TRUE



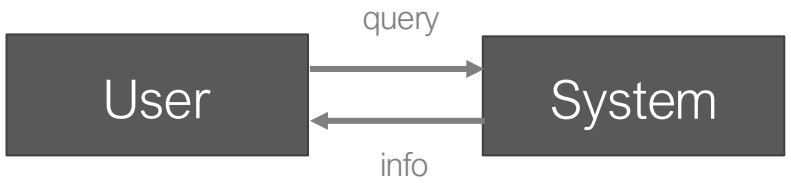
people drive both

The controller's own actions can disprove its model, but it never does.

Every action is already an experiment

Prediction System

LLM, music recommender, weather forecaster, radiology classifier

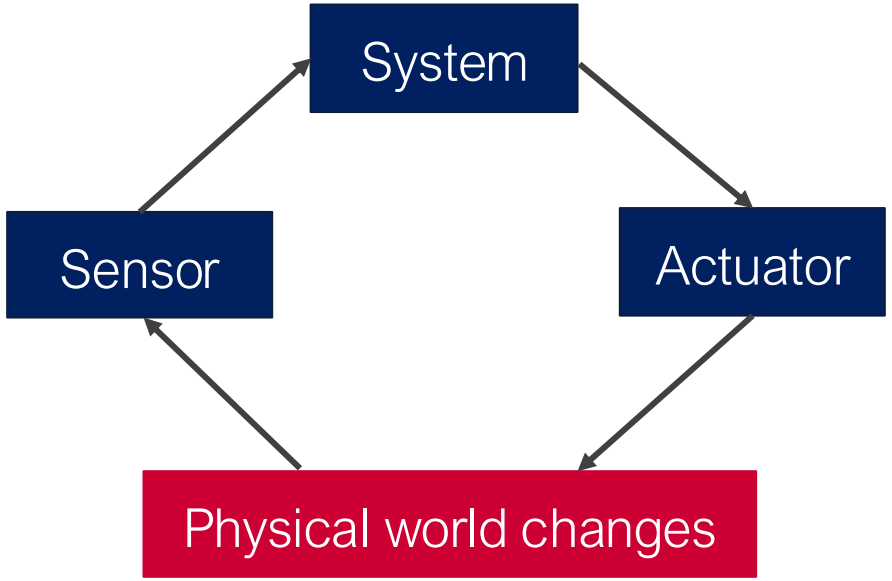


Physical world unchanged

Action semantics: prediction/recommendation

Control System

smart HVAC, autonomous vehicle, surgical robot, insulin pump

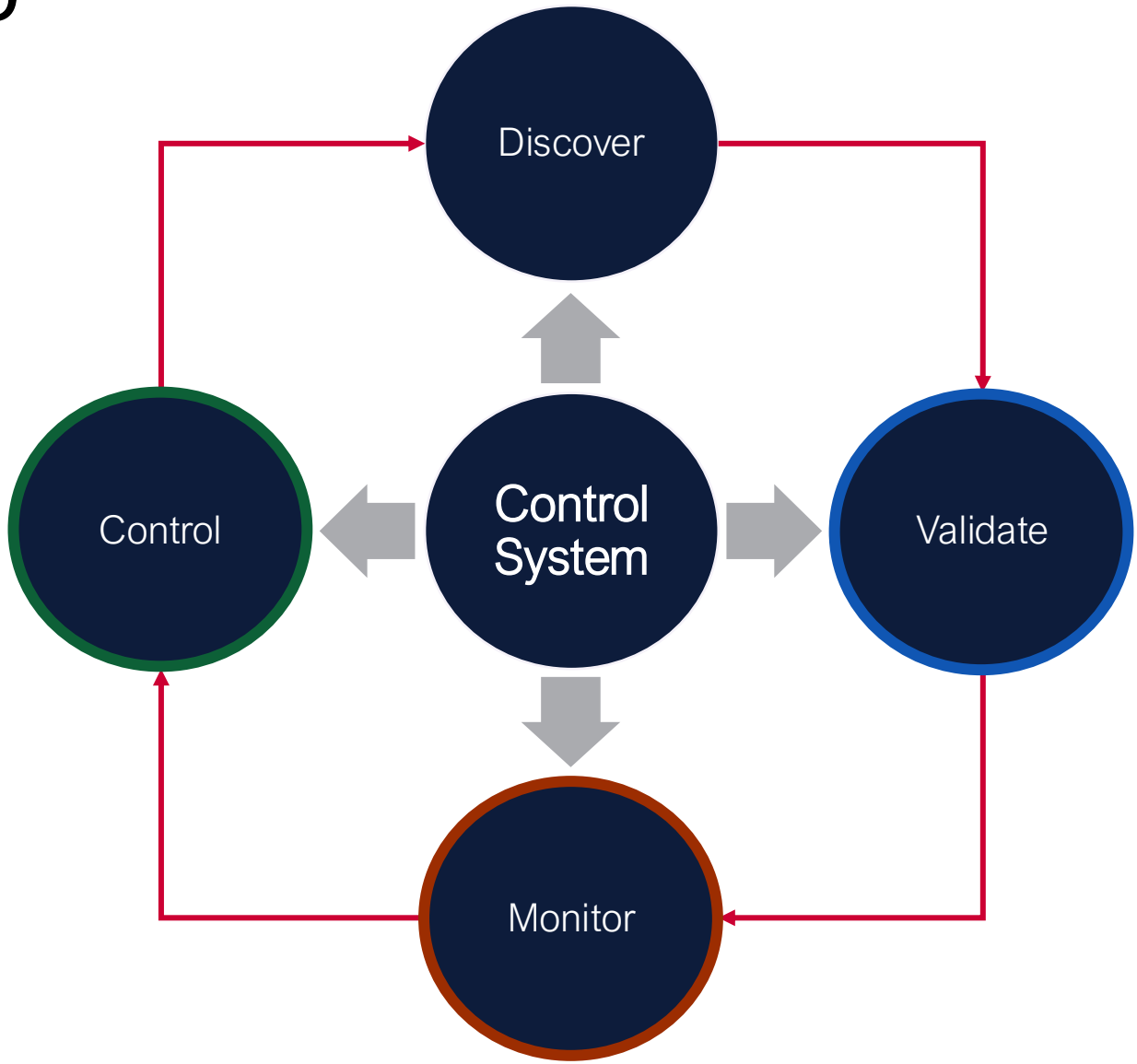


Action semantics: intervention on physical variables

A controller can test its model with the actions it already performs.

The Act-Validate-Adapt Loop

1. Can a controller test its own model by acting?
2. Does a tested model control better?
3. Can it tell which hidden condition changed?



PolicyGRID:

$$a_t^* = \operatorname{argmax}_a \; \boldsymbol{w_s(\varepsilon)}\, \mathrm{Sat}\left(\mathbf{a};\, \widehat{\mathcal{G}}, \boldsymbol{\beta}^{\hat{r}_t}\right) + \boldsymbol{w_e(\varepsilon)} \left(1 - \mathrm{Eng}\left(\mathbf{a};\, \widehat{\mathcal{G}}, \boldsymbol{\beta}^{\hat{r}_t}\right)\right)$$

where,

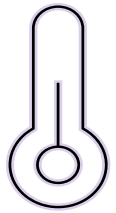
$$\widehat{\mathcal{G}} = \left\{ i \to j \;\in \mathcal{G}_{\cup}\text{: } t-\text{test on } \Delta_{ij} \text{ rejects } 0 \right\} \cup \mathcal{G}_{rescue}$$

$$\Delta_{ij} = \mathbb{E}\left[V_j \,\middle|\, do(V_i = v_{max})\right] - \mathbb{E}\left[V_j \,\middle|\, do(V_i = 0)\right]$$

$$\hat{r}_t = \operatorname{argmax}_r \, P(r \,|y_{1:t}) \; (posterior\ from\ model's\ prediction\ error)$$

