



# Challenges in Designing Control Strategies for Buildings

**Omid Ardakanian**

Assistant Professor, Department of Computing Science

December 9, 2022



**UNIVERSITY  
OF ALBERTA**

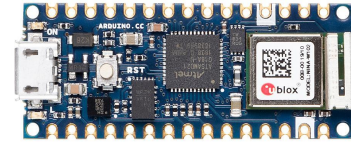
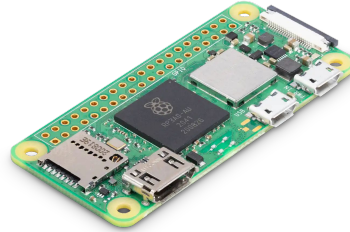
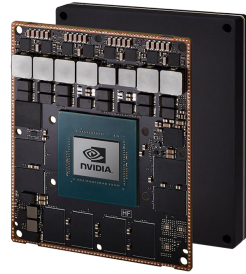
# Commercial buildings house thousands of sensors

They are large-scale distributed systems





# Processors and MCUs are prevalent in buildings



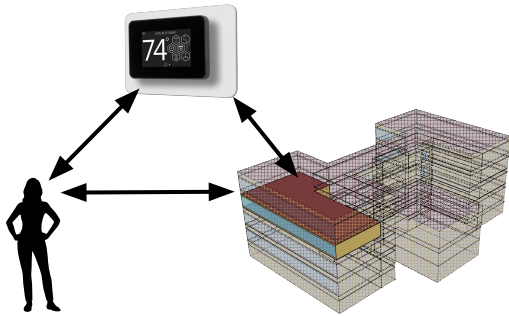
# The gap between the vision of smart buildings and the present reality

- **two thirds** of occupants are not comfortable
- buildings account for around **one third** of the final energy consumption
  - heating and cooling systems are controlled based on fixed schedules and maximum occupancy assumption
  - faults and control errors cannot be easily identified and fixed

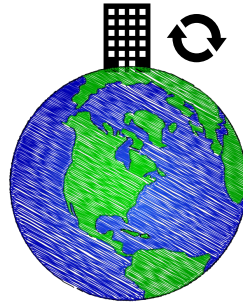


# Why is this gap so large?

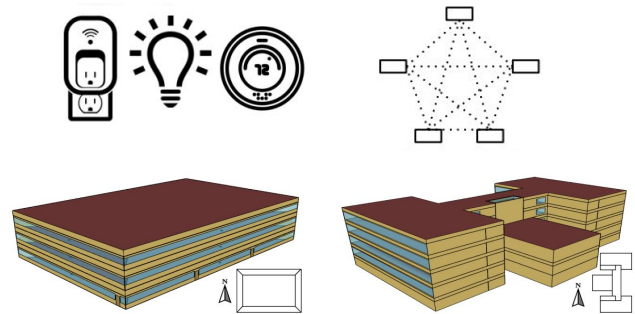
measurement challenges



modeling challenges



generalization challenges

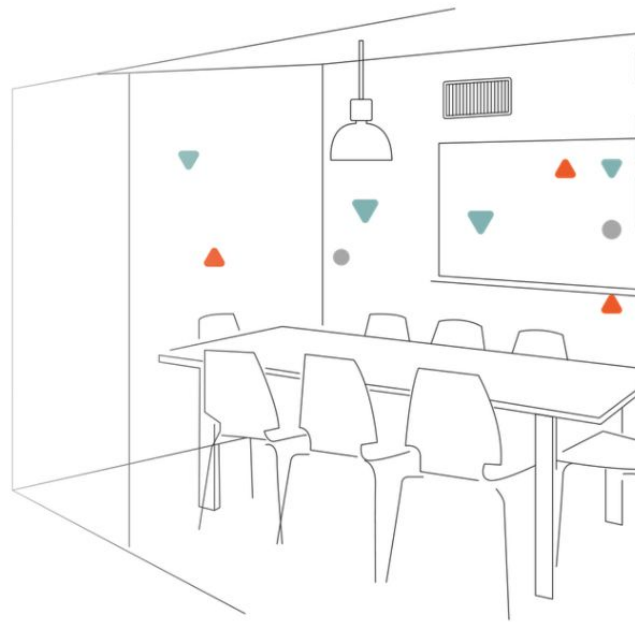


## **Close the loop among data generation, computation, and control in buildings**

- to reduce cost and emissions, and increase adaptability and reliability
- to improve human well-being, comfort, and performance

# Smart building research in my lab

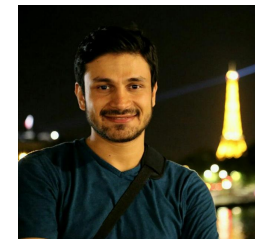
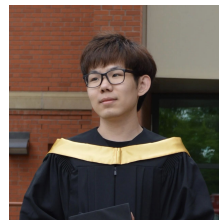
- addressing the measurement challenge
  - infrastructure-mediated sensing
  - higher-order and virtual sensors  
(control command  $\sim$  sensor data)
- addressing the modeling challenge
  - system identification
- addressing the generalization challenge
  - diversity-induced reinforcement learning
  - efficient policy evaluation





# Outline

- addressing the measurement challenge
- addressing the modeling challenge
- addressing the generalization challenge

