District cooling system control for grid services based on safe reinforcement learning

Hongcai Zhang
University of Macau
hczhang@um.edu.mo

In collaboration with:

Dr. Peipei YU

Lecturer, Shanghai University of Electric Power (since 2024)

PhD in electrical engineering, University of Macau (2024)

Master in applied mathematics, Zhejiang University (2019)

Bachelor in mathematics, Zhejiang University (2016)

Prof. Yonghua SONG

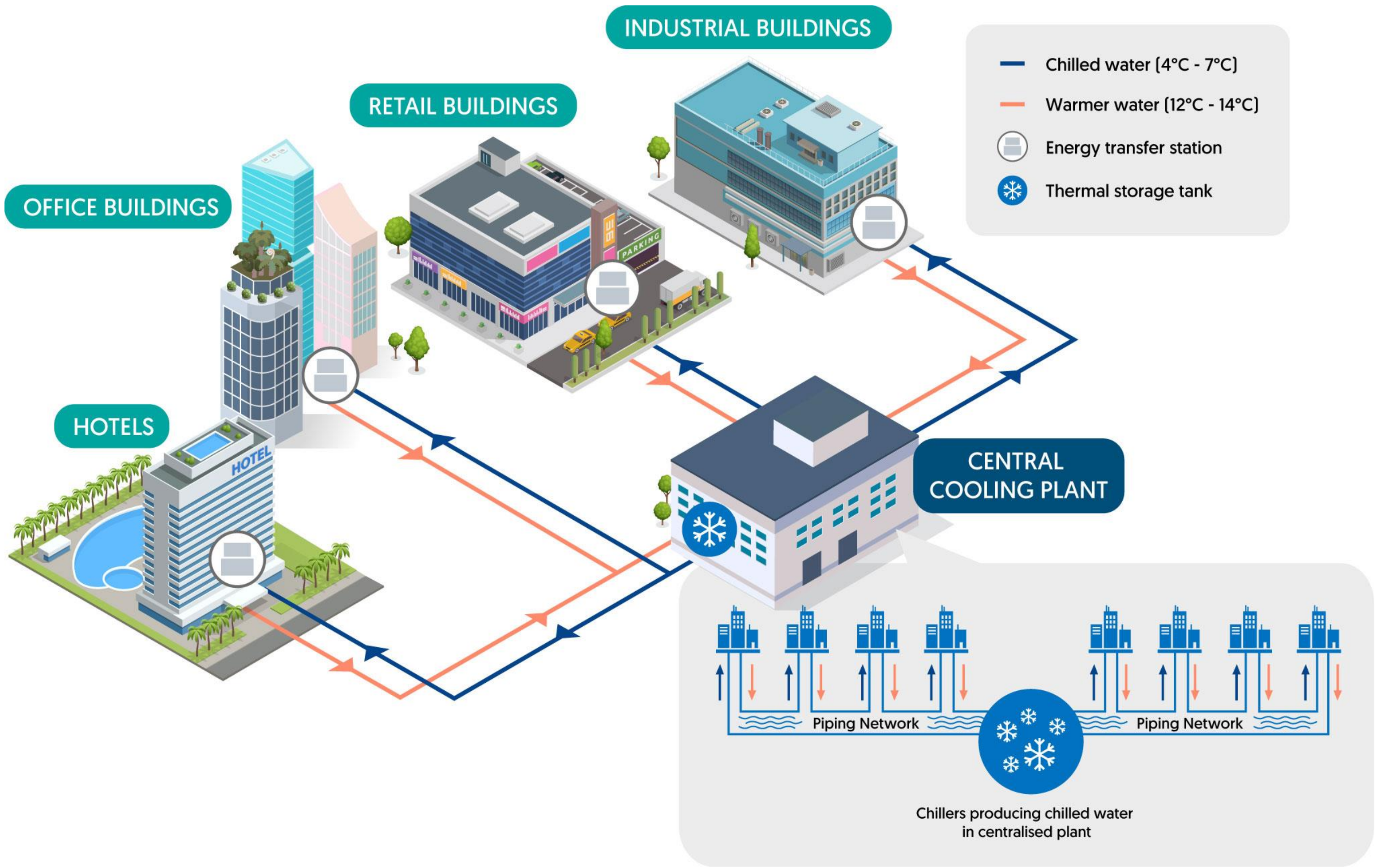
Rector of University of Macau

Director of State Key Laboratory of Internet of Things for Smart City

Fellow of IEEE

Fellow of Royal Academy of Engineering

District cooling system (DCS)



- Producing cooling in **centralized energy station**
- Supplying cooling to buildings in the neighborhood (up to **1.5 km away**) using water pipelines
- Total cooling power capacity of one station can be up to **150 MW** with equivalent electric power up to **30 MW**

Figure: https://www.spgroup-asia.com/solutions/district-cooling

District cooling system (DCS) has significant flexibility

- One district cooling system is **equivalent to thousands of household air conditioning systems** in terms of power capacity
- **No** demand for an **aggregator**
- Significantly enhanced flexibility if installed with **thermal energy storage**

