香 港 大 學
THE UNIVERSITY OF HONG KONG

Energy Digitalization Laboratory

Learning to Optimize for Urban Energy Systems

Yi Wang

yiwang@eee.hku.hk

Assistant Professor, EEE

The University of Hong Kong

June 2024

About Myself

Appointment

2021.9- Assistant Professor, The University of Hong Kong

2019.2-2021.8 Postdoc, ETH Zurich (Prof. Gabriela Hug)

Education

2014.9-2019.1 Ph.D., Tsinghua University (Prof. Chongqing Kang)

2017.3-2018.4 Visiting Student, University of Washington (Prof. Daniel Kirschen)

2010.9-2014.6 B.S., Huazhong University of Science and Technology

Research Interests

Data analytics for smart energy

Cyber-physical power and energy systems

Multi-energy systems integration

Research Group: EDL@HKU

- Energy Digitalization Laboratory at The University of Hong Kong (EDL@HKU) focuses on the digitalization of power and energy systems with an emphasis on the distribution and consumer side, including **data analytics, data privacy, cyber-physical-social systems, Internet-of-things**, etc. The overall goal is to make the distribution systems more adaptive to accommodate the high penetration of renewable energy toward a decarbonized future.

In addition to publishing research papers, we develop/provide:

- Software
- Hardware
- Technical reports
- Policy recommendations

Research Topics in EDL@HKU

- **Topic 1: Urban Energy Systems**
- Multi-energy systems
- Building energy systems
- Long-term Storage

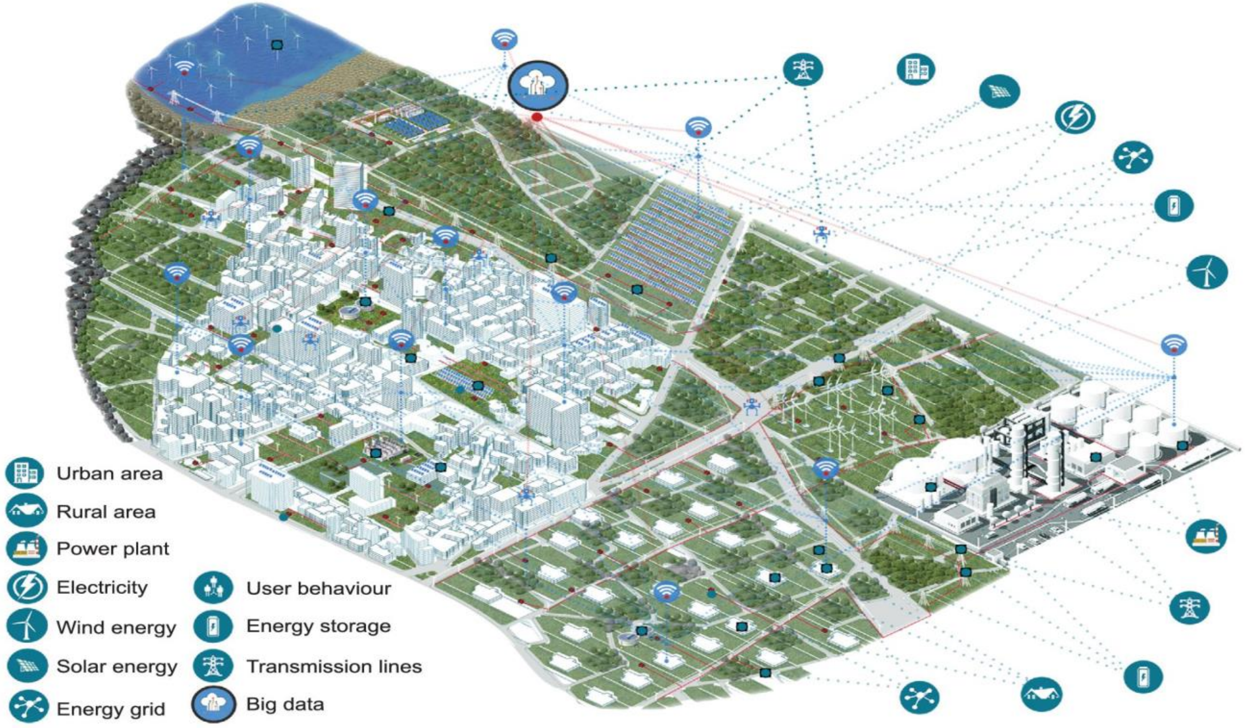
- **Topic 2: Demand Response**
- Virtual power plants
- Electric vehicles
- Internet data center/5G base station

- **Topic 3: Data Analytics in Energy Systems**
- Energy Forecasting
- Privacy-preserving analytics
- Data valuation and pricing

Contents

Introduction

- **Urban Energy Systems**
  - **Urban energy systems are multi-variant cyber-physical systems.**
  - **The system complexity increases by higher integration of decentralized renewable energy generation, making the energy flow more complex.**



Natl Sci Rev, Volume 8, Issue 3, March 2021, nwaa134, https://doi.org/10.1093/nsr/nwaa134

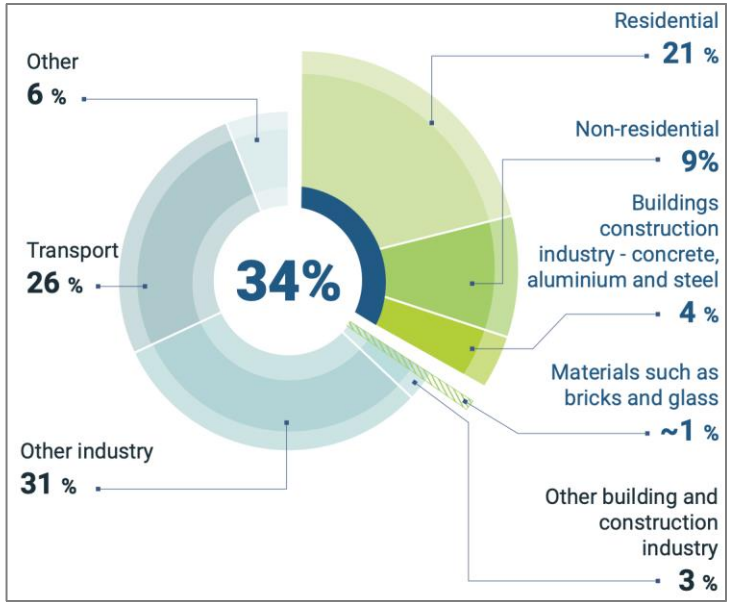
Introduction

- The Role of Building Energy Systems

➢ ***High energy consumption amounts***

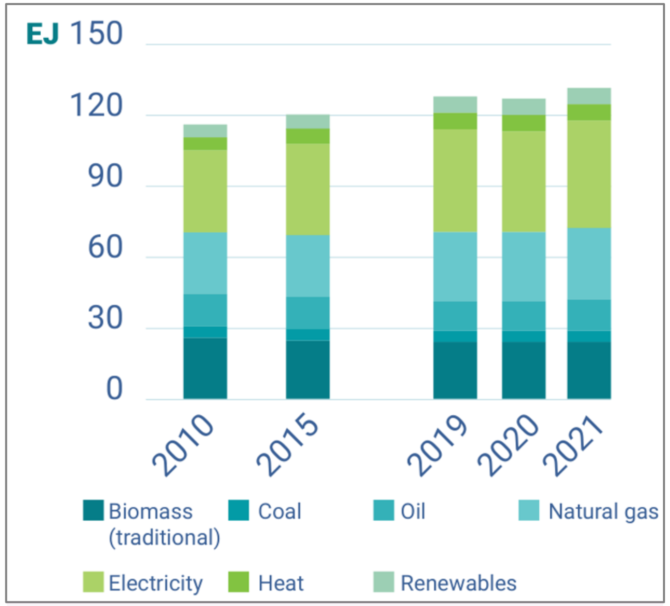
➢ ***High energy flexibility potential***

Buildings consume more than 30% of global final energy



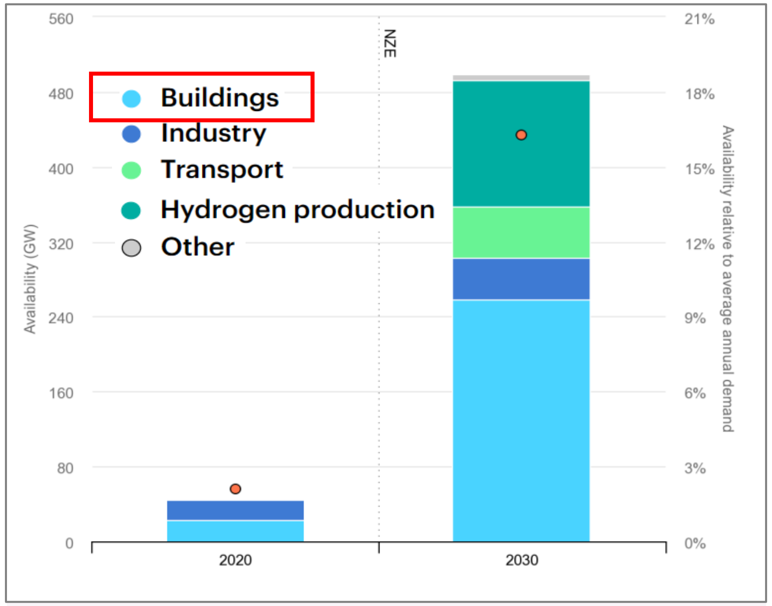
Global share of buildings and construction final energy demand, 2021

About 34% of energy consumption in buildings is through electricity



Energy consumption in buildings by fuel, 2010-2021

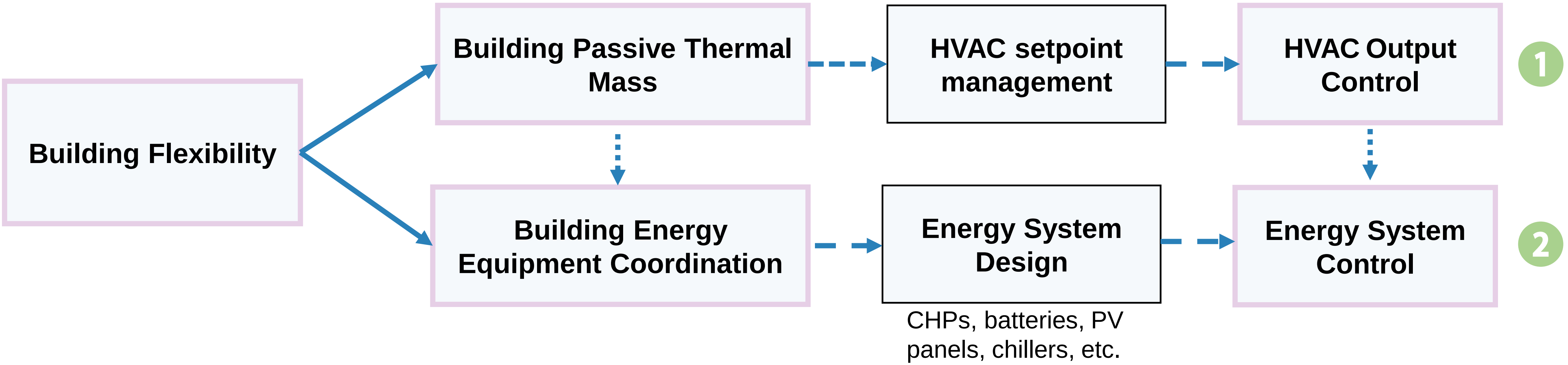
About 10% of total energy consumption are flexible loads



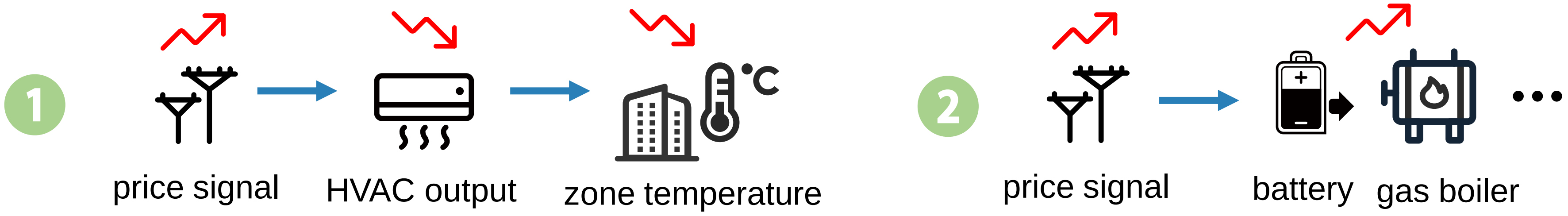
Demand response availability in the Net Zero Scenario, 2020 and 2030

Introduction

- Flexible Resources in Building Energy Systems



For example,



Introduction

○ Control Framework for Flexibility Utilization

Most widely-used due to comprehensive performance

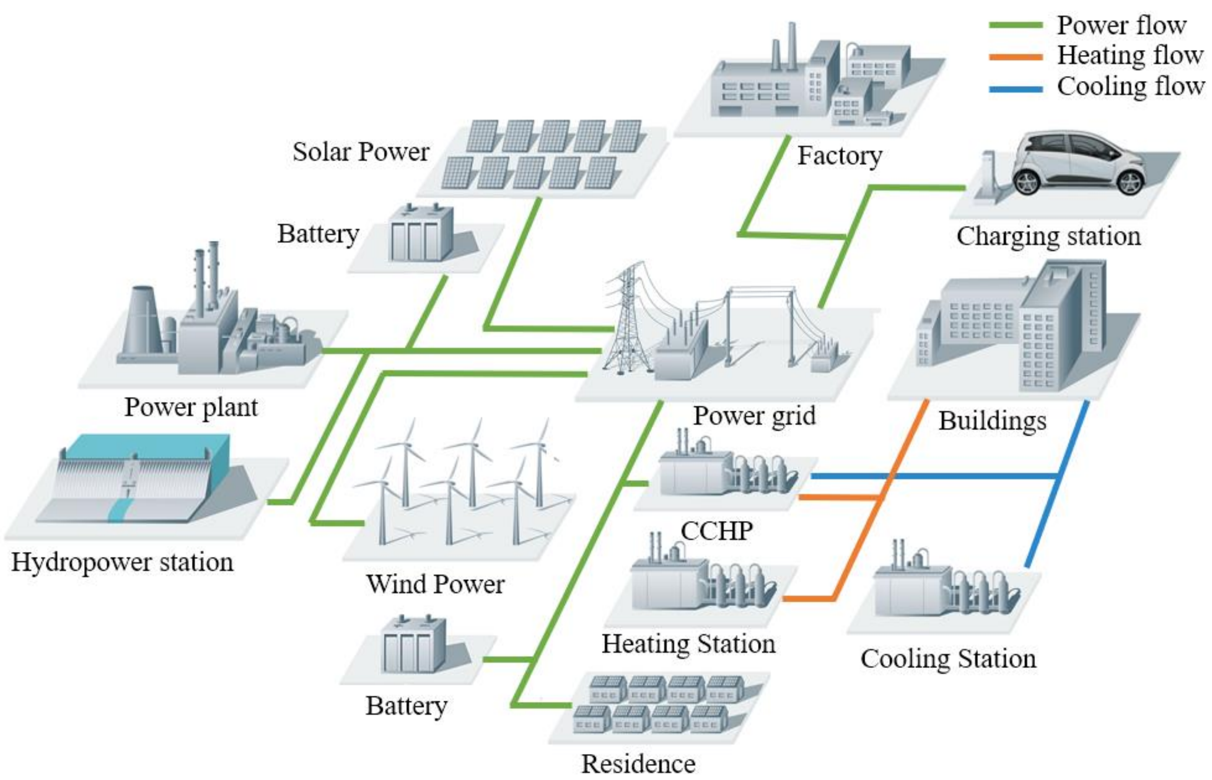
Model Predictive Control (MPC)	Black-Box Control (usually reinforcement learning)	Rule-Based Control (RBC) (most buildings used in practice)
• Robust • Model-Based • Computational Burden	• Easy to be implemented • Model-free • High demand for collected data	• Based on operator's expertise and knowledge • Model-free

