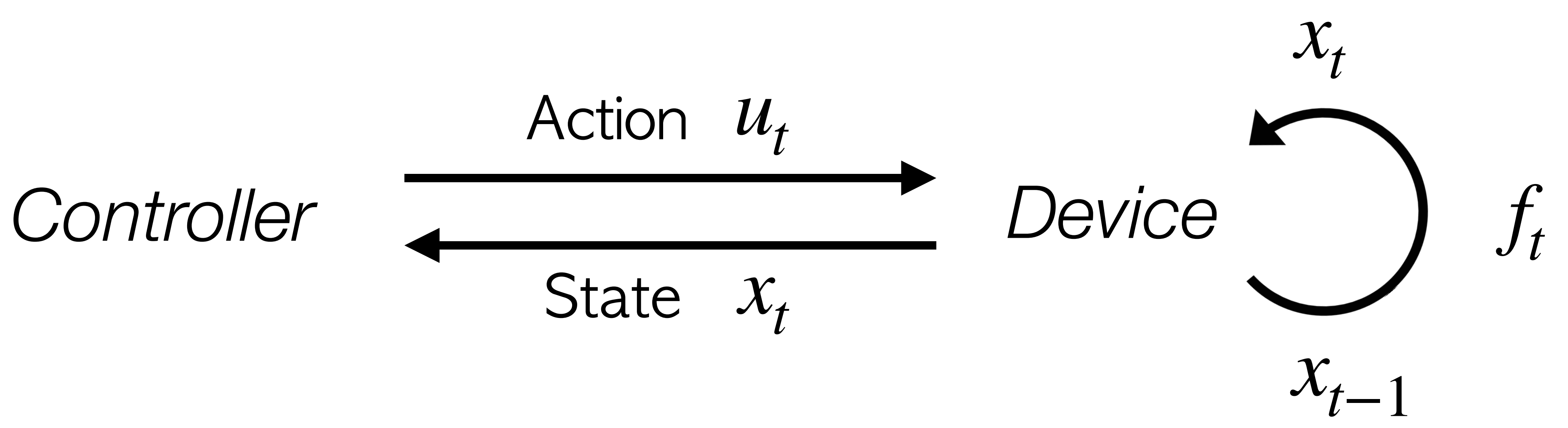
Information Aggregation for Constrained Online Control

Tongxin Li, Yue Chen, Bo Sun, Steven H. Low and Adam Wierman

Revisit Linear Control Systems

Classical Feedback Control



Simple well-studied setting:

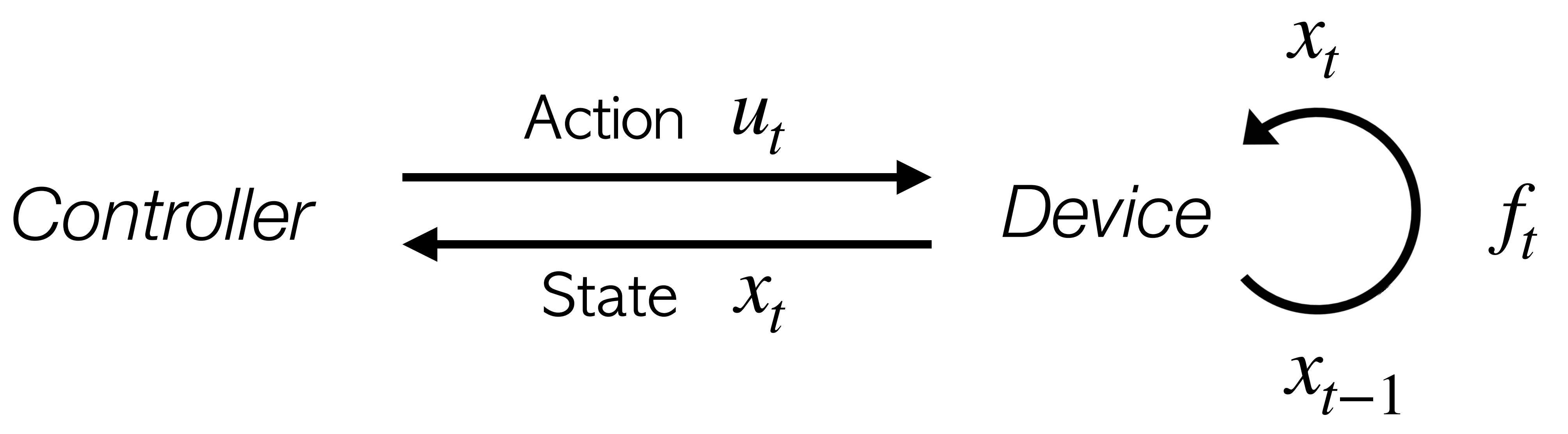
Dynamics	*Cost*	*Predictions*
$x_{t+1} = f_t(x_t, u_t) = Ax_t + Bu_t + w_t$	$\sum_{t=0}^{T-1} x_t^\top Q x_t + u_t^\top R u_t + x_T^\top Q_f x_T$	$w_t, \ldots, w_{t+k}$ at time t

Classic linear control theory — Optimal controller $x_t = -Ku_t$

Regret analysis with **predictions** — MPC is optimal [Yu 2020]

General Nonlinear Systems

Classical Feedback Control



Nonlinear setting:

Dynamics	*Cost*	*Predictions*
$x_{t+1} = f_t(x_t, u_t) = Ax_t + Bu_t + w_t$	$\sum_{t=0}^{T-1} x_t^\top Q x_t + u_t^\top R u_t + x_T^\top Q_f x_T$	$w_t, \ldots, w_{t+k}$ at time t
$x_{t+1} = f_t(x_t, u_t)$ Borel-measurable $x_t \in \mathsf{X}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t})$ $u_t \in \mathsf{U}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t})$	$\sum_{t=0}^{T-1} c_t(x_t, u_t)$ Lipschitz-continuous	$c_t, \ldots, c_{t+k}$ at time t Dynamics Constraints ?

General Nonlinear Systems

Classical Feedback Control

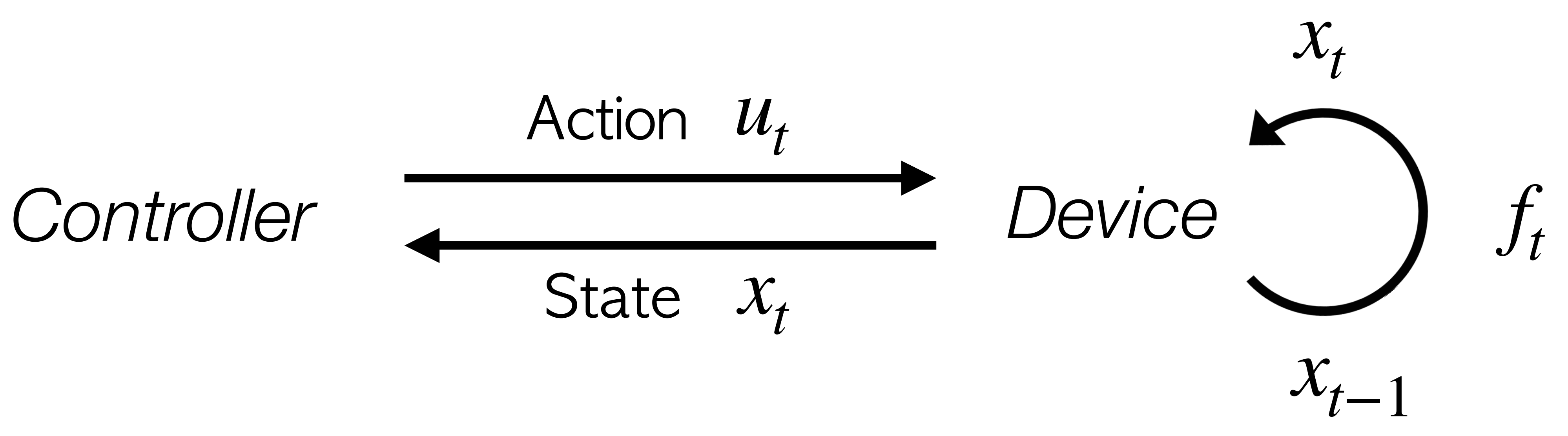
Controller $\underset{\text{State } x_t}{\overset{\text{Action } u_t}{\rightleftarrows}}$ *Device* $\circlearrowleft f_t$ ($x_{t-1} \to x_t$)

Nonlinear setting:

Dynamics	*Cost*	*Predictions*
$x_{t+1} = f_t(x_t, u_t) = Ax_t + Bu_t + w_t$	$\sum_{t=0}^{T-1} x_t^\top Q x_t + u_t^\top R u_t + x_T^\top Q_f x_T$	$w_t, \ldots, w_{t+k}$ at time t
$x_{t+1} = f_t(x_t, u_t)$ Borel-measurable $x_t \in \mathsf{X}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t})$ $u_t \in \mathsf{U}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t})$ Time-coupling	$\sum_{t=0}^{T-1} c_t(x_t, u_t)$ Lipschitz-continuous	$c_t, \ldots, c_{t+k}$ at time t $\mathsf{X}_t, \ldots, \mathsf{X}_{t+k}$ $\mathsf{U}_t, \ldots, \mathsf{U}_{t+k}$ $f_t, \ldots, f_{t+k}$ Feasibility Issues

General Nonlinear Systems

Classical Feedback Control

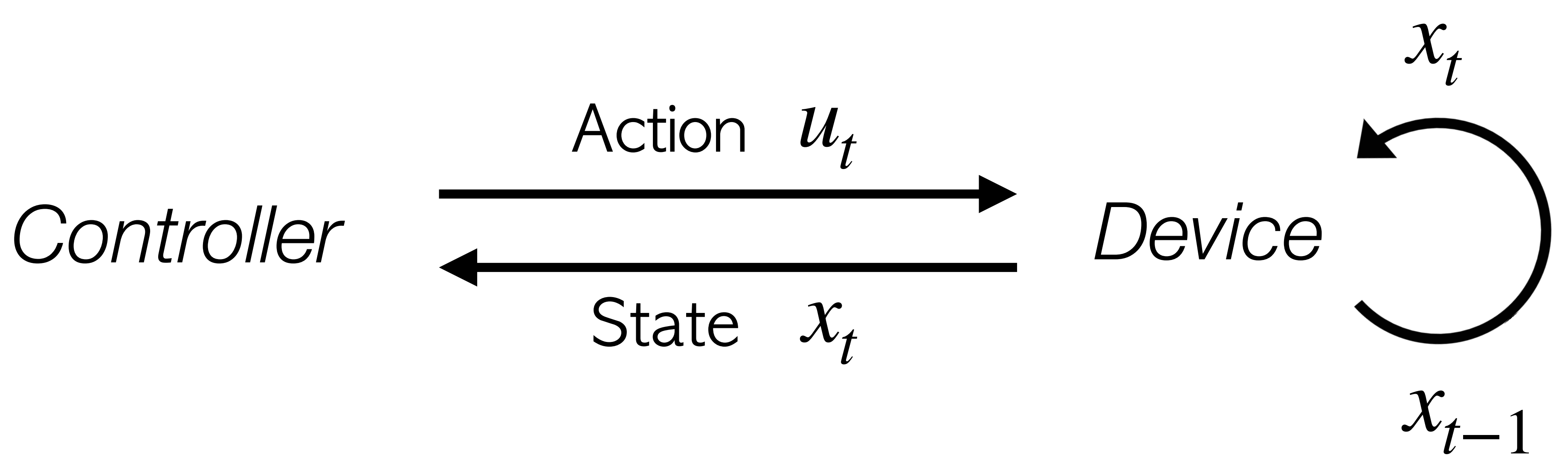


Nonlinear setting:

Dynamics	*Cost*	*Predictions*
$x_{t+1} = f_t(x_t, u_t) = Ax_t + Bu_t + w_t$	$\sum_{t=0}^{T-1} x_t^\top Q x_t + u_t^\top R u_t + x_T^\top Q_f x_T$	$w_t, \ldots, w_{t+k}$ at time t
$x_{t+1} = f_t(x_t, u_t)$ Borel-measurable $x_t \in \mathsf{X}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t})$ $u_t \in \mathsf{U}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t})$	$\sum_{t=0}^{T-1} c_t(x_t, u_t)$ Lipschitz-continuous	$c_t, \ldots, c_{t+k}$ at time t $p_t, \ldots, p_{t+k}$ ML-learned Feasibility feedback