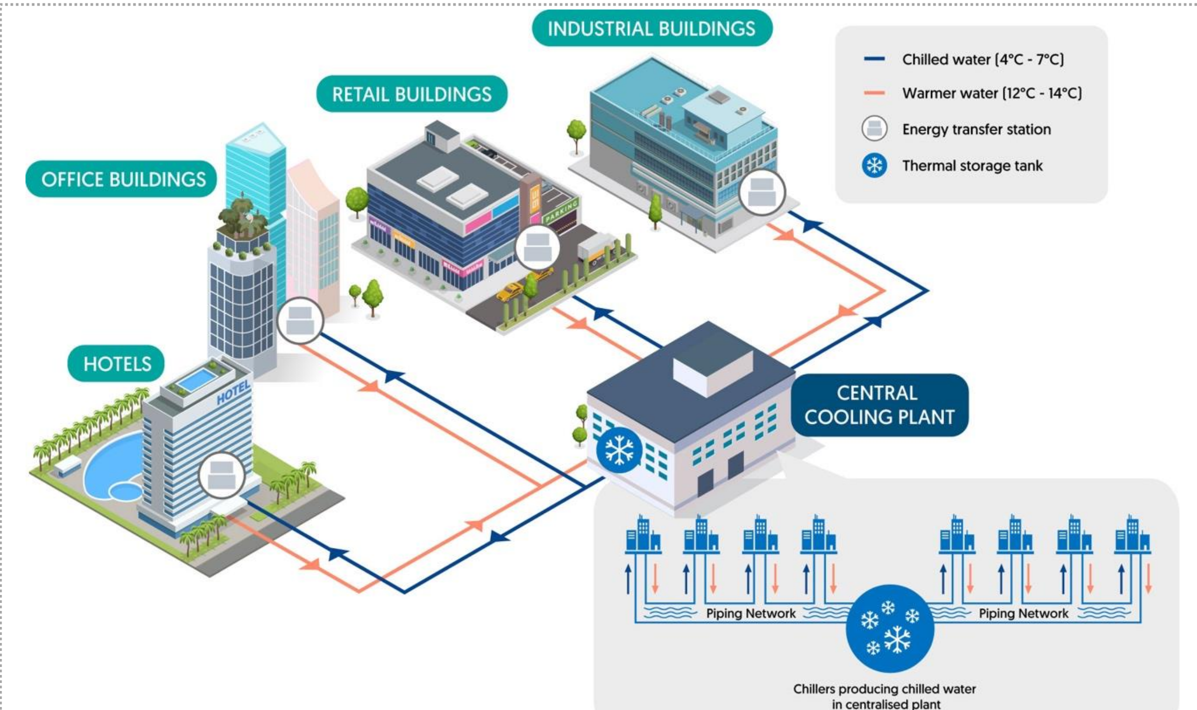
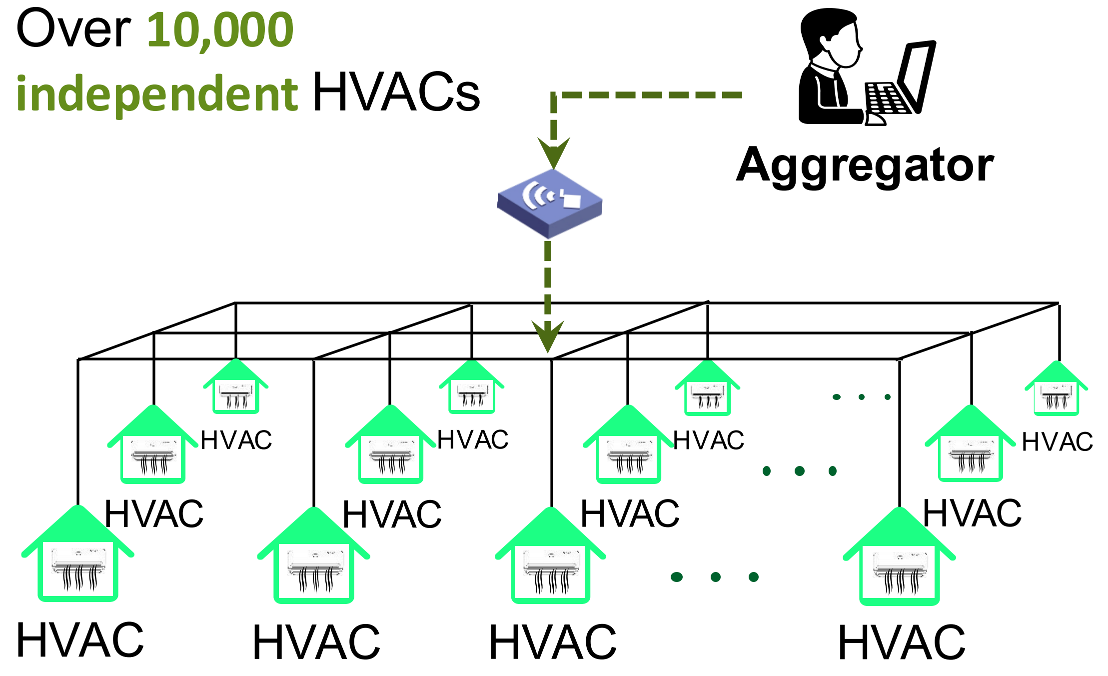


Figure: https://www.spgroup-asia.com/solutions/district-cooling

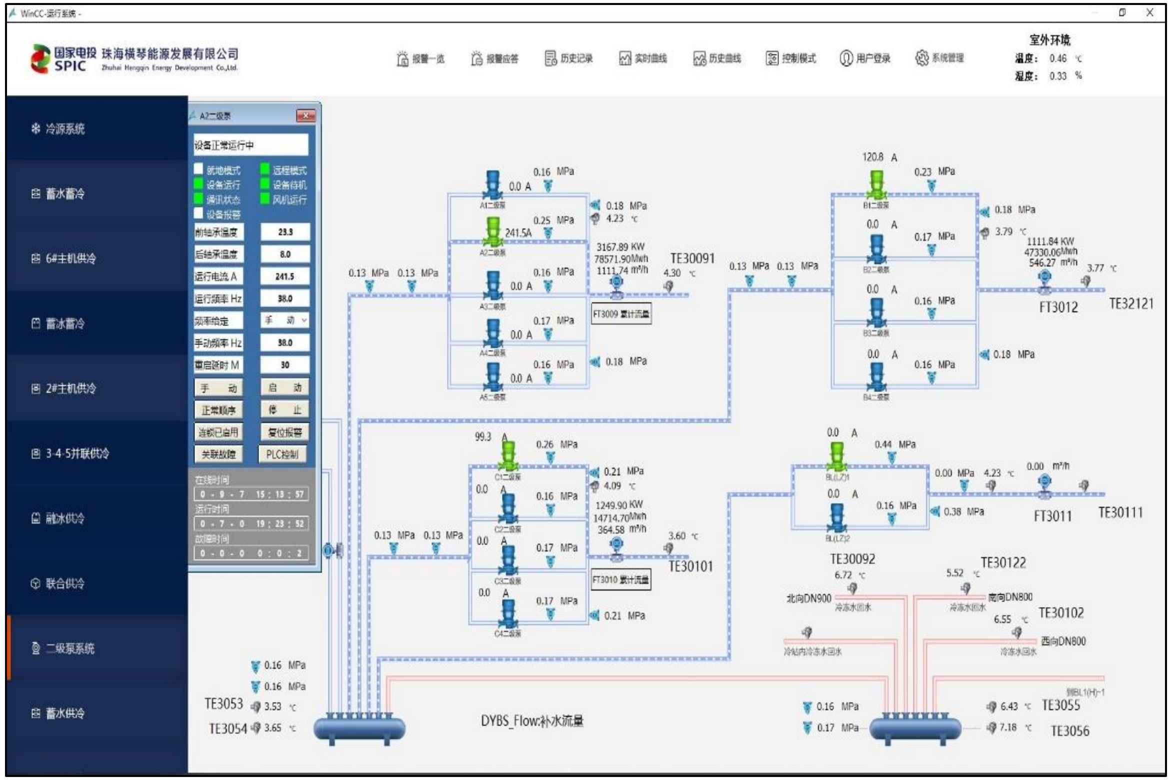
≈



District cooling system (DCS) – a real-world example

- Large-scale **interconnected DCS with ice-storage** in Hengqin, Zhuhai, China
- Arbitrage **time-of-use electricity price differences** (peak/valley>4.5) with **ice-storage**

