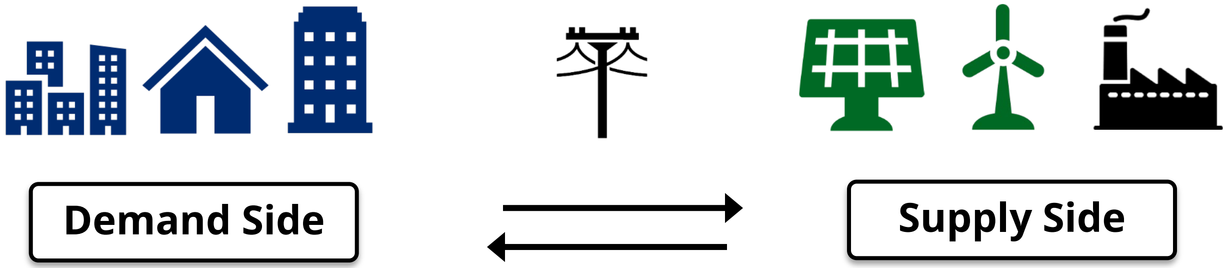
Towards Safe and Sample-efficient Autonomous Energy Systems via Differentiable Programming

Bingqing Chen

ML Research Scientist

BOSCH

It is important to combat climate change by 1) reducing energy consumption, and 2) increasing renewable energy penetration.

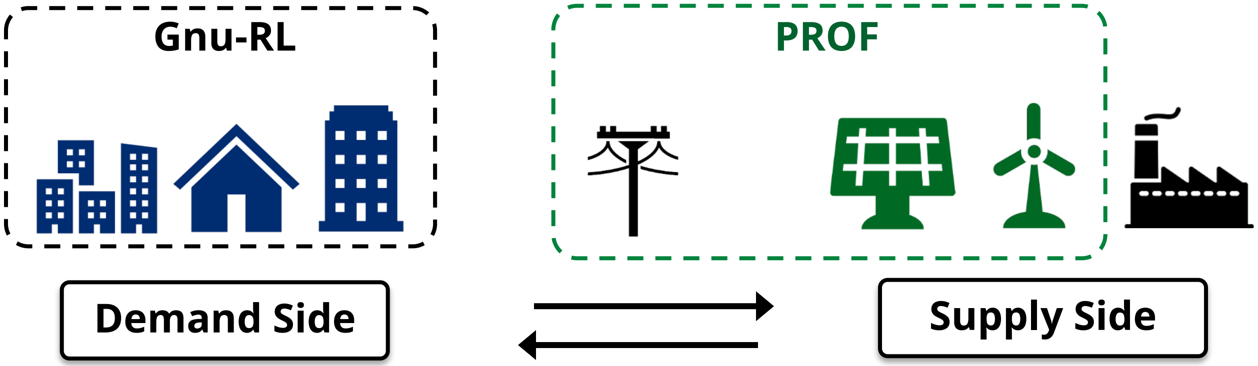


Energy-efficiency

Renewable energy integration

$\Big\}$ **Climate Change Mitigation**

We present two works that applies **differentiable programming** to reinforcement learning (RL) control of energy systems.



Energy-efficiency – Gnu-RL

Renewable energy integration – PROF

While there is growing interest in applying RL to energy systems, real-world applications are numbered due to the facts that:

RL agents generally take a long time to learn and reach acceptable performance.

The actions from an RL agent may not satisfy the safety constraints imposed by underlying physical system.

While there is growing interest in applying RL to energy systems, real-world applications are numbered due to the facts that:

Differentiable programming can tackle these challenges by incorporating domain knowledge on system dynamics and constraints.

Vision: Autonomous energy systems can learn safely and sample-efficiently.

☑ **S**afe

☑ **S**ample-efficient

Energy-efficiency

Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy

Bingqing Chen, Zicheng Cai, Mario Bergés

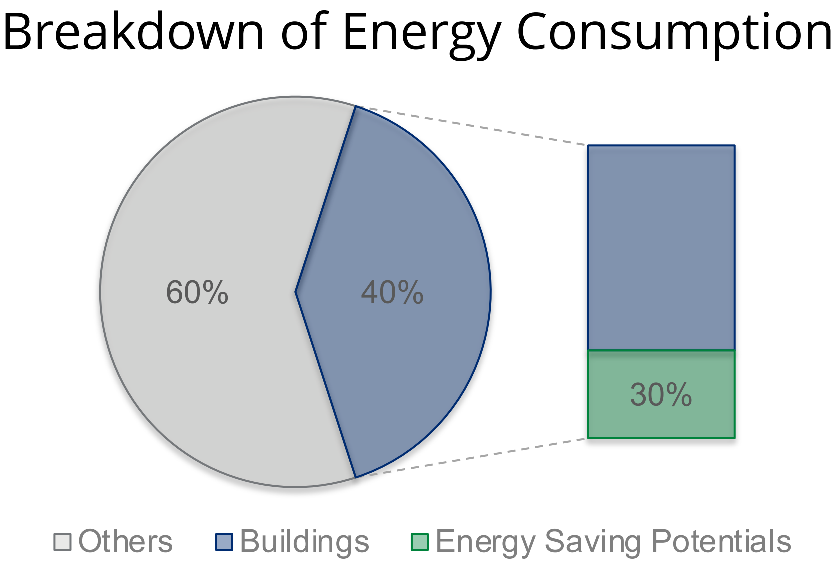
Best Paper Award at ACM BuildSys'19

Bingqing Chen, Zicheng Cai, and Mario Bergés. "Gnu-RL: A practical and scalable reinforcement learning solution for building HVAC control using a differentiable MPC policy." *Frontiers in Built Environment* (2020): 174.

Carnegie Mellon University

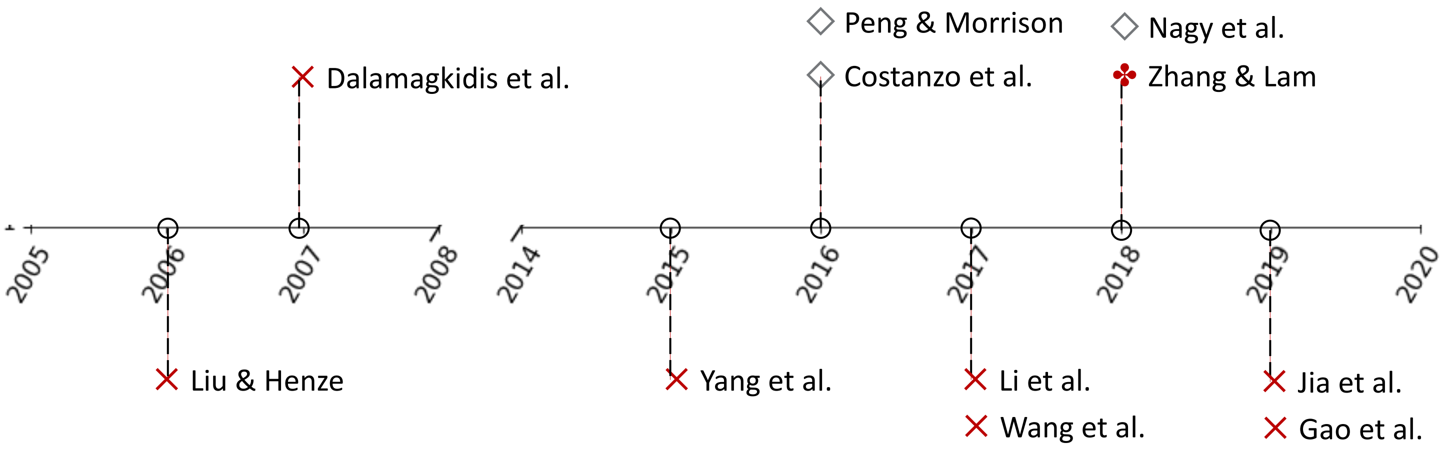
Background: Buildings present significant energy saving potentials.

Demand Side



Carnegie Mellon University

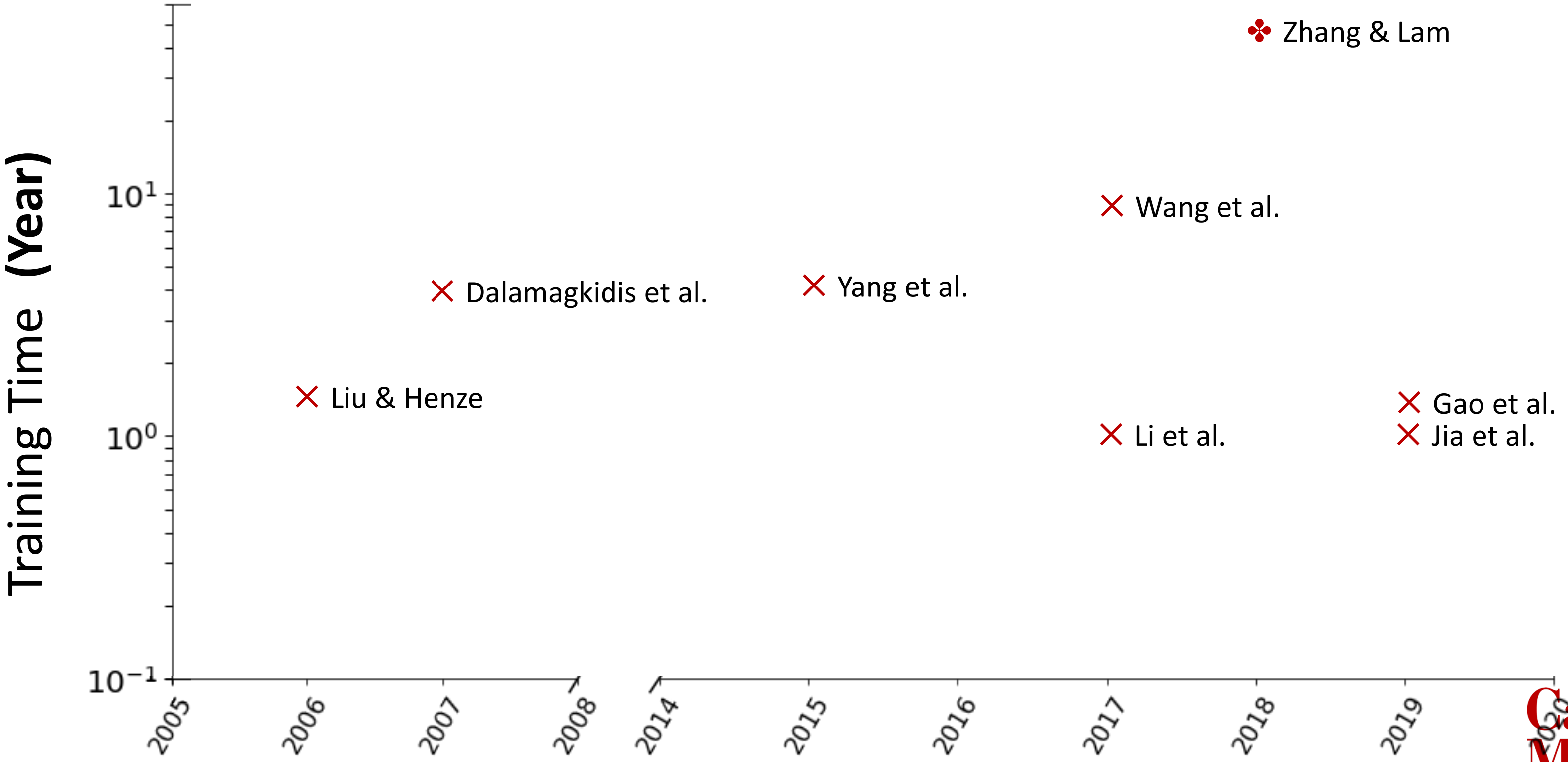
Training Time in Literature for RL Control of HVAC Systems



(◇ Evaluated on Grey-box Model)