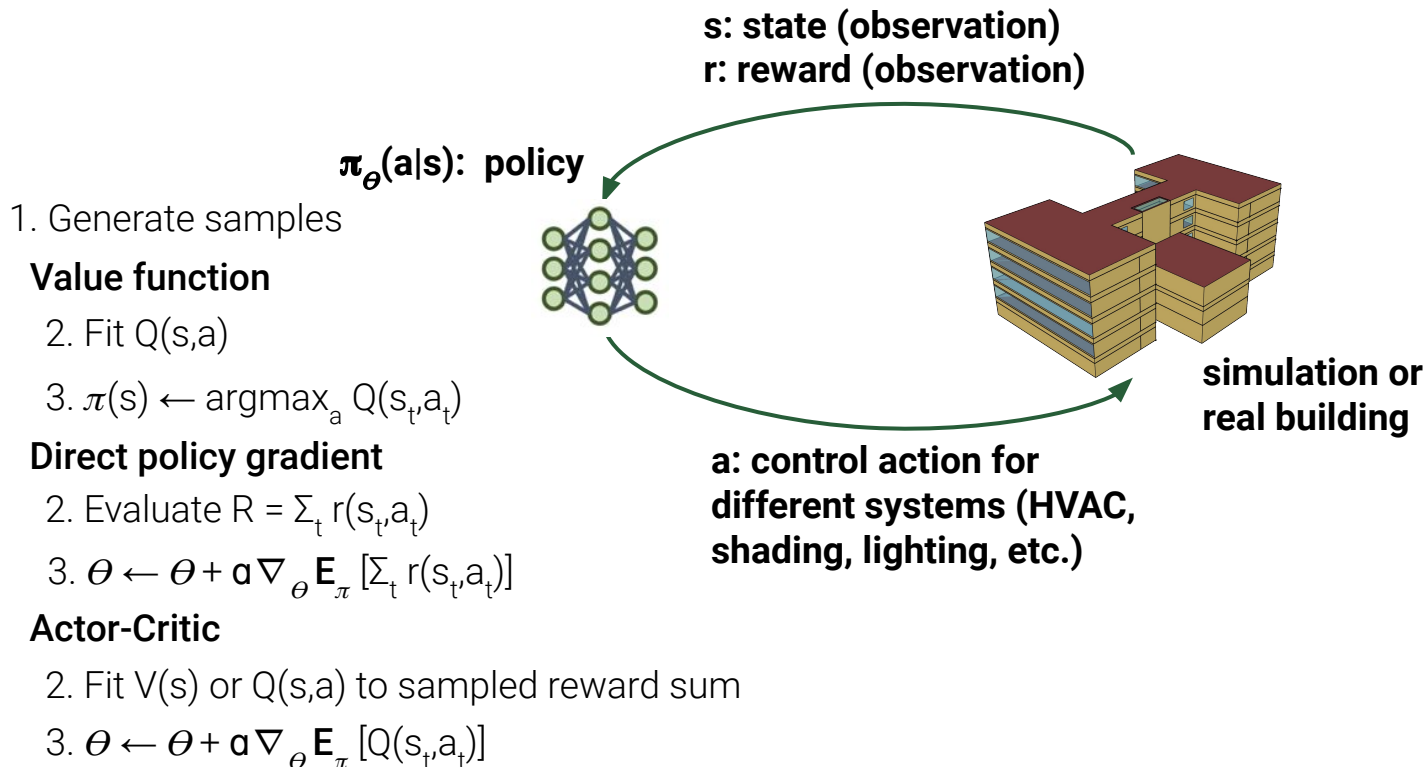




**How to learn control policies that can be used in every building?**

# Learning-based control of building systems

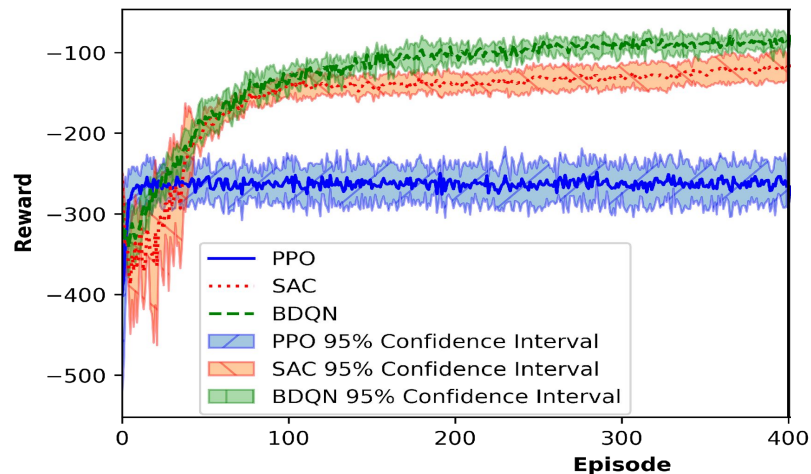
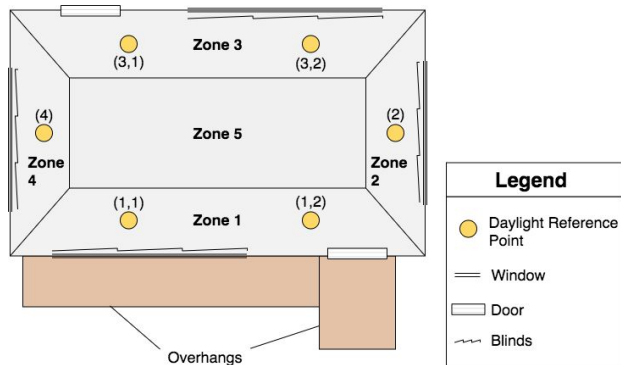
## Basic idea



# Practical challenges

- a large amount of data (>200 months) is required to learn a high-quality control policy despite using sample-efficient RL algorithms

floor plan of a 5-zone building  
simulated in EnergyPlus & COBS



Tianyu Zhang, Gaby Baasch, Omid Ardakanian, Ralph Evins, "On the Joint Control of Multiple Building Systems with Reinforcement Learning", In Proceedings of the 12th ACM International Conference on Future Energy Systems (ACM e-Energy), pp. 60-72, 2021.