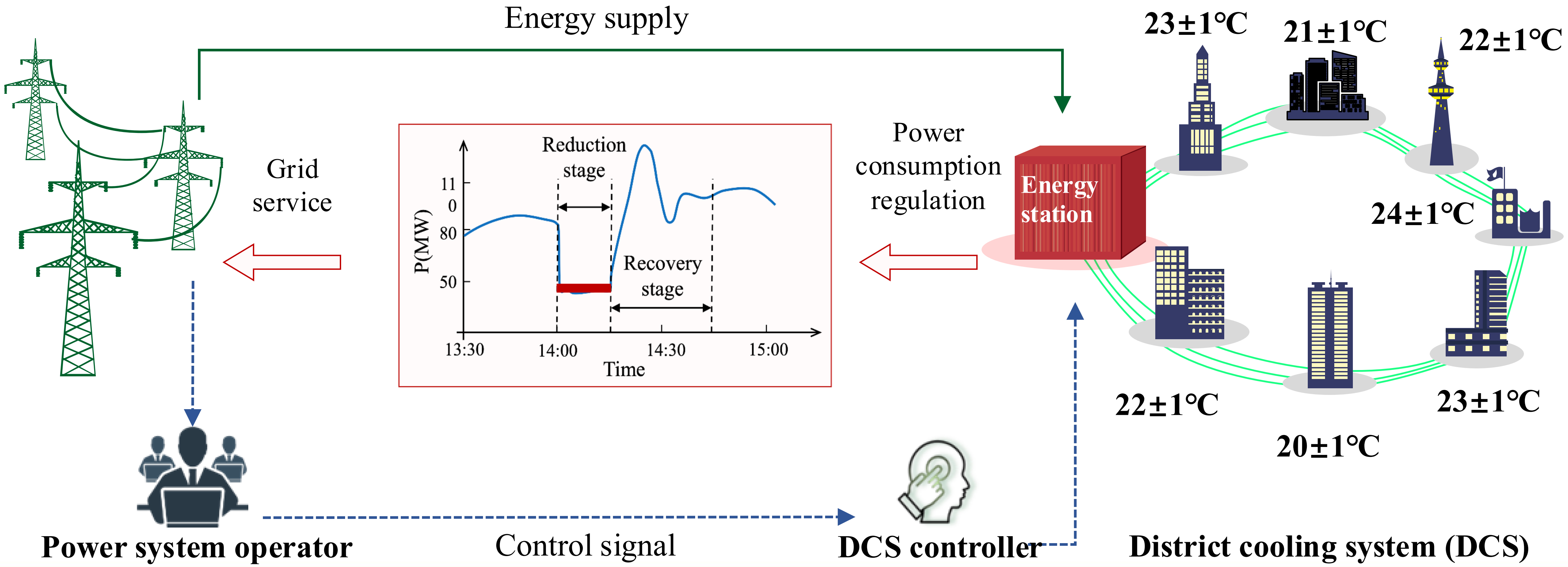
District cooling system planning in Hengqin, Zhuhai, China:
No. of energy stations: **10**, total cooling capacity: **450,000 RT**,
cooling power per station: **up to 150 MW**, electric power : **up to 30 MW**



Operation platform of the #3 DCS energy station in Hengqin:
Cooling capacity: **30,000 RT**, service radiation: **1.5 km**, cooling power: **107 MW**,
electric power : **up to 18 MW**, ice storage capacity: **182 MWh**

Controlling DCS for grid services – problem statement

- **Problem:** **Manipulate** power consumption of a DCS **following regulation signals** from a system operator subject to critical operational constraints:
  - Customers’ **temperature comfort** requirement
  - System operators’ **regulation performance** requirement

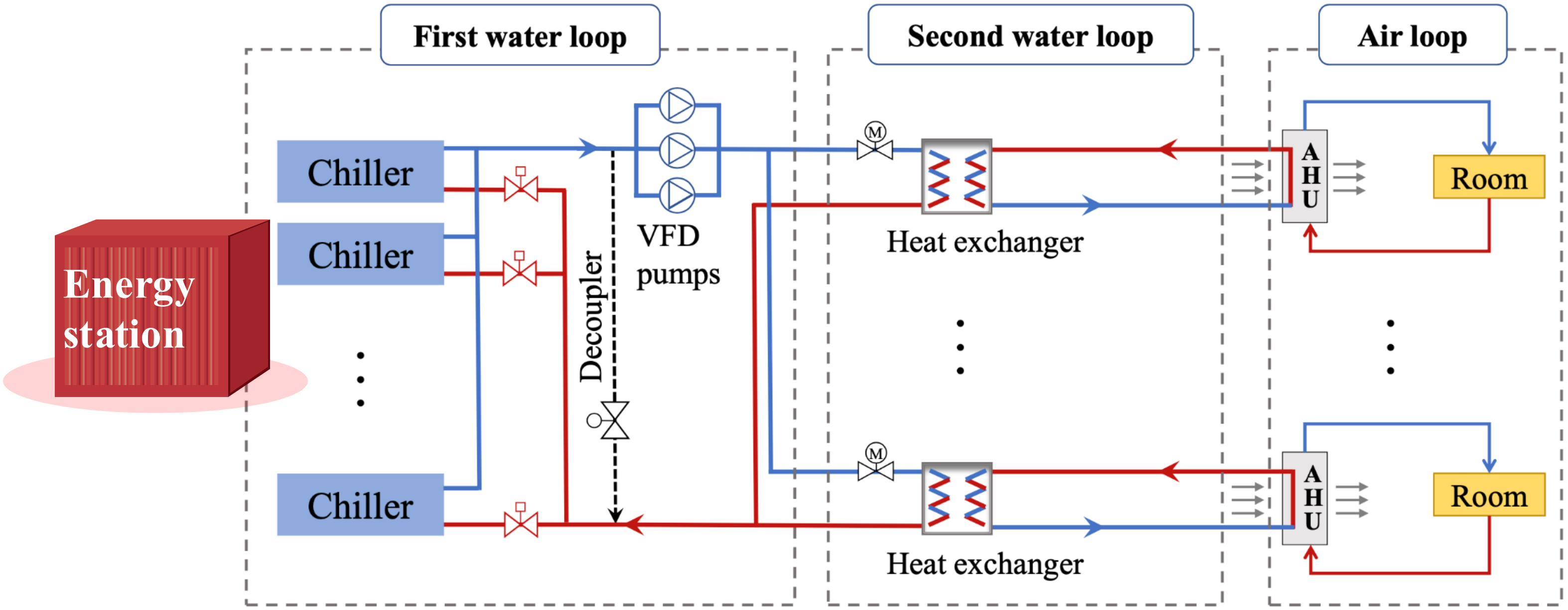


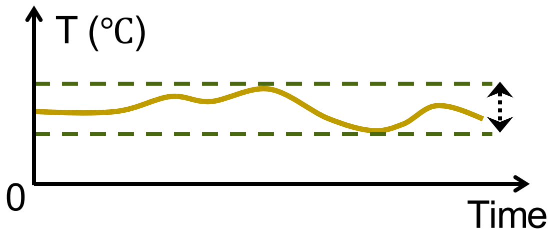
Controlling DCS for grid services is nontrivial

- A DCS's **thermal dynamics** is complex, and its power can only be controlled **indirectly** by adjusting
  - Mass flow rate
  - Supply water temperature

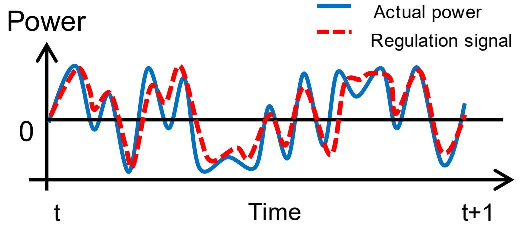
$$Q_{ch} = m_{ch} \cdot c_{water} \cdot (T_{ch\_return} - T_{supply})$$

- Cooling demands & grid regulation are both highly **stochastic**
- Ancillary grid services usually require **fast and accurate** responses