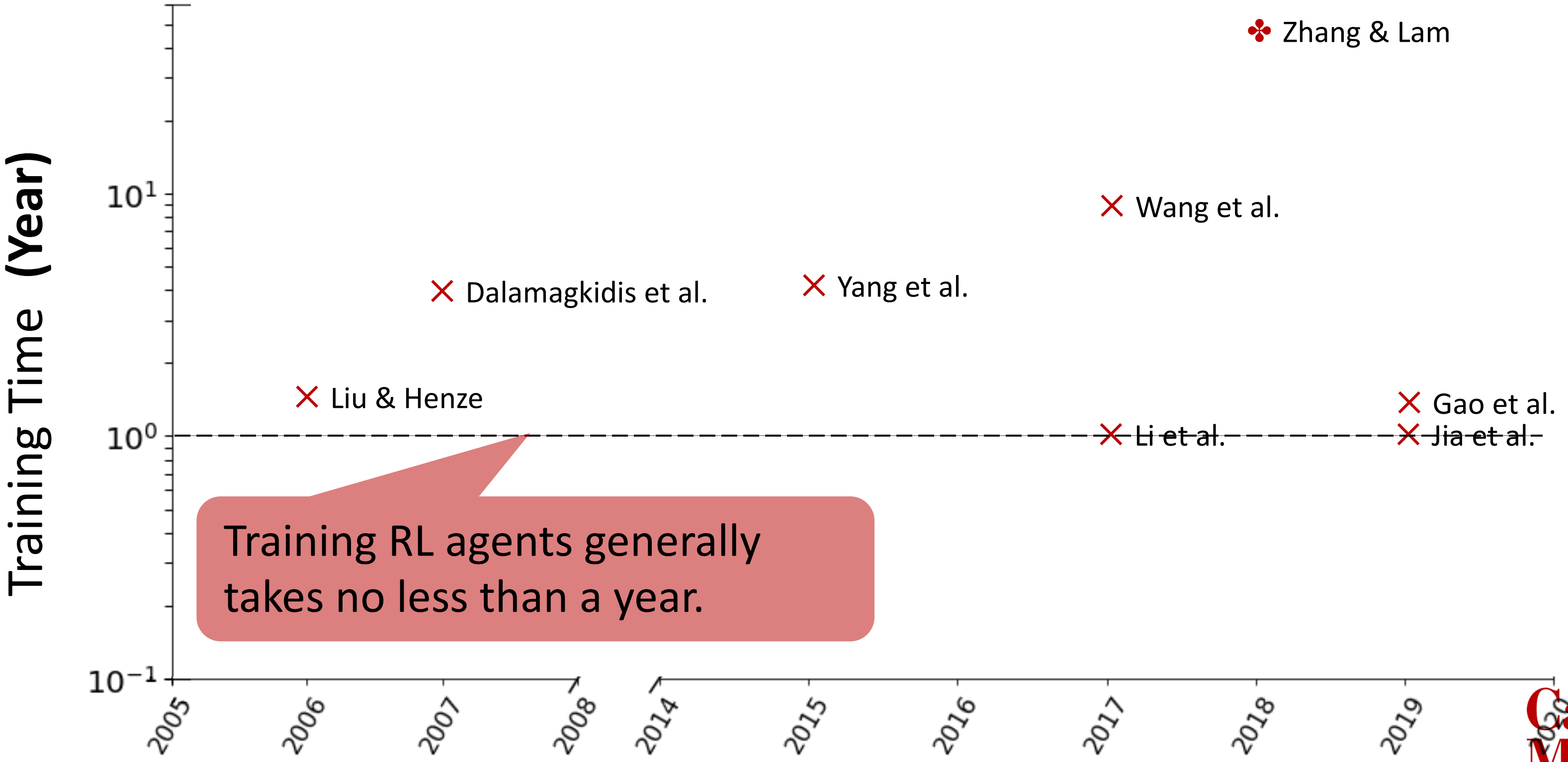
Carnegie Mellon University

Training Time in Literature for RL Control of HVAC Systems



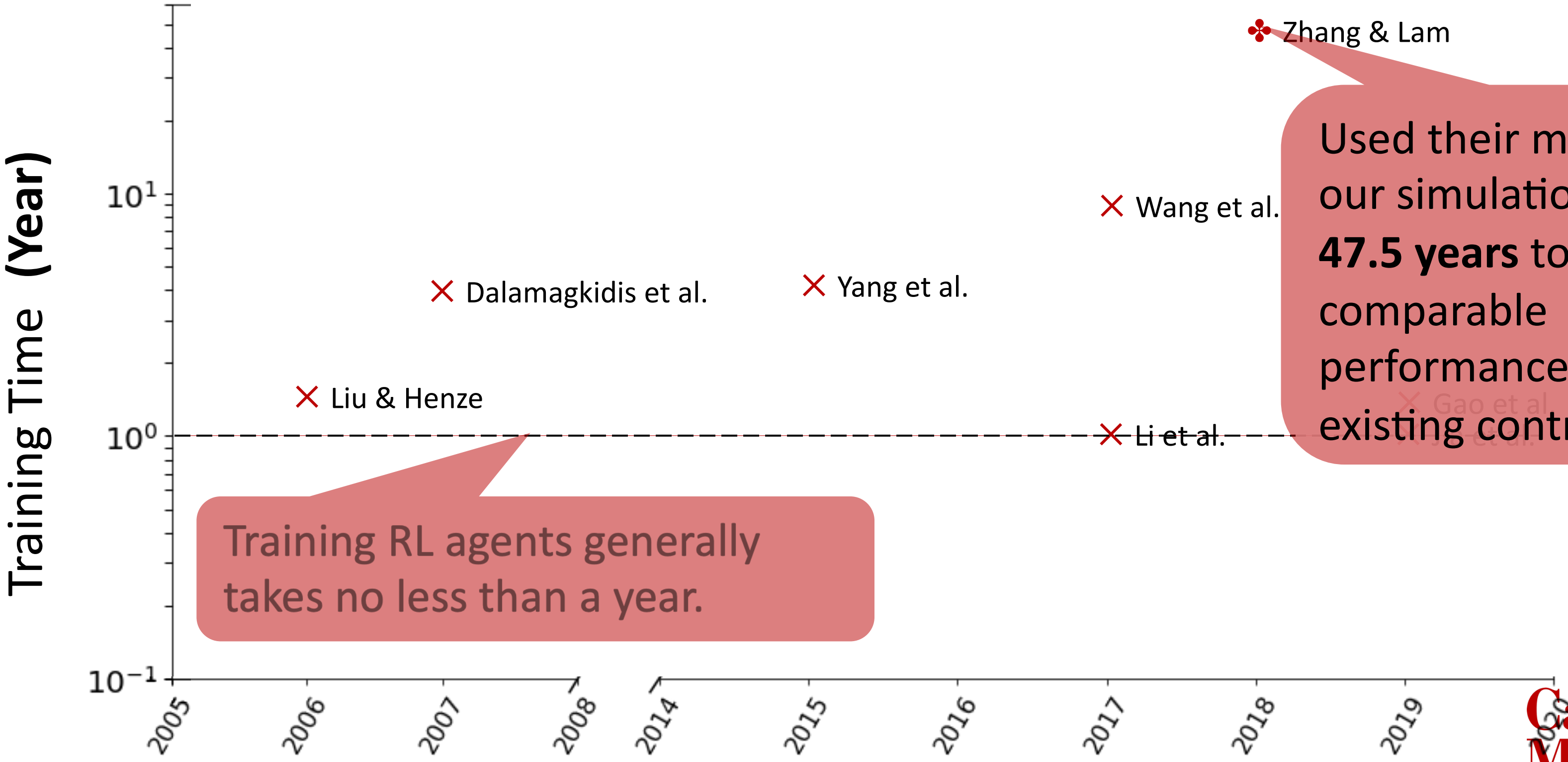
Carnegie Mellon University

Training Time in Literature for RL Control of HVAC Systems



Carnegie Mellon University

Training Time in Literature for RL Control of HVAC Systems

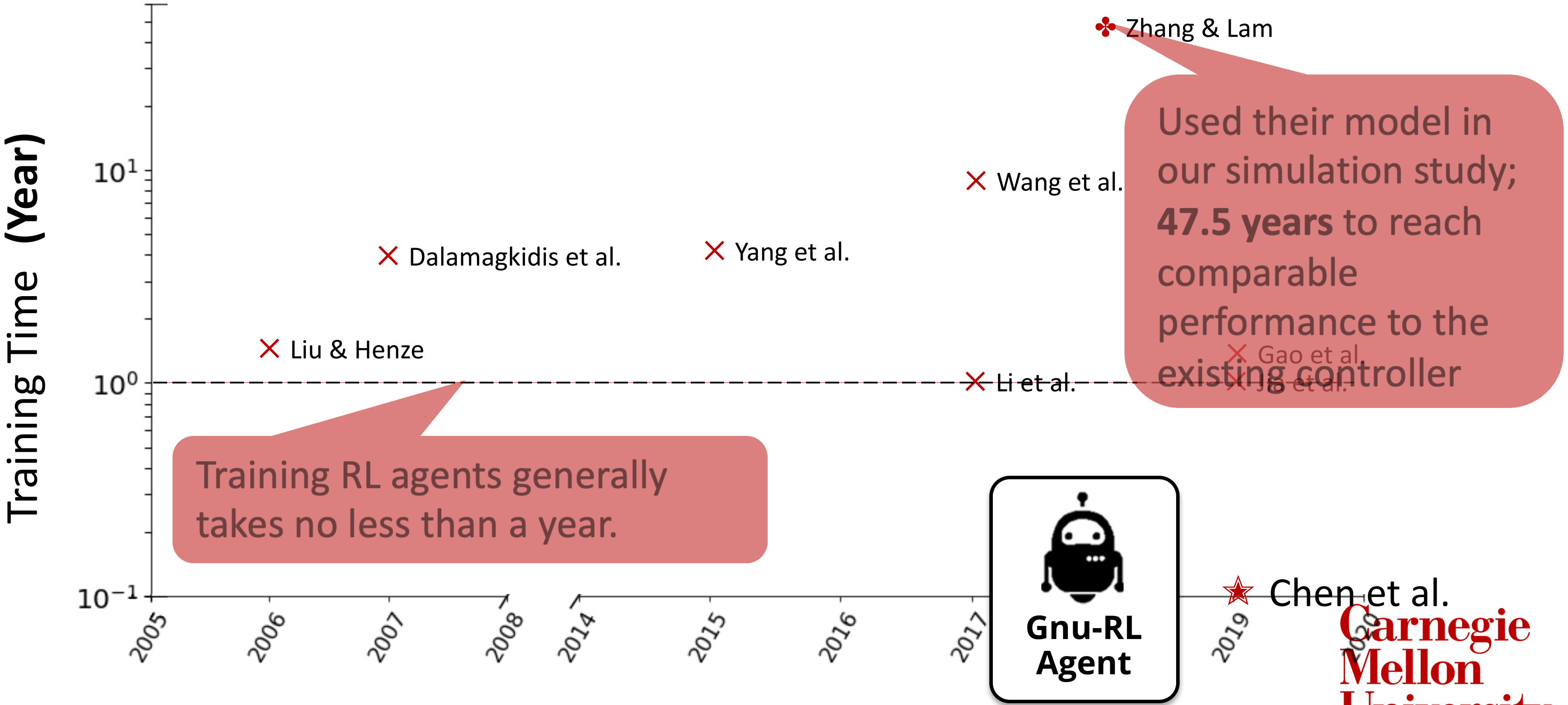


Used their model in our simulation study; **47.5 years** to reach comparable performance to the existing controller

Training RL agents generally takes no less than a year.