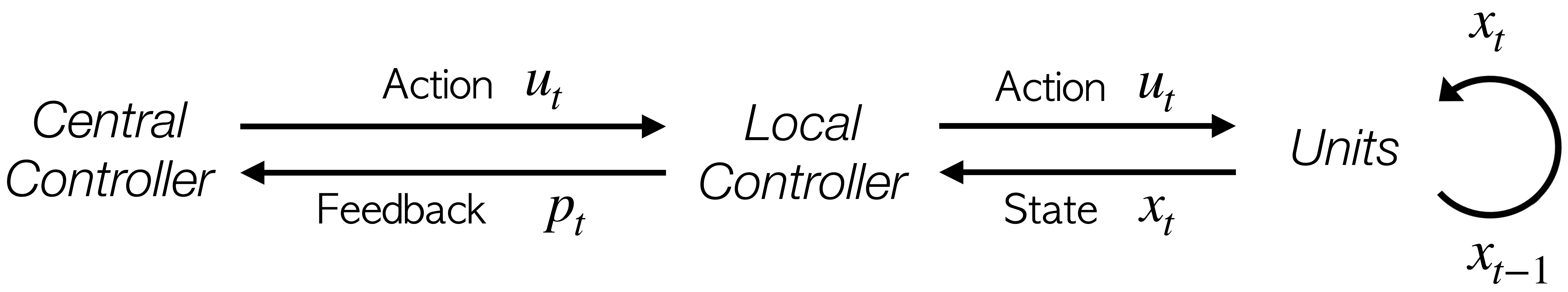
Control System Model

Classical Feedback Control

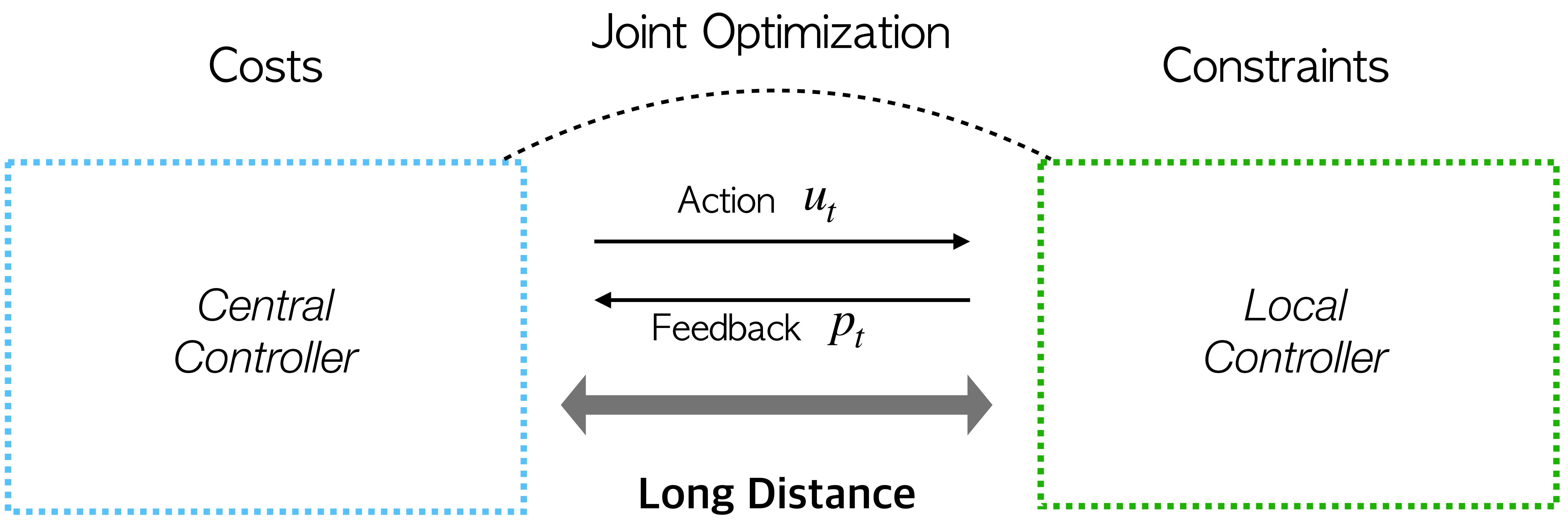


Large-scale Feedback Control

Two-controller System



Control System Model



Challenges

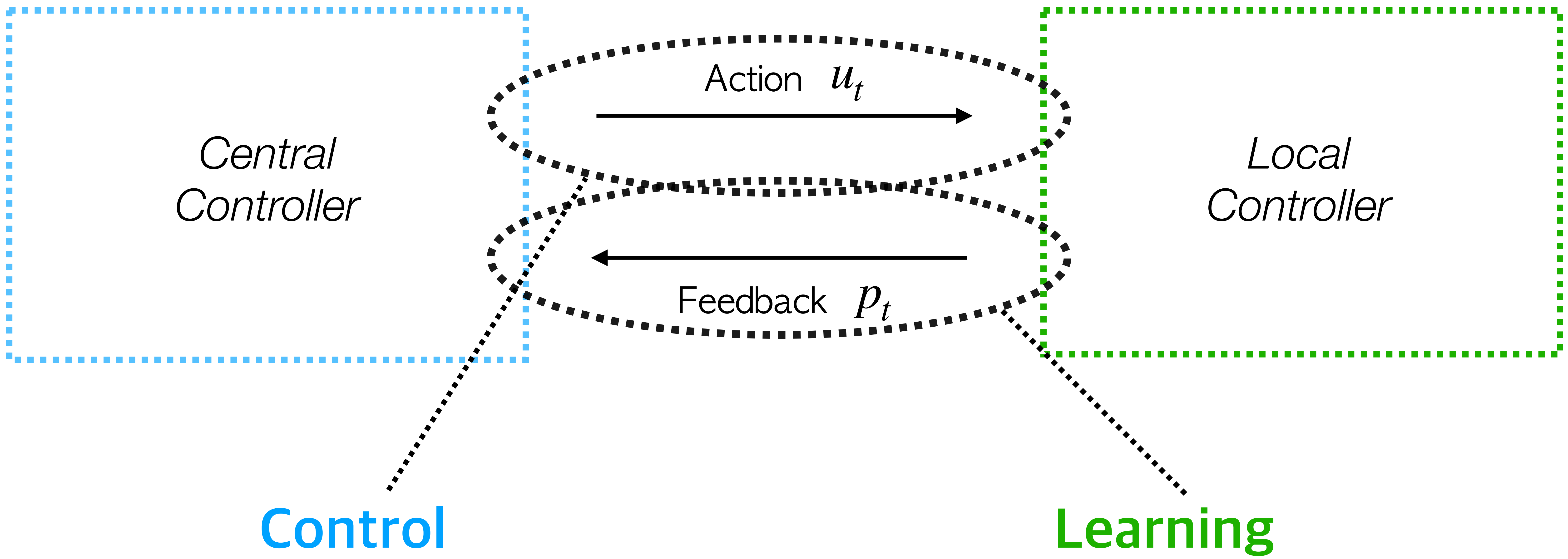
(1) Central controller cannot access dynamic and sub-system states

(2) Local controller does not know the central controller's objective

(3) Design of feedback

(4) Online; Uncertainty

This Work: Control + Learning



Assumptions

Assumptions:

(1) **Predictions** $p_t, \ldots, p_{t+w}, c_t, \ldots, c_{t+w}$ are available

(2) Action space is much smaller than state space $|\mathbb{U}| \ll |\mathbb{X}|$

(3) The dynamic $f_t(\cdot, \cdot) : \mathsf{X}_t \times \mathsf{U}_t \to \mathsf{X}_{t+1}$ is a *Borel measurable* function $\forall t \in [T]$

(4) The cost $c_t(\cdot) : \mathsf{U}_t \to \mathbb{R}_+$ are Lipschitz continuous $\forall t \in [T]$

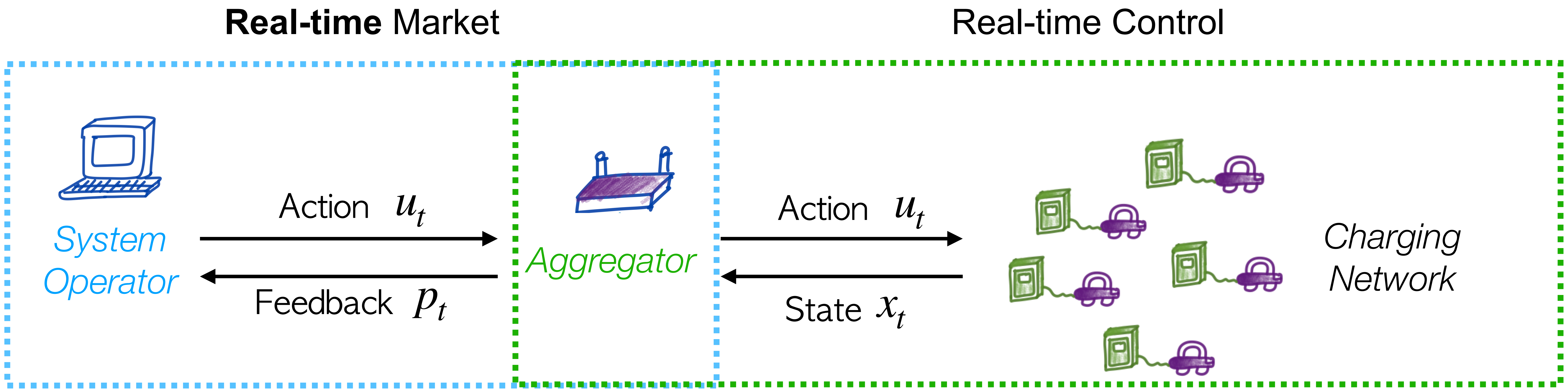
Joint Objective
of Central and Local Controllers:

$$\min_{\mathbf{u}} \sum_{t=1}^{T} c_t(u_t) \quad \text{s.t.} \quad \begin{aligned} & x_{t+1} = f_t(x_t, u_t) \\ & x_t \in \mathsf{X}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t}) \subseteq \mathbb{X} \\ & u_t \in \mathsf{U}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t}) \subseteq \mathbb{U} \end{aligned} \qquad \forall t = 0, \ldots, T$$

A Power System Example

Closed-loop Aggregator-Operator Coordination

Example



System Model

Goal: Minimize cumulative costs

$$\min_{\mathbf{u}} \sum_{t=1}^{T} c_t(u_t) \quad \text{s.t.} \quad \text{Constraints}(t), \forall t = 0, \ldots, T$$

Generic dynamical system

$$x_{t+1} = f_t(x_t, u_t)$$

$$x_t \in \mathsf{X}_t\left(\mathbf{x}_{<t}, \mathbf{u}_{<t}\right) \subseteq \mathbb{X}$$

$$u_t \in \mathsf{U}_t\left(\mathbf{x}_{<t}, \mathbf{u}_{<t}\right) \subseteq \mathbb{U}$$

Constraints(t)

System Operator Only Cares about Actions

Generic dynamical system

$$x_{t+1} = f_t(x_t, u_t)$$

$$x_t \in \mathsf{X}_t\left(\mathbf{x}_{<t}, \mathbf{u}_{<t}\right) \subseteq \mathbb{X}$$

$$u_t \in \mathsf{U}_t\left(\mathbf{x}_{<t}, \mathbf{u}_{<t}\right) \subseteq \mathbb{U}$$

$$\forall t = 0, \ldots, T$$

$$\mathsf{S} := \{\mathbf{u} \in \mathbb{U} : u_t \in \mathsf{U}_t \bigcap f^{-1}\left(\mathsf{X}_t\right), \forall t \in [T]\}$$

Set of feasible Actions

Information Aggregation (IA)

?

Maximum Entropy Feedback

Question: *How to do information aggregation in real-time?*

Our Solution

Decomposition

$$S \xrightarrow{\textit{Decomposition}} p_1, \ldots, p_T$$

Maximum Entropy-based
Real-time Aggregate Flexibility

$$\max_{p_1,\ldots,p_T} \sum_{t=1}^{T} \mathbb{H}\left(U_t \,|\, \mathbf{U}_{<t}\right)$$
$$\text{subject to } \mathbf{U} \in S$$

- Maximize the information encapsulated in $p_1, \ldots, p_T$
- The variables are conditional densities (feedback): $p_t := p_t(\cdot \,|\, \cdot) := \mathbb{P}_{X_t|X_{<t}}(\cdot \,|\, \cdot),\ t \in [T]$
- $\mathbf{U} = (U_1, \ldots, U_T) \in \mathbb{U}^T$ is a random variable distributed according to the joint distribution $\prod_{t=1}^{T} p_t$

Maximum entropy feedback

$$\mathsf{S} \xrightarrow{\textit{Decomposition}} p_1, \ldots, p_T$$

$$\max_{p_1,\ldots,p_T} \sum_{t=1}^{T} \mathbb{H}\left(U_t \,|\, \mathbf{U}_{<t}\right) \quad \text{subject to } \mathbf{U} \in \mathsf{S} \tag{1}$$

Definition (MEF)

Maximum entropy feedback (MEF) $p_1^*, \ldots, p_T^*$ if it is a unique optimal solution of (1)

- Time-based decomposition of the set of feasible action trajectories
- The variable $p_t^* = p_t^*(\cdot \,|\, \mathbf{u}_{<t})$ conditioning on previous actions is a *density function* on $\mathbb{U}$

Our Solution

$$\mathrm{S} \xrightarrow{\textit{Decomposition}} p_1, \ldots, p_T$$

$$\max_{p_1,\ldots,p_T} \sum_{t=1}^{T} \mathbb{H}\left(U_t \,|\, \mathbf{U}_{<t}\right) \quad \text{subject to } \mathbf{U} \in \mathrm{S} \tag{1}$$