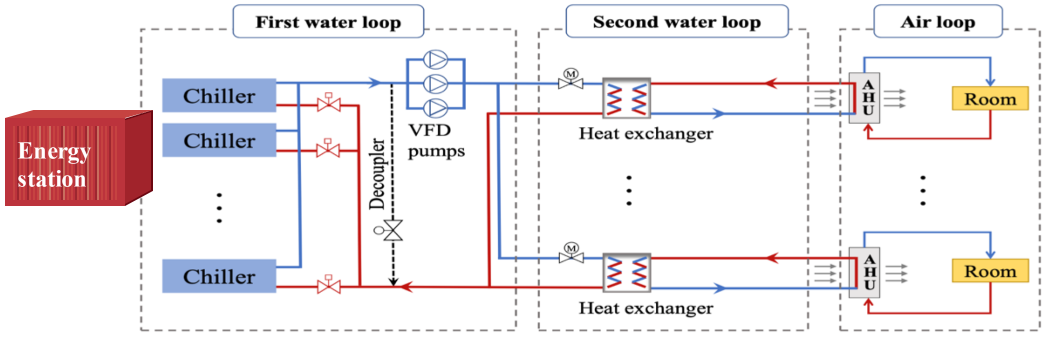
Uncertain cooling demand



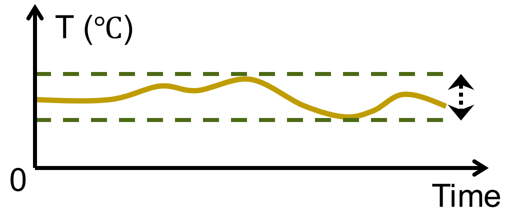
Uncertain grid regulation

Controlling DCS for grid services is nontrivial

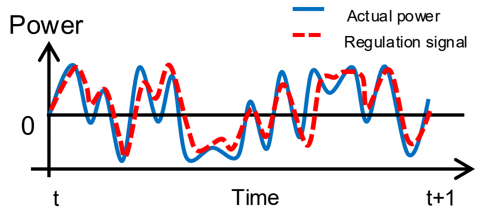
- State of the art in the industry & literature
  - **PI control:** only passively adjusts cooling output based on monitored building temperature
  - **Model predictive control:** requires accurate system model, which may be unavailable in practice
  - **Reinforcement learning (state-of-the-art):** requires no physical model & is adaptive to changing environment, but has critical safety concerns if online training is needed



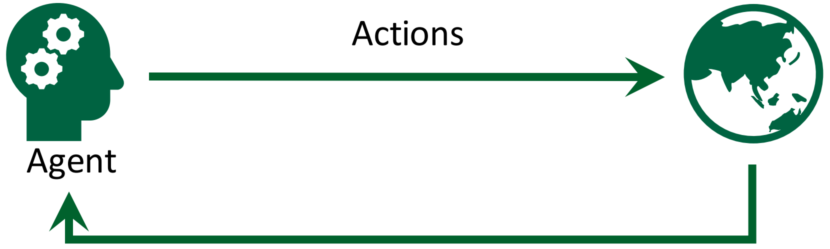
Complex DCS thermal dynamics



Uncertain cooling demand



Uncertain grid regulation



Reinforcement learning (RL)

Trial and error for training

Random interaction with real power grids can be **unsafe!**

Controlling DCS for grid services based on safe RL

- **Safe reinforcement learning (safe RL)***
  - does not require accurate physical model
  - is adaptive to changing environment
  - ensure satisfaction of critical constraints during training/application

- **Our research in today's talk**
  - Scenario 1: **With explicit formula of critical constraints**: Safe layer-based RL control
  - Scenario 2: **With partial formula of critical constraints**: Barrier function-based RL control
  - Scenario 3: **Without formula of critical constraints**: CVaR-based RL control

*Survey on safe RL applications in power systems:

P. Yu, **H. Zhang**, Y. Song, et. al., "Safe Reinforcement Learning for Power System Control: A Review," ***Renewable and Sustainable Energy Reviews***, vol. 223, p. 116022, 2025.

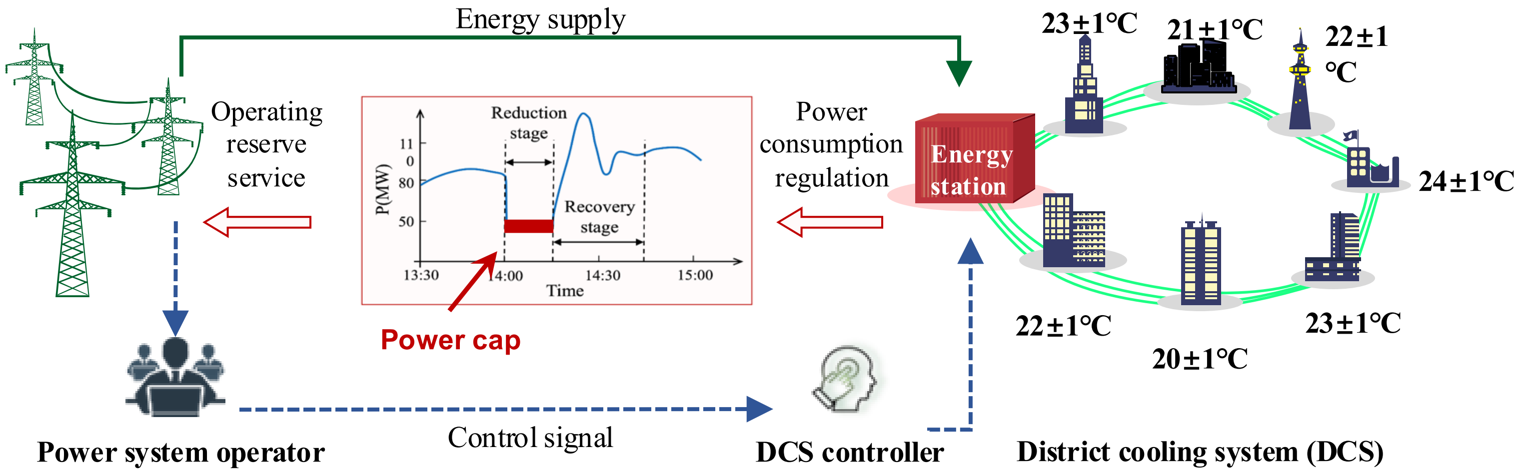
Scenario 1: Control problem with explicit formula of critical constraints – problem statement

- **Objective**: Provide **demand response** by actively adjust DCS power consumption