Carnegie Mellon University

Training Time in Literature for RL Control of HVAC Systems



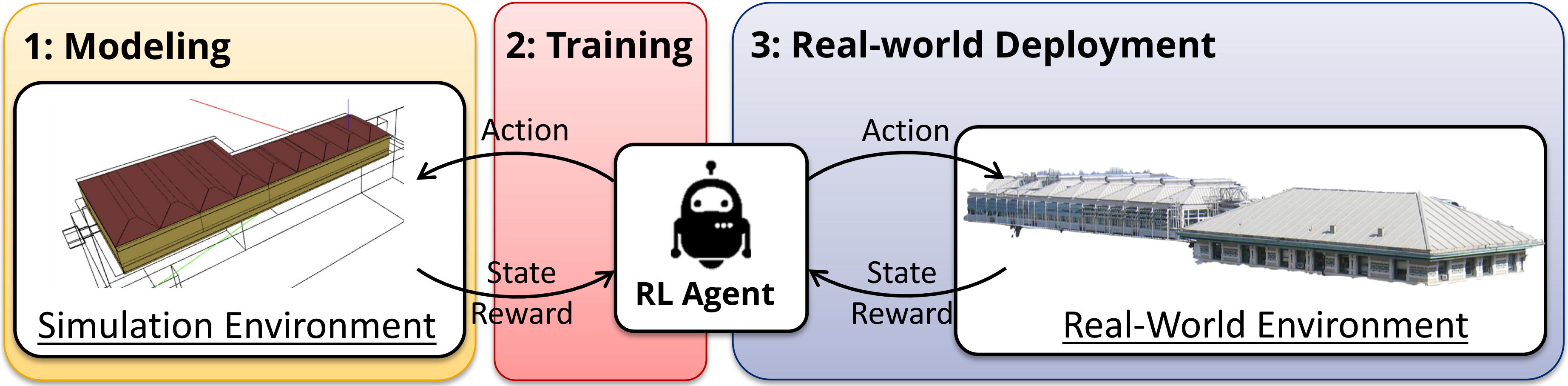
Carnegie Mellon University

Gnu-RL: a precocial reinforcement learning solution for HVAC control

[**Precocial**:= capable of a high degree of independent activity from birth]

(From Adobe Stock)

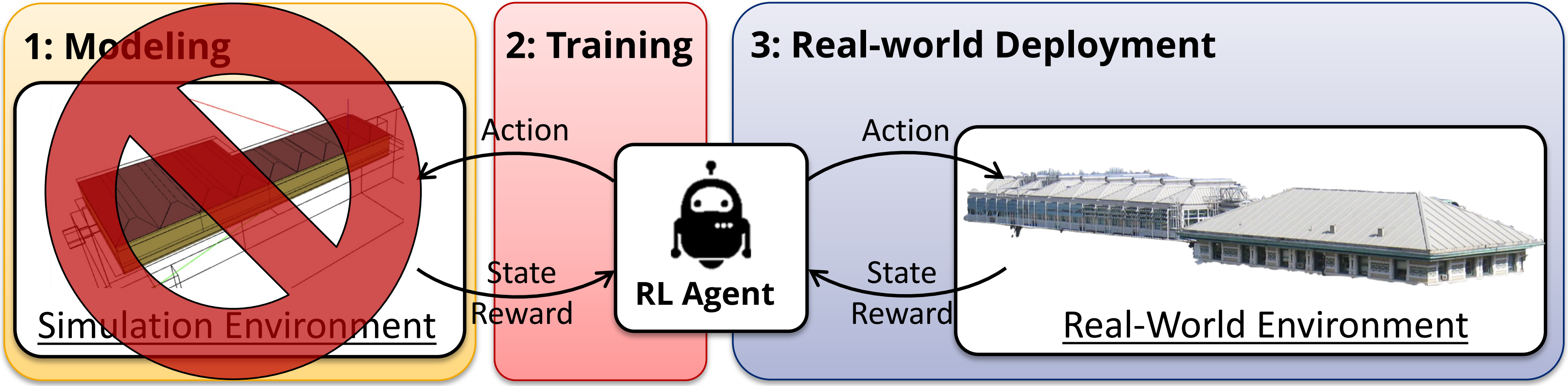
Typical Framework



(Modified from Zhang & Lam, 2018)

Carnegie Mellon University

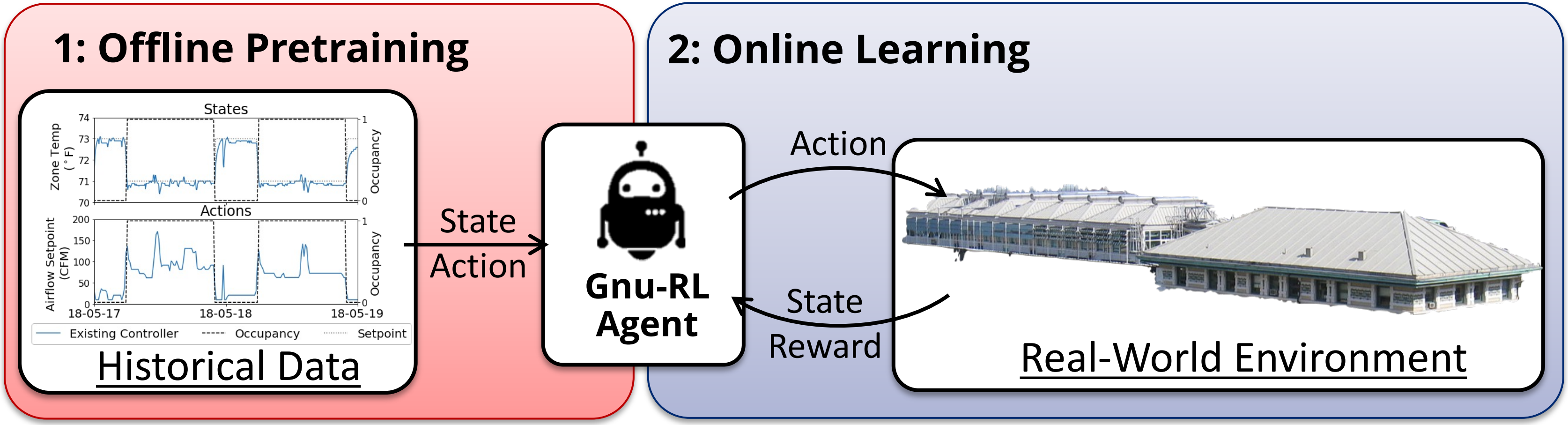
Typical Framework



(Modified from Zhang & Lam, 2018)

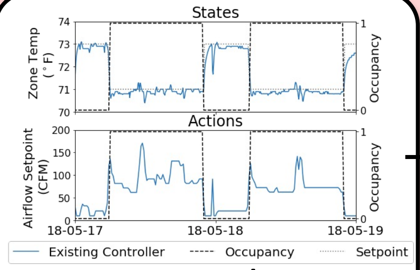
Carnegie Mellon University

Our Framework



We expedite the training by imitating the existing controller.

1: Offline Pretraining



Historical Data

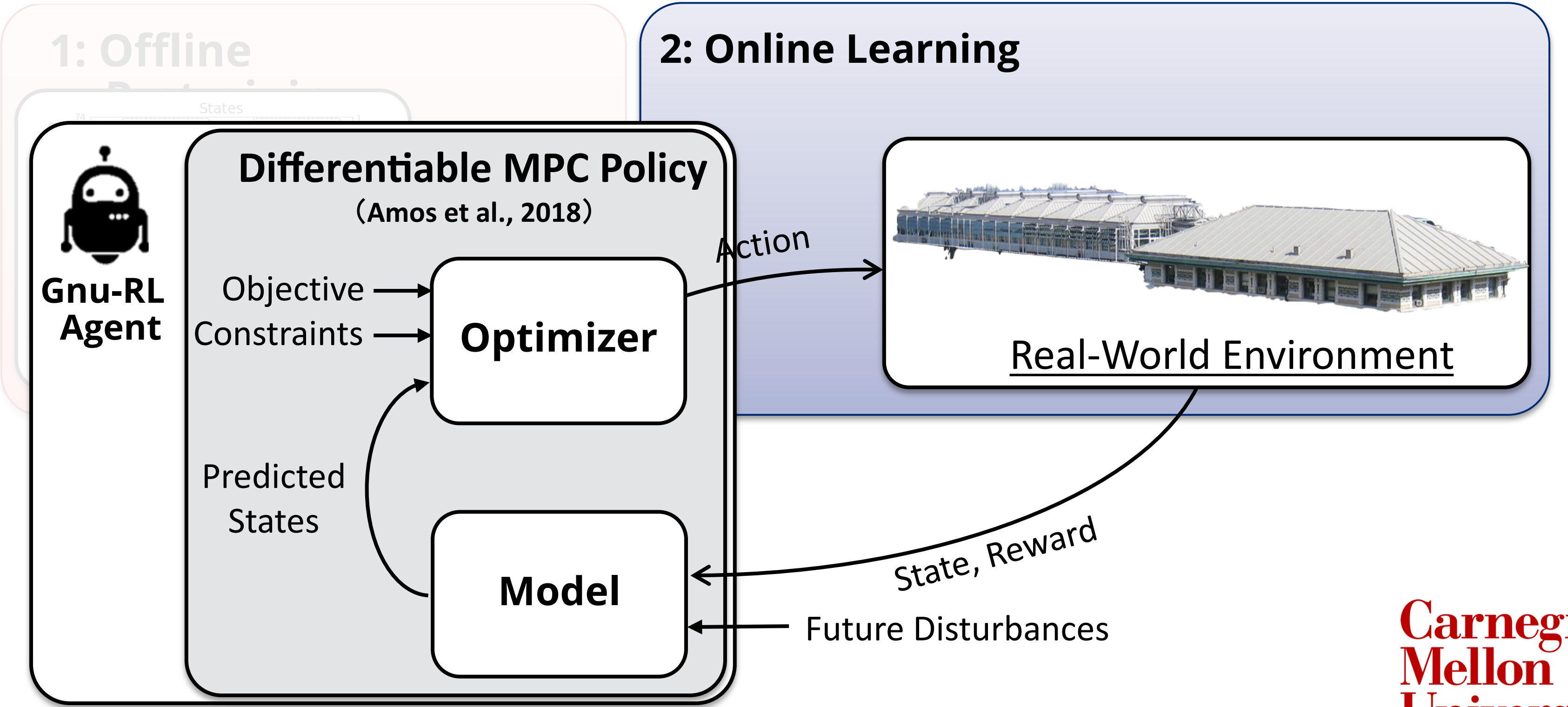
State

Action

Gnu-RL Agent

2: On

We expedite the training by using a policy that encodes knowledge on system dynamics and control.



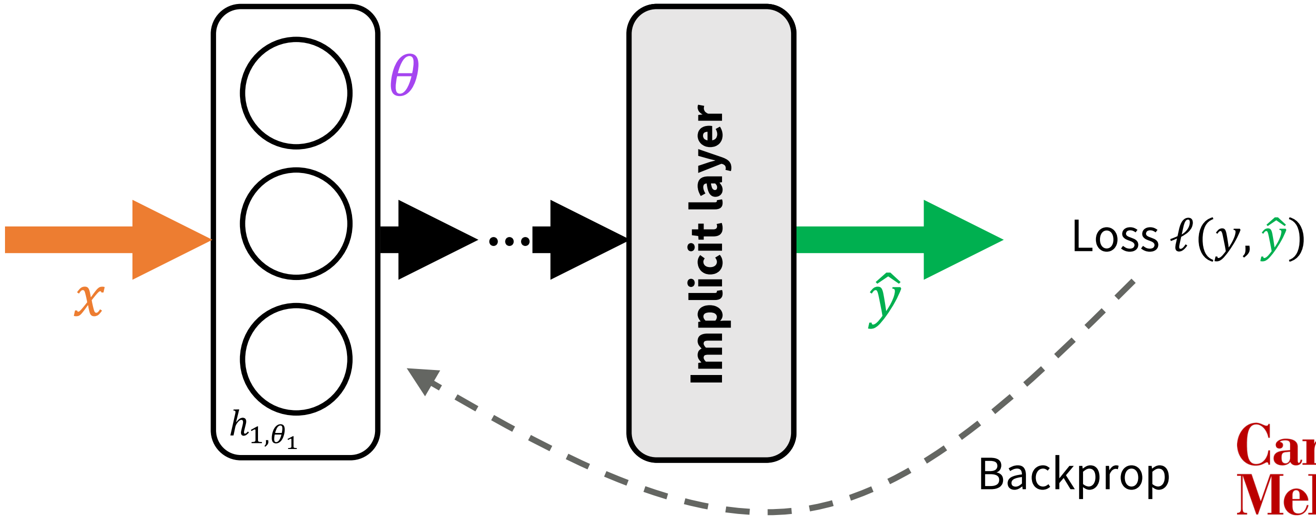
Carnegie Mellon University

Background: Neural Networks and Differentiable Learning

Neural network = composition of non-linear, parameterized, differentiable functions

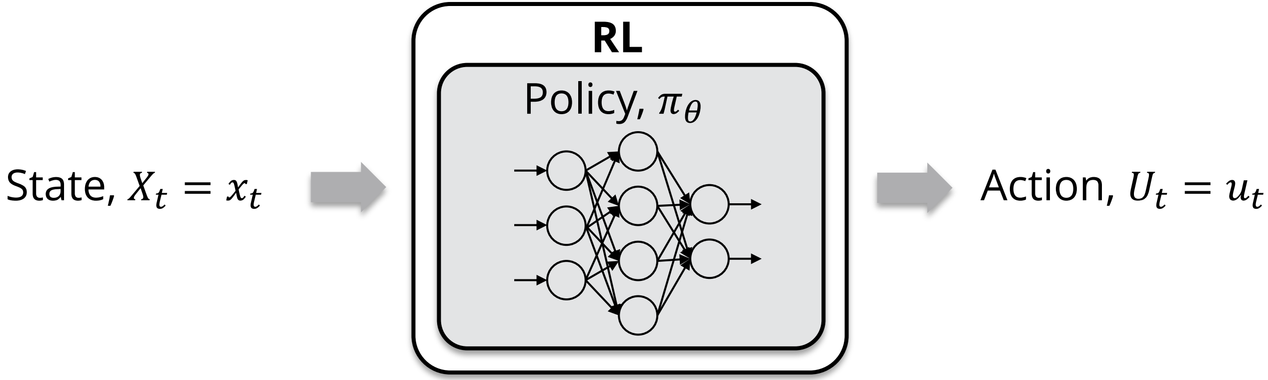
$$h_\theta = h_{1,\theta_1} \circ \cdots \circ h_{L,\theta_L}$$

Recent interest in enriching the set of functions that can be accommodated (“implicit layers”), such as **optimization problems** and **physical equations**.