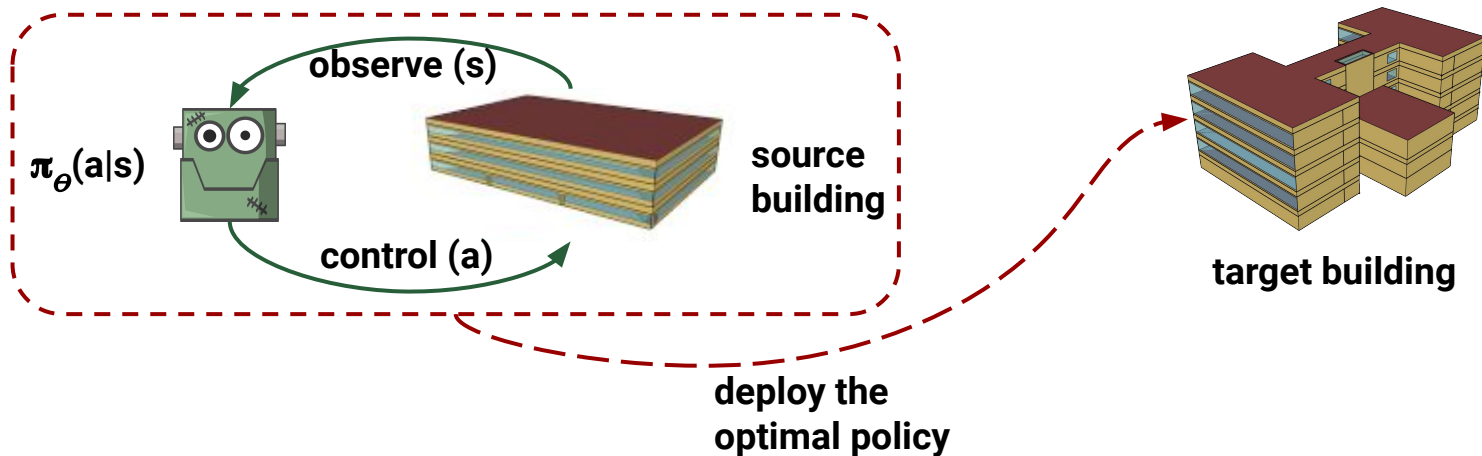
# Practical challenges

- if RL agent is trained in the real building, comfort violation and significant energy consumption are likely to happen in the early stage of training
- training the RL agent by interacting with simulation building has its own limitation
  - simulation and real buildings can be quite different
- how to reduce the training cost of RL?  
**train on a prototype building or controlled environment, then transfer to the target building and retrain**

# How to reduce training cost of RL agents?

Transfer learning can help

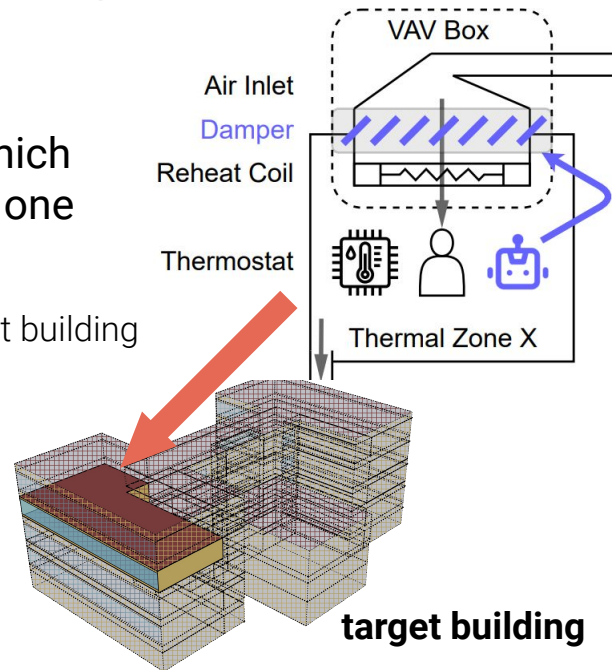
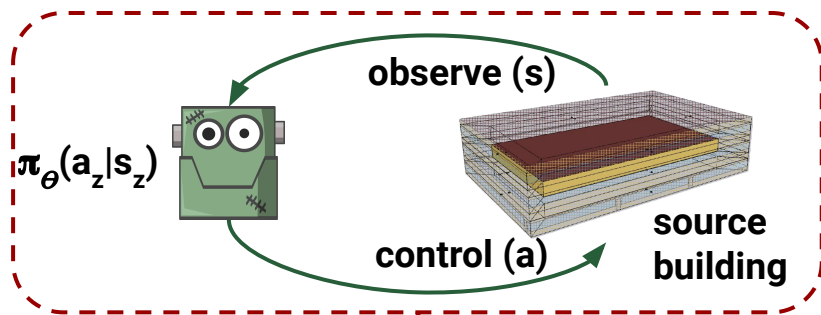
- train RL agents on the source building and deploy to the target building (**novel** environment)
  - commercial buildings are custom built so the control problem is solved for different MDPs



# How to reduce training cost of RL agents?

Transfer learning can help

- use the **multi-agent reinforcement learning** framework in which each agent is responsible for controlling the environment of one zone
  - the optimal policy in the source building may perform poorly in the target building

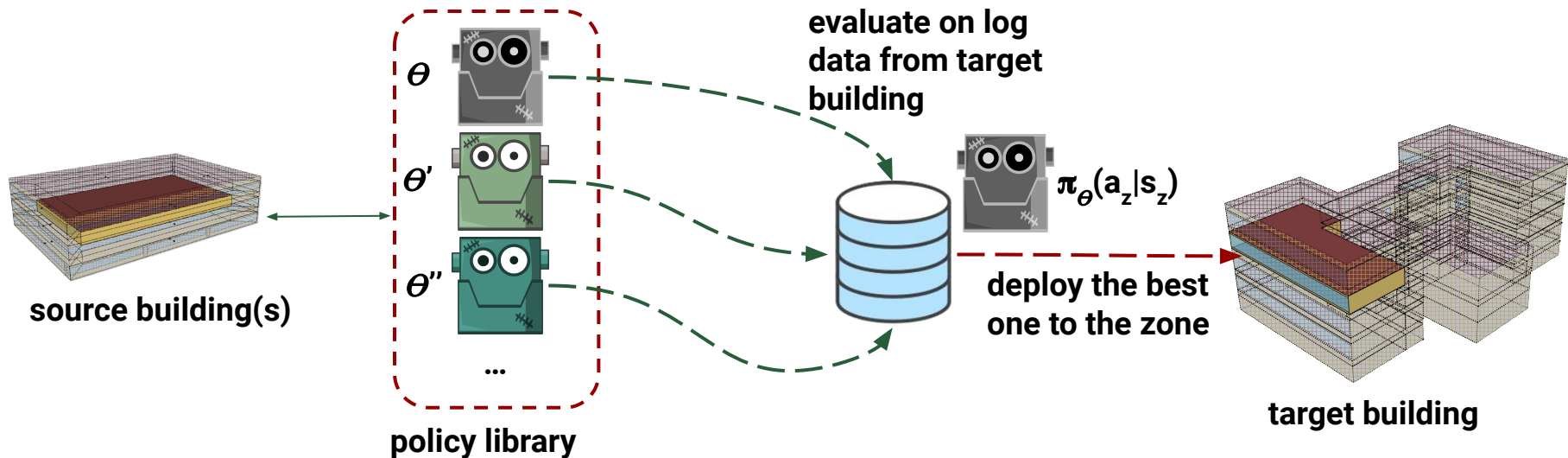


deploy the optimal policy to the corresponding zone

# How to reduce training cost of RL agents?

## Diversity for transfer learning

- train a **diverse** population of RL agents and assign the **most suitable** one to each zone





# Policy Diversity and Evaluation

- why is diversity useful for transfer learning?
- how is it defined?
- how to efficiently identify the best policy given a diverse population for each zone?