Definition (MEF)

Maximum entropy feedback (MEF) $p_1^*, \ldots, p_T^*$ if it is a unique optimal solution of (1)

- *What are the properties of the MEF?*
- *How can we use the MEF to minimize cumulative costs?*

MEF Properties

Proposition (Feasibility)

For any trajectory $\mathbf{u} = (u_1, \ldots, u_T)$, *if* $p_t^*(u_t \mid \mathbf{u}_{<t}) > 0, \forall t \in [T]$, *then* $\mathbf{u} \in \mathsf{S}$

Proposition (Self-Interpretability)

For all $u_t, u_t' \in \mathbb{U}$ *at each time* $t \in [T]$, *if* $p_t^*(u_t \mid \mathbf{u}_{<t}) \geq p_t^*(u_t' \mid \mathbf{u}_{<t})$, *then*

$$\mu\left(\mathsf{S}((\mathbf{u}_{<t}, u_t))\right) \geq \mu\left(\mathsf{S}((\mathbf{u}_{<t}, u_t'))\right)$$

where $\mu(\cdot)$ *is the Lebesgue measure and the set of subsequent feasible trajectories is*

$$\mathsf{S}(\mathbf{u}_{\leq t}) := \left\{\mathbf{v} \in \mathsf{S} : \mathbf{v}_{\leq t} \equiv \mathbf{u}_{\leq t}\right\}$$

MEF Properties

Proof of the Propositions:

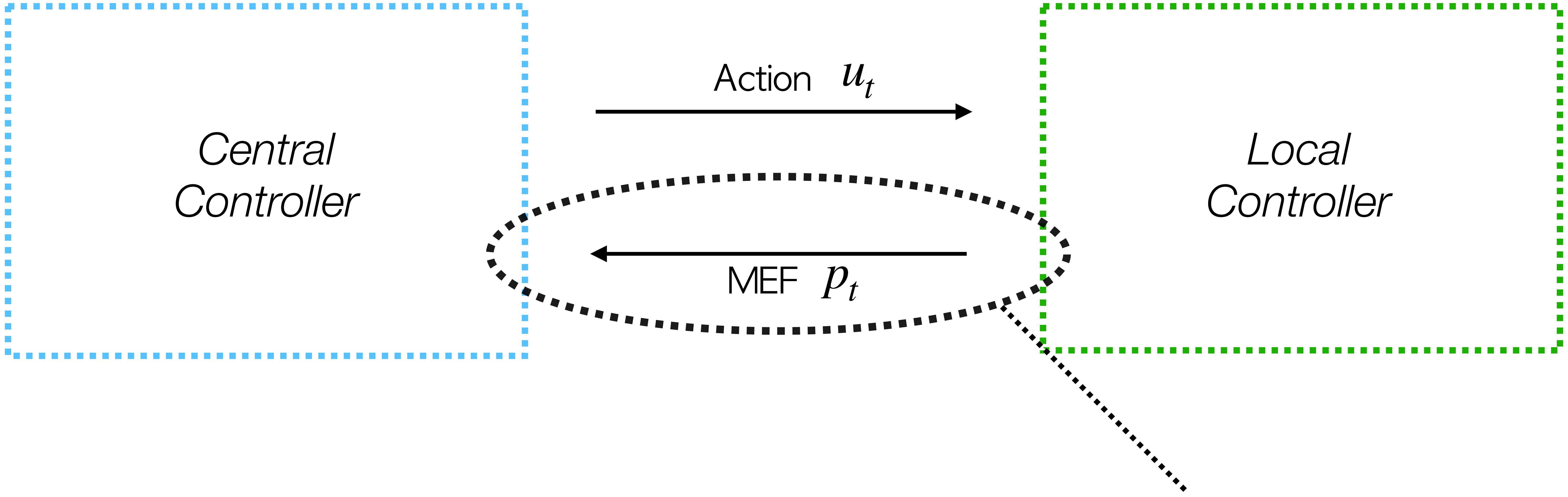
Lemma

The maximum entropy feedback (MEF) is given by

$$p_t^*(u \mid \mathbf{u}_{<t}) \equiv \frac{\mu\left(\mathsf{S}((\mathbf{u}_{<t}, u))\right)}{\mu\left(\mathsf{S}(\mathbf{u}_{<t})\right)}, \quad \forall u \in \mathbb{U} \textbf{ and } \mathbf{u}_{<t} \in \mathbb{U}^{t-1}.$$

(Assume feasible sets are atomic)

This Work: Control + Learning

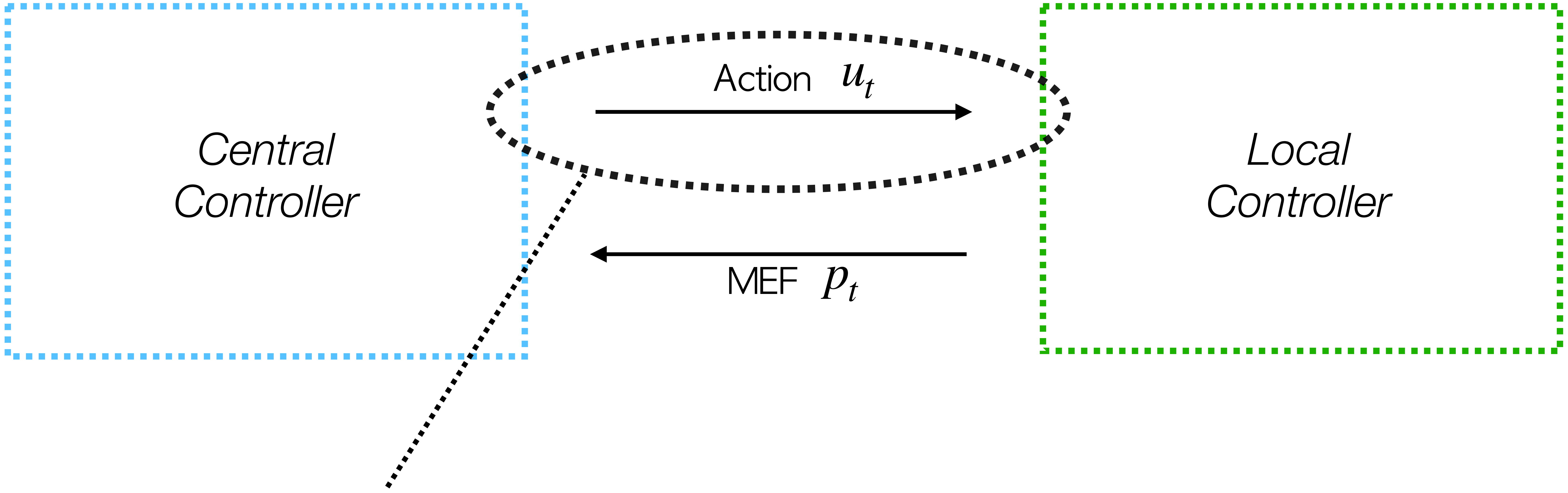


Control

Learning

Reinforcement Learning

This Work: Control + Learning



Control

Learning

Penalized Predictive Control

Penalized Predictive Control

Model Predictive Control (time τ)

$$\min_{\mathbf{u}_{t:t+w}} \sum_{\tau=t}^{t+w} c_t(u_t)$$

s.t. $\forall \tau = t, \ldots, t+w$

$$x_{\tau+1} = f_\tau(x_\tau, u_\tau)$$

$$x_\tau \in \mathsf{X}_\tau\left(\mathbf{x}_{<\tau}, \mathbf{u}_{<\tau}\right) \subseteq \mathsf{X}$$

$$u_\tau \in \mathsf{U}_\tau\left(\mathbf{x}_{<\tau}, \mathbf{u}_{<\tau}\right) \subseteq \mathsf{U}$$

VS

Penalized Predictive Control (time τ)

$$\min_{\mathbf{u}_{t:t+w}} \sum_{\tau=t}^{t+w} c_t(u_t) - \beta \log p_\tau^*(u_\tau \,|\, \mathbf{u}_{<\tau})$$

s.t. $\forall \tau = t, \ldots, t+w$

$$u_\tau \in \mathsf{U}$$

Tuning Parameter $\beta > 0$

Need to know the dynamics, constraints and the states

Only need to know the feedback

Need to solve a large-scale optimization

Variable Bits: $\Omega(w \log |\mathbb{X}|)$

$|\mathbb{U}| \ll |\mathbb{X}|$

Variable Bits: $O(w \log |\mathbb{U}|)$

Penalized Predictive Control

Theorem (Feasibility)

For any predication window size $w \geq 1$*, the sequence of actions* $\mathbf{u} = (u_1, \ldots, u_T)$ *generated by the penalized predictive control always satisfies* $\mathbf{u} \in \mathsf{S}$