Carnegie Mellon University

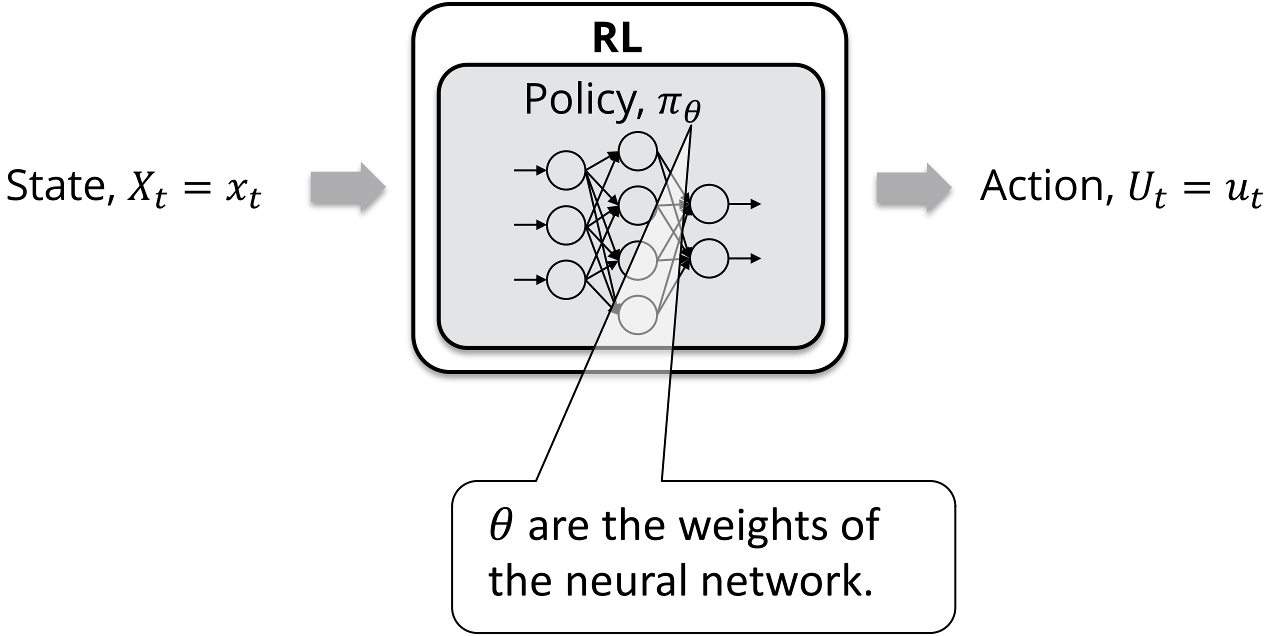
We optimize the differentiable MPC policy end-to-end with a policy gradient algorithm.



Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017).
Proximal Policy Optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Carnegie Mellon University

We optimize the differentiable MPC policy end-to-end with a policy gradient algorithm.



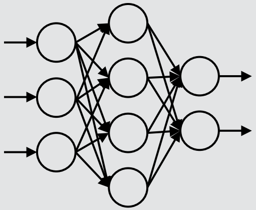
Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). **Proximal Policy Optimization** algorithms. *arXiv preprint arXiv:1707.06347*.

We optimize the differentiable MPC policy end-to-end with a policy gradient algorithm.

State, $X_t = x_t$

RL

Policy, π_θ



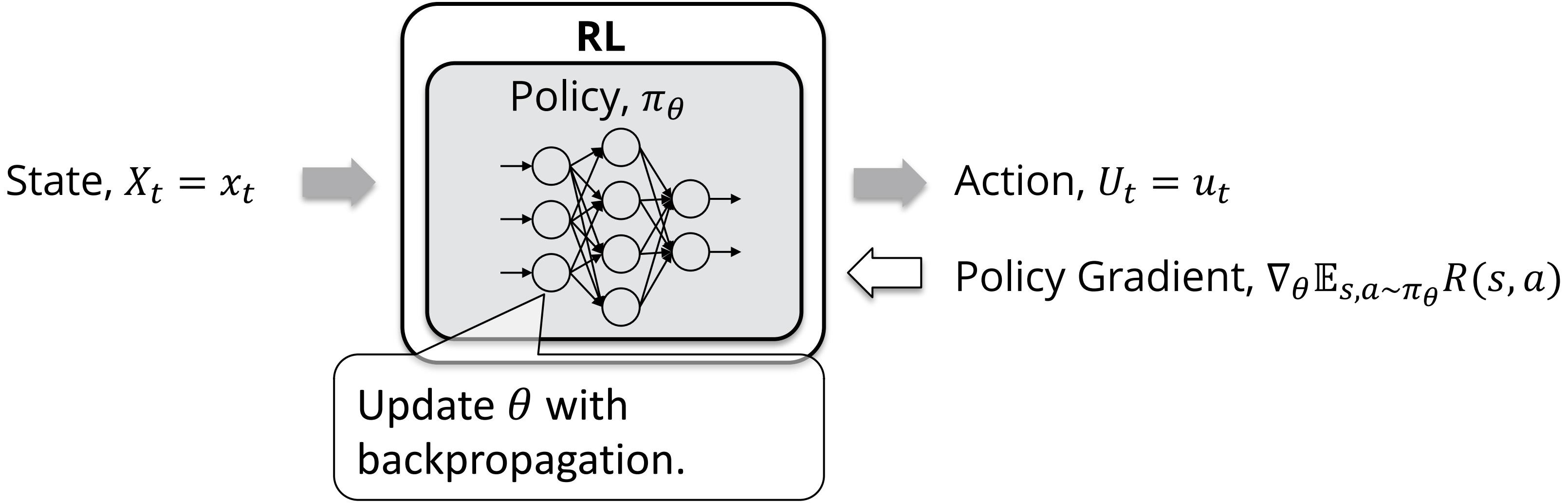
Action, $U_t = u_t$

Policy Gradient, $\nabla_\theta \mathbb{E}_{s,a \sim \pi_\theta} R(s,a)$

The policy gradient optimizes the expected reward.

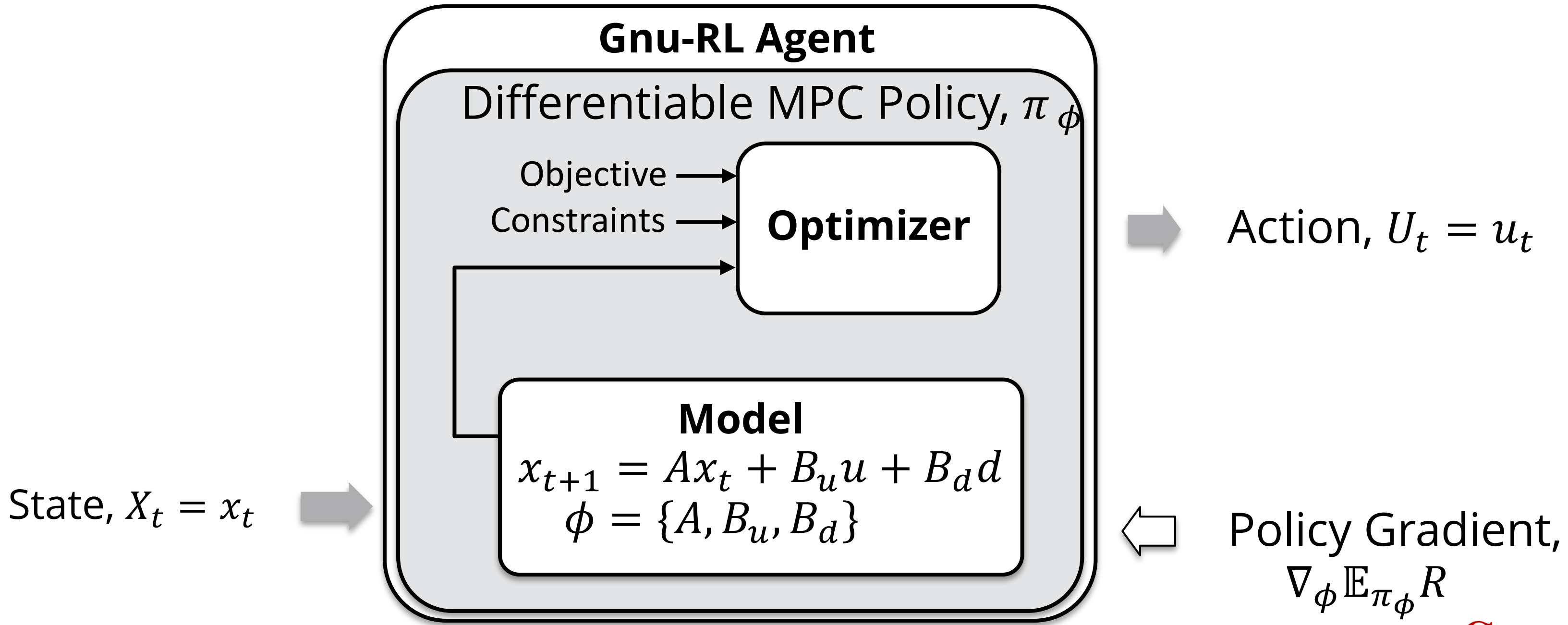
Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017).
Proximal Policy Optimization algorithms. *arXiv preprint arXiv:1707.06347*.

We optimize the differentiable MPC policy end-to-end with a policy gradient algorithm.

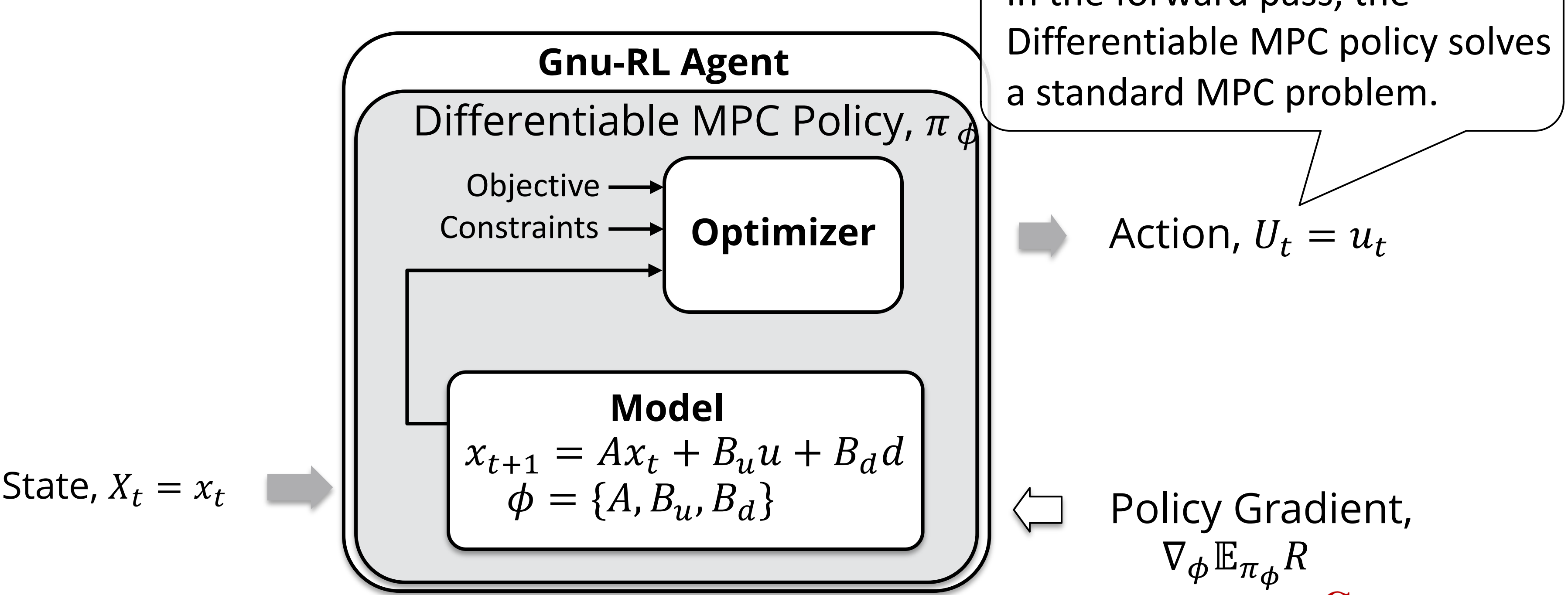


Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017).
Proximal Policy Optimization algorithms. *arXiv preprint arXiv:1707.06347*.