A. Reduce power consumption by promised MWh for a given period following grid signal

B. Balance the temperature among buildings

C. Prevent power rebound after the service

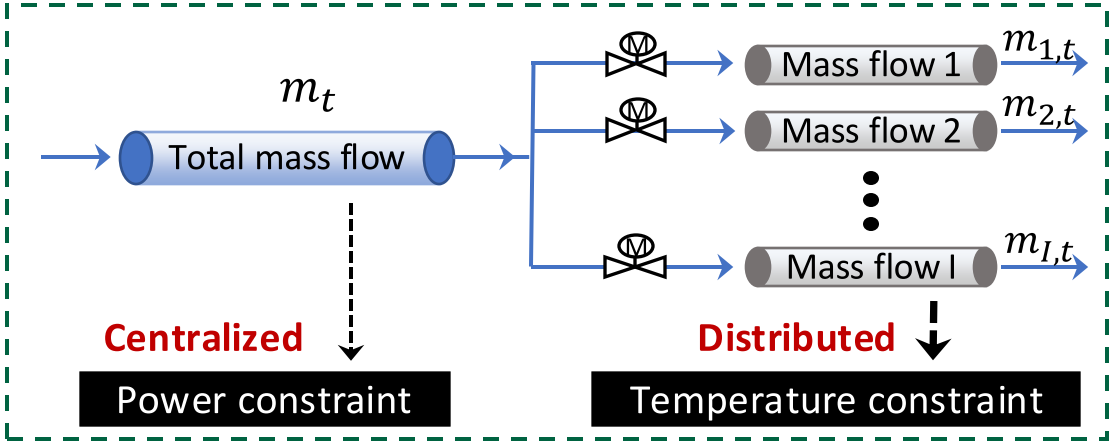


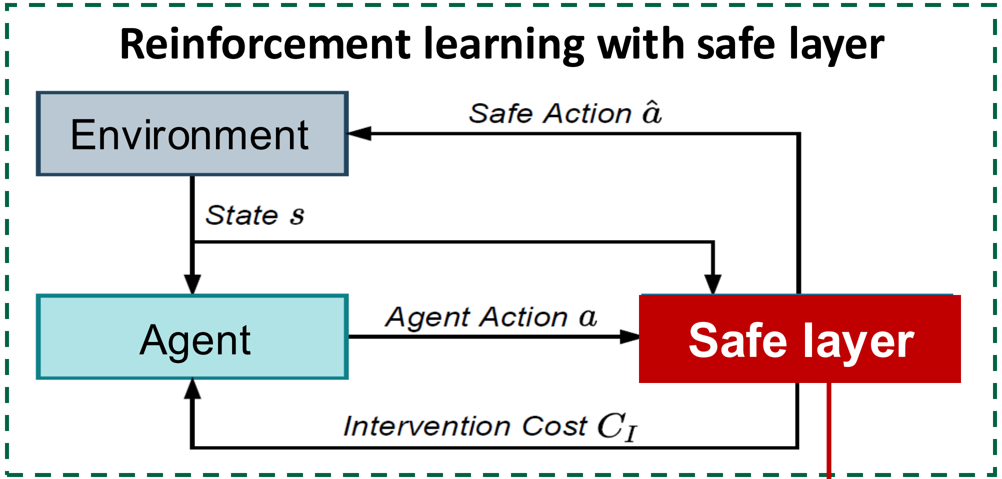
- **Action**: mass follow rates m_t (total) and $m_{i,t}$ (buildings) (cannot directly adjust power)

- **Critical constraints**:
  - Indoor thermal comforts (temperature)
  - Power reduction shall satisfy the grid operator's requirement $P_t = f(m_t) = f(\Delta m_t + m_{t-1}) \leq P^{\max}$
  - Assumption: with accurate power consumption—mass flow relationship, building dynamics are known

Scenario 1: Safe layer-based RL (safe-DRL) control for problem with explicit formula of critical constraints

- **Solution:** Adopt **model-based safe layer (safe-DRL)** to map unsafe actions to safe ones





Keep the proportion of **the ordinary mass flow**

Satisfy the power **constraint**

$$(m_t) = \mu_t \cdot \Delta m_t + v_t \cdot m_{t-1}$$

Keep the proportion of **the action**

Map unsafe actions to safe ones based on linear optimization

$$\Delta\tilde{\boldsymbol{m}}^{\mathrm{I}}_t \leftarrow \Delta\boldsymbol{m}^{\mathrm{I}}_t + \mu_t \Delta\boldsymbol{m}^{\mathrm{I}}_t + \upsilon_t \boldsymbol{m}^{\mathrm{I}}_t, \quad \forall t \in \mathcal{T},$$

$$\max_{\mu_t, \upsilon_t} \ \mu_t + \upsilon_t,$$

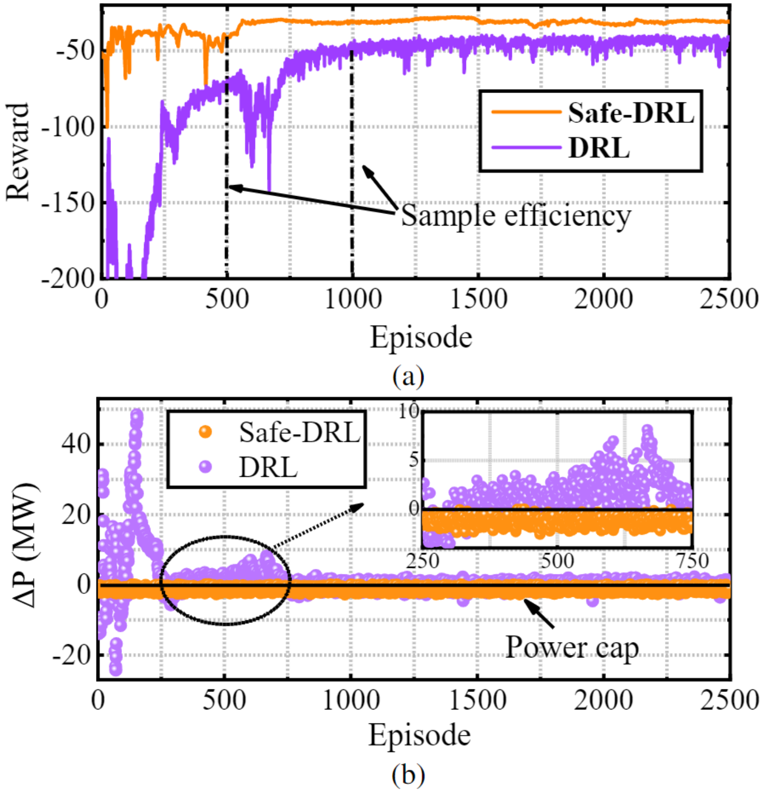
$$\text{s.t.:} \ \sum_{i\in\mathcal{I}} (\mu_t \Delta m^{\mathrm{I}}_{i,t} + \upsilon_t m^{\mathrm{I}}_{i,t})\Theta_t \le P^{\mathrm{cap}}, \quad \forall t \in \mathcal{T},$$

$$\underline{m}^{\mathrm{I}}_i \le \mu_t \Delta m^{\mathrm{I}}_{i,t} + \upsilon_t m^{\mathrm{I}}_{i,t} \le \overline{m}^{\mathrm{I}}_i, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T},$$

$$\mu_t, \upsilon_t, \le 0, \quad \forall t \in \mathcal{T},$$

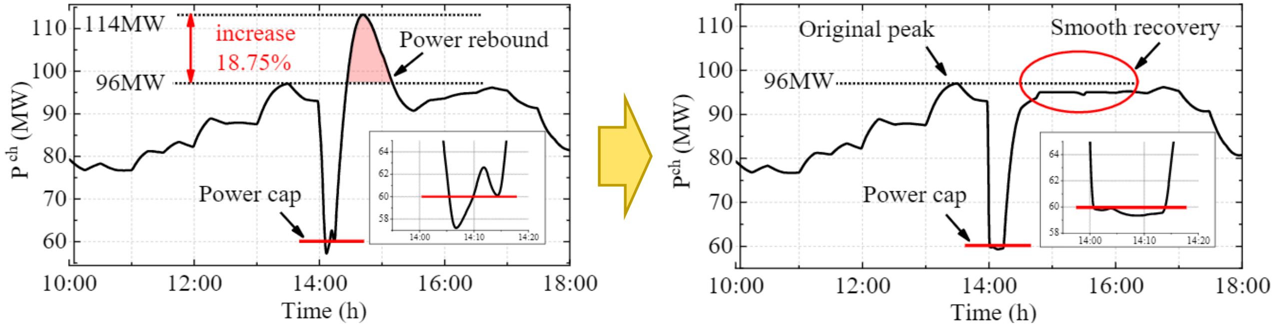
Scenario 1: Safe layer-based RL (safe-DRL) control for problem with explicit formula of critical constraints