# Intuition

- the skill learned for one task may not be useful in another, but having a diverse set of skills could be helpful in a new task
- “skills” are policies in the **policy library** and “task” is the optimal control problem

# Different kinds of diversity

- augmenting the loss function of a policy gradient RL algorithm (**policy diversity**)
  - the policy must be different from previously learned policy in addition to maximizing the expected return

Diversity weight

$$\mathcal{L} = \mathcal{L}_{RL} + [w] \mathcal{L}_{diversity}$$


- learning policies on multiple environments (**environment diversity**)
  - the more environments we see during training, the lower would be our uncertainty
  - policies learned through interaction with these environments will be added to the policy library

# Defining a measure of diversity

- policy  $\pi$  is considered different from  $\pi'$  if for every  $(s, a)$  visited in a trajectory taken under  $\pi$ , probabilities  $\pi(a|s)$  and  $\pi'(a|s)$  are sufficiently different but the two policies agree about the value of state  $s$ 
  - we can use  $\pi(a|s)/\pi'(a|s)$  or  $\pi'(a|s)/\pi(a|s)$  (whichever is greater), to measure the difference between the two policies for a given tuple  $(s, a)$
  - this ratio can be unbounded, **so we clip it**
- since several  $(s, a)$  pairs are visited in a trajectory taken under  $\pi$ , we use the sum of these clipped ratios as a measure of diversity
- how to make sure a policy is different from  $N$  policies that were previously learned?



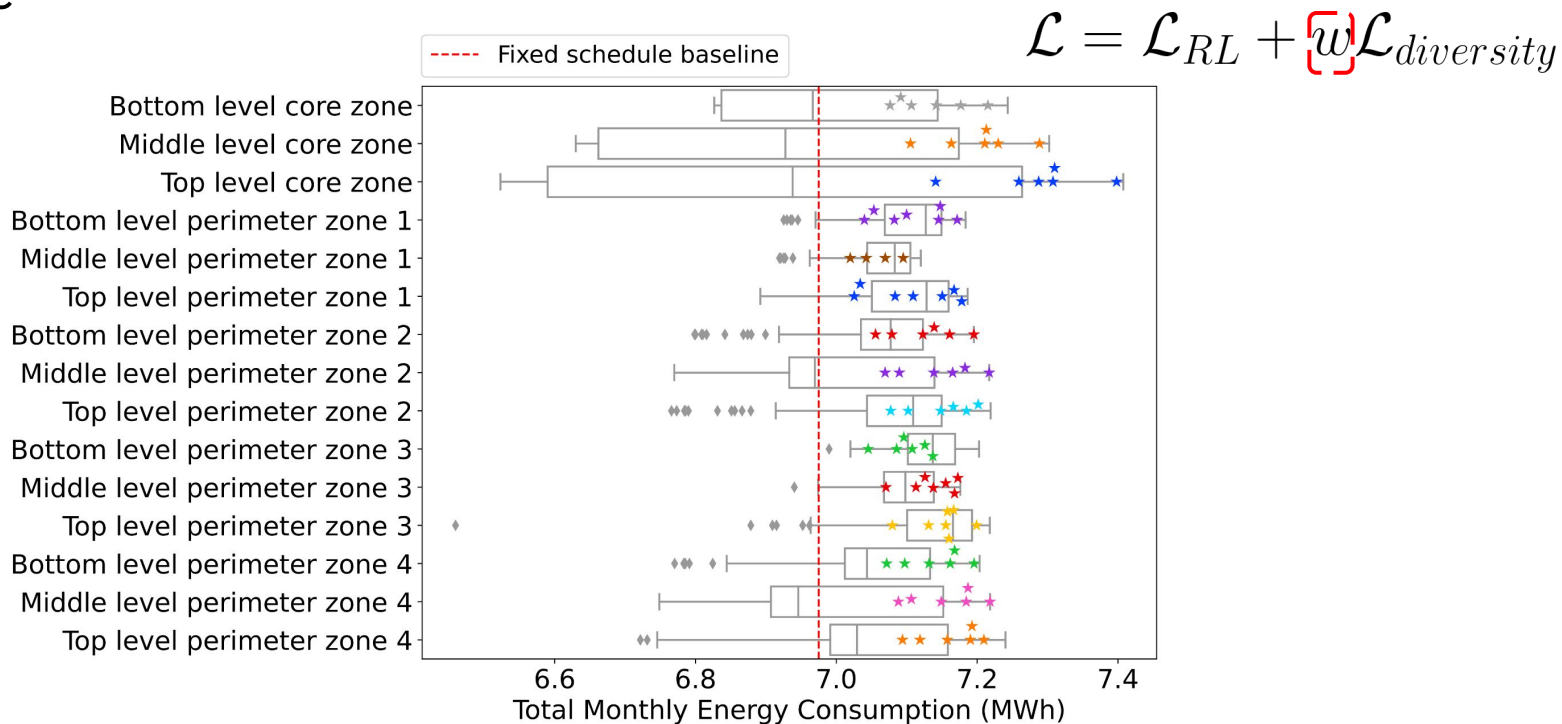
# Diversity loss

Set of previously learned policies

$$\mathcal{L}_{diversity} = \frac{\sum_{\pi' \in \Pi_{learned}} \sum_{(s,a) \in \text{exp}} \frac{\max\left(\frac{\pi(a|s), \pi'(a|s)}{\min(\pi(a|s), \pi'(a|s))}, \bar{\rho}\right)}{|G^{\text{exp}}(s) - V^{\pi'}(s)|}}{|\Pi_{learned}|}$$

# Diversity is useful in transfer learning

## Evidence



# Policy evaluation/selection methods

- **brute force**: run every policy in the library on each zone of the target building and use energy use and comfort under this policy as evaluation metrics
  - too expensive
- **off-policy evaluation** (OPE): evaluate the performance of policies learned in a training environment without online interaction with the target environment
  - IPW and SNIPW use importance and rejection sampling to reweight the rewards according to the **log data** obtained from the target environment
  - Gaussian Kernel (GK) uses kernel density estimation instead of rejection sampling for continuous actions

# Policy evaluation/selection methods

- **zero-cost proxy** (ZCP) is studied in the domain of neural architecture search (NAS) to rank different architectures at initialization (i.e., before training)
  - gradnorm (GN) and single-shot network pruning (SNIP) compute the loss and its gradient for a minibatch of data
- since policy is represented by a neural network, ZCP can be used to rank policies!