We optimize the differentiable MPC policy end-to-end with a policy gradient algorithm.

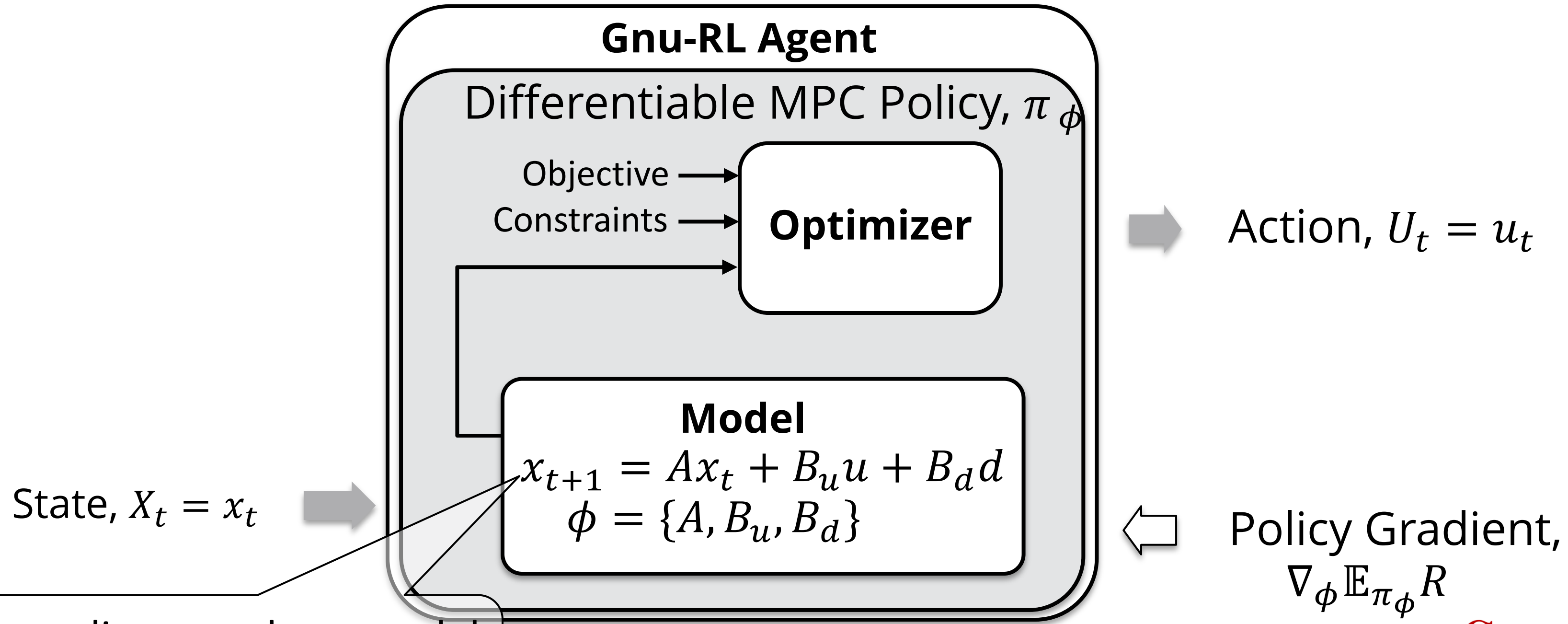


We optimize the differentiable MPC policy end-to-end with a policy gradient algorithm.



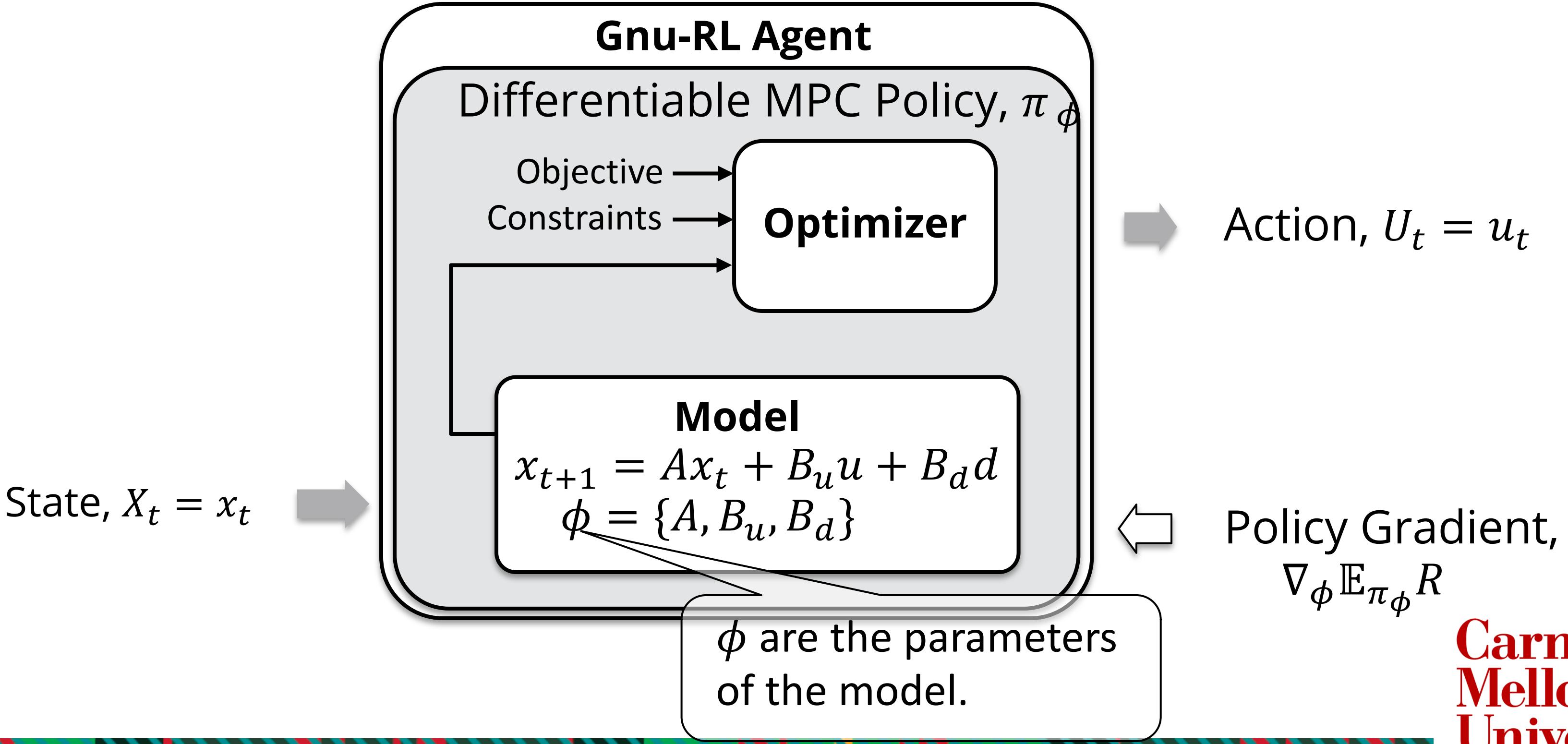
Carnegie Mellon University

We optimize the differentiable MPC policy end-to-end with a policy gradient algorithm.

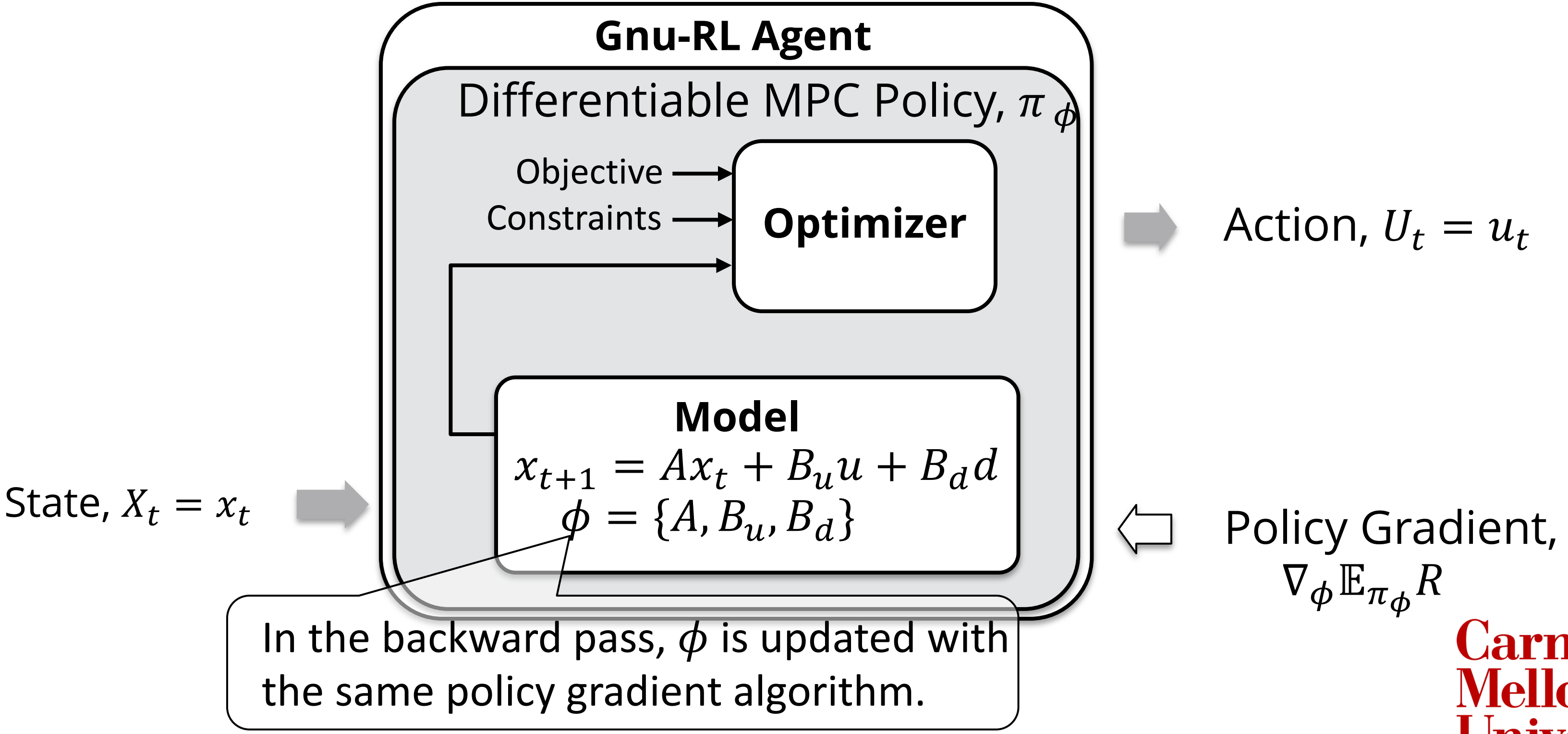


The policy encodes a model of system dynamics.

We optimize the differentiable MPC policy end-to-end with a policy gradient algorithm.



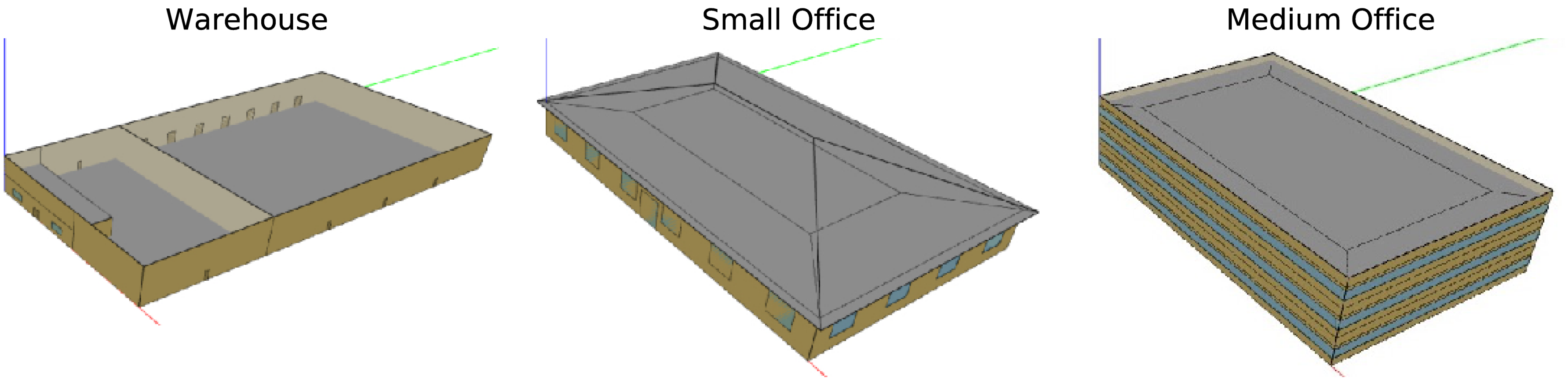
We optimize the differentiable MPC policy end-to-end with a policy gradient algorithm.