- **Results:** Enhanced training efficiency & regulation performance with **strict satisfaction of critical constraints**



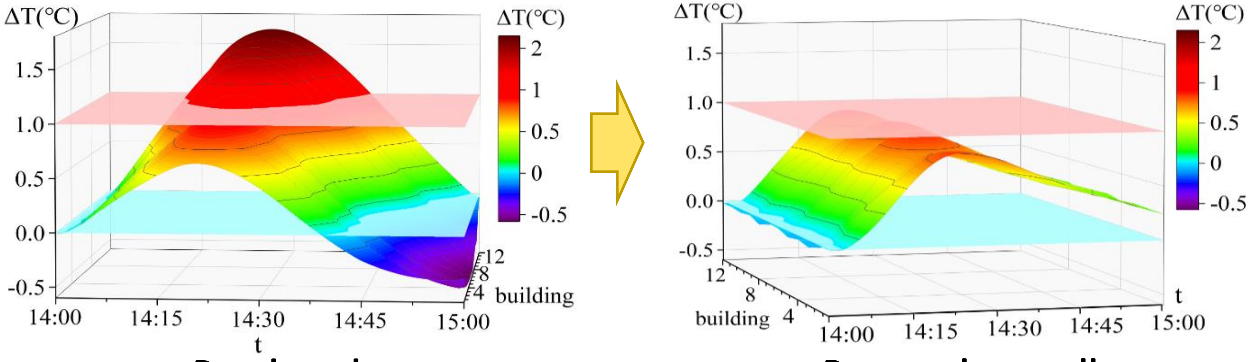
Training process of the DDPG-CBF agent
(a) Reward, (b) Constraint violations

Accurate reserve service control without significant rebound



Benchmark | Proposed controller

Guaranteed & fair building temperature comfort



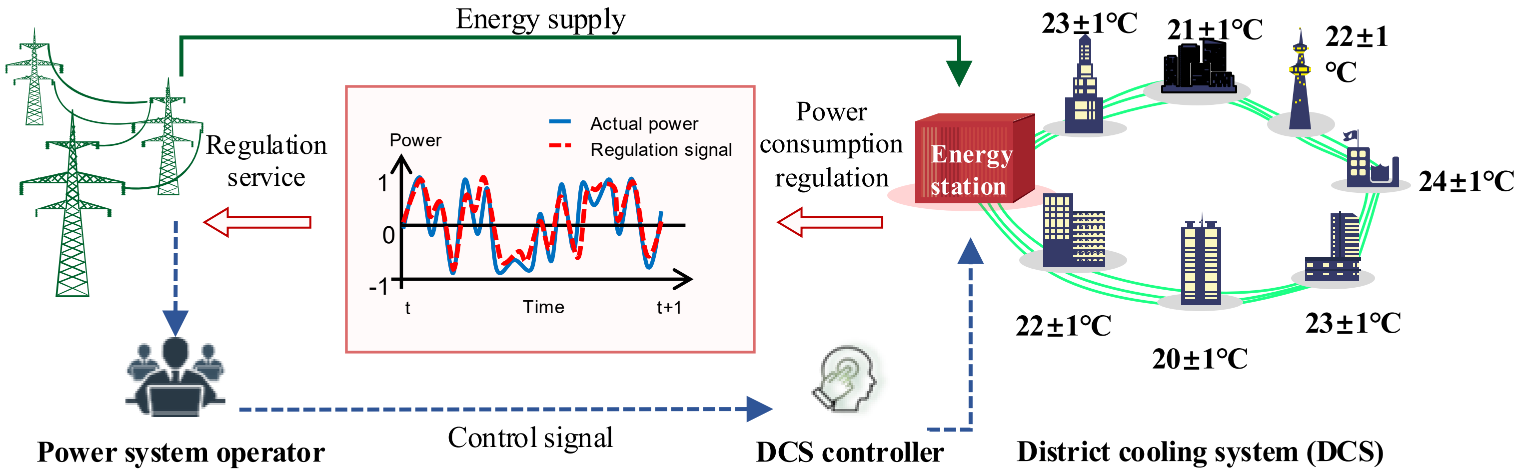
Benchmark | Proposed controller

Scenario 2: Control problem with partial formula of critical constraints – problem statement

- **Objective**: Provide **regulation services** by actively adjust DCS power consumption

  A. Adjust power consumption following high frequent regulation signals

  B. Balance the temperature influence among heterogenous buildings



- **Action**: mass follow rates m_t (total) and $m_{i,t}$ (buildings) (cannot directly adjust power)

- **Critical constraints**:
  - Indoor thermal comforts (temperature) – without accurate power consumption—mass flow relationship, building dynamics are also unknown
  - Regulation performance – without explicit formula

Scenario 2: Barrier function-based RL (DDPG-CBF) control for problem with partial formula of critical constraints

- **Solution:** Safe reinforcement learning with **control barrier function** (CBF) & **Gaussian process estimation (DDPG-CBF)** that handles **general unobservable critical constraints**

Key idea 1 (partial formula)

Control-affine deterministic system dynamics

$$s_{t+1} = \underline{f(s_t)} + \underline{g(s_t)}a_t + \underline{d(s_t)}$$

Known system formula

Unknown system formula

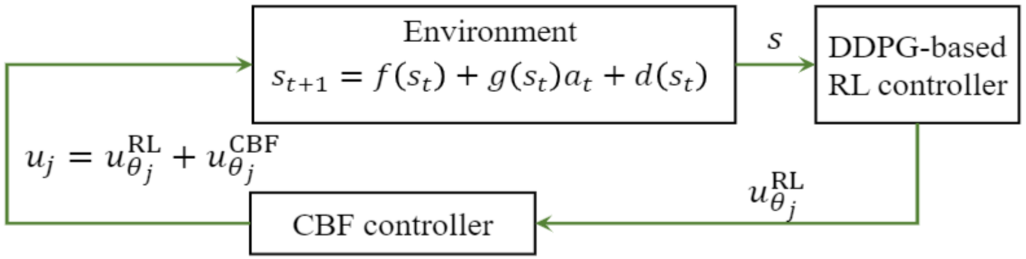
Gaussian process estimation

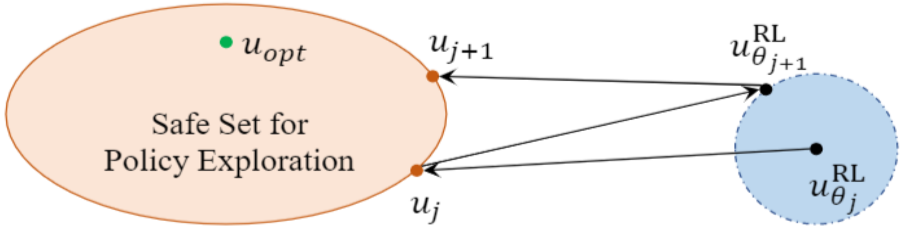
$$\boxed{\mu_d(s) - k_\delta \sigma_d(s) \leq d(s) \leq \mu_d(s) + k_\delta \sigma_d(s)}$$

Key idea 2 (general constraint)