# Contributions

- we use the proposed methodology to build the policy library and identify the best candidate policy for each zone of a novel building
- we show that the selected policies together save **up to 30.4%** on energy consumption in the unseen building during a winter month, **without sacrificing thermal comfort**

# MARL

## State as perceived by each agent

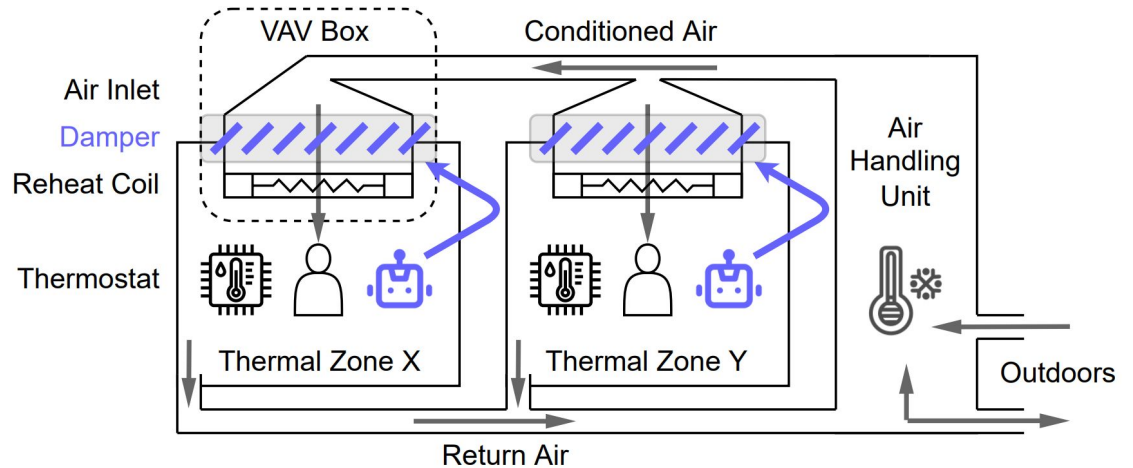
- zone temp.
- zone humidity
- zone occupancy
- outdoor temp.
- solar radiation
- hour of the day

## Action of each agent

- VAV minimum damper position

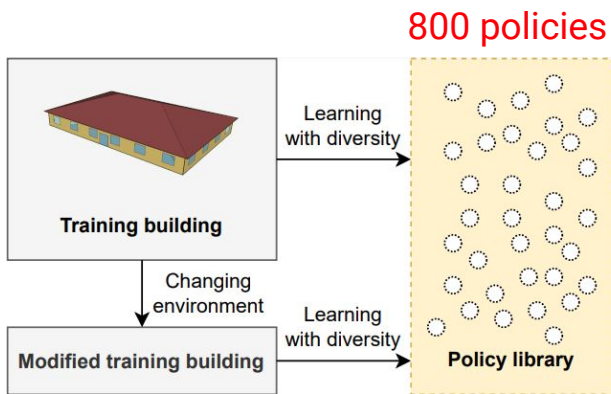
## Reward provided to each agent

- energy use of the respective VAV system



# The big picture

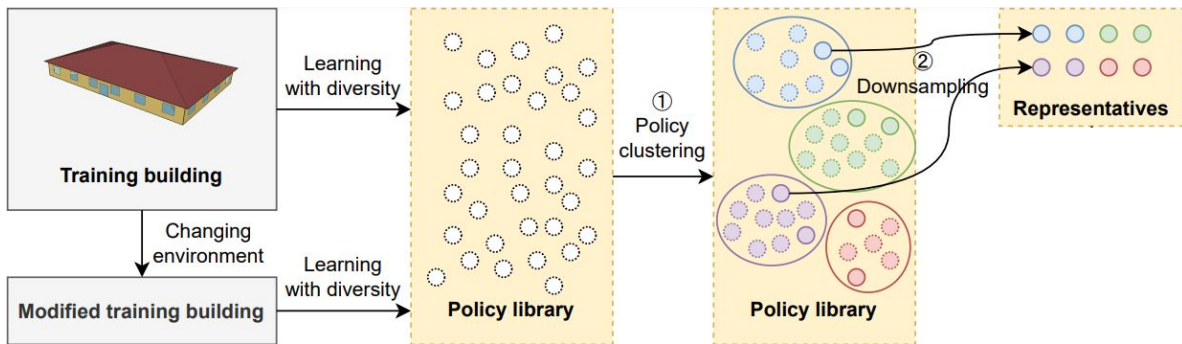
- PPO is selected for training the control agents