Experiment 1: Simulation Study on Commercial Reference Buildings

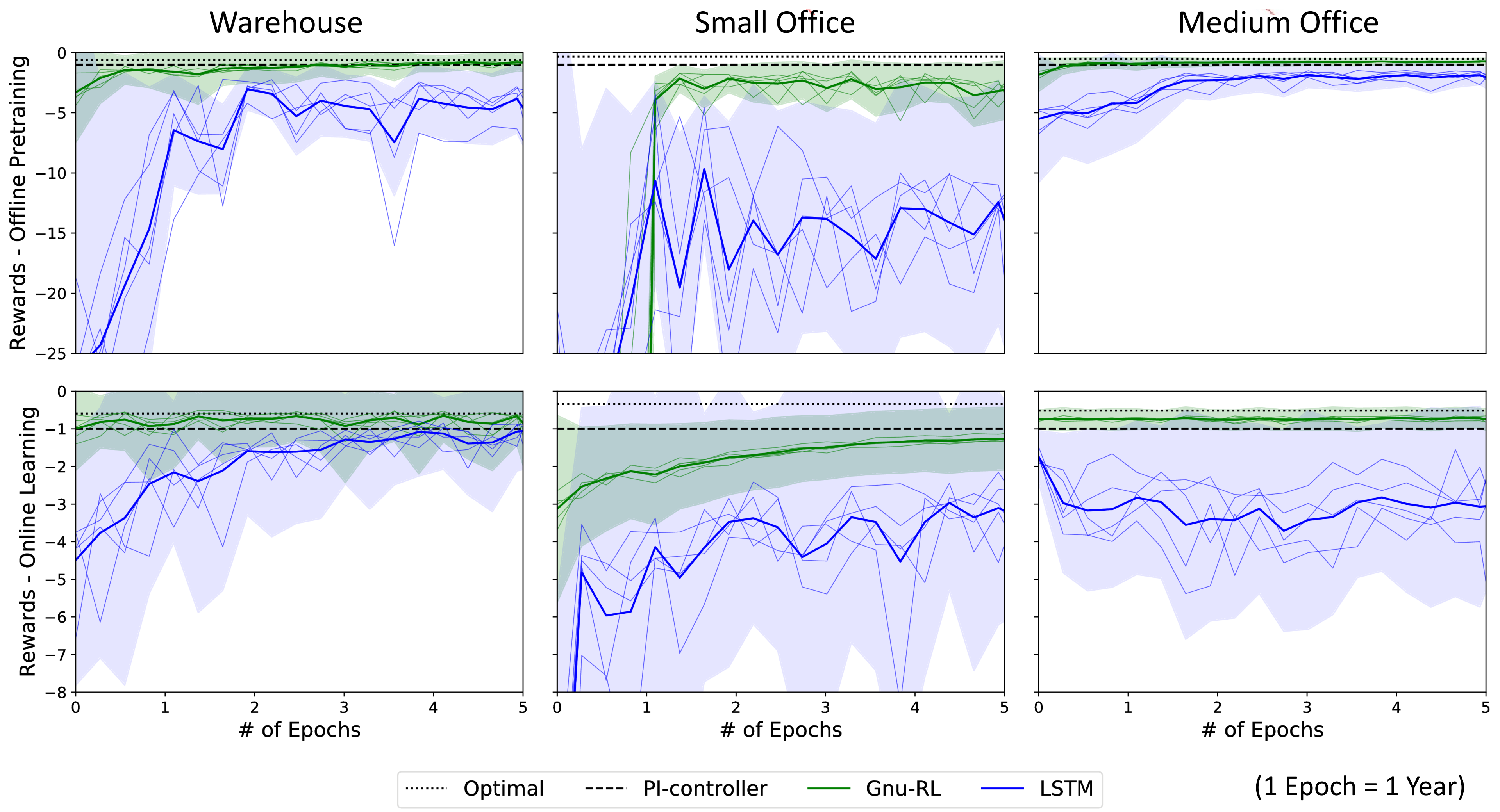
We compared our proposed approach to:

- Optimal solution
- PI-Controller
- An RL agent with LSTM policy



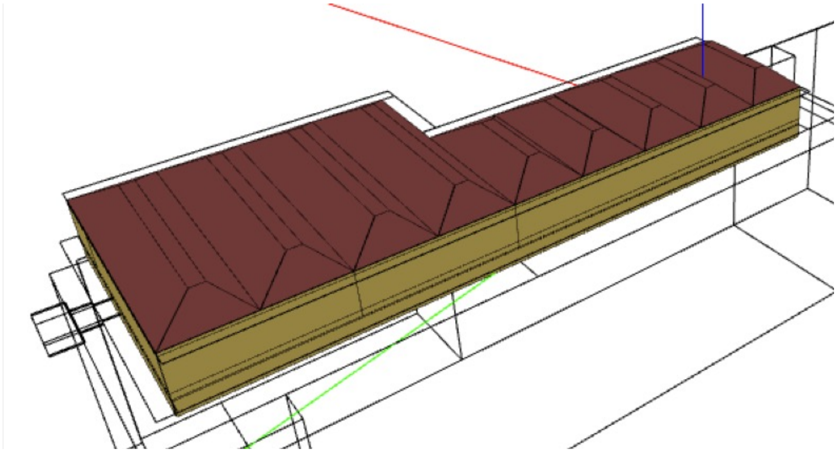
Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., ... & Crawley, D. (2011). US Department of Energy commercial reference building models of the national building stock.

Carnegie Mellon University

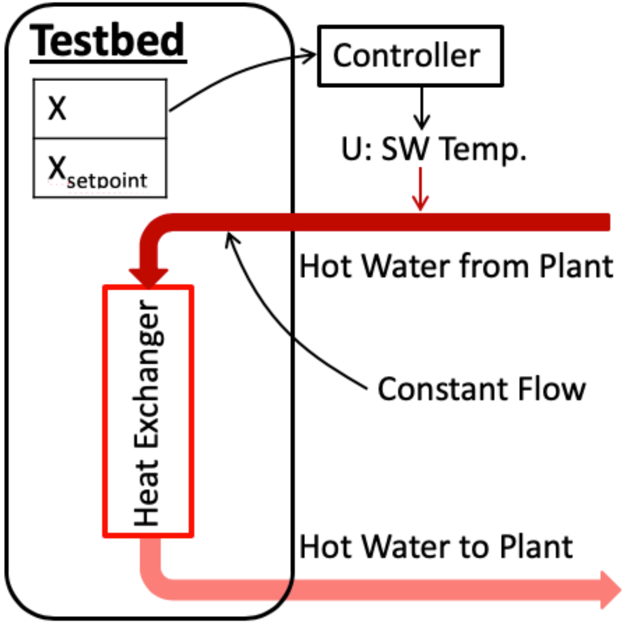


Warehouse
Small Office
Medium Office
Rewards - Offline Pretraining
Rewards - Online Learning
# of Epochs
# of Epochs
# of Epochs
Optimal
PI-controller
Gnu-RL
LSTM
(1 Epoch = 1 Year)

Experiment 2: Simulation Study on Intelligent Workplace



Intelligent Workplace
Margaret Morrison Hall, 4th Floor
(✤ Zhang & Lam, 2018)



HVAC Schematic

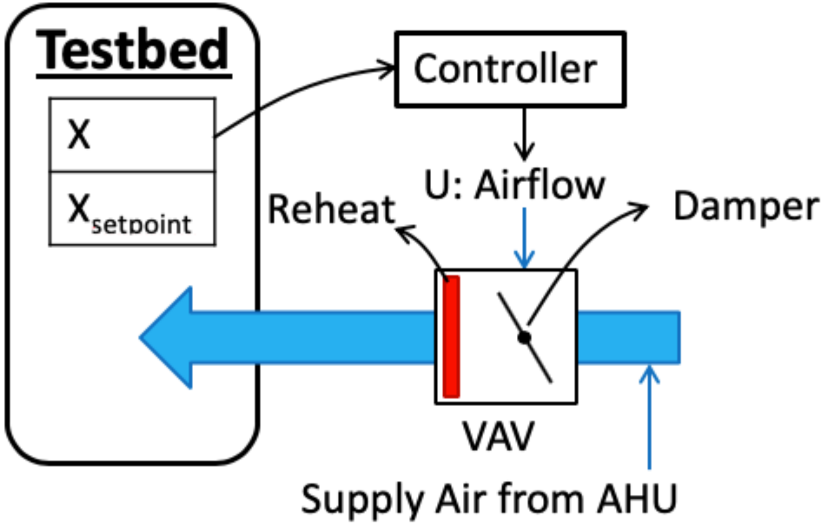
Gnu-RL achieved significant energy savings without compromising thermal comfort.

	Total Heating Demand (kWh)	Predicted Percentage Dissatisfied	
		Mean (%)	STD (%)
Existing Controller	43709	**9.45**	5.59
Agent #6 (✤ Zhang & Lam, 2018)	37131	11.71	3.76
Gnu-RL	**34678**	9.56	6.39

Gnu-RL achieved **20.6%** energy savings compared to the existing controller and **6.6%** energy savings compared to the best published RL result in the same environment.

Experiment 3: Real-world Deployment

6404, Gates & Hillman Center



HVAC Schematic

Carnegie Mellon University

Gnu-RL can be directly deployed in real-world environment with no prior information other than historical data.

- Specifically, we only used a month of historical data for pretraining.
- In the three-week experiment, the Gnu-RL agent continuously improved its policy.
- By the end of experiment, the Gnu-RL agent could track the temperature setpoint well, despite the complex room usage pattern.

Results: Summary Statistics

		Cooling Demand (kWh)	Zone Temp. RMSE (°F)
Existing Controller	Jun. 2017	169.4	2.4
	Jun. 2018	130.7	2.7
	Normalized	99.4	/
Gnu-RL		**82.8**	**1.02**

Gnu-RL achieved **16.7%** energy savings compared to the existing controller, while maintaining zone temperature better.

Carnegie Mellon University

Summary

Gnu-RL is a practical and scalable RL solution for building control.

Leverage the differentiable MPC policy, Gnu-RL is significantly more sample-efficient than prior works.

Gnu-RL can be directly deployed in real-world environment with no prior information other than historical data.

Network Constraints

Enforcing Policy Feasibility Constraints through Differentiable Projection for Energy Optimization

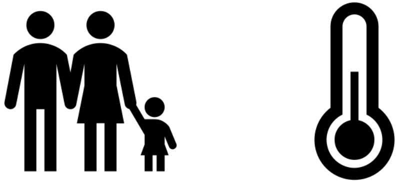
Bingqing Chen*, Priya L. Donti*, Kyri Baker, J. Zico Kolter, and Mario Bergés

Best Paper Runner-up at ACM e-energy’21

Motivation: Autonomous energy systems should be safe.

Application 1: An autonomous agent that learning to control buildings energy-efficiently, while ensuring the occupants' thermal comfort requirements are satisfied.

subject to



Application 2: An autonomous agent that learning to operate variable renewable energy resources, without violating constraints in the distribution network.

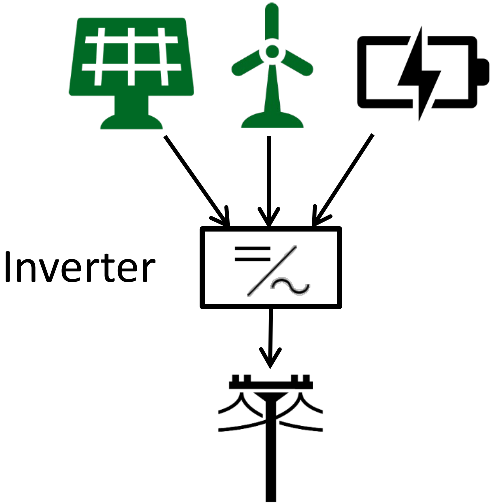
subject to

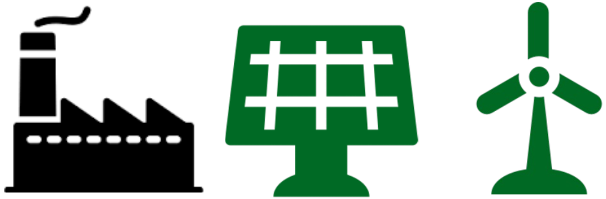
Carnegie Mellon University

Experiment 2: Inverter Control

subject to

As the penetration of renewable energy resources continues to grow, the grid of the future will become inverter-dominated.