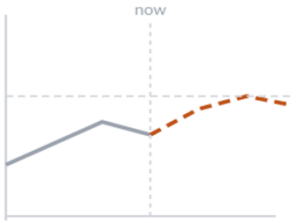
DATA-DRIVEN CONTROL

Rule-Based



fixed setpoints

MPC/DDPC



predict + optimize

Neural Nets

learn + optimize

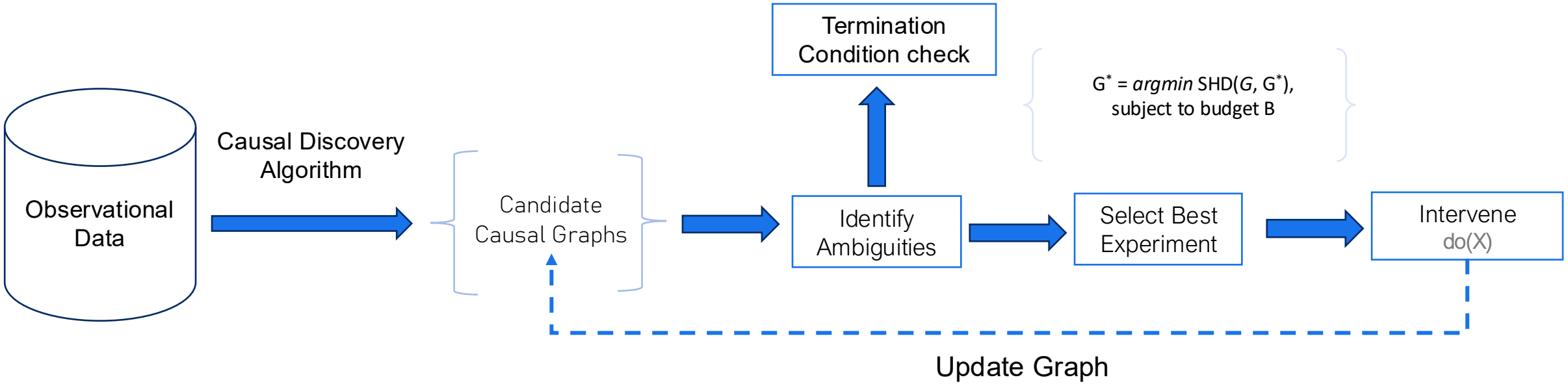
- ✓ Multi-objective optimization
- ✓ Real-world deployment
- ✗ Learns only from observational logs
- ✗ Cannot distinguish causation from spurious correlation
- ✗ Fails silently under distribution shifts



$$\pi^* = \textit{argmin} \; \Sigma \, J(v_t, a_t)$$

such that $v_{\{t+1\}} = \mathbf{f}(v_t, a_t)$

ACTIVE CAUSAL DISCOVERY



Examples:

GIES
Hauser & Bühlmann 2012

DCDI
Brouillard et al. 2020

ABCI
Toth et al. 2022

CAASL
Annadani et al. 2024

✓ Runs interventions

✗ Stale graph with distribution shifts

✗ No multi-objective optimization

✗ No real-world deployment

Positioning **PolicyGRID**

Method	Interventional Data	Edge Validation	Regime Detection	Multi-objective Control	Physical Evaluation
Passive discovery (PC, FCI)	✗	✗	✗	✗	✗
Active discovery (GIES, CAASL, DCDI)	✓	✗	✗	✗	✗
Regime-change discovery (FANTOM)	✗	✗	✓	✗	✗
Meta-causal agent (Zhao et al.)	✓	✗	✓	✗	✗
Online causal RL (Cai et al.)	✓	✓	✗	✗	✗
Causal model-based RL (C-MBPO)	✗	✗	✗	✓	✗
Data-driven control (MPC, DDPC)	✗	✗	✗	✓	✓