Penalized Predictive Control

Dynamic Regret

$$\text{Regret}(\mathbf{u}) := \sup_{\mathbf{c}\in\mathsf{C}} \sup_{\mathsf{S}} C_T(\mathbf{u}) - \inf_{\mathbf{u}\in\mathsf{S}} C_T(\mathbf{u})$$

$$C_T(\mathbf{u}) := \sum_{t=1}^{T} c_t(u_t)$$

Theorem (Dynamic Regret; Informal)

Without assumptions, any sequence of actions generated by a deterministic online policy that can access the sets $\{\mathsf{U}_t : t \in [T]\}$ *and* $\{\mathsf{X}_t : t \in [T]\}$ *with a prediction window size* $w \geq 1$ *gives* $\text{Regret}(\mathbf{u}) = \Omega(d(T-w))$ *where* d *is the diameter of* U

Penalized Predictive Control

$\mathsf{S}_w(\mathbf{u}) :=$ *Set of length-w subsequent actions given* $\mathbf{u}$

Causal Invariance

There exists constants $\delta, \lambda > 0$ *such that*

(1) For any t and $\mathbf{u}_{\leq t}, \mathbf{v}_{\leq t}$

(Hausdorff distance) $$d_{\mathsf{H}}\left(\mathsf{S}_w(\mathbf{u}_{\leq t}), \mathsf{S}_w(\mathbf{v}_{\leq t})\right) \leq \delta \left(\frac{\left|\mu(\mathsf{S}(\mathbf{u}_{\leq t})) - \mu(\mathsf{S}(\mathbf{v}_{\leq t}))\right|}{\mu\left(\mathscr{B}\right)} \right)^{1/((T-t)m)} \quad (\mathsf{U} \subseteq \mathbb{R}^m)$$

(Unit ball)

(2) For any t and $\mathbf{u}_{\leq t}$

$$\frac{\mu\left(\mathsf{S}\left(\mathbf{u}_{\leq t}\right)\right)}{\mu\left(\mathsf{S}\left(\overline{\mathbf{u}}_{\leq t}\right)\right)} \leq \lambda \left(\frac{\mu\left(\mathsf{S}\left(\left(\mathbf{u}_{\leq t}, \overline{\mathbf{u}}_{t+1:t+w}\right)\right)\right)}{\mu\left(\mathsf{S}\left(\overline{\mathbf{u}}_{\leq t+w}\right)\right)} \right)^{\frac{T-t}{T-t-w}} \quad \left(\overline{\mathbf{u}}_{t+1:t+w} := \operatorname{argsup}_{\mathbf{u} \in \mathsf{U}^w} \mu\left(\mathsf{S}_w\left(\mathbf{u}_{\leq t}, \mathbf{u}\right)\right)\right)$$

Theorem (Dynamic Regret; Informal)

Under causal invariance assumptions, the sequence of actions generated by the penalized predictive control always satisfies $\mathsf{Regret}(\mathbf{u}) = O(dT/w^{1/4})$

$$w = \omega(1) \implies \mathsf{Regret}(\mathbf{u}) = o(T)$$

Penalized Predictive Control

Inventory constraints

$$\sum_{t=1}^{T} ||u_t||_2^2 \leq \gamma \qquad u_t \in \mathbb{R}^m$$

The inventory constraints are causally invariant with $\delta = \lambda = 1$

Tracking constraints

$$\sum_{t=1}^{T} |u_t - y_t|^p \leq \sigma \qquad u_t, y_t \in \mathbb{R}, p \geq 2$$

The tracking constraints are causally invariant with $\delta = \frac{2}{\sqrt{\pi}}, \lambda = 1$

Penalized Predictive Control

Intuition

$$\boxed{\begin{aligned}\min_{\mathbf{u}} \sum_{t=1}^{T} c_t(u_t)\\ \text{subject to} \quad \mathbf{u} \in \mathsf{S}\end{aligned}} \iff \min_{\mathbf{u}} \sum_{\tau=1}^{T} c_\tau(u_\tau) - \beta \log p^*_\tau(u_\tau \mid \mathbf{u}_{<\tau})$$

Action-value function

$$\min_{\mathbf{u}} \sum_{\tau=t}^{t+w} c_\tau(u_\tau) \boxed{- \beta \log p^*_\tau(u_\tau \mid \mathbf{u}_{<\tau})}$$

Experiments

Adaptive Charging Network (ACN)

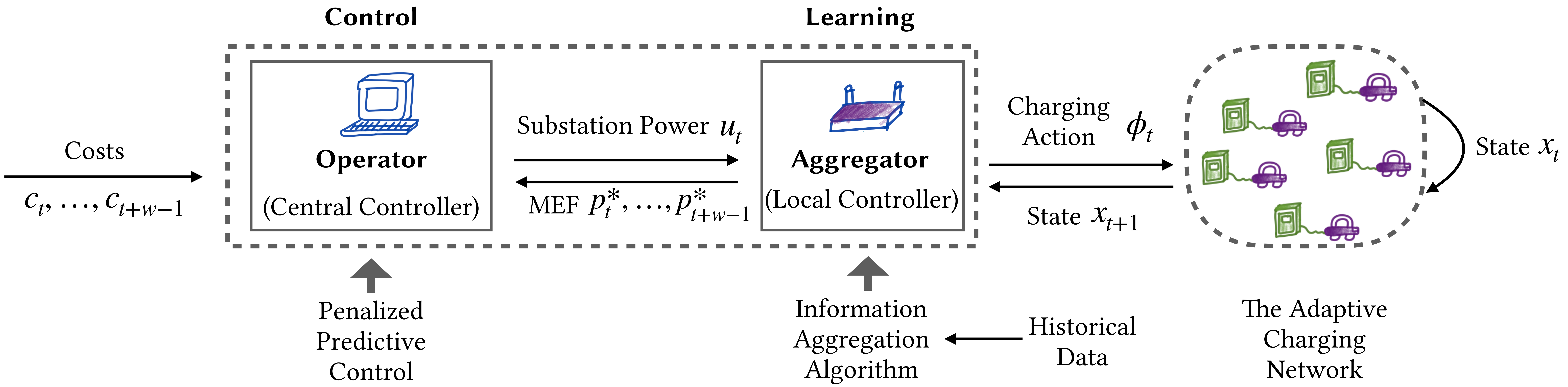
ev.caltech.edu

State Space Dimension (Number of EVSEs)	
Caltech Garage:	54
JPL Workplace:	52

Action Space Dimension (Number of Power levels)
10

Experiments

Closed-loop Aggregator-Operator Coordination

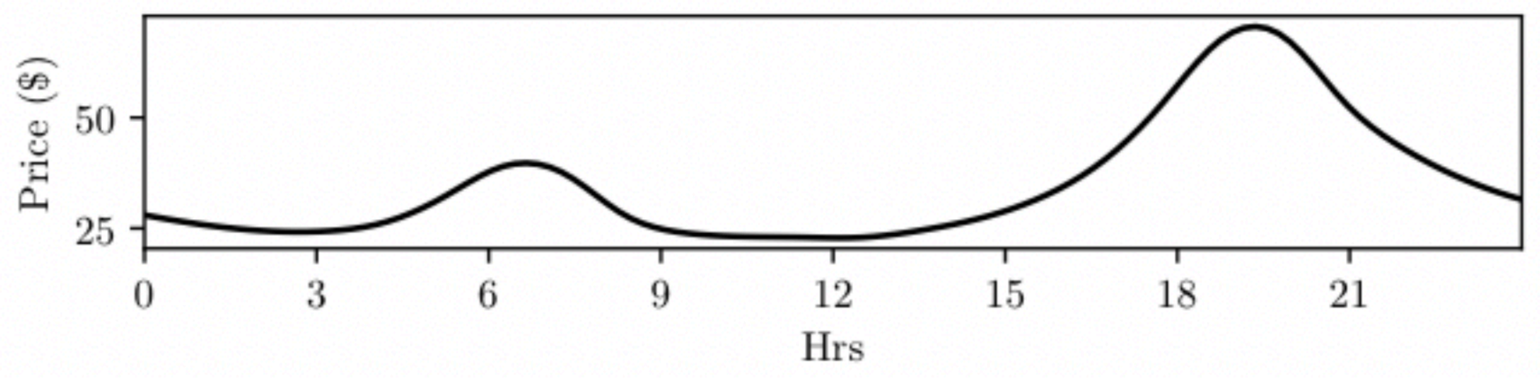


Experiments

$\phi_t(j)$	*Scheduling*
$a(j)$	*Arriving time*
$d(j)$	*Departure time*
$e(j)$	*Energy to be delivered*
$r(j)$	*Charging rate limit*
c_t	*Electricity price at each time*
n	*Number of accepted EVs*