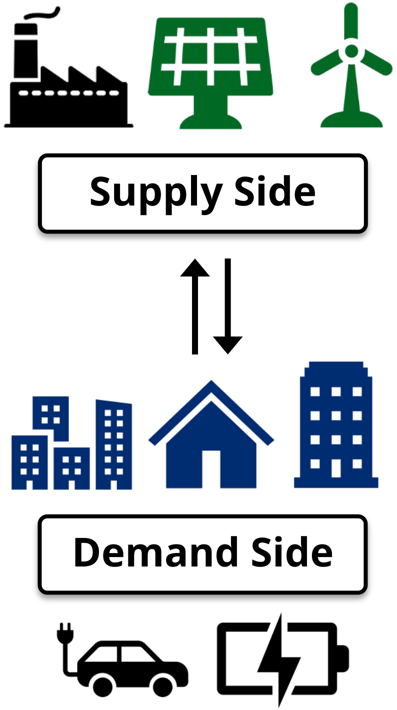


Kroposki, B., Johnson, B., Zhang, Y., Gevorgian, V., Denholm, P., Hodge, B. M., & Hannegan, B. (2017). Achieving a 100% renewable grid: Operating electric power systems with extremely high levels of variable renewable energy. *IEEE Power and Energy Magazine*, 15(2), 61-73.

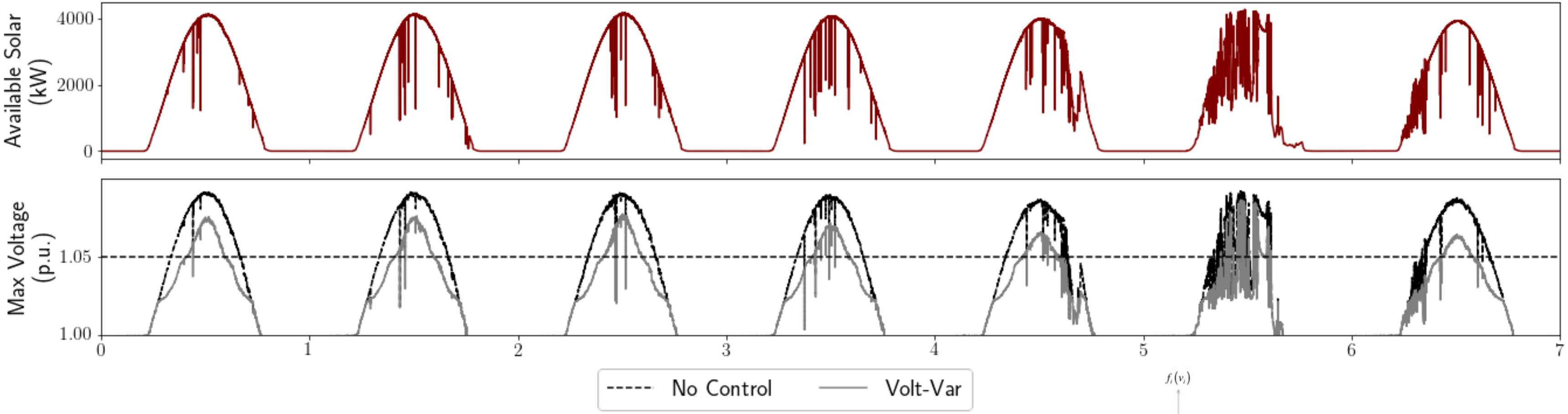
Background: Increasing penetration of renewable generation can introduce unintended challenges for grid operators.

The intermittent and variable nature of renewable generation makes it difficult to **balance supply and demand** of energy in the power grid.

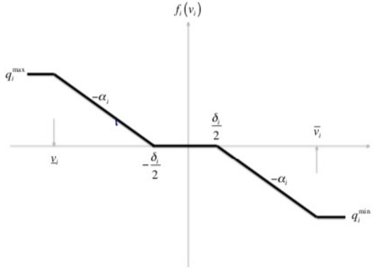
High penetration of renewable generation without proper control results in **constraint violations** in the distribution network.



Over-voltage has become a common occurrence in areas with high renewable penetration.



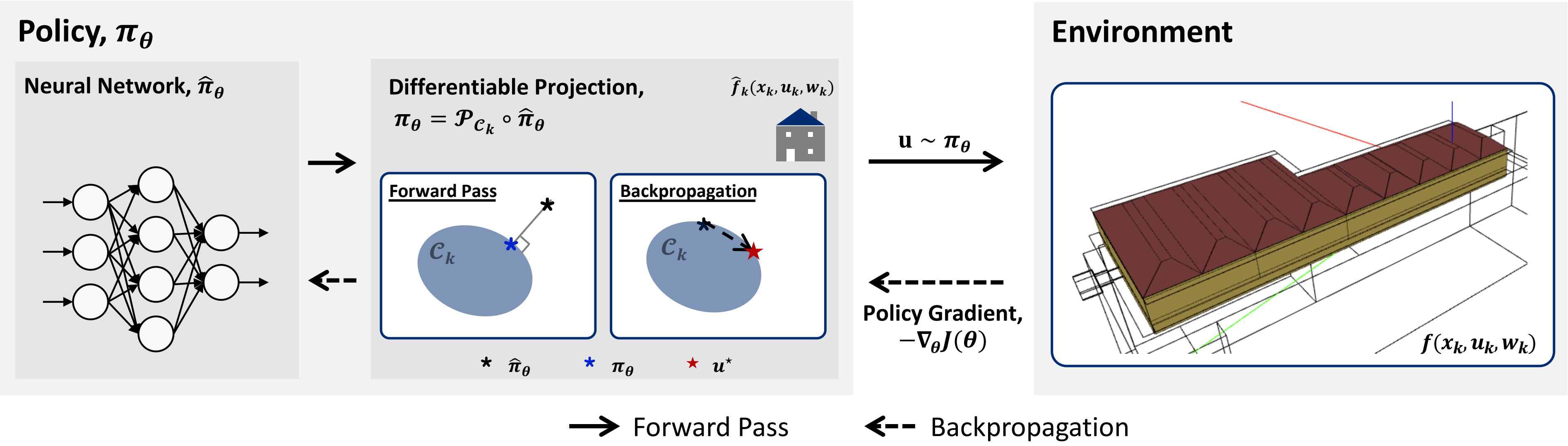
Volt-Var control is recommended in IEEE 1547.8-2018.



Volt-Var Control Curve

Carnegie Mellon University

PROF: Projected Feasibility



Differentiable Projection Layer

Consider the projection operation

$$\mathcal{P}_{\mathcal{C}}(\hat{u}) = \underset{u \in \mathcal{C}}{\operatorname{argmin}} \frac{1}{2} ||u - \hat{u}||_2^2$$

For linear constraints $\mathcal{C} = \{u : Au = b, Gu \leq h\}$, can write and differentiate through KKT conditions

$$\begin{gathered} \operatorname{diag}(\lambda^\star)(Gu^\star - h) = 0 \\ Au^\star - b = 0 \\ u^\star - \hat{u} + A^T\nu^\star + G^T\lambda^\star = 0 \end{gathered} \quad \rightarrow \quad \begin{gathered} \mathrm{d}u^\star - \mathrm{d}\hat{u} + \mathrm{d}A^T\nu^\star + A^T\mathrm{d}\nu^\star + \mathrm{d}G^T\lambda^\star + G^T\mathrm{d}\lambda^\star = 0 \\ \mathrm{d}Au^\star + A\mathrm{d}u^\star - \mathrm{d}b = 0 \\ \operatorname{diag}(Gu^\star - h)\mathrm{d}\lambda + \operatorname{diag}(\lambda^\star)(\mathrm{d}Gu^\star + G\mathrm{d}u^\star - \mathrm{d}h) = 0 \end{gathered}$$

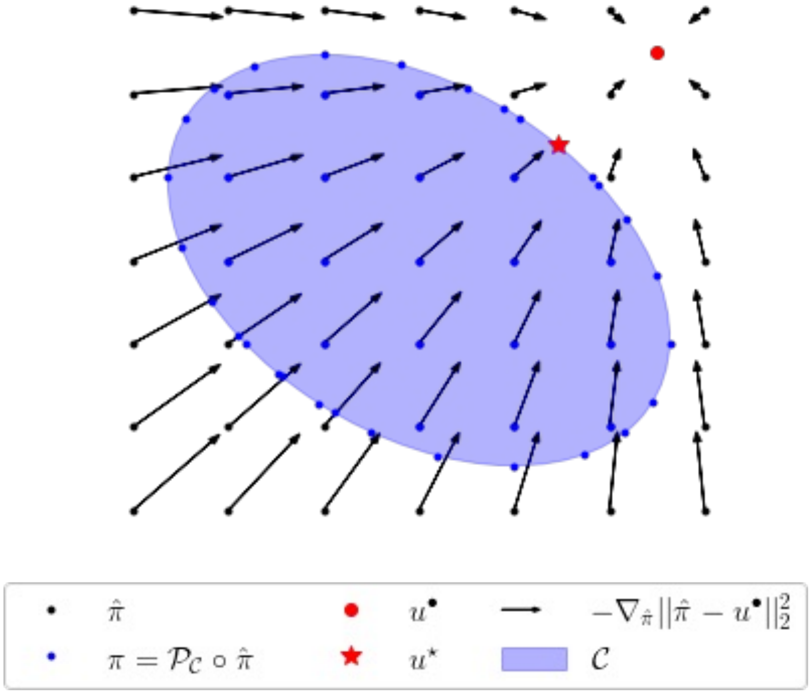
Note: Can also differentiate through general convex projections

See also:

- Amos, B., Kolter, J.Z. (2017). OptNet: Differentiable Optimization as a Layer in Neural Networks. *ICML*.
- Agrawal, A., Amos, B., Barratt, S., Boyd, S., Diamond, S., & Kolter, J.Z. (2019). Differentiable convex optimization layers. *NeurIPS*.

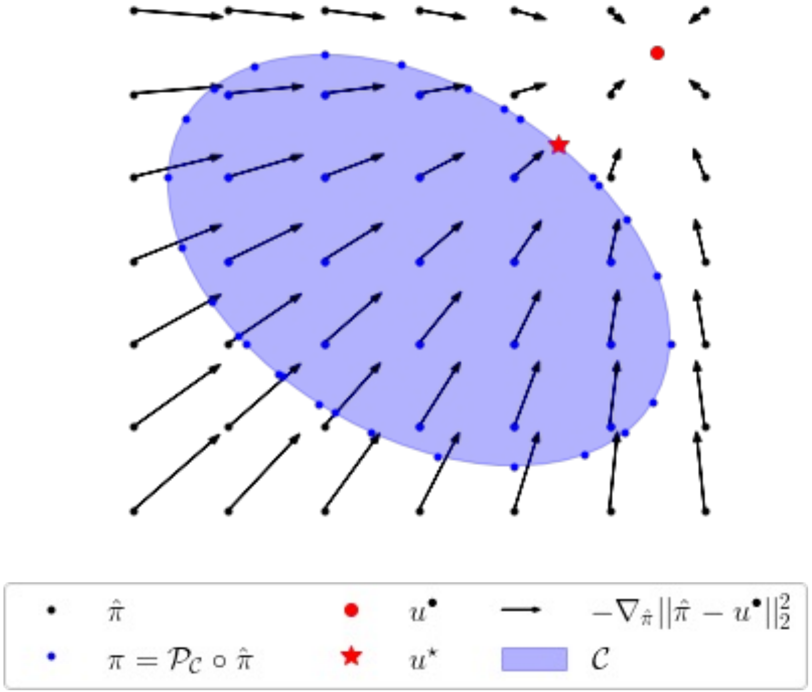
Carnegie Mellon University

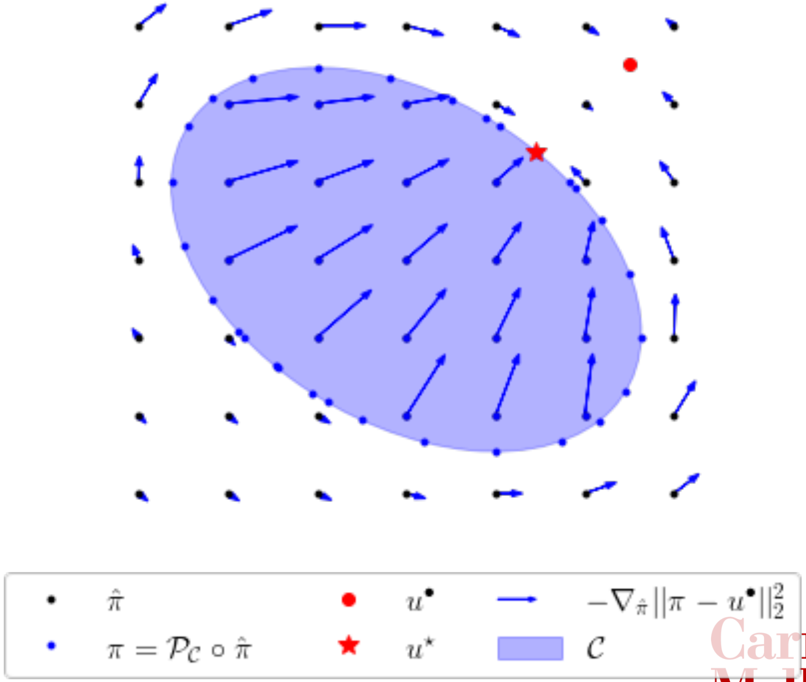
"Post-hoc" projections enforce constraints, but in a way that is hidden from the neural network during learning.