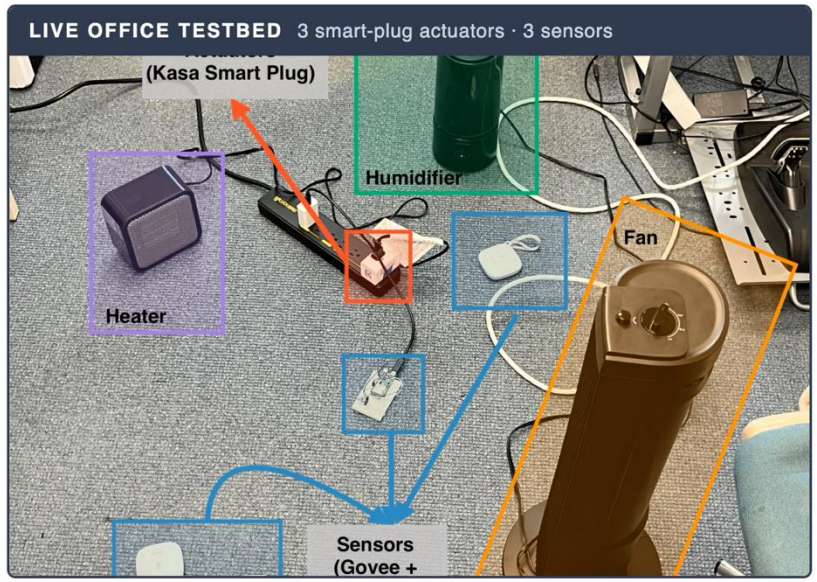
Positioning **PolicyGRID**

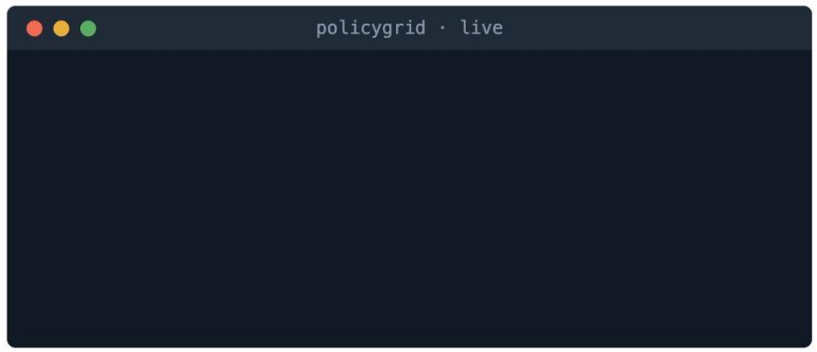
Method	Interventional Data	Edge Validation	Regime Detection	Multi-objective Control	Physical Evaluation
Passive discovery (PC, FCI)	✗	✗	✗	✗	✗
Active discovery (GIES, CAASL, DCDI)	✓	✗	✗	✗	✗
Regime-change discovery (FANTOM)	✗	✗	✓	✗	✗
Meta-causal agent (Zhao et al.)	✓	✗	✓	✗	✗
Online causal RL (Cai et al.)	✓	✓	✗	✗	✗
Causal model-based RL (C-MBPO)	✗	✗	✗	✓	✗
Data-driven control (MPC, DDPC)	✗	✗	✗	✓	✓
PolicyGRID	✓	✓	✓	✓	✓

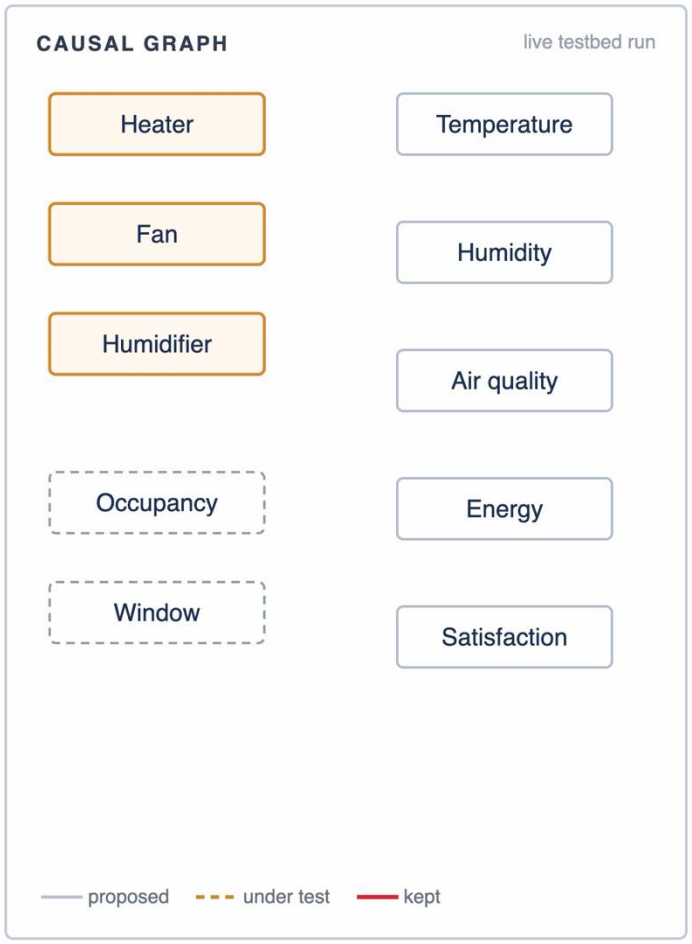
PolicyGRID in action

PolicyGRID Demo Link

PHASE 1 Hypothesis generation | PHASE 2 Intervention & validation | PHASE 3 Regime monitoring | PHASE 4 Policy control