Operator Objective

$$\min_{\mathbf{u}} \sum_{t=1}^{T} c_t \times u_t$$



EV Constraints

$$\phi_t(j) = 0 \ , \ t < a(j), \ \ j = 1,\ldots,n,$$ *(EV arrives)*

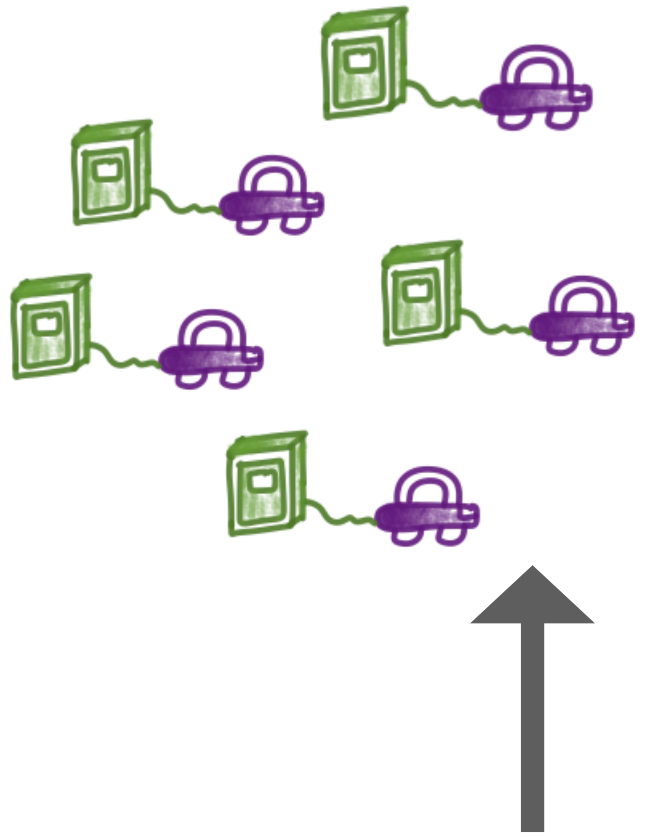
$$\phi_t(j) = 0 \ , \ t > d(j), \ \ j = 1,\ldots,n,$$

$$\sum_{j=1}^{n} \phi_t(j) = u_t, \ \ t = 1,\ldots,T,$$ *(Assigning Aggregate Power)*

$$\sum_{t=1}^{T} \phi_t(j) = e(j), \ \ j = 1,\ldots,n,$$ *(Charging) (Linear and lossless)*