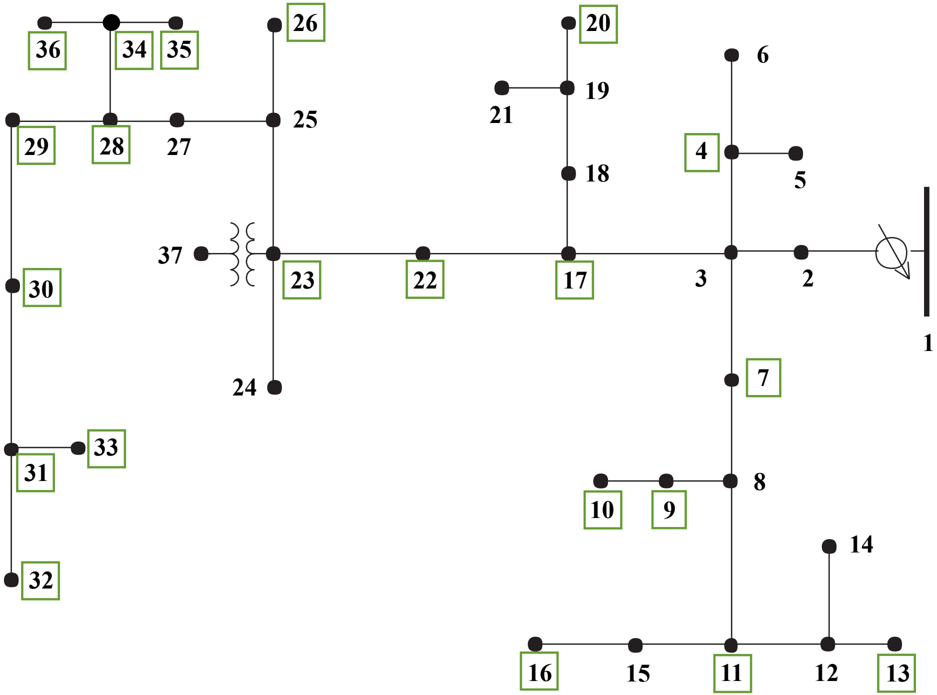
Carnegie Mellon University

Backpropagating through the differentiable projection layer makes the neural network cognizant of the constraints in its learning.





Simulation Testbed: IEEE 37-bus Feeder System



PV Panels

Baker, K., Bernstein, A., Dall'Anese, E., & Zhao, C. (2017). Network-cognizant voltage droop control for distribution grids. *IEEE Transactions on Power Systems*, *33*(2), 2098-2108.

Problem Formulation

Control the active and reactive power at each inverter p_i, q_i in order to

Minimize Curtailment

$$C(\theta) = \min_{\mathbf{p},\mathbf{q}} \sum_{i=1}^{N} [p_i - p_{av,i}]_+, \quad \text{where } [\mathbf{p}, \mathbf{q}] = \pi_\theta$$

Subject to system-level and device-level constraints

$$\mathcal{X} = \{v \mid 0.95 \leq v \approx Hu \leq 1.05\},$$

$$\mathcal{U}_i(k) = \{(p_i, q_i) : 0 \leq p_i \leq p_{av,i}(k), p_i^2 + q_i^2 \leq s_i^2\}$$

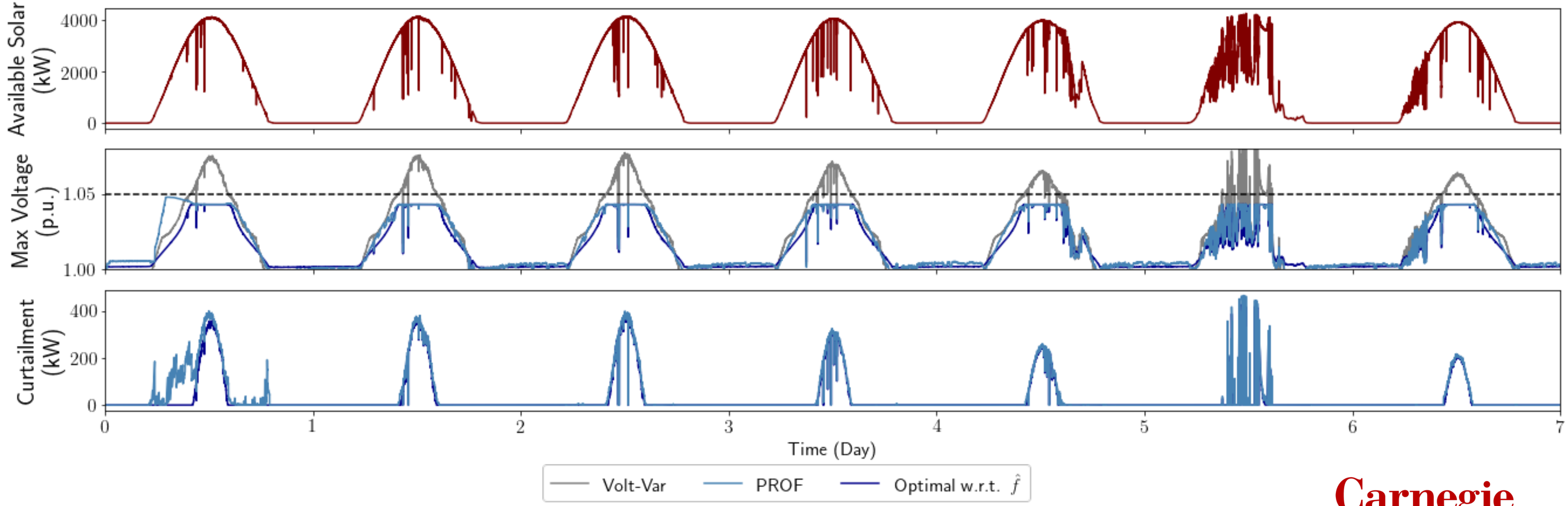
$$\mathcal{U}(k) := \mathcal{U}_1(k) \times \cdots \times \mathcal{U}_N(k).$$

Given a linearized model of the distribution network

Bolognani, S., & Dörfler, F. (2015, September). Fast power system analysis via implicit linearization of the power flow manifold. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (pp. 402-409). IEEE.

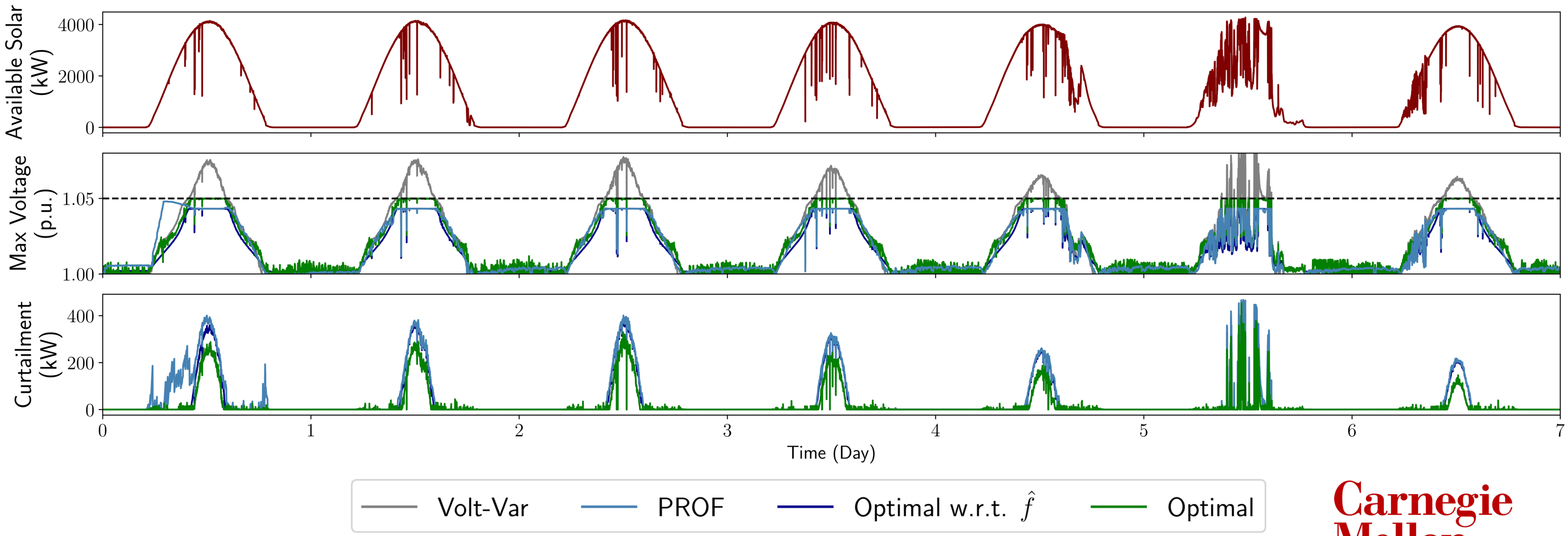
Carnegie Mellon University

PROF satisfies voltage constraints throughout the experiment and learns to minimize curtailment as well as possible within its conservative safety set.



Carnegie Mellon University

Due to the conservativeness of its safe set, PROF curtails more energy than the optimal solution, which is expensive to compute.



Carnegie Mellon University

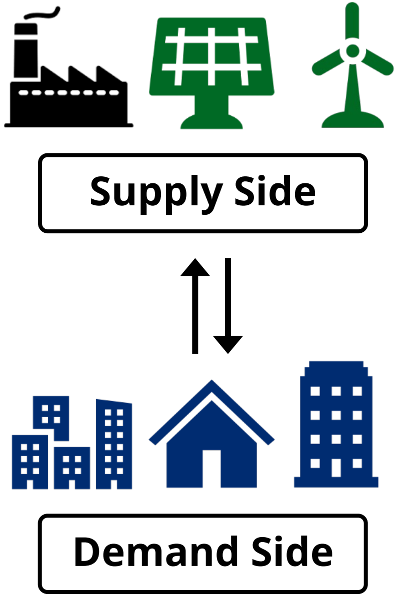
Summary

PROF is a method that enforces convex operational constraints within neural policies with a differentiable projection layer.

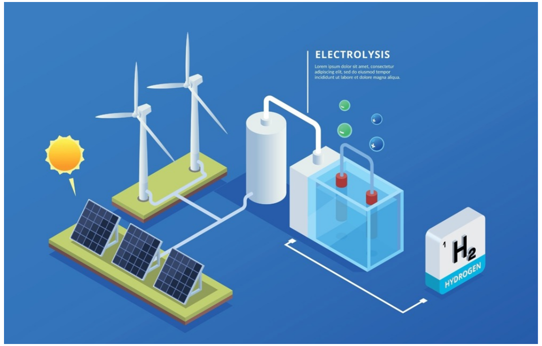
The result is a powerful neural policy that can flexibly optimize performance on the true dynamics, while satisfying constraints.

In the inverter control setting, PROF satisfies the constraints 100% of the time over more than half a million time steps.

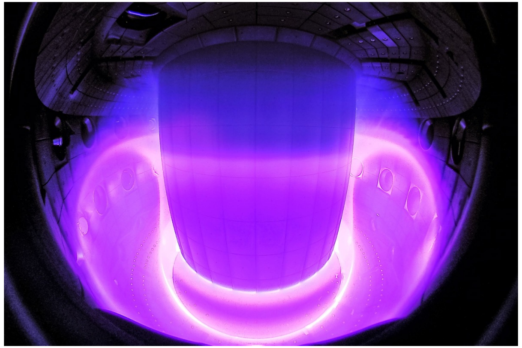
Outlook: Differentiable programming can find other applications in the design and operation of energy systems.



Fuel Cell



Electrolysis



Nuclear Fusion

...