each policy is represented  
in an M-dimensional space

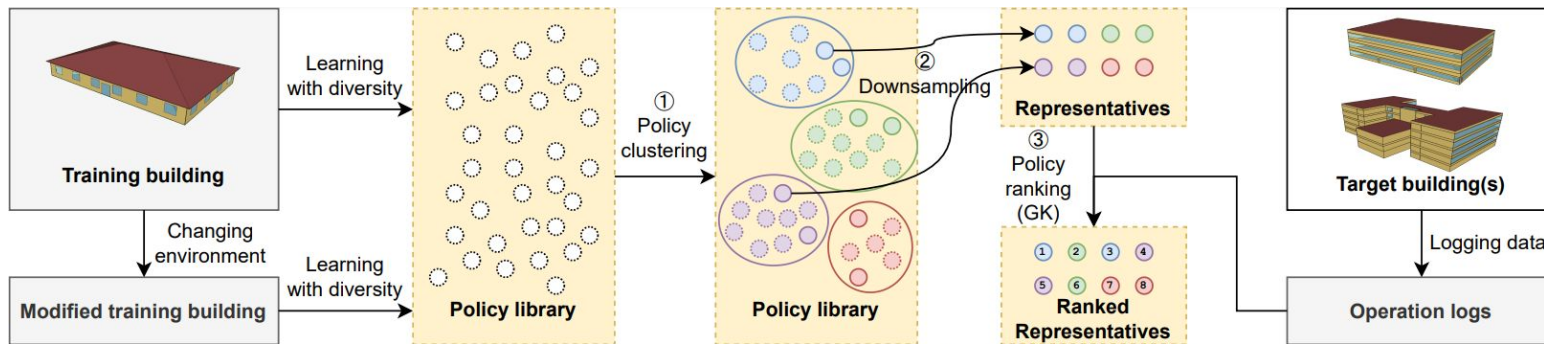
# The big picture

- PPO is selected for training the control agents



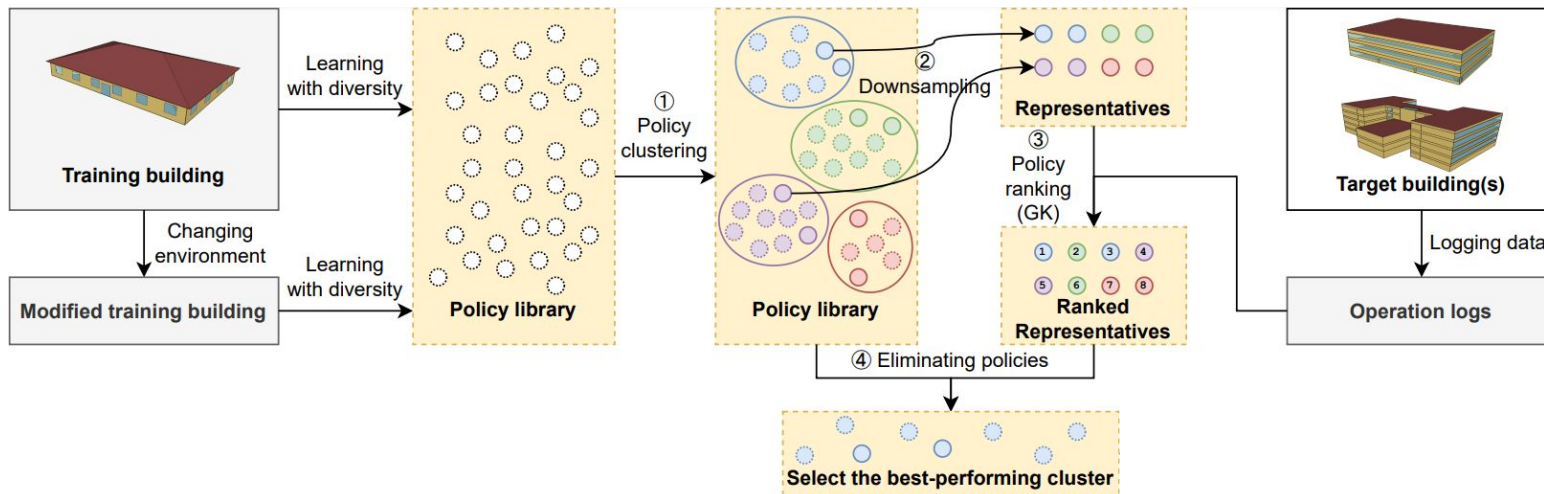
# The big picture

- PPO is selected for training the control agents



# The big picture

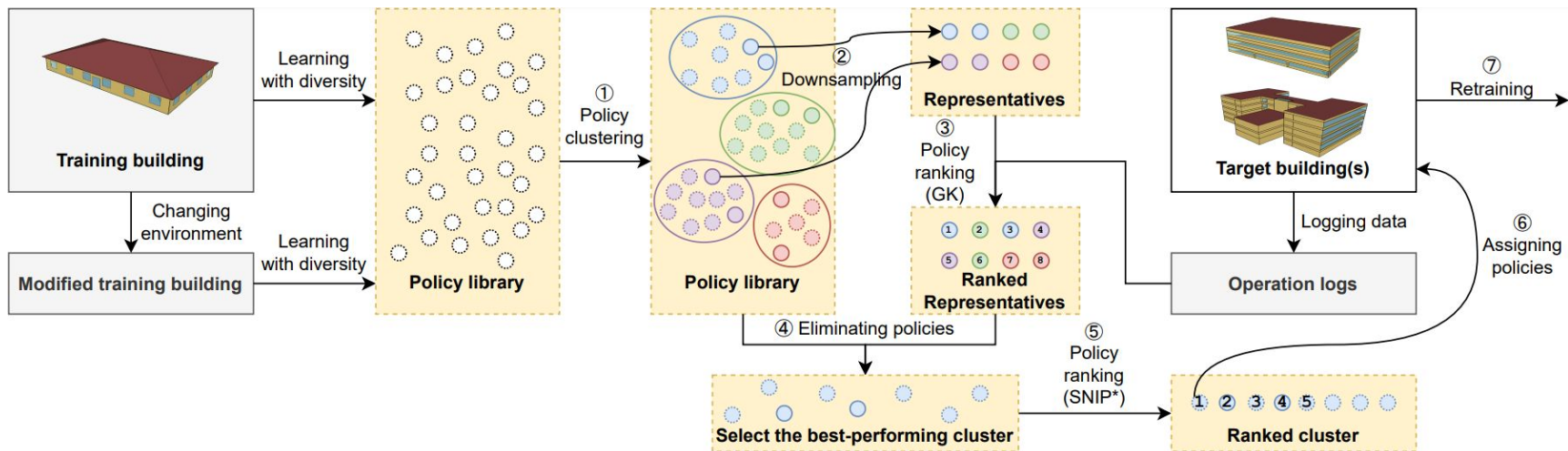
- PPO is selected for training the control agents