$$0 \le \phi_t(j) \le r(j), \ \ t = 1,\ldots,T$$ *(Rate Limit)*

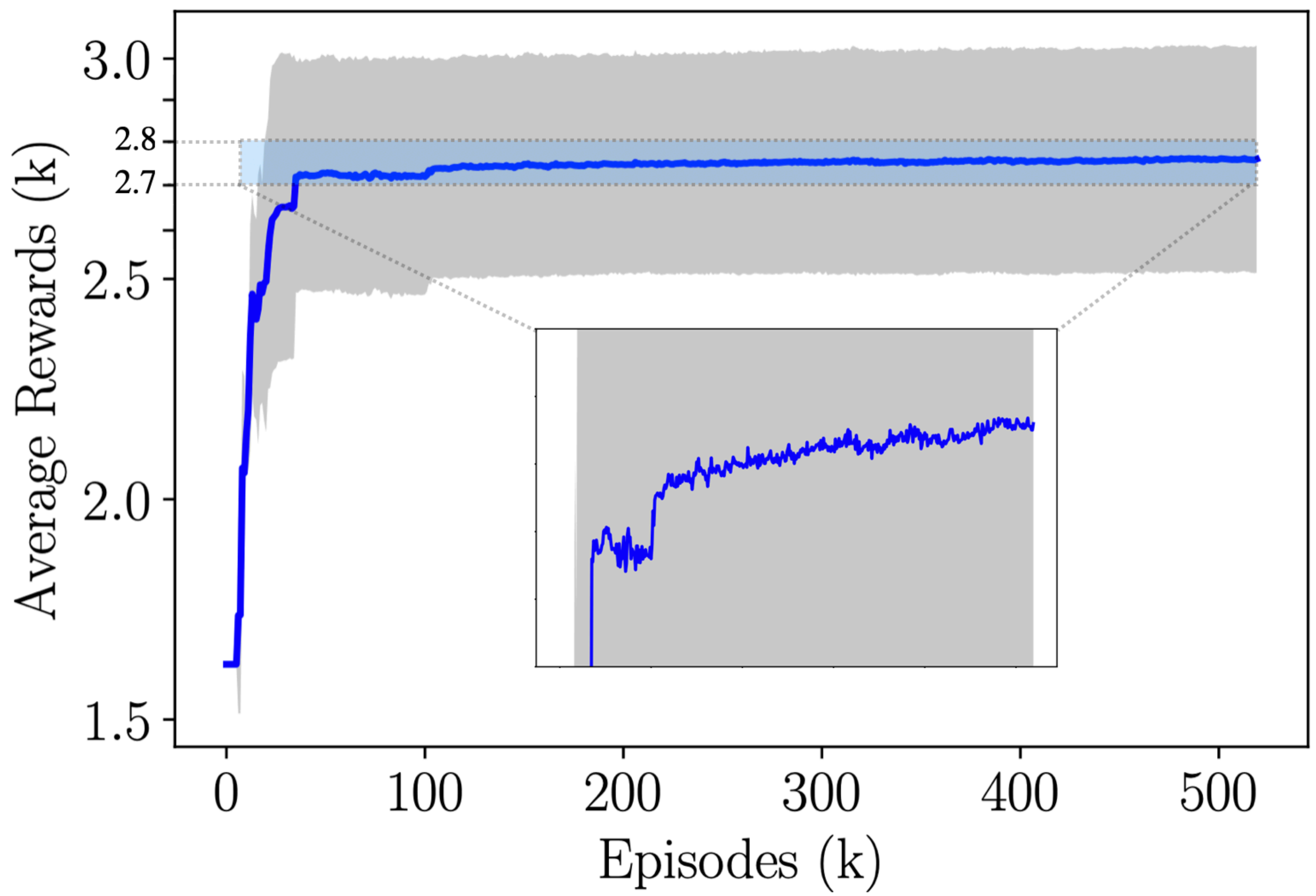
Experiments



The Adaptive Charging Network

Training: Nov. 18 - Nov. 19

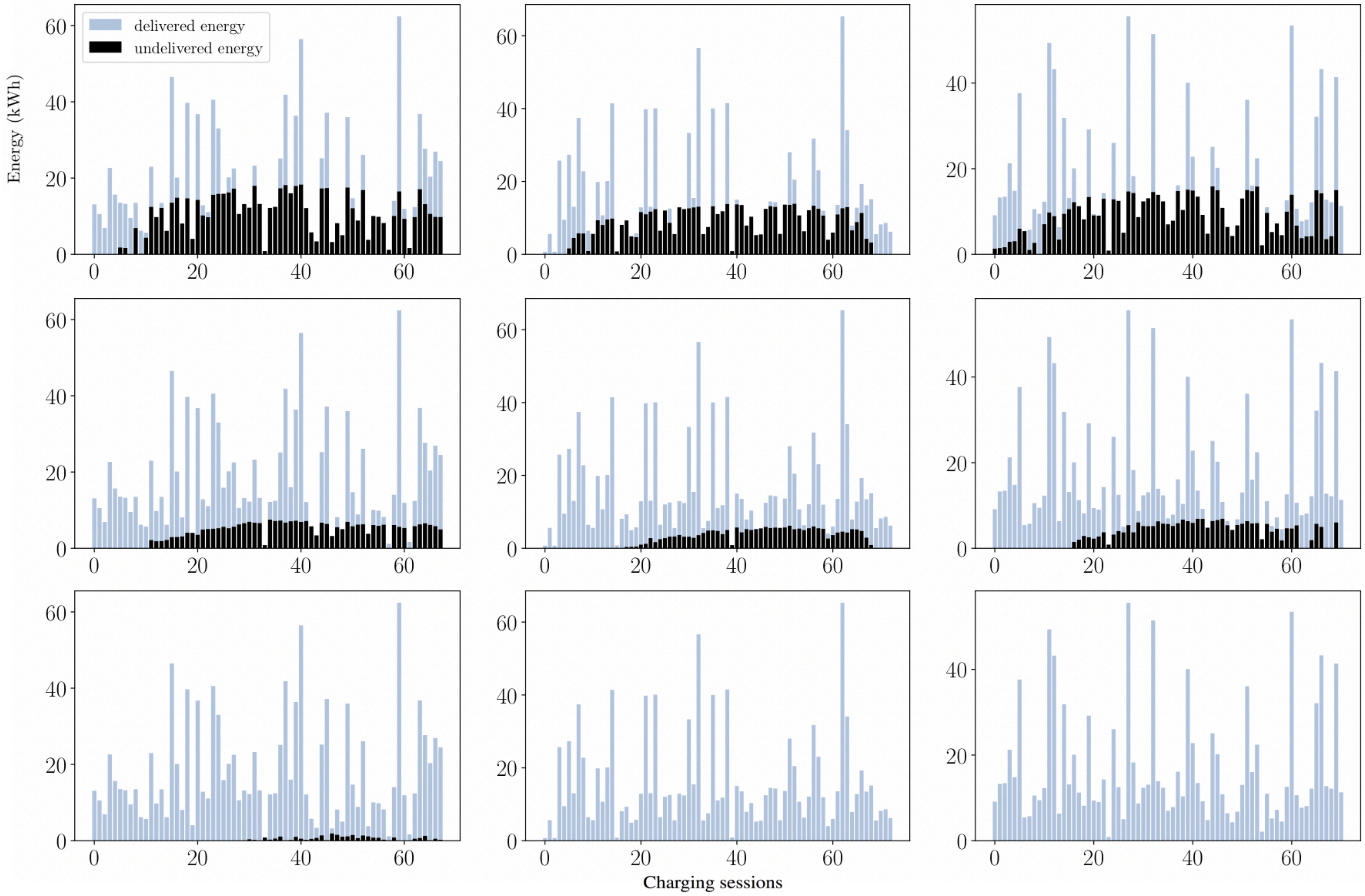
Testing: Dec. 19 - Jan. 20



RL Reward function

MEF Definition + Charging Performance + Match Operator's Action

Experiments

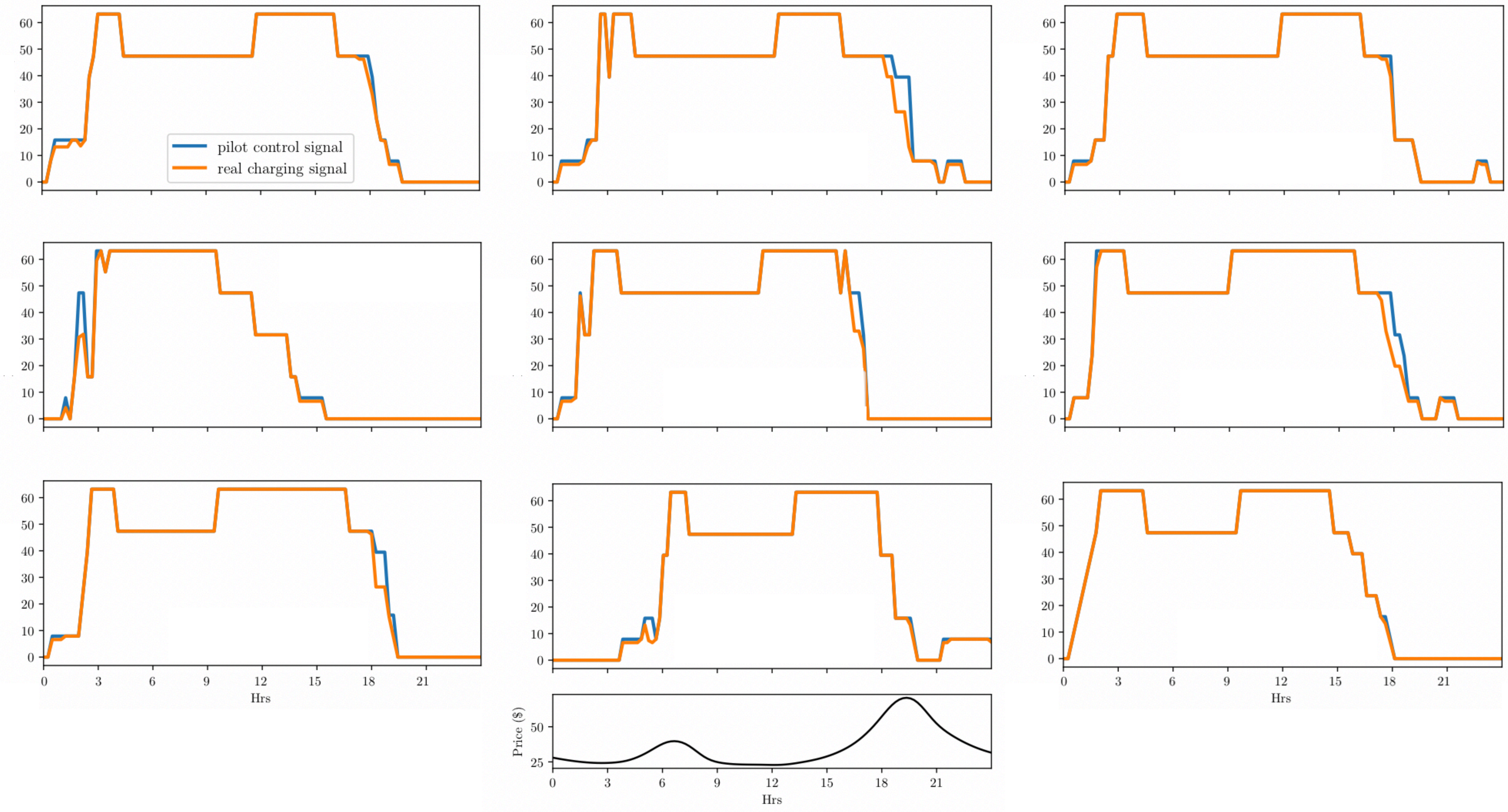


$\beta = 2 \times 10^3$

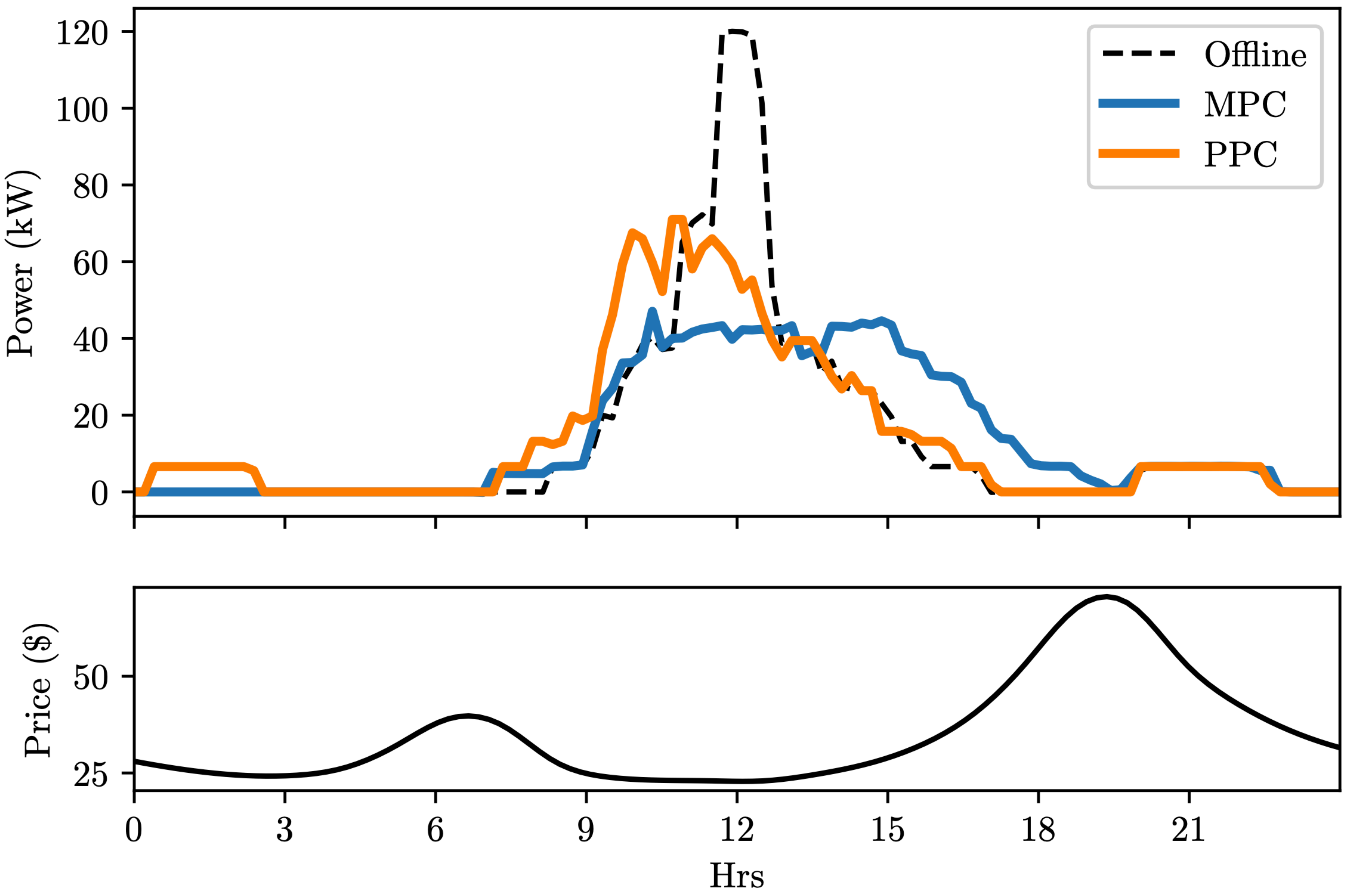
$\beta = 4 \times 10^3$