Required models

Thermal Model			Energy Model	
White-Box Models	Simulation software	• Complicated • Detailed	Detailed Modeling	• Close to reality • Less generalized
Grey-Box Models	Resistance–capacitance (RC)	• Robust • Additional information required • Not sufficiently accurate		
Black-Box Models	Traditional neural network	• More accurate • Less robust • Demand for a rich dataset	Generalized Modeling (Energy Hub model)	• More generalized • Omit HVAC details

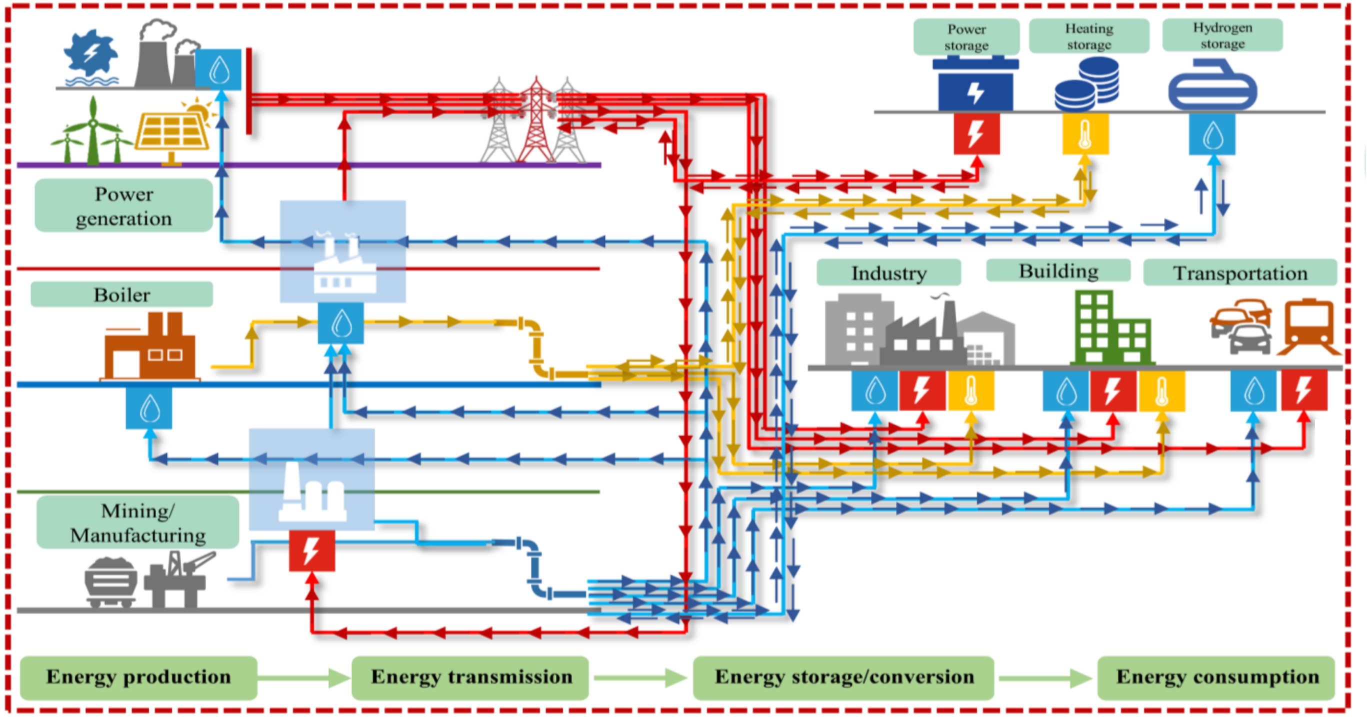
Introduction

- **Multi-energy Systems (MES)**
  - The integration of the generation, transmission, storage and consumption of electricity, heat, cooling and gas and other energy subsystems.
  - Overall energy efficiency can be enhanced, and cross-sector flexibility can be explored within MES.



Introduction

- **Data Barrier in Multi-energy Systems**
  - Power, gas, and heat/cooling load data are probably owned by different system operators **separately**.
  - These data owners tend to prioritize their **own economic benefits** over social benefits when making decisions.
  - Only by reasonably valuing the data, will they be willing to share their data set .



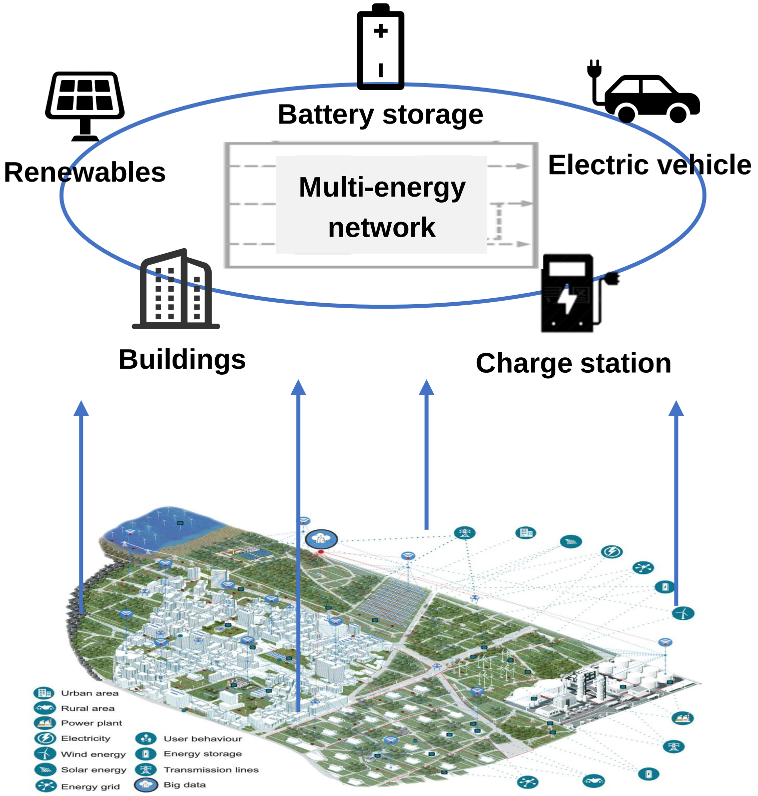
Introduction

Learning and Optimization

- **Learning and optimization are typical and powerful tools that are widely adopted in urban energy systems.**

- Topology identification, capacity evaluation, optimal power flow…
- Consumption behavior modeling, solar energy forecasting…
- Load forecasting, thermal dynamics modeling, system identification…

Typical research with Learning



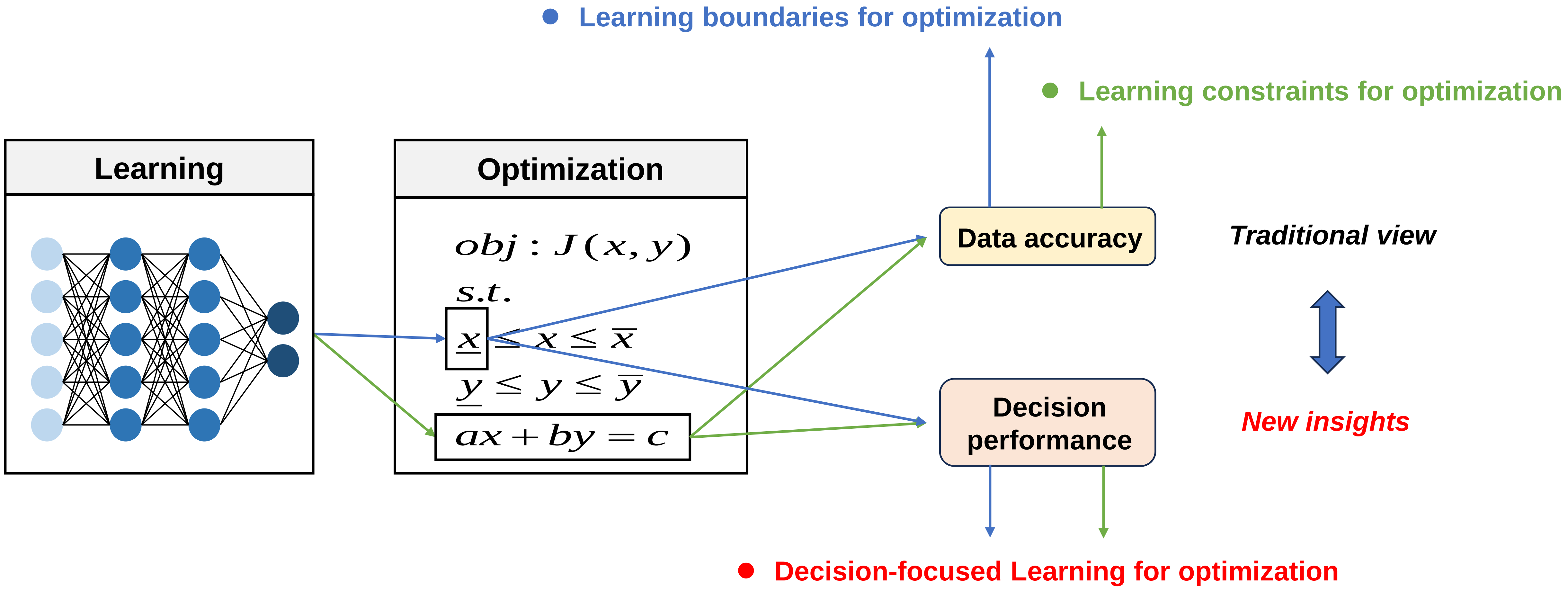
Urban energy systems

- Bilevel energy trading, flexibility region aggregation…
- Peer-to-peer trading, net-zero building, storage arbitrage…
- Demand response, model predictive control, multi-energy optimization…

Typical research with Optimization

Introduction

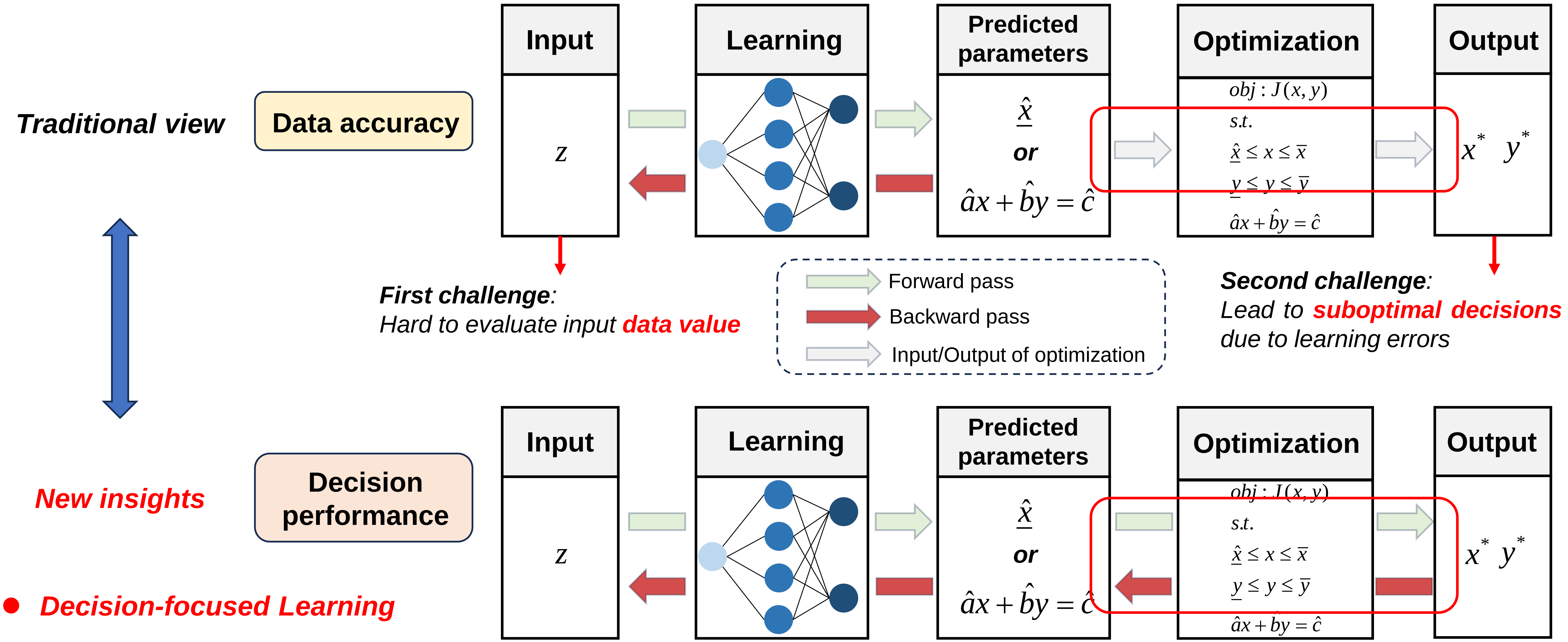
- **Insights Between Learning and Optimization**



The goal of learning is not to minimize the error from statistical perspectives (RMSE, MAPE), but to minimize the decision-making costs in real world.

Introduction

- Challenges From Traditional View



- ***Decision-focused Learning***

Introduction

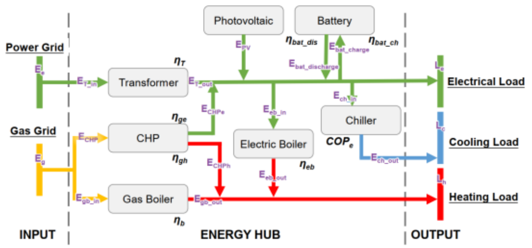
○ Research Work

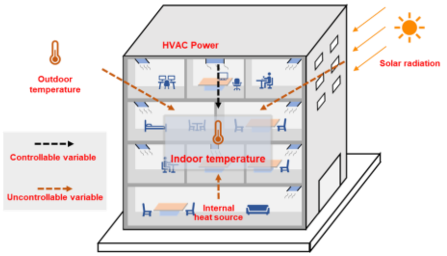
Focus on two challenges

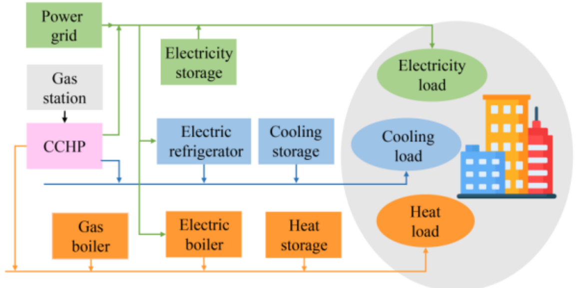
Decision errors minimization

Data valuation

Building energy systems	Building energy systems	Multi-energy systems
Traditional learning to optimize	Decision-focused constraint learning	Decision-focused boundary condition learning
Neural ODEs-Based Model Predictive Control for Building Energy Management [1]	**Decision-oriented modeling of building thermal dynamics [2]**	**Data Valuation in Multi-Energy Systems: An End-to-End Approach [3]**







[1] Boyu He, Ning Zhang, Chen Fang, Yun Su, and Yi Wang, "Flexible Building Energy Management with Neural ODEs-Based Model Predictive Control," IEEE Transactions on Smart Grid, in press.