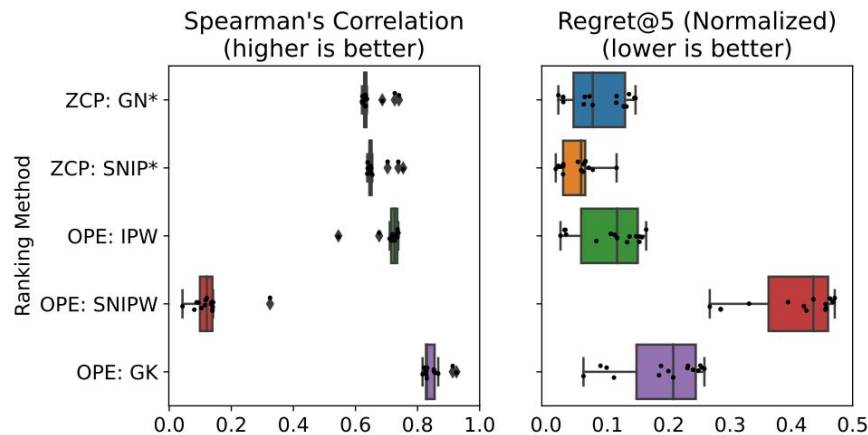


# The big picture

- PPO is selected for training the control agents



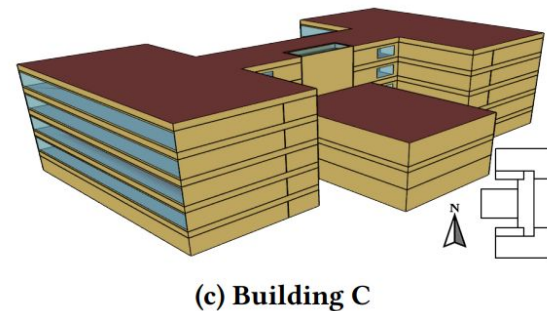
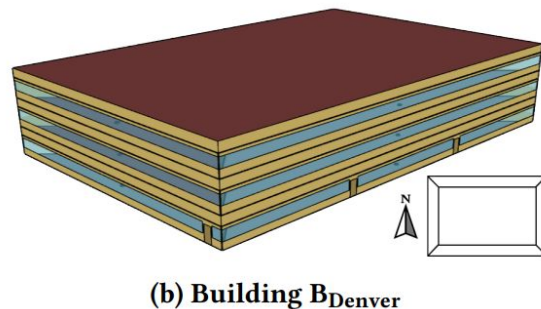
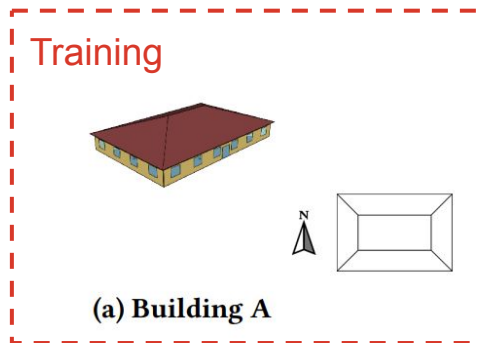
# Choosing the policy ranking method



- GK yields a better overall rank estimation so it is used to identify the cluster that contain high-quality policies
- SNIP yields a better rank estimation for top policies so it is used to identify the best policy within the top ranked cluster

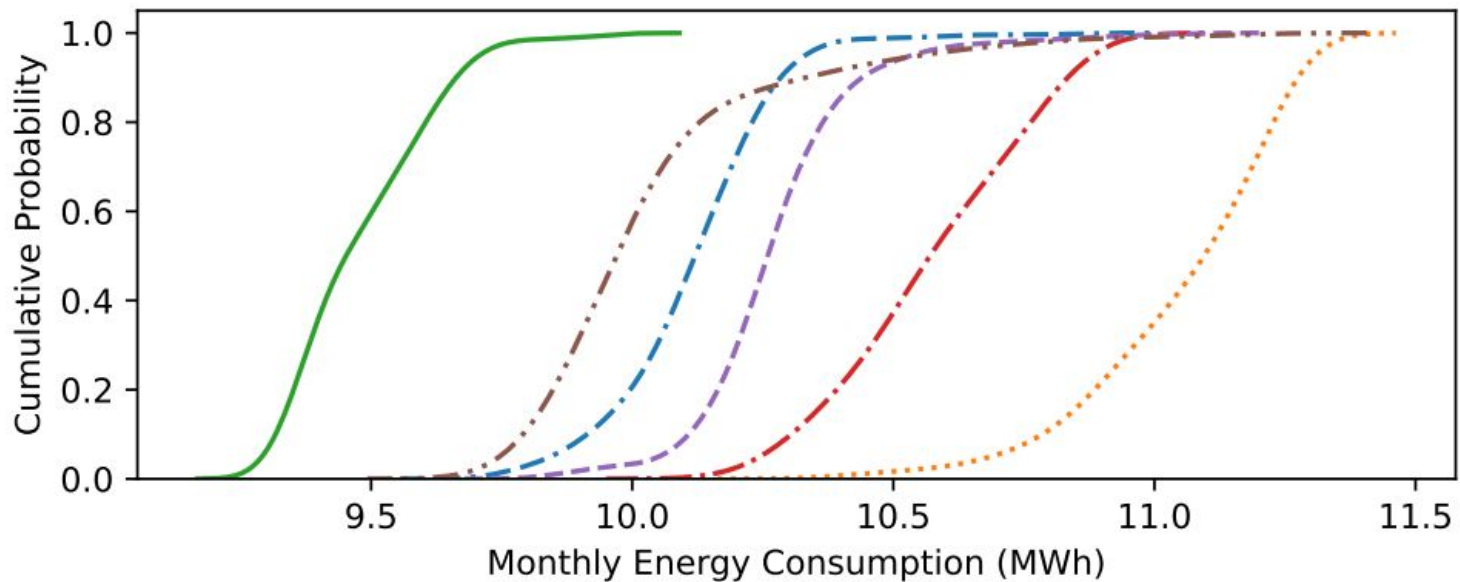
# Source and target buildings

- Building A: 1-story 5-zone small office (511.16 m<sup>2</sup>), located in Denver, US
  - Building B: 3-story 15-zone medium office (4,982.19 m<sup>2</sup>)
  - Building C: 5-story 26-zone real building (5,051 m<sup>2</sup>), located in San Francisco, US
- all experiments were run in COBS\*