$\beta = 6 \times 10^3$

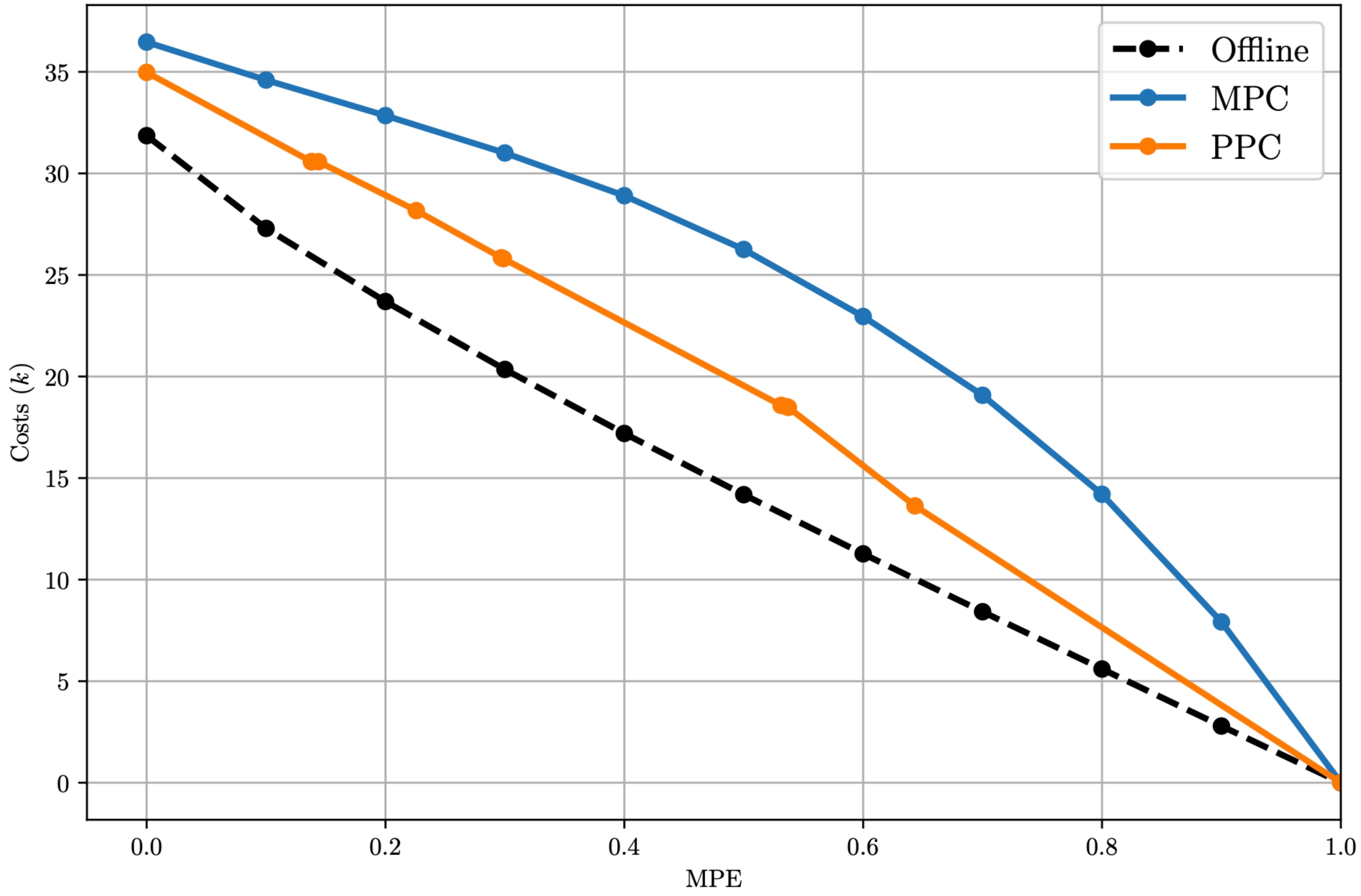
Experiments



Experiments



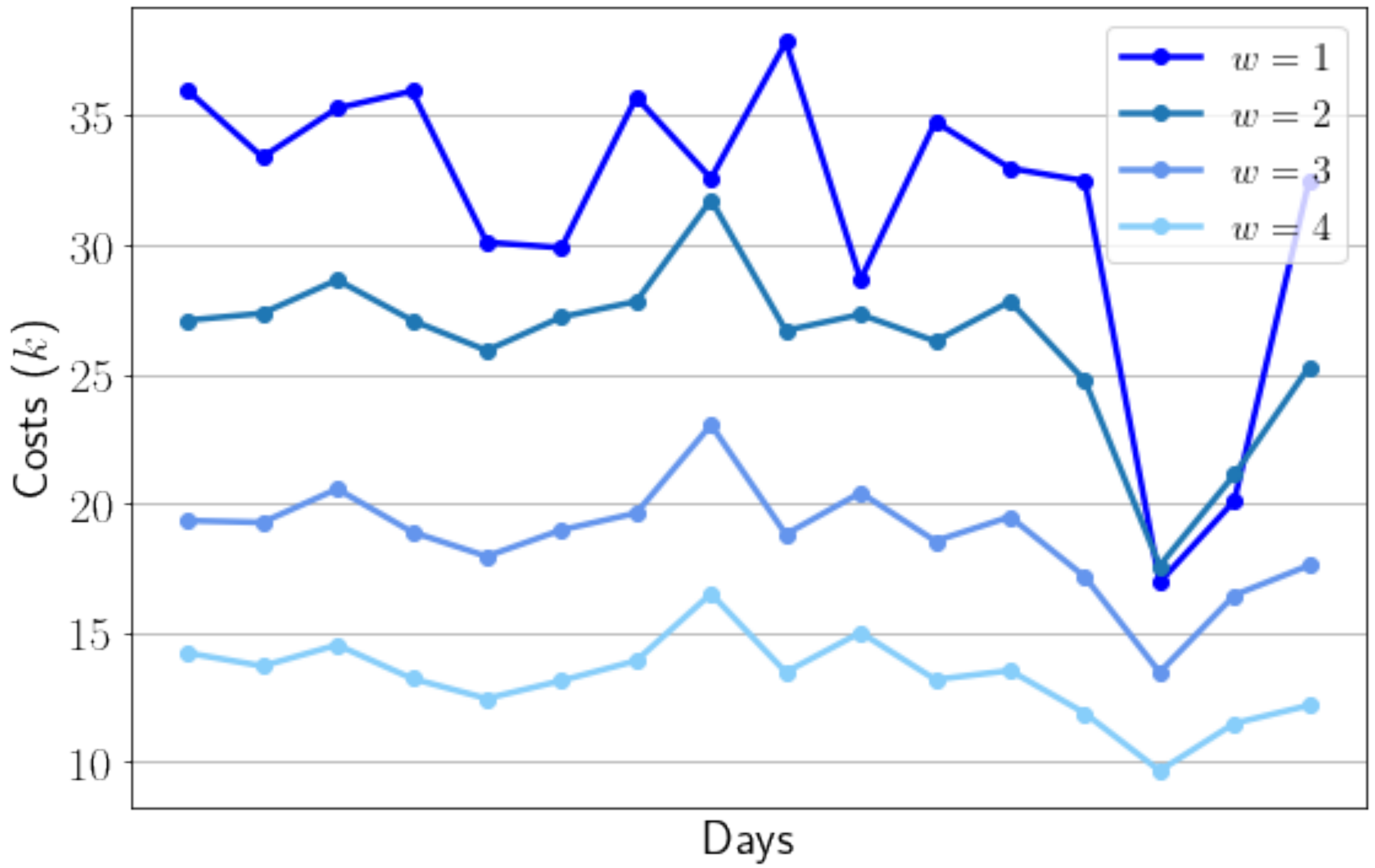
Experiments



$$\mathrm{MPE}(\phi, \mathbf{x}) := 1 - \sum_{k=1}^{L} \sum_{t=1}^{T} \sum_{j=1}^{N} \phi_t^{(k)}(j) \Big/ \left((L \times T) \cdot \sum_{j=1}^{N} e(j) \right) \times \%$$

Charged Energy *Total Energy*

Experiments



Cumulative costs for different prediction window sizes (16 Days)

Thank You

For Your Attention

tongxin@caltech.edu