[2] Xueyuan Cui, Jean-Francois Toubeau, Francois Vallee, and Yi Wang, "Decision-Oriented Modeling of Thermal Dynamics within Buildings," IEEE Transactions on Smart Grid, in press.

[3] Yangze Zhou, Qingsong Wen, Jie Song, Xueyuan Cui, and Yi Wang, "Load Data Valuation in Multi-Energy Systems: An End-to-End Approach," IEEE Transactions on Smart Grid, in press.

Contents

Problem Statement

○ Research Purpose

1. **Balance reliability and accuracy** with **limited training data** for building thermodynamics modeling

2. Reliably control buildings while simultaneously utilizing both the **thermal capacity** and the **flexibility resources of the energy system**.

We propose a system with…

Control Framework	Building Thermal Model	Building Energy Model
*The most **reliable** and commonly used control framework*	*1. Utilize the **reliability** of the RC model and the **accuracy** of the black-box model* *2. establish a **continuous digital twin** between the building and the model*	*A **generalized** energy framework*
MPC	**Neural ODEs (Neural Ordinary Differential Equations)**	**Energy Hub**

MPC + **Neural ODEs (Neural Ordinary Differential Equations)** + **Energy Hub**

Methodology

○ Neural ODEs-Based Thermal Model

● Building thermal dynamics can be represented as:

$$\frac{\mathrm{d}\mathbf{h}(t)}{\mathrm{d}t} = f(\mathbf{h}(t), t) \qquad (1)$$

▲ Substituting f with a neural network

$$\frac{\mathrm{d}\mathbf{h}(t)}{\mathrm{d}t} = f(\mathbf{h}(t), t, \boldsymbol{\theta}) \qquad (2)$$

■ Solving it with numerical integration methods (Euler method)

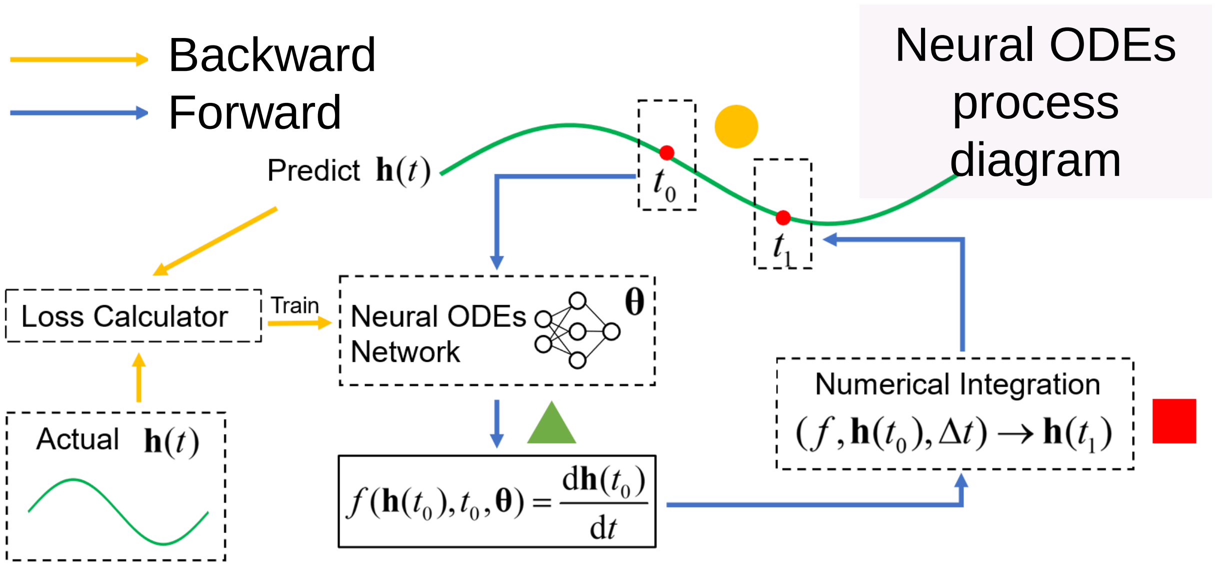
$$\mathbf{h}_{t+1} = \mathbf{h}_t + f(\mathbf{h}_t, t, \boldsymbol{\theta}) \cdot \Delta t \qquad (3)$$

Symbols

$\mathbf{h}(t)$ → the building states (zone temperature)

$\boldsymbol{\theta}$ → neural network parameters

$f(\cdot)$ → states derivative. Can be replaced by a neural network



Methodology

○ Neural ODEs-Based Thermal Model

The recursion formula of the system is

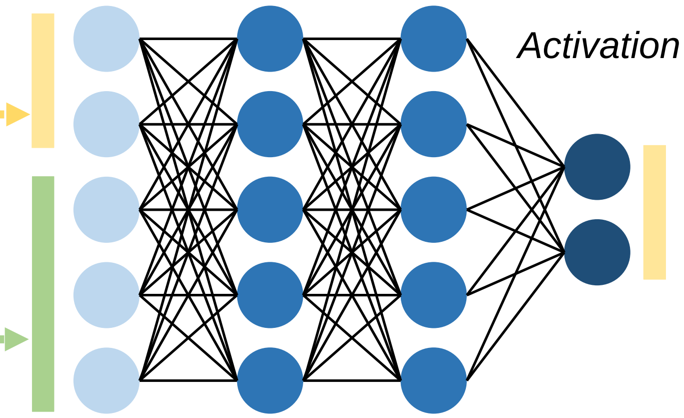
$$\mathbf{h}_{t+1} = \mathbf{h}_t + \boxed{f(\mathbf{h}_t, \mathbf{l}_t)} \cdot \Delta t$$

Building States ($\mathbf{h}_t$)
($e.g.$, zone temperature)

Network Input ($\mathbf{l}_t$)
($e.g.$, outdoor temperature, solar radiation, outdoor humidity, zone heating rate)

f is the neural ODEs network with the following structure:

NeuralODEs Structure

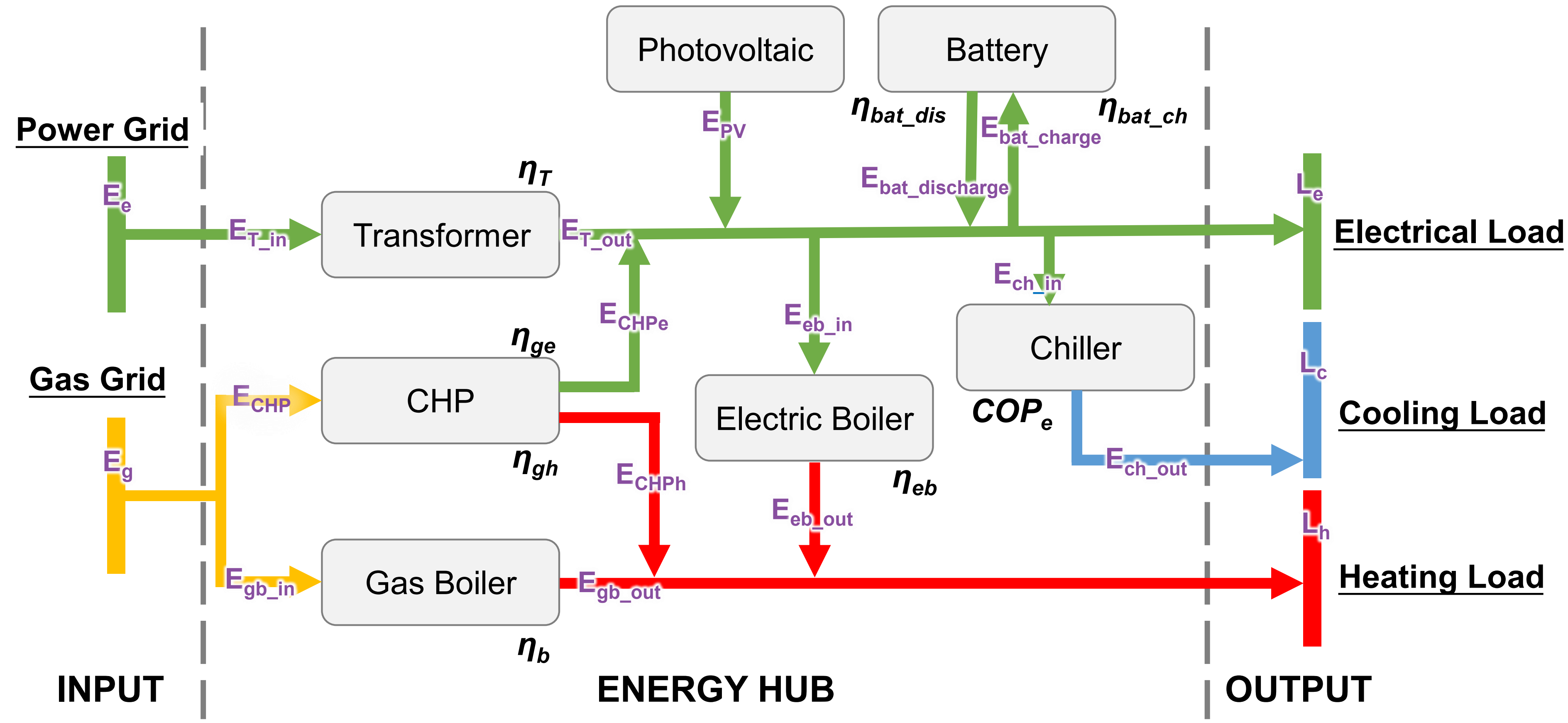


Derivative of building states $\mathrm{d}\mathbf{h}_t/\mathrm{d}t$

- ✓ By modeling the "differential dynamic" of the building, Neural ODEs has **less demand on data** and is **more robust on performance** **than a traditional black-box model**. Specifically, If the number of the hidden layers reduce to 0, Neural ODEs **degenerates into a RC model** (state-space representation).
- ✓ By introducing neural networks, Neural ODEs is **more accurate** than a **RC model**.

Methodology

- Energy Hub modeling



Energy Hub diagram

Source

- power
- gas

Load

- electrical load
- cooling load
- heating load

Equipment

- CHP
- gas boiler
- electric boiler
- battery
- PV
- chiller

Methodology

○ MPC Formulation

Objective Function Minimize the total building energy cost

Constraints:

- ✓ Energy Hub structure constraint

Energy Hub embedded in

- ✓ Thermal balance constraint
- ✓ Battery constraint
- ✓ Energy flow direction constraint
- ✓ Equipment capacity constraint

- ✓ Building state constraint

Neural ODEs embedded in

Methods to Handle Nonlinear Neural Network

$$a = \mathrm{ReLU}(z)$$

$$\Updownarrow$$

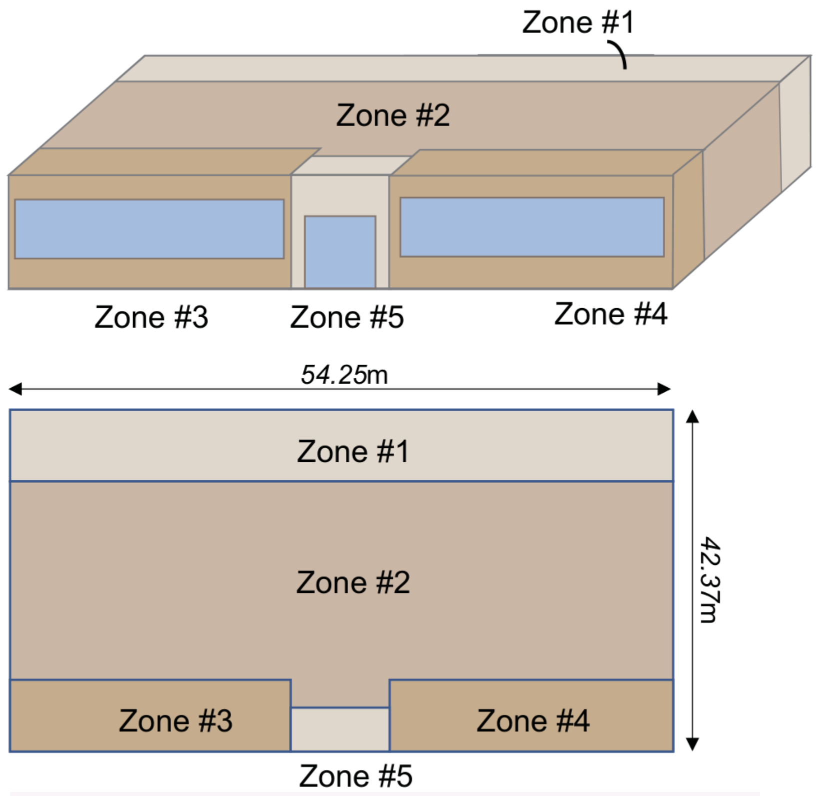
$$a = \max(0, z) \quad \Leftrightarrow \quad \begin{cases} a \geq 0 \\ a \geq z \\ a \leq M(1 - u_1) \\ a \leq z + M(1 - u_2) \\ u_1 + u_2 \geq 1 \\ u_1 \in \{0,1\} \\ u_2 \in \{0,1\} \end{cases}$$

The whole nonlinear optimization is transformed into an MILP!

Case Study

- **Setting**

Basic	Simulation Period	January 1st to February 1st
	Solver	Gurobi 9.5.1
	Simulation Software	EnergyPlus
MPC	MPC default timestep	30 min
	MPC prediction horizon	12 h
Building	Area	2294 m^2
	Height	6.1 m
	Zone	5 zones, with Zone #5 unconditioned
Price	Gas	remain constant
	Electricity	time-of-use signal (off-peak, mid-peak, on-peak)



DOE prototype commercial building: stand-alone retail building

Case Study

Network Structure

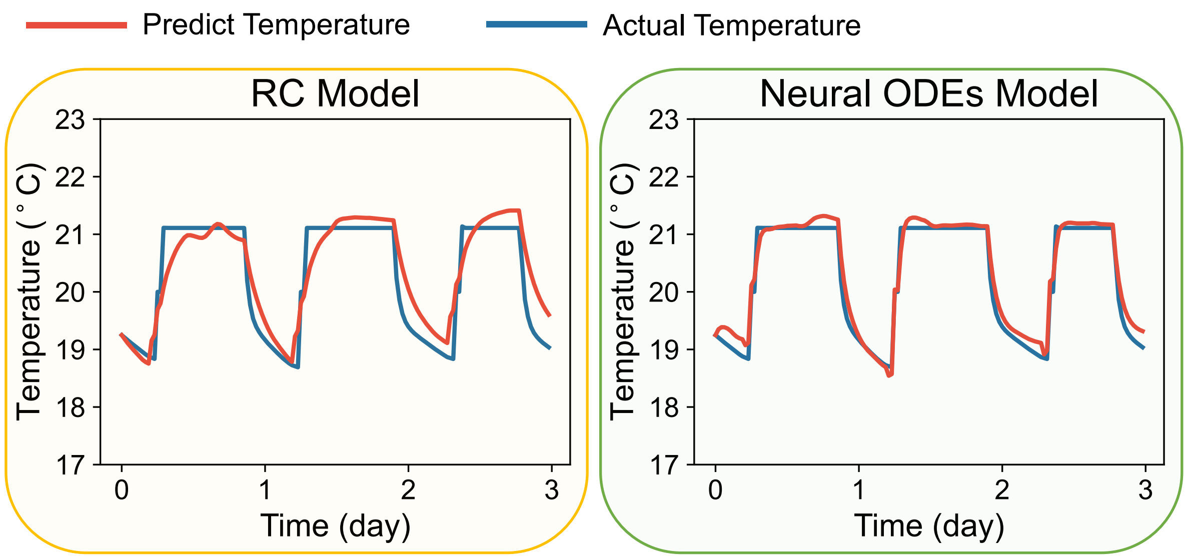
Network Structure Comparison

Name	Structure	RMSE after a Certain Time Length (°C) 0.5 h	3 h	12 h	1 day	3 days
Training Set						
Network 1	13×5	0.2534	0.6396	0.6467	0.6373	0.6853
Network 2	13×15×5	0.1794	0.3840	0.4888	0.5197	0.4920
Network 3	13×15×15×5	0.1588	0.1901	0.2176	0.2566	0.2899
Network 4	13×15×15×15×5	0.1288	0.1576	0.2003	0.2060	0.2555
Validation Set						
Network 1	13×5	0.2648	0.5615	0.7420	0.6893	0.7497
Network 2	13×15×5	0.1705	0.4749	0.6724	0.6461	0.6961
Network 3	13×15×15×5	0.1918	0.3010	0.4325	0.4757	0.4930
Network 4	13×15×15×15×5	0.3392	0.7515	0.6267	0.6656	0.7337