# Experiment result

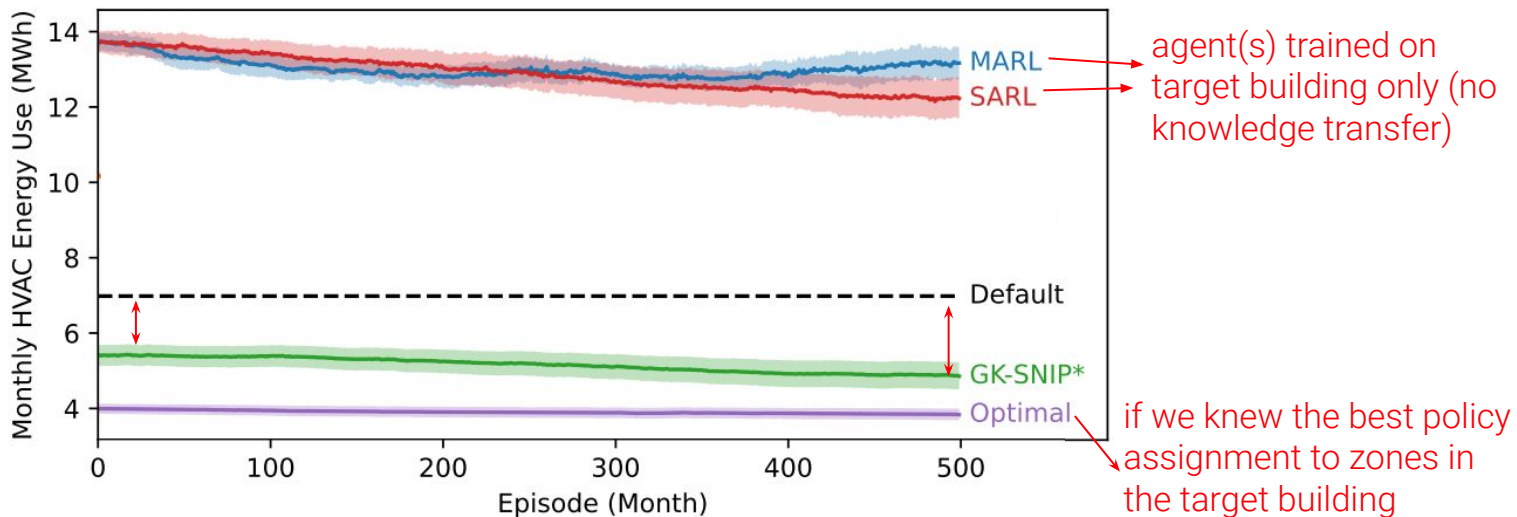
Importance of policy clustering



# Experiment result

Transfer to Building  $B_{\text{Denver}}$  (same city)

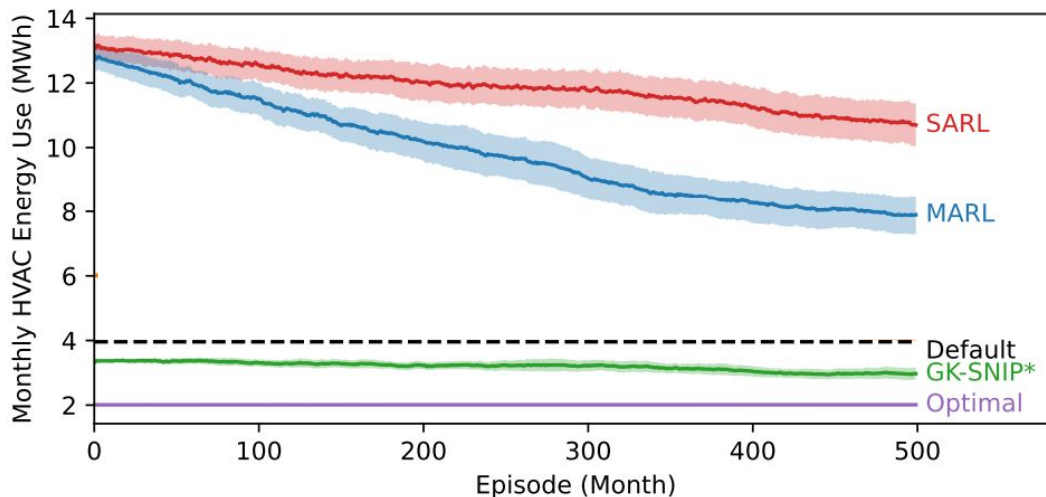
Policy evaluation/selection is done using 15 days of log data from Building  $B_{\text{Denver}}$



# Experiment result

Transfer to Building  $B_{\text{SanFrancisco}}$

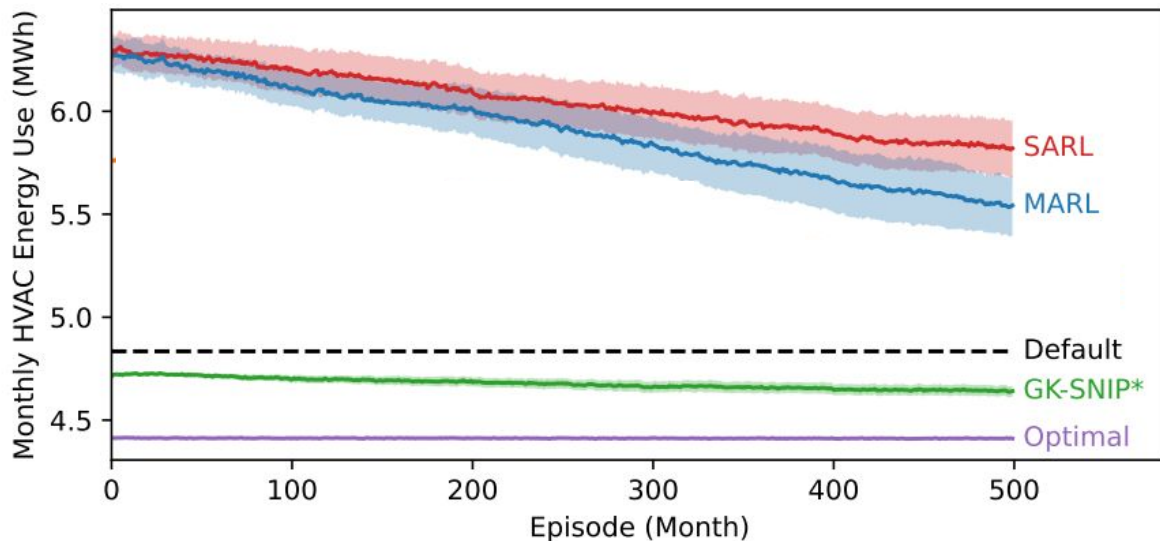
Policy evaluation/selection is done using 15 days of log data from Building  $B_{\text{SanFrancisco}}$



# Experiment result

Transfer to a real building (Building C)

Policy evaluation/selection is done using 15 days of log data from Building C





# Takeaways

- Having access to a set of diverse policies might be better than learning just an optimal policy when it comes to policy transfer to **novel environments!**
- High-quality policies in the policy library can be **efficiently** identified using the proposed policy evaluation and clustering technique.
- **Up to 30.4%** energy savings can be achieved early on using the proposed methodology on three buildings in different locations/climates. The RL-based controller outperforms the default rule-based controller, **even without retraining** in the target building.



[ardakanian@ualberta.ca](mailto:ardakanian@ualberta.ca)  
<https://cs.ualberta.ca/~oardakan/>