Ensure safety by **compensating CBF controller**

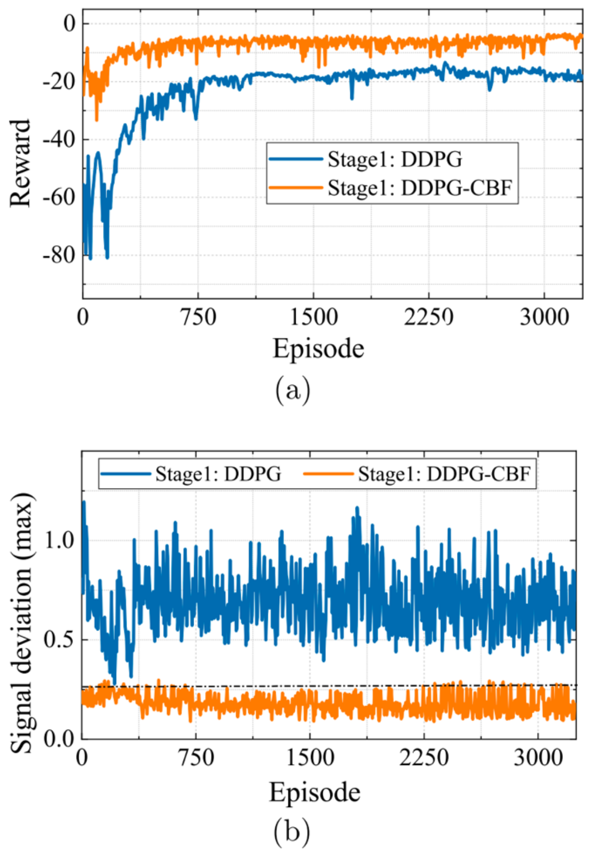
$$u_j(s_t) = u^{\mathrm{RL}}_{\theta_j}(s_t) \boxed{+\, u^{\mathrm{CBF}}_j(s_t, u^{\mathrm{RL}}_{\theta_j})}$$



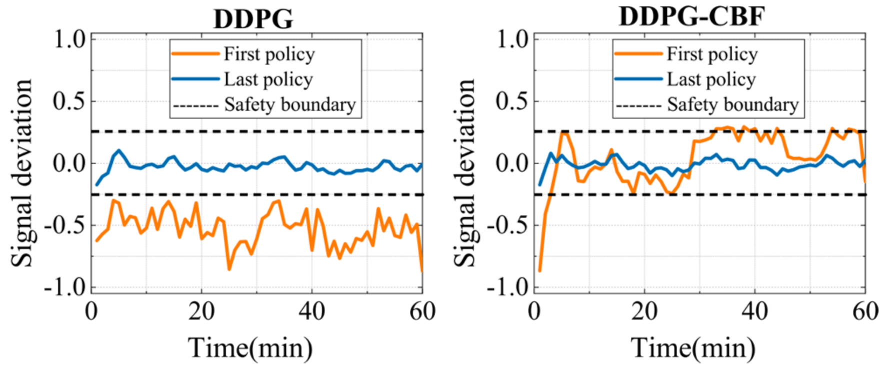


Scenario 2: Barrier function-based RL (DDPG-CBF) control for problem with partial formula of critical constraints

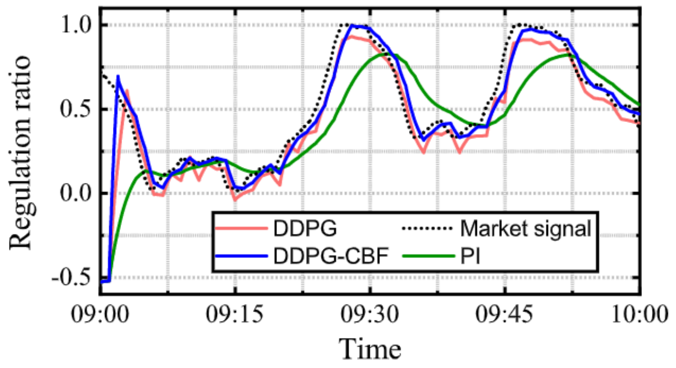
- **Results:** High-performance regulation services with **guaranteed control quality**



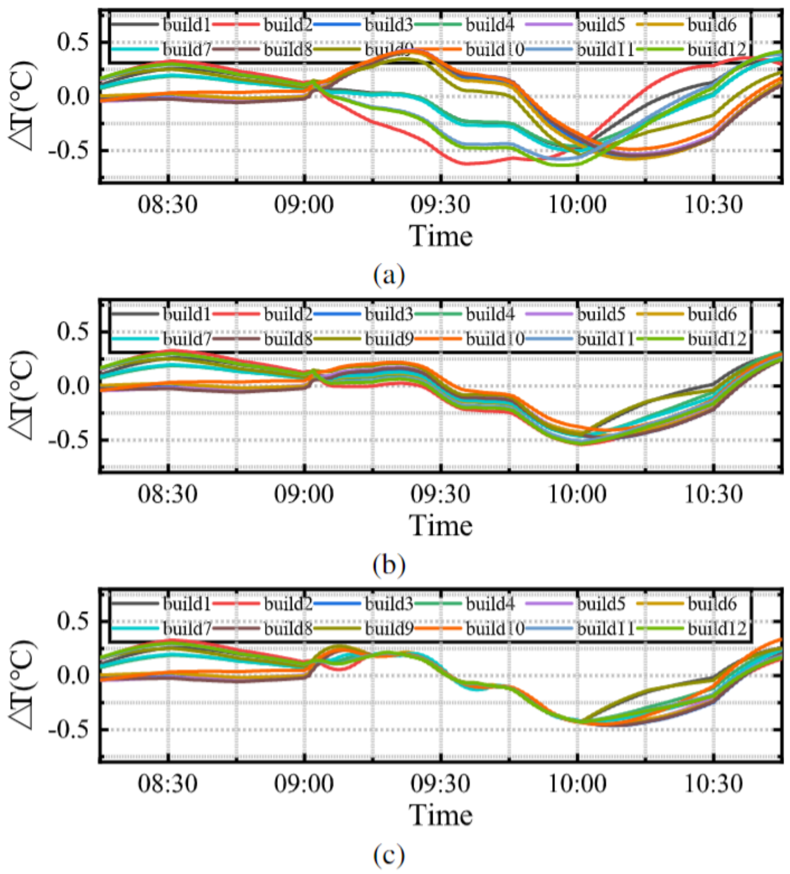
Training process of CBF-based DDPG (a) Reward, (b) Constraint violations



DDPG-CBF has better initial & trained policies



Regulation signal following performance

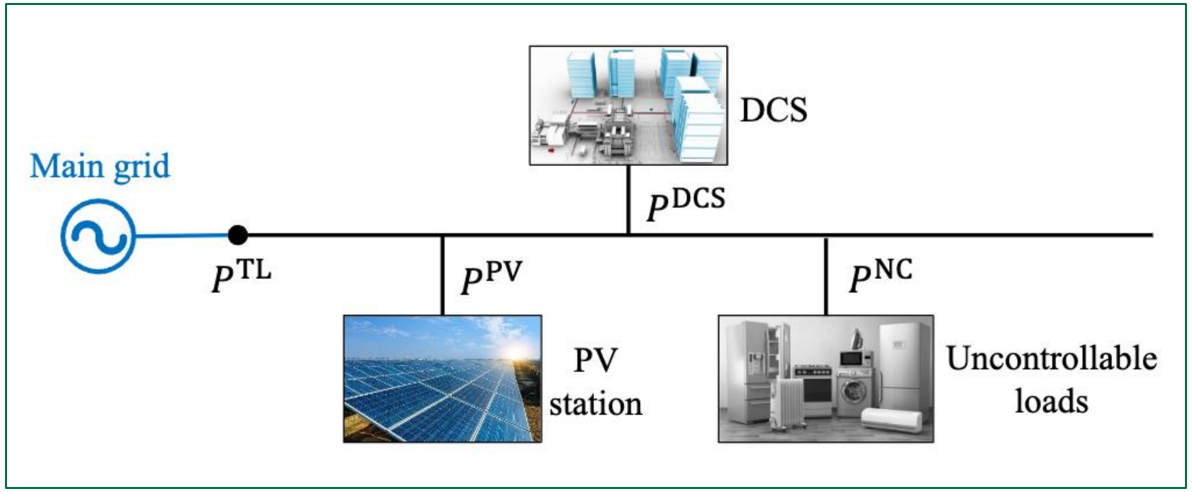


Building temperature deviations (a) PI, (b) DDPG and (c) CBF-based DDPG

Scenario 3: Control problem without formula of critical constraints – problem statement