Network 1 is a RC model

Network 2 is our selected model

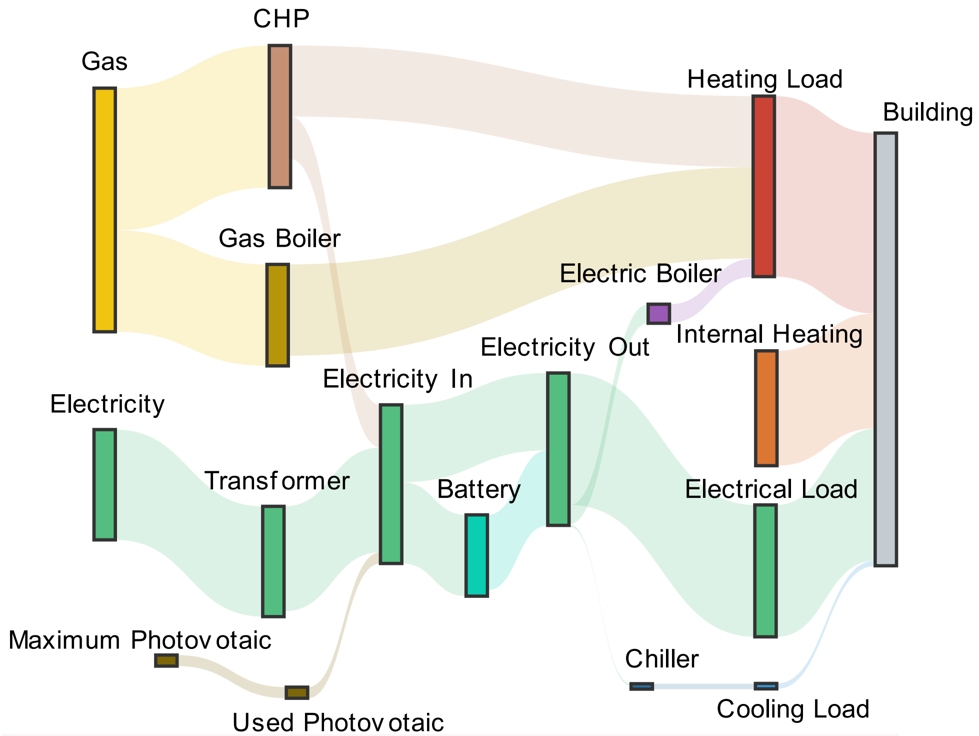
Model prediction result comparison



- ✓ Neural ODEs is **more accurate** than a **RC model**
- ✓ Complexed network (Network4) may face **overfitting** problem

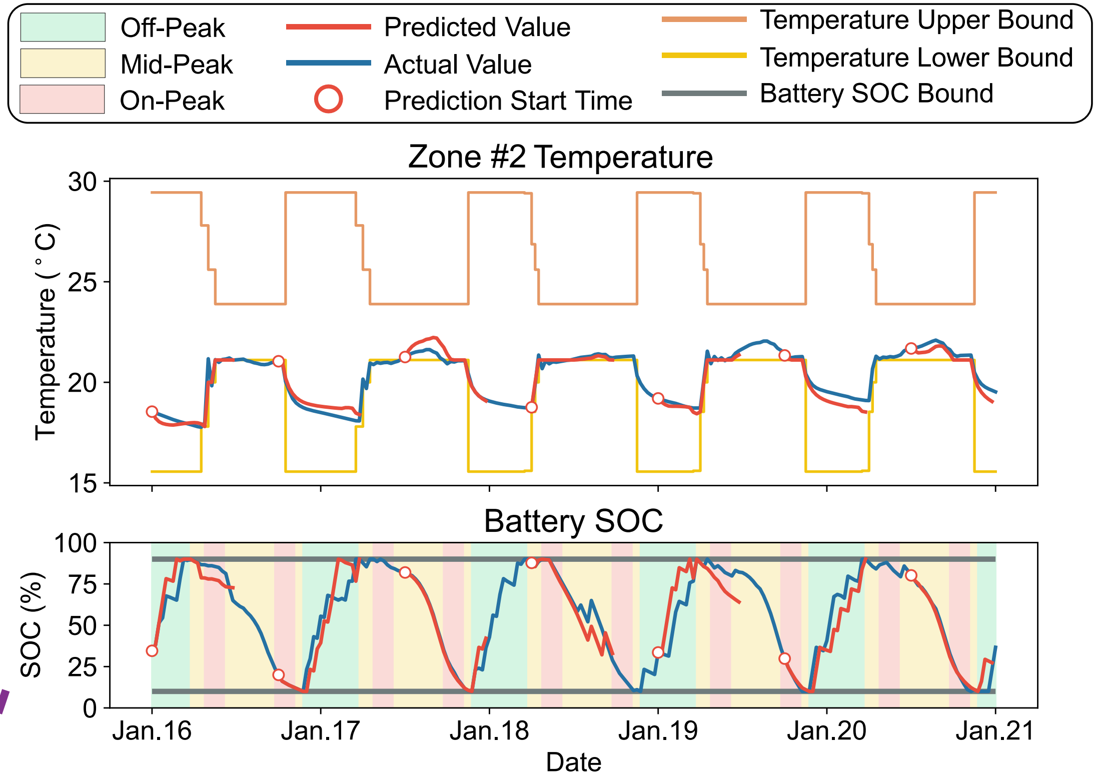
Case Study

○ Energy Flow



Total energy flow in the Energy Hub

○ MPC Prediction

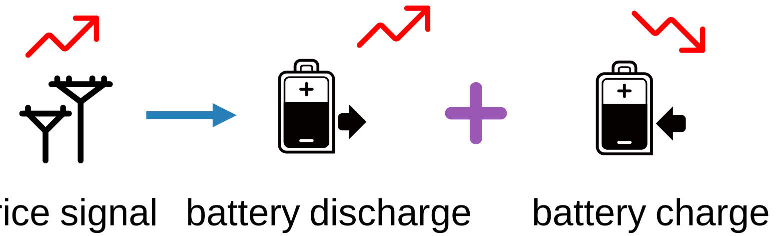


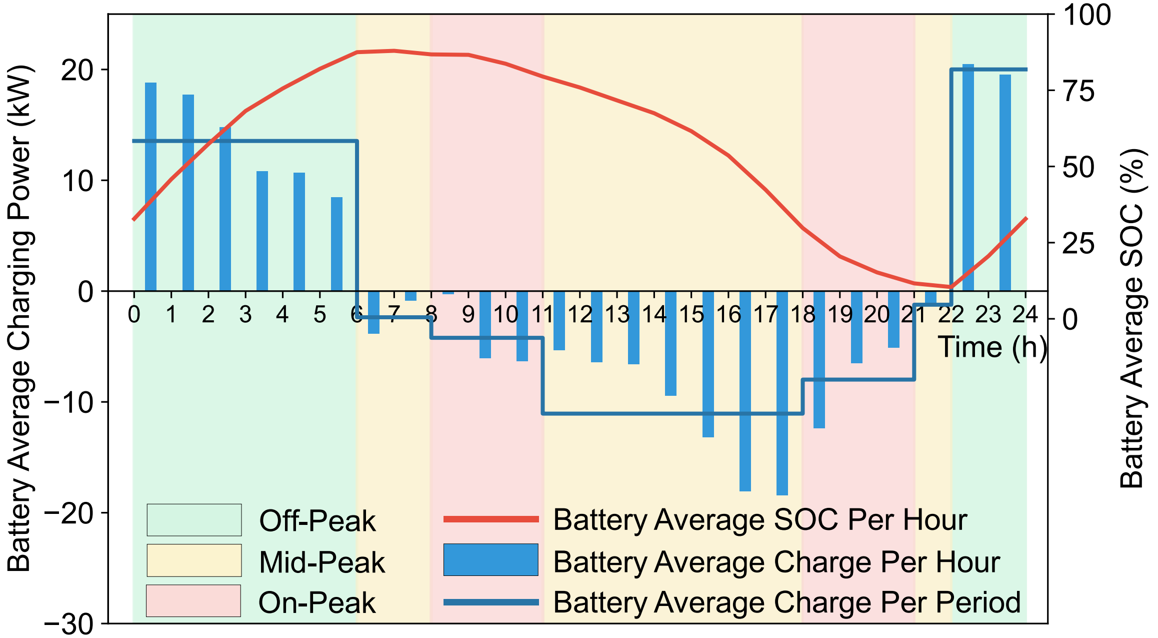
MPC prediction of the temperature of Zone #2 and the battery SOC versus actual data

➢ The **high accuracy** between predicted and actual values verifies that the trained model can be incorporated into MPC for state prediction.