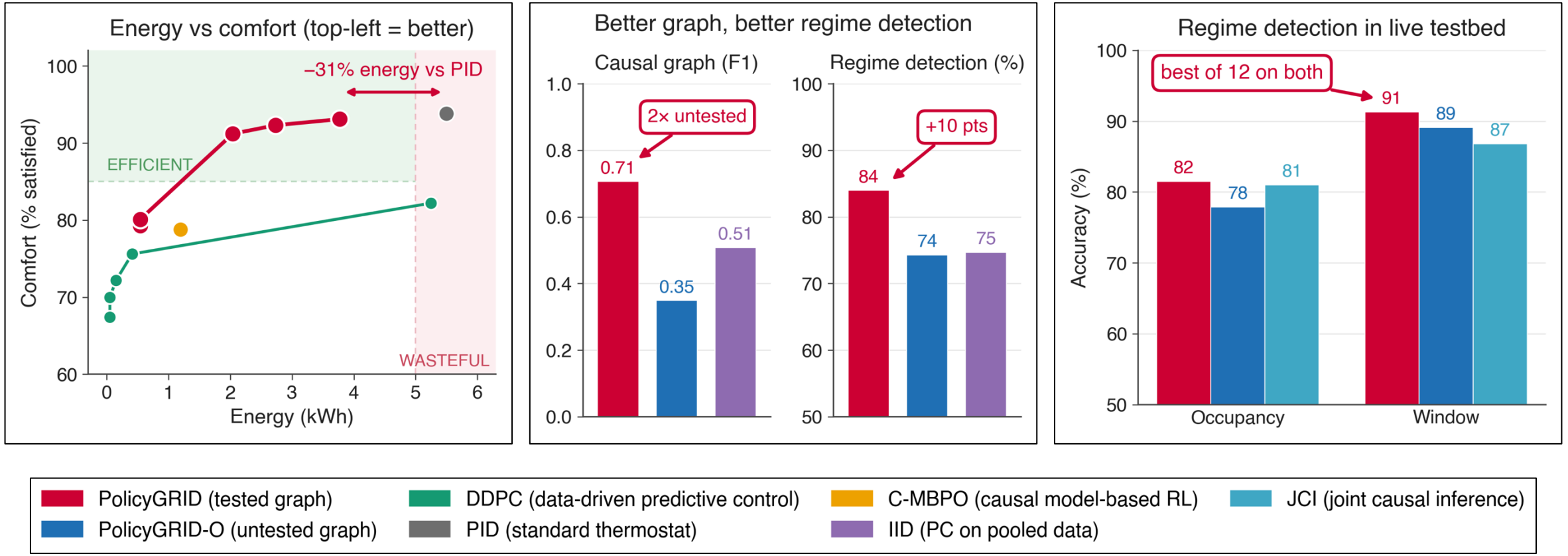






PolicyGRID closes the causal control loop, end-to-end

Closed-Loop **Results**



One tested graph, three payoffs: **better energy–comfort trade-offs**, more accurate regime shift detection, and same loop transfers to real hardware.

Key Takeaways

1. Can a controller test its own model by acting?

   Yes! Every actuation is an experiment — a controller can test its own model by acting.

2. Does a tested model control better?

   Yes! A tested model controls better — **it traces the energy–comfort trade-off, with 21% less energy at the same comfort target.**

3. Can it tell which hidden condition changed?

   Yes! It can tell which latent condition changed — right 81% of the time vs, **61% without testing.**



Taqiya Ehsan

PhD Candidate, Rutgers University

taqiya.ehsan@rutgers.edu

Poster# 153013

Thu, Dec. 10, 2026
4:30 PM – 7:30 PM EST
Hall C1

Project Page