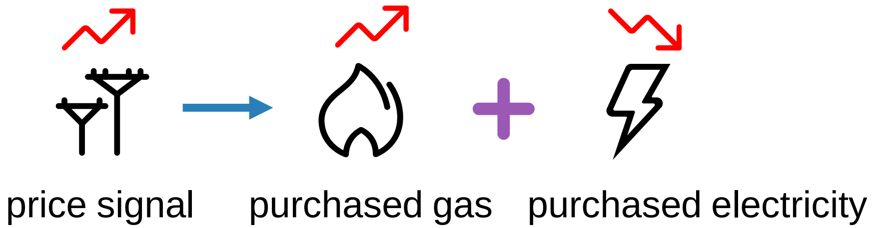
Case Study

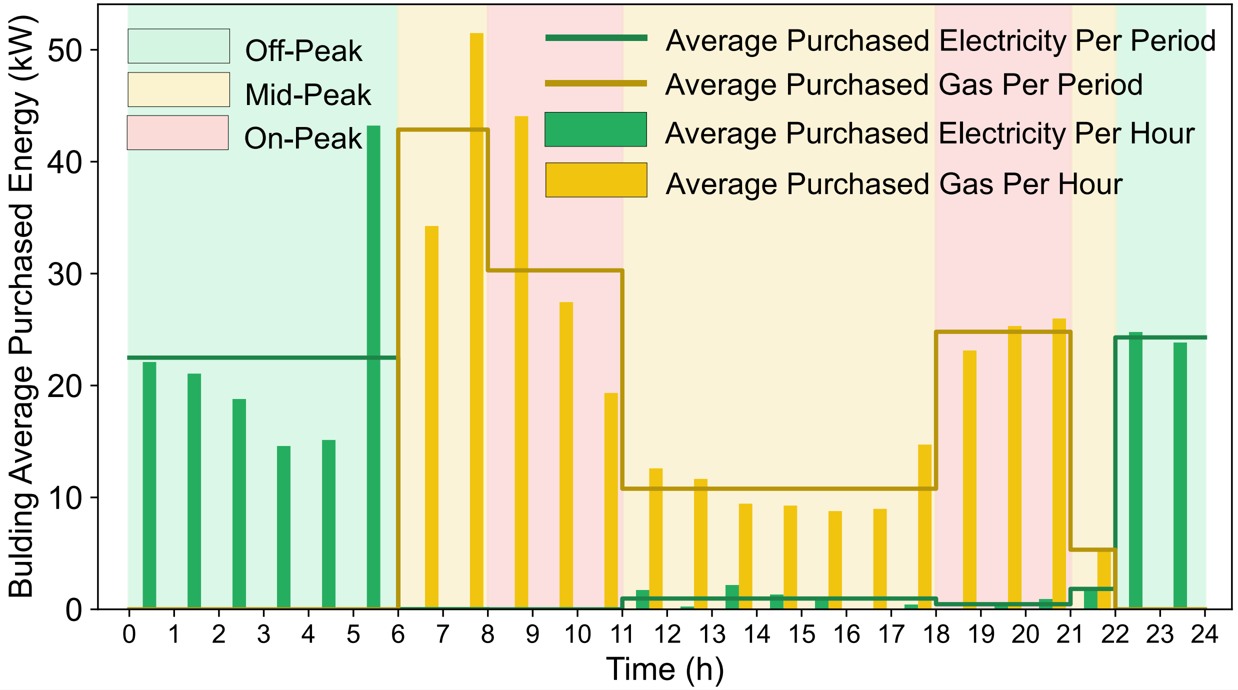
- Demand Response





Average battery SOC and charging rate in different time periods

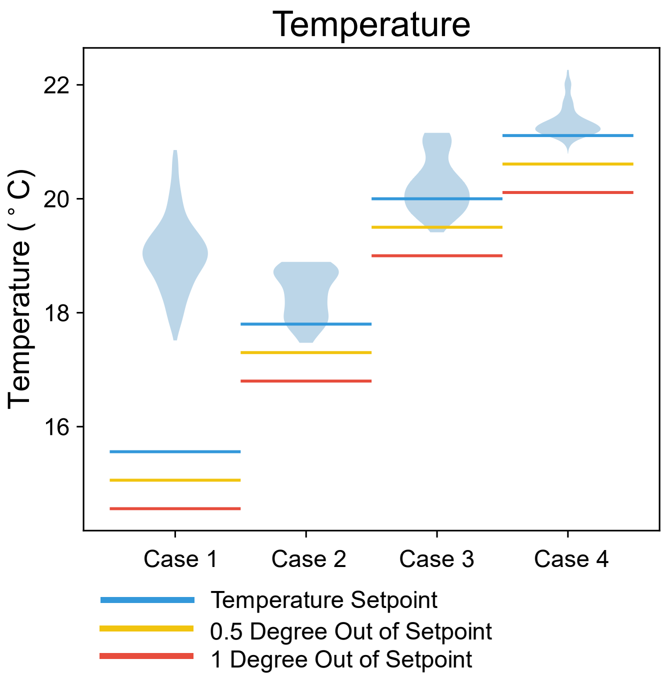


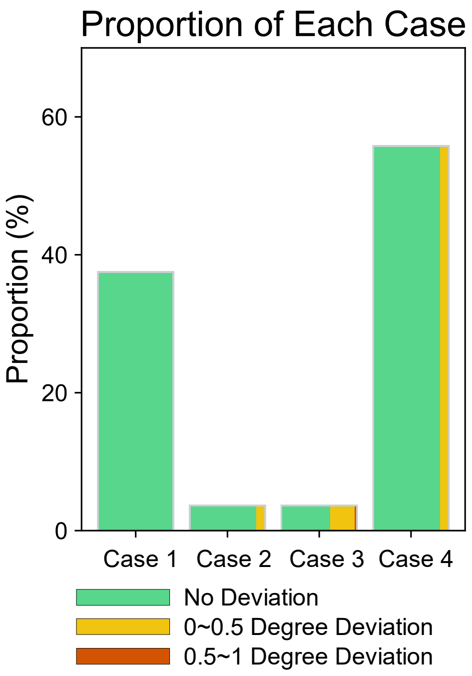


Average purchased building energy in different time periods

Case Study

○ Thermal Comfort Performance

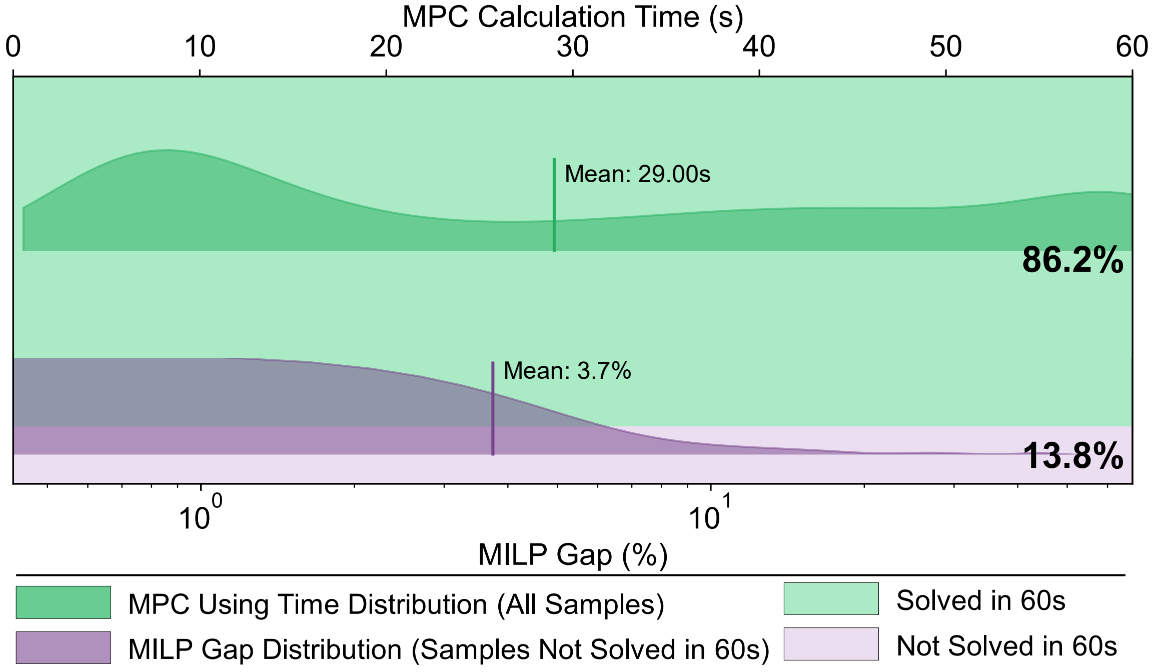




Temperature distribution of Zone #2 in January

- Set 4 different temperature setpoints as Cases 1-4
- Temperature falls within the dual setpoints: **92.31%**
- Deviations of less than 0.5 ℃: **7.62%**
- Deviations between 0.5 and 1 ℃: **0.07%**

○ Calculation Time Performance



MPC calculation time and MILP gap of unsolved samples

- Average calculation time **29 s**

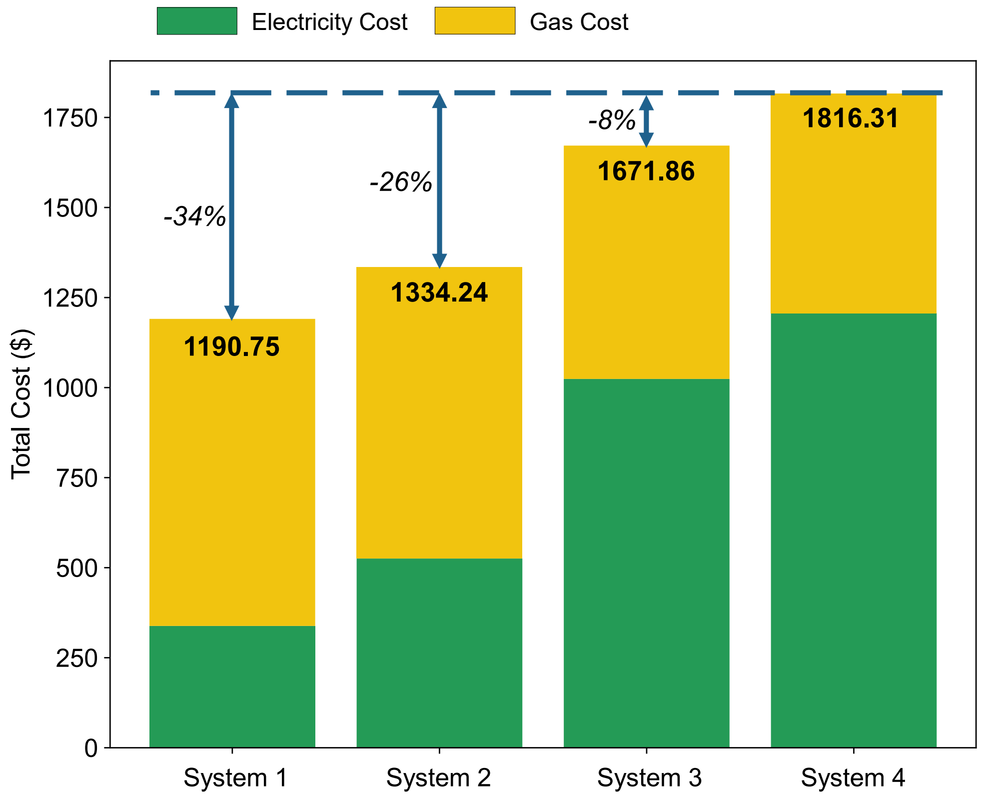
Case Study

○ System Comparison

Compared System Information

	System 1	System 2	System 3	System 4
Thermal Model	✓	✗	✓	✗
HVAC Control	Instruction	PID	Instruction	PID
Control Framework	MPC	MPC	MPC	None
Energy Component	Whole	Whole	No battery and CHP	No battery and CHP
Feasibility	✓	✗	✓	✓

- (orange) normal building + Energy Hub + thermal model **(our proposed method)**
- (yellow) normal building + Energy Hub **(not practical in reality)**
- (blue) normal building + thermal model
- (green) normal building



System simulation results

- ✓ Through **building thermal mass utilization**, System 3 achieves **8%** cost reduction
- ✓ By **implementing the proposed Energy Hub**, System 2 achieves **26%** cost reduction
- ✓ Our proposed method gets a total **34%** cost reduction

Contents

Problem Statement

○ Thermal dynamics modeling for optimization

HVAC power *Solar radiation*

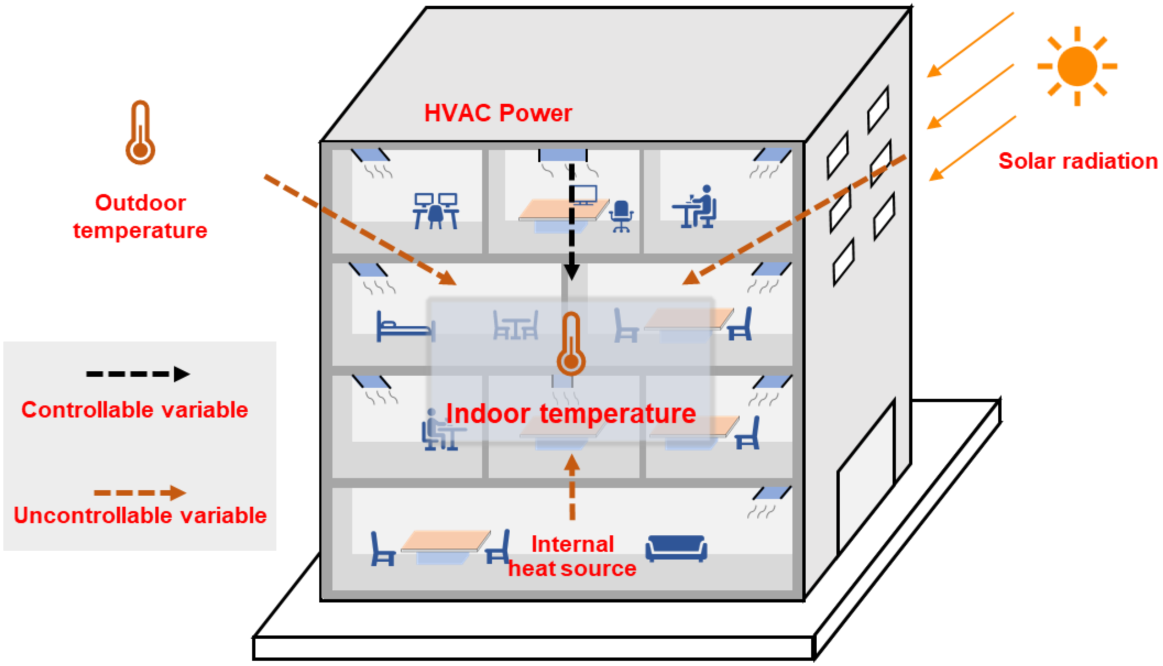
$$\boldsymbol{X}_i = [\tau_i, q_i, \tau^{\text{out}}, q_i^{\text{rad}}, q_i^{\text{occ}}]$$

Influence features *Indoor temperature* *Outdoor temperature* *Occupant*

Model definition: Ordinary Differential Equation-based

$$\dot{\tau}_i = \mathcal{M}(\boldsymbol{X}_i; \psi)$$

Model parameters



Decision variable τ_i, q_i

Boundary parameter $\tau^{\text{out}}, q_i^{\text{rad}}, q_i^{\text{occ}}$

Original way in optimization

$$\min_{\boldsymbol{v}} \; \mathcal{C}^{\text{opt}}(\boldsymbol{v} \in \mathcal{V}; \mathcal{M}(\boldsymbol{X}_i; \widehat{\psi}))$$

$$s.t. \;\; \widehat{\psi} \in \arg\min_{\psi} \|\boldsymbol{\tau}_i - \tilde{\boldsymbol{\tau}}_i\|_2^2 + \varepsilon$$

Data accuracy-based learning

Decision-focused way

$$\min_{\boldsymbol{v} \in \mathcal{V}, \psi} \; \mathcal{C}^{\text{opt}}(\boldsymbol{v}; \mathcal{M}(\boldsymbol{X}_i; \psi))$$

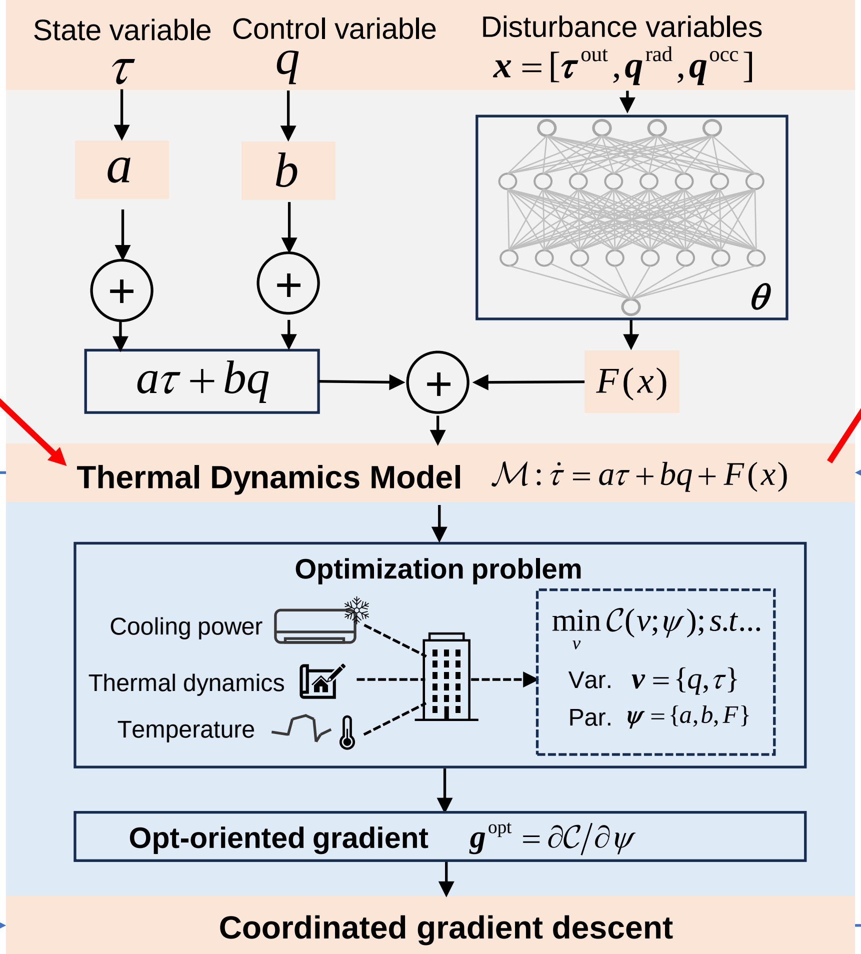
Decision performance-based learning

Methodology

○ Decision-oriented modeling framework

1. Backbone model structure

$$\dot{\tau}_i = \mathcal{M}(\boldsymbol{X}_i; \boldsymbol{\psi})$$



$$\mathcal{M}: \ \dot{\tau} = a \cdot \tau + b \cdot q + F(\boldsymbol{x})$$

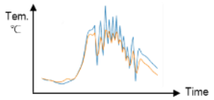
$$F(\boldsymbol{x}) = \Psi(\boldsymbol{x}; \boldsymbol{\theta})$$

- linear relationship for decision variables;
- Black-box (NN) representation for complex disturbances.

2. Forward: loss functions

Forward with model parameters

Data accuracy

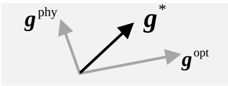


Auxiliary gradient

$\boldsymbol{g}^{\text{phy}} = \partial\mathcal{L}/\partial\psi$

3. Backward: training strategy

Backward with updated gradients



Gradient update

Methodology

- **Forward: Optimization-oriented loss function**

  - Downstream optimization problem $\longrightarrow$ (obj.: loss function)

Obj: energy consumption and temperature violation

$$\min \ \mathcal{C}^{\text{opt}} = \sum_{t \in \mathcal{T}} c_t p(t) + c^{\text{U}} e^{\text{U}}(t) + c^{\text{L}} e^{\text{L}}(t)$$

Subject to: model; power limits, comfort range

$$\tau(t+\Delta t) = \tau(t) + \Delta t \cdot (a\tau(t) + bq(t) + F(\boldsymbol{x}(t)))$$

$$q(t) = \eta p(t) \qquad \underline{\tau}_t - e^{\text{L}}(t) \leq \tau(t) \leq \overline{\tau}_t + e^{\text{U}}(t)$$

$$\underline{p} \leq p(t) \leq \overline{p} \qquad 0 \leq e^{\text{L}}(t), e^{\text{U}}(t) \leq \overline{e}$$

  - Gradient derivation (obj. w.r.t. model parameters)

$$\begin{aligned} \min \ & \mathcal{C}^{\text{opt}}(\boldsymbol{v}) \\ \text{s.t.} \ & \boldsymbol{f}(\boldsymbol{v}, \boldsymbol{\psi}) = 0 \\ & \boldsymbol{h}(\boldsymbol{v}, \boldsymbol{\psi}) \leq 0 \end{aligned}$$

Decision variables

Parameters

$$\xrightarrow{\text{Lagrangian function}} \quad \mathcal{L}(\boldsymbol{v}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathcal{C}^{\text{opt}}(\boldsymbol{v}) + \boldsymbol{\lambda}^{\text{T}} \boldsymbol{f}(\boldsymbol{v}, \boldsymbol{\psi}) + \boldsymbol{\mu}^{\text{T}} \boldsymbol{h}(\boldsymbol{v}, \boldsymbol{\psi}) \quad \xrightarrow{\text{KKT condition}} \quad \mathcal{K} = \begin{bmatrix} \frac{\partial \mathcal{L}(\boldsymbol{v}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \boldsymbol{\psi})}{\partial \boldsymbol{v}} \\ \boldsymbol{f}(\boldsymbol{v}, \boldsymbol{\psi}) \\ \boldsymbol{\mu}^{\text{T}} \boldsymbol{h}(\boldsymbol{v}, \boldsymbol{\psi}) \end{bmatrix} = 0$$

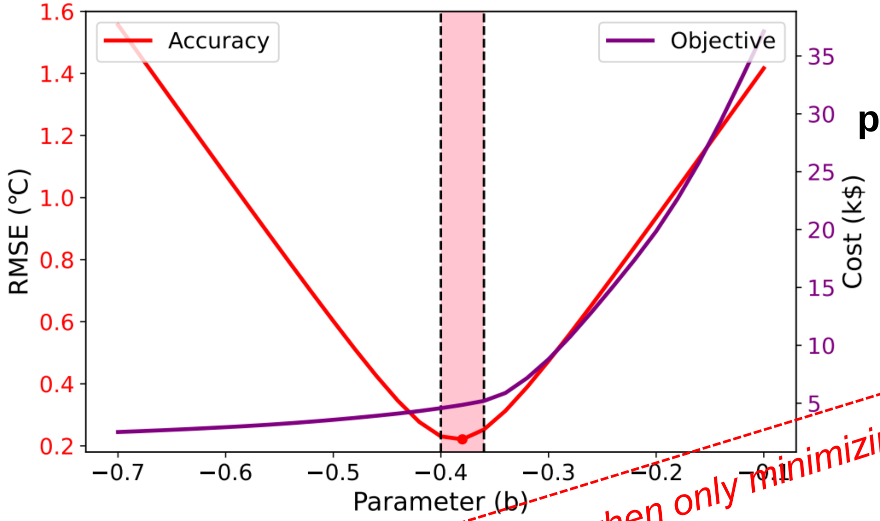
Implicit function theorem

$$\frac{\partial \mathcal{K}}{\partial \boldsymbol{\psi}} + \frac{\partial \mathcal{K}}{\partial \boldsymbol{v}} \cdot \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\psi}} = 0 \quad \xrightarrow{\text{Variable w.r.t. parameters}} \quad \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\psi}} = -\left[\frac{\partial \mathcal{K}}{\partial \boldsymbol{v}}\right]^{-1} \frac{\partial \mathcal{K}}{\partial \boldsymbol{\psi}} \quad \xrightarrow{\text{Chain rule}} \quad \boldsymbol{g}^{\text{opt}} = \boxed{\frac{\partial \mathcal{C}^{\text{opt}}}{\partial \boldsymbol{\psi}}} = \frac{\partial \mathcal{C}^{\text{opt}}}{\partial \boldsymbol{v}} \cdot \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\psi}}$$

B. Amos, et al, "Optnet: Differentiable optimization as a layer in neural networks," in International Conference on Machine Learning. PMLR, 2017, pp. 136–145.

Methodology

- **Forward: Physics-informed auxiliary loss function**
  - Only minimizing the obj. will deviate from real physical characteristics (accuracy);
  - The deviation is hard to correct in a “unsupervised” way (without ground-true model).



The training direction when only minimizing the obj.

Corrected with physical characteristics

Physical characteristics → ***Equivalent penalty terms***

Stability issue: $a < 0$

Power correlation: $\frac{\partial \tau(t+\Delta t)}{\partial q(t)} \leq 0$

Disturbance correlation: $\frac{\partial \tau(t+\Delta t)}{\partial \tau^{\text{out}}(t)} \geq 0, \frac{\partial \tau(t+\Delta t)}{\partial q^{\text{rad}}(t)} \geq 0, \frac{\partial \tau(t+\Delta t)}{\partial q^{\text{occ}}(t)} \geq 0$

$$
\begin{cases}
a \leq 0 \Rightarrow \min \ [a]^+ \\
b \leq 0 \Rightarrow \min \ [b]^+ \\
\frac{\partial F}{\partial \tau^{\text{out}}} \geq 0 \Rightarrow \min \ [-\frac{\partial F}{\partial \tau^{\text{out}}}]^+ \\
\frac{\partial F}{\partial q^{\text{rad}}} \geq 0 \Rightarrow \min \ [-\frac{\partial F}{\partial q^{\text{rad}}}]^+ \\
\frac{\partial F}{\partial q^{\text{occ}}} \geq 0 \Rightarrow \min \ [-\frac{\partial F}{\partial q^{\text{occ}}}]^+
\end{cases}
$$

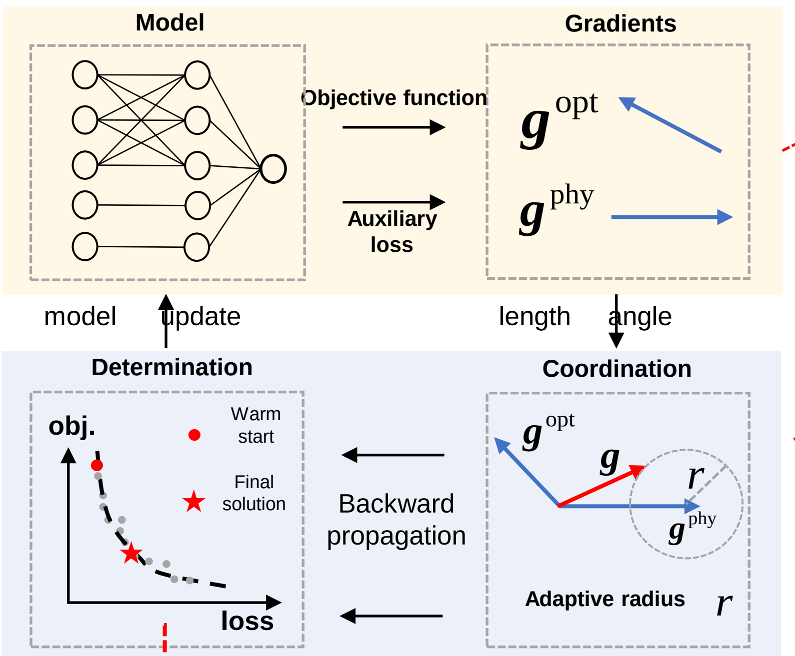
Physics-informed loss function: MSE & penalty terms

$$
\mathcal{L}^{\text{phy}}(\boldsymbol{\tau}, \tilde{\boldsymbol{\tau}}; \boldsymbol{\psi}) = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \|\tau(t) - \tilde{\tau}(t)\|_2^2 + \omega^{\text{R}} \|\boldsymbol{R}(t)\|_2^2
$$

$$
\boldsymbol{R} = \left[[a]^+, [b]^+, [-\frac{\partial F}{\partial \tau^{\text{out}}}]^+, [-\frac{\partial F}{\partial q^{\text{rad}}}]^+, [-\frac{\partial F}{\partial q^{\text{occ}}}]^+ \right]^{\text{T}}
$$

Methodology

○ Backward: Coordinated gradient descent



- The given two gradient vectors could be contradictory

$$\langle \boldsymbol{g}^{\text{phy}}, \boldsymbol{g}^{\text{opt}} \rangle = (\boldsymbol{g}^{\text{phy}})^{\mathrm{T}} \cdot \boldsymbol{g}^{\text{opt}} < 0$$

- Determine a coordinated gradient vector $\boldsymbol{g}$ that minimizes the conflict degree between the given two gradients

$$\max_{\boldsymbol{g}} \ \min \{\langle \boldsymbol{g}^{\text{phy}}, \boldsymbol{g} \rangle, \langle \boldsymbol{g}^{\text{opt}}, \boldsymbol{g} \rangle\}$$
$$\text{s.t.} \ \|\boldsymbol{g} - \boldsymbol{g}^{\text{phy}}\| \le r \|\boldsymbol{g}^{\text{phy}}\|$$

The objective is to minimize the conflict degree (i.e., *maximize inner product) by finding the new vector **g**.*

- Find the optimal training epoch by selecting the optimal obj. within the preset error threshold.

$$\min_k \left\{ \mathcal{L}_k^{\text{opt}} \,\middle|\, \text{s.t.} \left| \frac{\mathcal{L}_k^{\text{phy}} - \mathcal{L}_0^{\text{phy}}}{\mathcal{L}_0^{\text{phy}}} \right| \le \epsilon, (\mathcal{L}_k^{\text{phy}}, \mathcal{L}_k^{\text{opt}}) \in \mathbb{P} \right\}$$

B. Liu, et al, “Conflict-averse gradient descent for multi-task learning,” Advances in Neural Information Processing Systems, vol. 34, pp. 18 878–18 890, 2021

Case Study

○ Simulation setup

○ Data preparation

- Building prototypes: 6-zone, 10-zone, 18-zone
- Simulation software: Energyplus
- Training period (01/06-31/07) and test period (01/08-31/08)

○ Parameter setting

Neural ODE

Optimization

	Parameter	Value	Parameter	Value
Model	hidden layers	3	η	3.6
	hidden units	48	$\underline{p}$, $\overline{p}$	0, 60 (kW)
	activation	ReLU	c^{L}	0.5 (\$/°C)
	optimizer	Adam	c^{U}	0.8 (\$/°C)
	batch size	16	$\overline{e}$	0.5 (°C)
	/	/	ω^{R}	5.0
Algorithm	K	30	ϵ	0.05

Coordinated gradient vector

Case Study

○ Evaluation of operation costs

○ Procedure

Identify thermal dynamics models → *Solve optimal daily decisions* → *Simulate the real building operation*

Proposed method and comparison | Commercial solvers | Energyplus

Quantify the actual ex-post operation performance of the model

○ Cost comparison

$$\min\ \underbrace{C^{\mathrm{opt}}}_{\text{Sum}} = \sum_{t\in\mathcal{T}} \underbrace{c_t p(t)}_{\text{Power}} + \underbrace{c^{\mathrm{U}} e^{\mathrm{U}}(t) + c^{\mathrm{L}} e^{\mathrm{L}}(t)}_{\text{Tem}}$$

MTO: modeling-then-optimization

Buildings	Costs	Training MTO	Training Proposed	Training *vs MTO*	Test MTO	Test Proposed	Test *vs MTO*
6-zone	Power	316.97	314.57	*-0.76%*	156.70	155.88	*-0.52%*
	Tem	26.35	11.99	*-54.50%*	24.51	14.89	*-39.25%*
	Sum	**343.32**	**326.56**	***-4.88%***	**181.21**	**170.77**	***-5.76%***
10-zone	Power	911.13	893.06	*-1.98%*	481.80	473.97	*-1.63%*
	Tem	51.07	47.93	*-6.15%*	25.11	26.21	*4.38%*
	Sum	**962.20**	**940.99**	***-2.20%***	**506.91**	**500.18**	***-1.33%***
18-zone	Power	2960.45	2917.53	*-1.45%*	1448.79	1441.24	*-0.52%*
	Tem	63.62	52.15	*-18.03%*	28.31	30.26	*6.89%*
	Sum	**3024.07**	**2969.68**	***-1.80%***	**1477.10**	**1471.50**	***-0.38%***

- The costs are mainly reduced in the term corresponding to temperature violations:
  - The temperature violation is affected by all factors, while the power consumption is mainly caused by the cooling power factor;
  - During the training process, the temperature violation part has thus a larger improvement space.

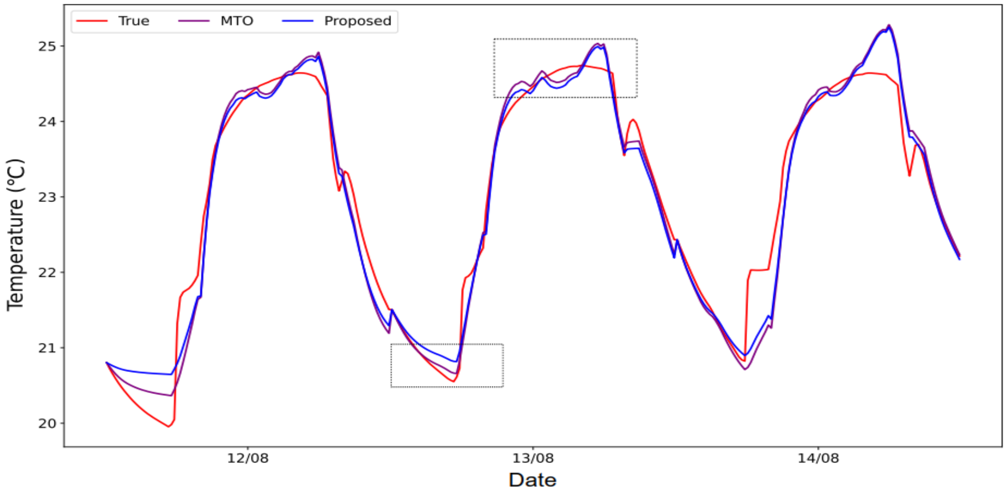
Case Study

- **Statistical accuracy**
  - Statistical metrics

Buildings	Dataset	RMSE			MAE			R2		
		MTO	Proposed	*vs MTO*	MTO	Proposed	*vs MTO*	MTO	Proposed	*vs MTO*
6-zone	Train	0.2022	0.2045	*1.14%*	0.1542	0.1561	*1.23%*	0.9841	0.9837	*-0.04%*
	Test	0.2710	0.2765	*2.03%*	0.1892	0.2026	*7.08%*	0.9729	0.9718	*-0.11%*
10-zone	Train	0.3867	0.3928	*1.58%*	0.3151	0.2945	*-6.54%*	0.8734	0.8694	*-0.46%*
	Test	0.4294	0.4533	*5.57%*	0.3195	0.3252	*1.78%*	0.8219	0.8015	*-2.48%*
18-zone	Train	0.2990	0.3059	*2.31%*	0.2241	0.2309	*3.03%*	0.8828	0.8773	*-0.62%*
	Test	0.3322	0.3326	*0.12%*	0.2665	0.2667	*0.08%*	0.8293	0.8290	*-0.04%*

- The MTO purely pursues minimizing MSE losses;
- The proposed method sacrifices some accuracy in pursuit of the operation cost minimization.

  - Details in temperature curves

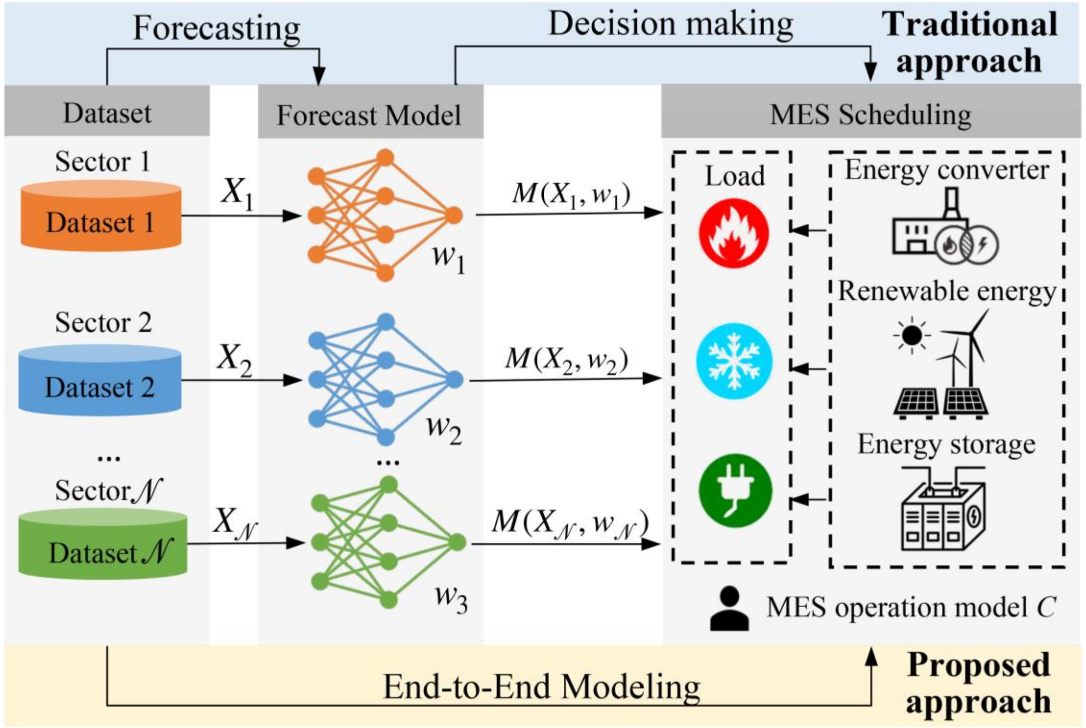


Daily temperature variation of the 6-zone building

- The "conservative" nature of the temperature data generated by the proposed method;
- Compared with MTO, the data are generally lower in the peak period and higher in the valley period;
- This conservativeness tends to reduce temperature violations part in decision costs.

Contents

Problem Statement



X_n: Input feature of sector *n*

w_n: Model parameters of sector *n*

$M_n(X_n, w_n)|_{n\in N}$: Load forecasts of sector *n*

$$\min_{z} \; C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}})$$

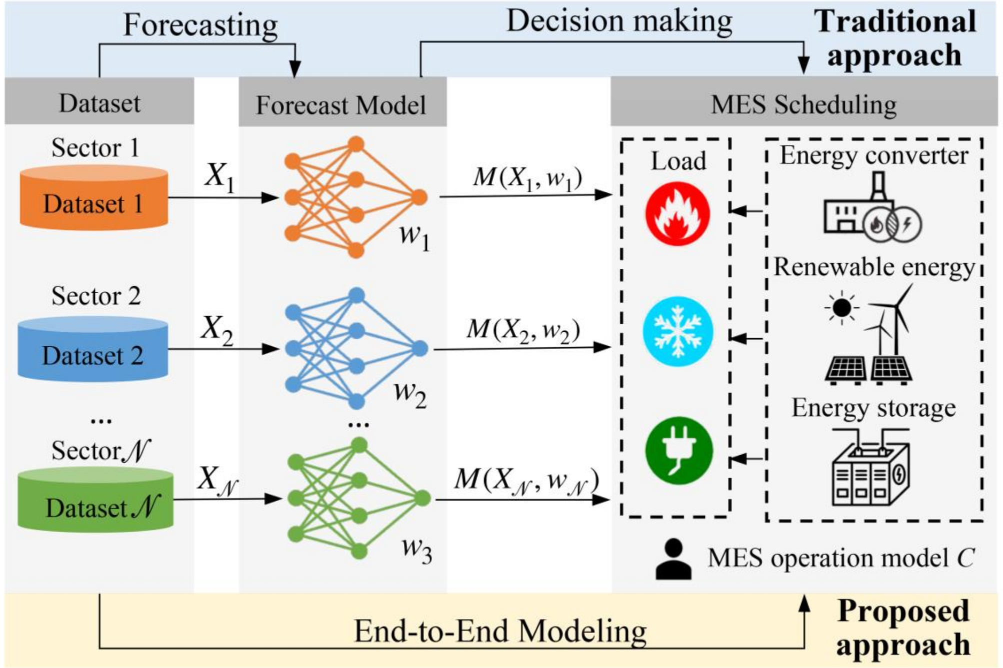
where *C* and z are the cost and decision variables for the scheduling of MES.

FTO: w_n and z are **determined sequentially**.

- Cross-sector data/information has not been shared and fully utilized to reduce operation costs.
- The forecasting and decision-making processes are treated separately so that data cannot directly serve final decision-making in MES.

Problem Statement

$$\min_{z;w_n|_{n\in\mathcal{N}}} C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}})$$



C_N : The operation costs If all sectors cooperate with the MES operator, which means the sectors **share their data X_n with the operator indirectly**.

End-to-End approach: w_n and z are **optimized as a whole.**

To encourage sectors to participate in the end-to-end model, the value of the data owned by various sectors should be quantified:

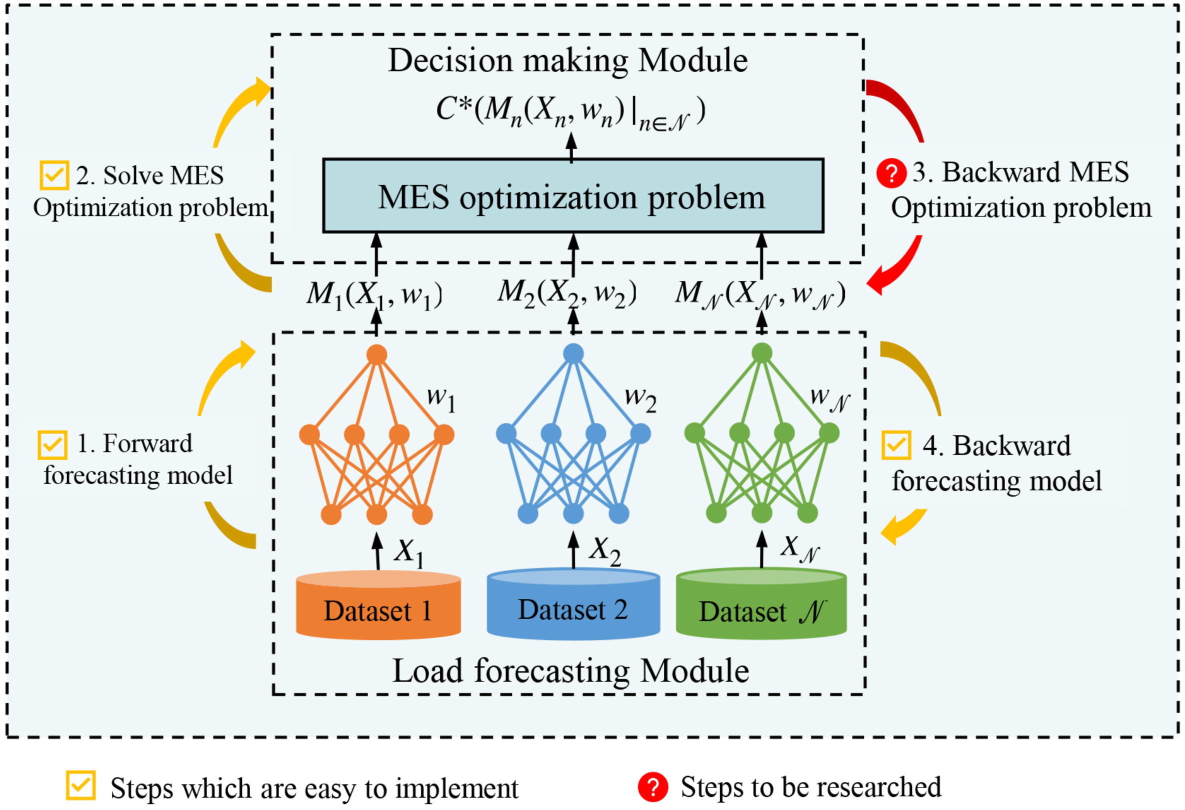
1) How many additional profits $V(N)$ can be derived from data sharing of various sectors in MES ?

2) How to make a fair plan $\{v_1, v_2, \cdots, v_N\}$ to allocate the profits $V(N)$ to each sector ?

End-to-End Optimization

- **How to optimize w_n and z as a whole?**

➢ An intuitive idea to train the end-to-end model is **forward and backward propagation**, as used for traditional neural network training.



How to obtain the gradient of cost *C* over load forecasts *M* ?

Optimization differentiable neural network (OptNet)

End-to-End Optimization

How to obtain the gradient cost *C* over load forecasts *M* ?

Chain principle $\frac{dC}{dM} = \frac{dC}{dz}\frac{dz}{dM}$

The MES optimization problem can be abstracted:

$$\min \quad C(z, M)$$
$$s.t. \quad f(z, M) \leq 0, h(z, M) = 0$$

The Lagrange function of the optimization problem

$$\mathcal{L}(z, \lambda, \mu, M) = C(z, M) + \lambda^T f(z, M) + \mu^T h(z, M)$$

The KKT condition of $\mathcal{L}(z, \lambda, \mu, M)$:

$$\begin{cases} f(z, M) \leq 0 \\ h(z, M) = 0 \\ \lambda_i \geq 0, i \in \{1,2,\cdots,q\} \\ \lambda_i f_i(z, M) = 0, i \in \{1,2,\cdots,q\} \\ \nabla_z \mathcal{L}(z, \lambda, \mu, M) = 0 \end{cases}$$

Implicit function:

$$G(\tilde{z}, M) = \begin{bmatrix} \nabla_z \mathcal{L}(z, \lambda, \mu, M) \\ \lambda f(z, M) \\ h(z, M) \end{bmatrix}$$

The gradient of $\tilde{z}$ over *M* can be obtained by the differential principle of implicit function:

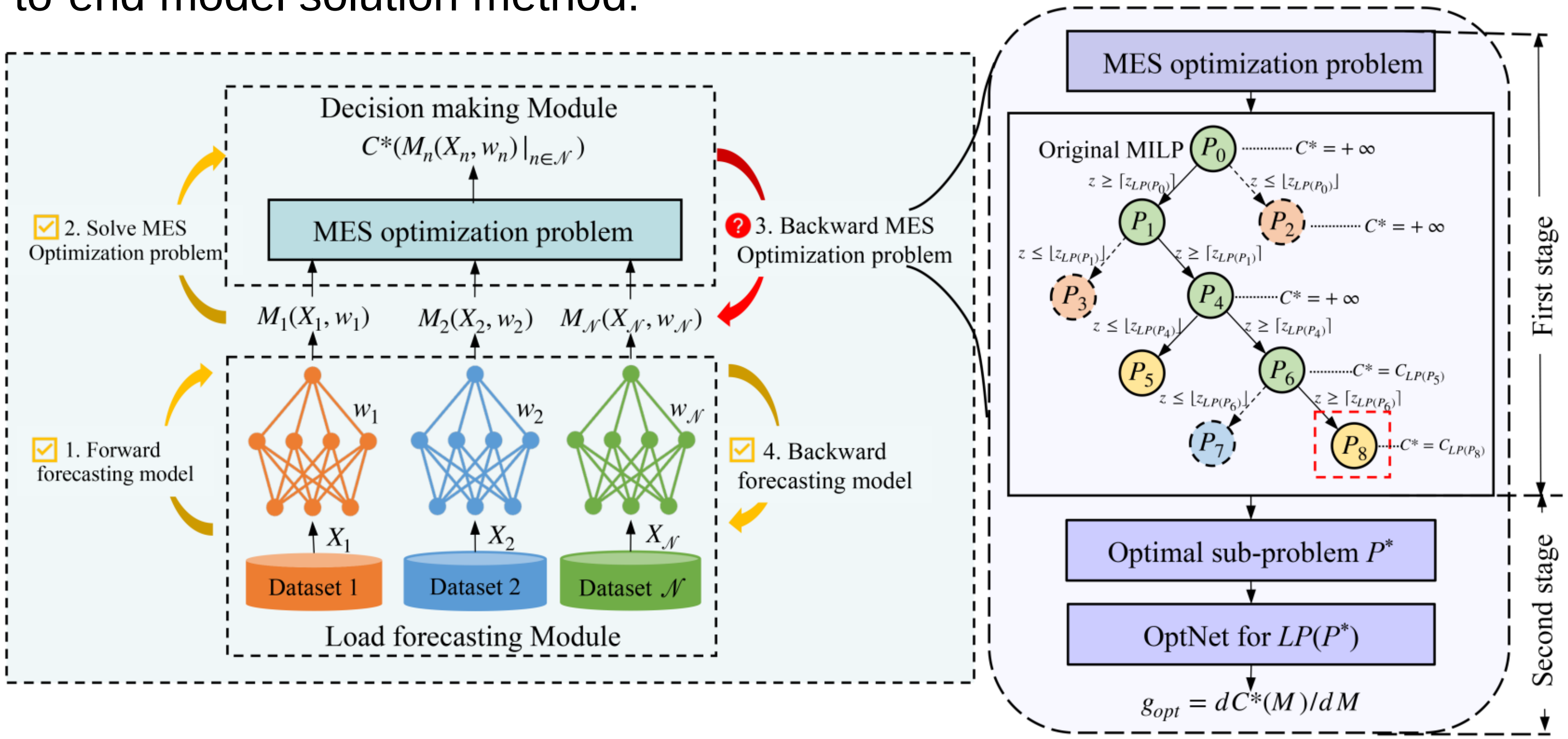
$$\frac{d\tilde{z}}{dM} = G_{\tilde{z}}^{-1}(\tilde{z}, M) G_M(\tilde{z}, M)$$

where

$$\frac{d\tilde{z}}{dM} = \begin{bmatrix} \boxed{\frac{dz}{dM}} & \frac{d\lambda}{dM} & \frac{d\mu}{dM} \end{bmatrix}^T$$

End-to-End Optimization

- However, OptNet is designed for LP/QP problem, if What if there are integer variables in the optimization problem?
- Two-stage end-to-end model solution method:



- OptNet-embedded branch and bound method： How about incorporating OptNet into the branch and bound search process (Construct OptNet for each yellow node)?

 Higher computational complexity and storage requirements

Data Valuation Framework

○ Additional Profit Quantification

The reduced operation costs can be regarded as the additional profits derived from the data sharing.

$$V(N) = C_{\emptyset} - C_N$$

End-to-End data valuation:

1) Each sector $n \in N$ utilizes its own data to develop the basic forecasting model M_n.

2) Computing operation costs of the traditional FTO approach $C_{\emptyset}$.

3) Integrating the forecasting model with the MES optimization problem for end-to-end model training.

4) Forward-propagating the end-to-end model to calculates the operation costs C_N.

Basic model development:

for *each sector* $n \in \mathcal{N}$ **do**

Random initialize parameters $w_n|_{n\in\mathcal{N}}$

for $k \in [0, E_1]$ **do**

$M_n = M_n(X_n; w_n^{(k)})$

$g_n = \text{Backward}(L_{MSE}(M_n, M_n^{\text{real}}))$

$w_n^{(k+1)} = w_n^{(k)} - lr \cdot g_n$

Return $M_n|_{n\in\mathcal{N}}$

End-to-End data valuation:

$C_{\emptyset}$ **Calculation:**

$C_{\emptyset} = \min_z C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}})$

End-to-End modeling:

for $k \in [0, E_2]$ **do**

$M = M_n(X_n, w_n)|_{n\in\mathcal{N}}$

$P^* =$ Optimal sub-problem of $P(z, M)$

Construct OptNet for $LP(P^*)$

for *sector* $n \in \mathcal{N}$ **do**

$g_{\text{opt},n} = \text{Backward(OptNet)}$

$g_n = \text{Backward}(g_{\text{opt},n})$

$w_n^{(k+1)} = w_n^{(k)} - lr \cdot g_n$

$C_{\mathcal{N}}$ **Calculation:**

$C_{\mathcal{N}} = \min_z C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}})$

Additioanl profits quantification:

$V(\mathcal{N}) = C_{\mathcal{N}} - C_{\emptyset}$

Return $V(\mathcal{N})$

Data Valuation Framework

○ Additional Profit Allocation

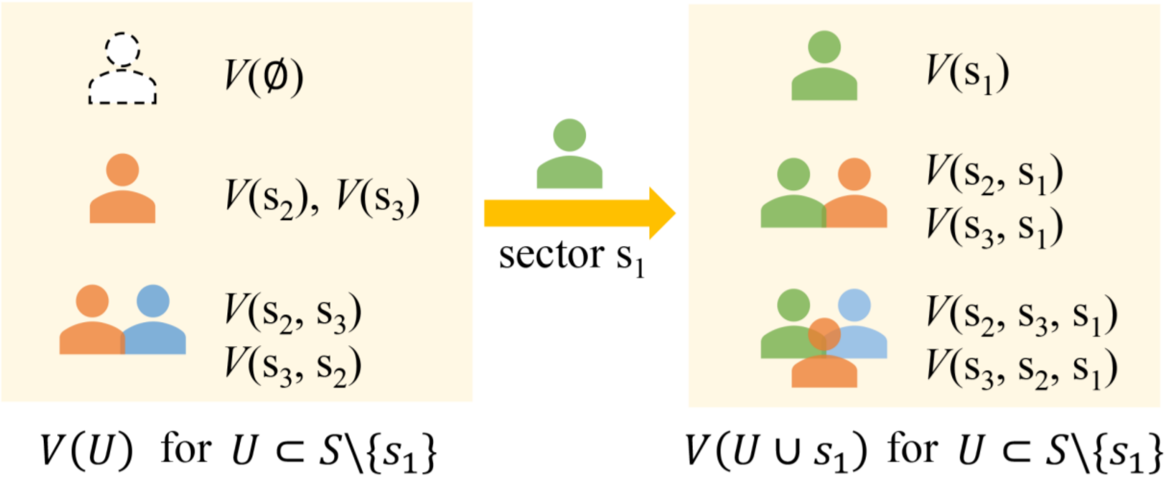
Shapley value may be negative

Shapley value has been widely adopted to measure the members' contributions to the collaboration earning.

Zero-Shapley value:
$$v_n = \frac{1}{|N|} \sum_{S \subseteq N \setminus \{n\}} \frac{1}{\binom{|N|-1}{|S|}} [V(S \cup \{n\}) - V(S)]^+$$

$V(S)$: the value of the cooperation formed by union S

$[\cdot]^+ = \max\{0, \cdot\}$



Two remaining question:

- When some sectors within the MES do not participate in the end-to-end modeling, how to measure $V(S)$?
- The zero-Shapley value does not satisfy the budget balance property:

$$\Gamma(v_n) = \frac{v_n}{\sum_{i \in N} v_i} (V(N) - V(\emptyset))$$

Data Valuation Framework

○ Additional Profit Allocation

When only sectors in U participate in the cooperation, how to measure $V(U)$?

End-to-End data valuation:
- **$C_\varnothing$ Calculation:**
  - $C_\varnothing = \min_z C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}})$
- **End-to-End modeling:**
  - **for** $k \in [0, E_2]$ **do**
    - $M = M_n(X_n, w_n)|_{n\in\mathcal{N}}$
    - $P^* =$ Optimal sub-problem of $P(z, M)$
    - Construct OptNet for $LP(P^*)$
    - **for** *sector* $n \in \mathcal{N}$ **do**
      - $g_{\text{opt},n} = \text{Backward(OptNet)}$
      - $g_n = \text{Backward}(g_{\text{opt},n})$
      - $w_n^{(k+1)} = w_n^{(k)} - lr \cdot g_n$
- **$C_\mathcal{N}$ Calculation:**
  - $C_\mathcal{N} = \min_z C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}})$
- **Additioanl profits quantification:**
  - $V(\mathcal{N}) = C_\mathcal{N} - C_\varnothing$
- **Return** $V(\mathcal{N})$

$$\min_{z, w_n|_{n\in N}} C_n(z, M_n(X_n, w_n)\big|_{n\in N})$$

$$\min_{z, w_n|_{n\in U}} C_n(z, M_n(X_n, w_n)\big|_{n\in U}, M_n(X_n, w_n)\big|_{n\in N\setminus U})$$

C_U denotes the operation costs of the **"partially integrated"** end-to-end model.

- ➢ Only sector $n \in U$ will update their model.
- ➢ Sectors in $N\setminus U$ will remain their **model parameters unchanged** (denoted as w_n) and **only submit final forecasts** $M_n(X_n, \overline{w}_n)|_{n\in N\setminus U}$ to the operator.

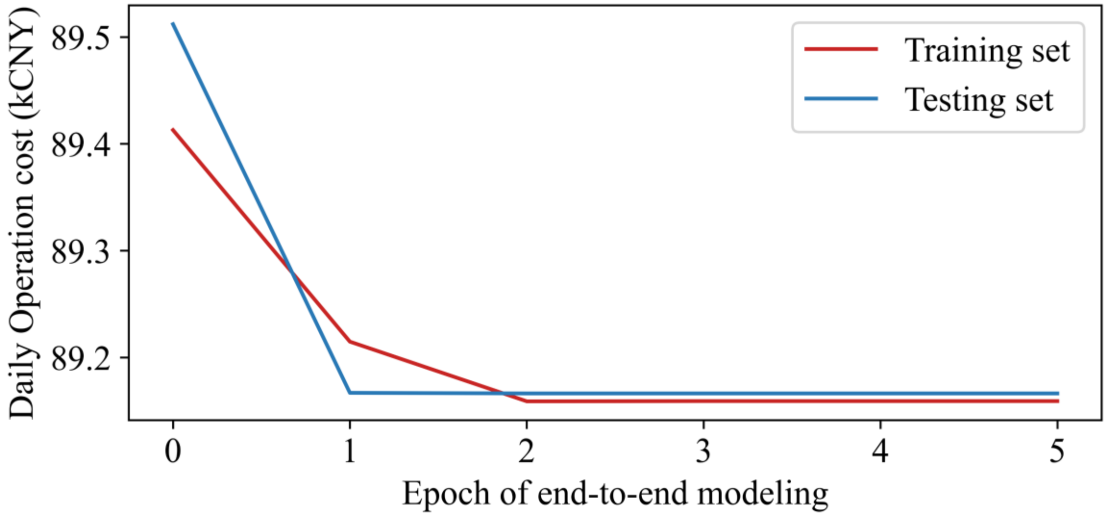
Case Study

Experiment result

- **Forecasting performance & convergence properties**

LOAD FORECASTING PERFORMANCE OF THREE SECTORS IN MES

	Model	MAE	RMSE	MAPE
Electricity sector	Benchmark	82.592	113.102	3.565
	End-to-End	82.383	112.951	3.562
	Accuracy Variation	0.25%	0.13%	0.00%
Heat sector	Benchmark	120.967	178.530	9.074
	End-to-End	121.331	181.348	9.216
	Accuracy Variation	-0.30%	-1.58%	-0.14 %
Cooling sector	Benchmark	487.406	626.781	12.895
	End-to-End	477.677	615.956	12.799
	Accuracy Variation	2.00%	1.73%	0.10%



Daily average cost of the training/testing dataset in the end-to-end modeling process

- Compared to FTO, The end-to-end approach has little effect on the forecasting accuracy.
- The proposed method possesses favorable convergence properties.

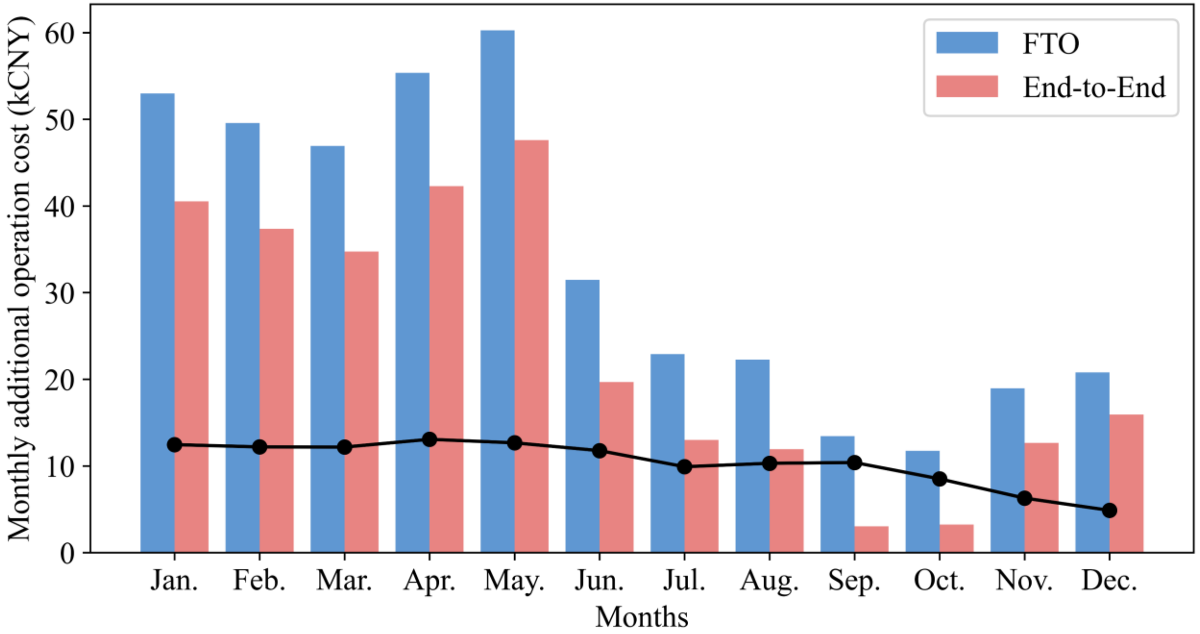
Case Study

○ Experiment result

- **Operation cost**

Daily operation costs for 12 months in 2017 (kCNY)

	FTO	End-to-End	Ideal	Improvement
Jan.	88.517	88.115	86.808	0.454 %
Feb.	87.500	87.064	85.730	0.498 %
Mar.	86.687	86.294	85.174	0.453 %
Apr.	88.015	87.579	86.170	0.495 %
May	88.282	87.873	86.338	0.463 %
Jun.	92.834	92.442	91.786	0.423 %
Jul.	95.141	94.821	94.402	0.336 %
Aug.	93.729	93.373	92.961	0.379 %
Sep.	92.823	92.477	92.375	0.373 %
Oct.	87.576	87.301	87.197	0.314 %
Nov.	86.505	86.231	85.681	0.316 %
Dec.	85.328	85.140	84.527	0.220 %



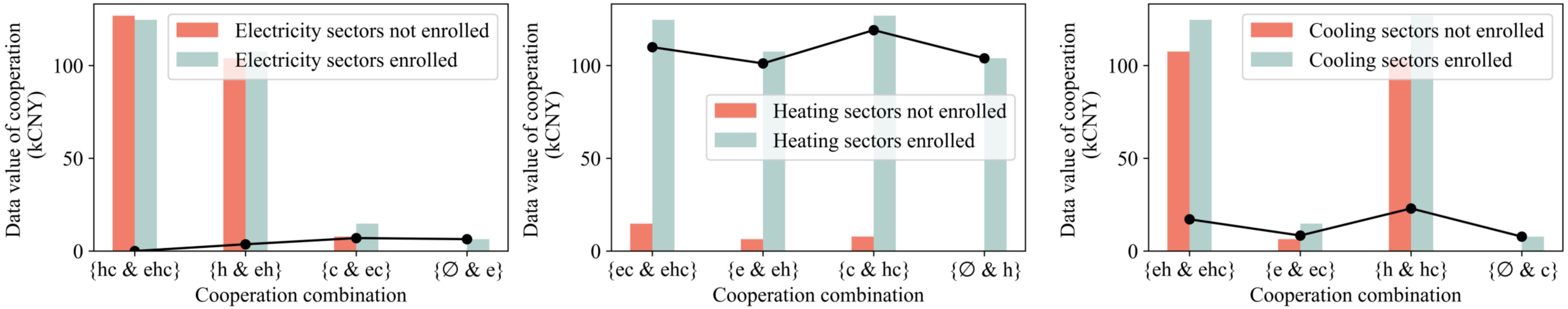
Monthly additional operation cost of FTO and end-to-end model compared to ideal cost (kCNY)

Deal: 31012.06 kCNY
FTO: 31418.71 kCNY (101.31% ideal cost)
End-to-End: 31294.04 kCNY (100.91% ideal cost)

➢ Operation achieves a **0.40%** reduction, resulting in annual cost savings of **124.66 kCNY**.

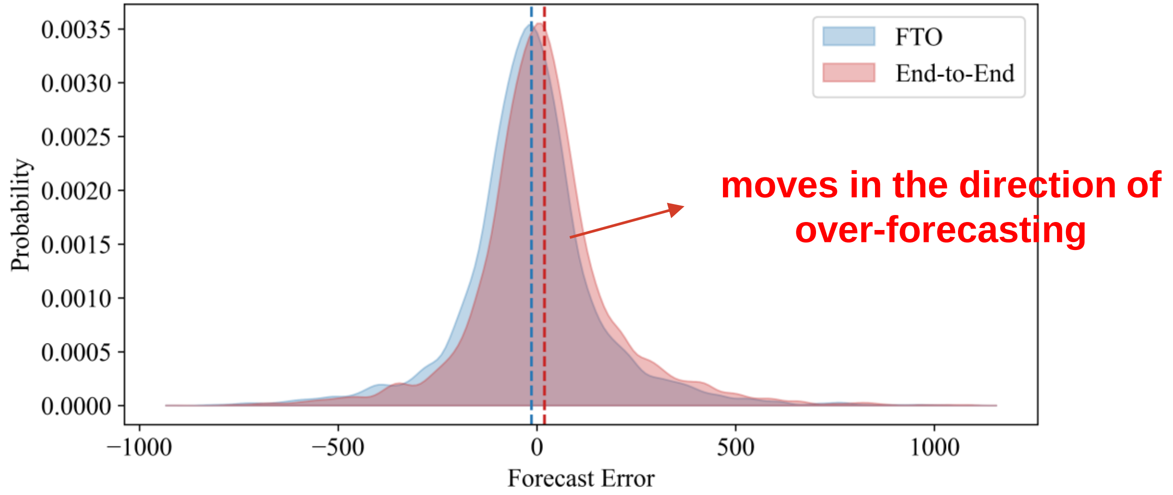
Case Study

○ Experiment result



(a) **(b)** **(c)**

- ➢ Electricity sector makes **little contribution.**
  - Accuracy of the electricity sector is relatively high.
  - The deviation of the electricity price in intra-day and day-ahead is relatively small.
- ➢ Heat sector cooperates with the MES operator can **markedly improve** additional profits.



The distribution of the forecasts error of heat sector

Contents

Conclusions

- **We proposes new insights into exploring deeper integration learning with optimization in urban energy systems:**

  - Propose Neural ODEs-based model structure in model predictive control for building energy management. The proposed learning-based method balances robust and accurate requirements in thermal dynamics modeling. Adaptive MPC mechanism is adopted to improve energy dispatch efficiency, supported by continuous modeling characteristics.
  - Proposes decision-oriented modeling method of building thermal dynamics. The proposed method achieves lower operation costs than the traditional accuracy-oriented modeling methods; the proposed model has properly learned to avoid decision spaces leading to expensive costs.
  - Presents an end-to-end framework designed to quantify data value by integrating forecasting and decision processes. A profit allocation strategy based on contribution to cost savings is investigated, encouraging data sharing in MES.

Thanks for your attention