- **Objective**: **Smooth tie-line power** by actively adjust DCS power consumption in distribution network

  A. Adjust power consumption to smooth tie-line power flow
  B. Balance the temperature influence among heterogenous buildings



- **Action**: mass follow rates m_t (total) and $m_{i,t}$ (buildings) (cannot directly adjust power)

- **Critical constraints**:
  - Indoor thermal comforts (temperature) – without accurate power consumption—mass flow relationship, , building dynamics are also unknown
  - Power flow constraint violation – without explicit formula

Scenario 3: CVaR-based RL (RSAC) control for problem without formula of critical constraints

- **Solution:** Risk-averse reinforcement learning with **conditional value-at-risk constraints – risk-averse soft actor-critic (RSAC)**

Conventional constrained Markov decision process (CMDP)

$$\textbf{P1:}\ \max_{\pi}\ J(\pi) = \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}\left[\sum\nolimits_t^{\infty} \gamma^t r(s_t,a_t)\right]$$

$$\text{s.t.:}\ \boxed{D(\pi) = \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}\left[\sum\nolimits_t^{\infty} \gamma^t c(s_t,a_t)\right] \le d,}$$

Expectation of critical constraints

- cannot consider constraint variance

Risk-averse constrained Markov decision process (risk-averse CMDP)

$$\textbf{P2:}\ \max_{\pi}\ J(\pi)$$

$$\text{s.t.:}\ \boxed{\begin{aligned}&\Gamma_\pi(s,a,\alpha) \doteq CVaR_\alpha \le d,\\ &\Gamma_\pi(s,a,\alpha) = Q_\pi^c(s,a) + \alpha^{-1}\phi(\Phi^{-1}(\alpha))\sqrt{V_\pi^c(s,a)},\\ &G_\pi^c(s,a) = \sum\nolimits_{t=0}^{\infty} \gamma^t c(s_t,a_t) \sim \mathcal{N}(Q_\pi^c(s,a), V_\pi^c(s,a)).\end{aligned}}$$

Probabilistic critical constraints based on conditional value-at-risk (CVaR)

- can consider constraint variance
- trade-off between optimality & risk

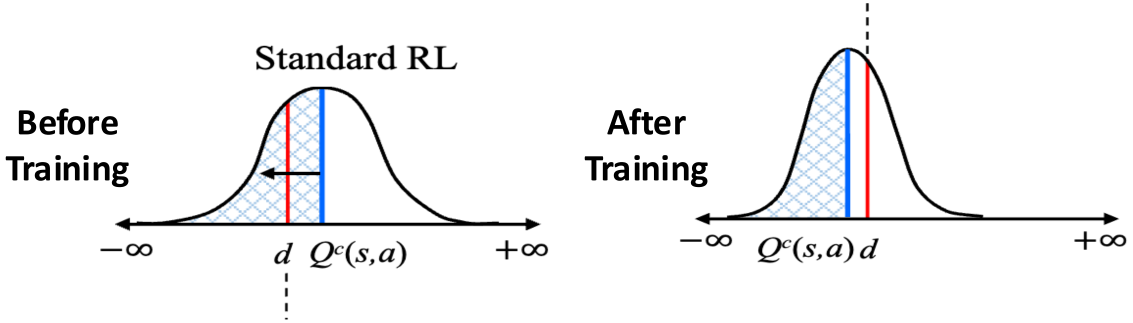
Scenario 3: CVaR-based RL (RSAC) control for problem without formula of critical constraints

- **Solution:** Risk-averse reinforcement learning with **conditional value-at-risk constraints – risk-averse soft actor-critic (RSAC)**

Conventional constrained Markov decision process (CMDP)

$$\textbf{P1:}\ \max_{\pi}\ J(\pi) = \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}\left[\sum\nolimits_t^{\infty} \gamma^t r(s_t,a_t)\right]$$

$$\text{s.t.:}\ \boxed{D(\pi) = \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}\left[\sum\nolimits_t^{\infty} \gamma^t c(s_t,a_t)\right] \le d,}$$

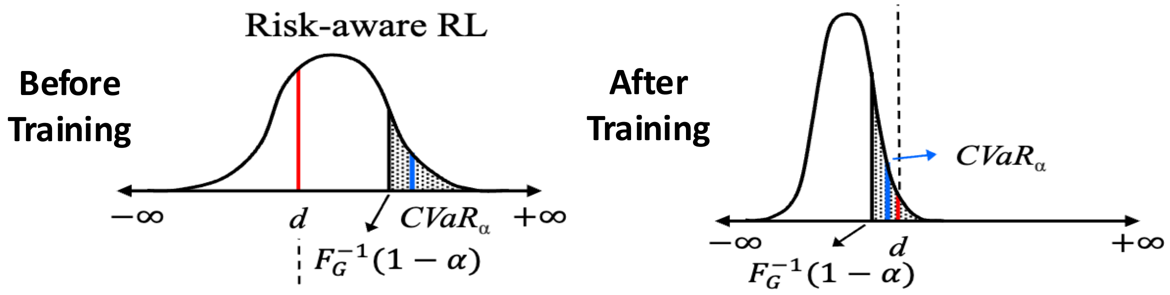


Standard RL with expectation safety

Risk-averse constrained Markov decision process (risk-averse CMDP)

$$\textbf{P2:}\ \max_{\pi}\ J(\pi)$$

$$\text{s.t.:}\ \boxed{\Gamma_\pi(s,a,\alpha) \doteq CVaR_\alpha \le d,}$$



Risk-aware RL with conditional value-at-risk safety

Scenario 3: CVaR-based RL (RSAC) control for problem without formula of critical constraints

- **Solution:** Risk-averse reinforcement learning with **conditional value-at-risk constraints – risk-averse soft actor-critic (RSAC)**

Risk-averse constrained Markov decision process (risk-averse CMDP)

P2: $\max_{\pi} \; J(\pi)$

s.t.:
$$\Gamma_\pi(s,a,\alpha) \doteq CVaR_\alpha \le d,$$
$$\Gamma_\pi(s,a,\alpha) = Q_\pi^c(s,a) + \alpha^{-1}\phi(\Phi^{-1}(\alpha))\sqrt{V_\pi^c(s,a)},$$
$$G_\pi^c(s,a) = \sum_{t=0}^{\infty} \gamma^t c(s_t,a_t) \sim \mathcal{N}(Q_\pi^c(s,a), V_\pi^c(s,a)).$$

Probabilistic critical constraints based on conditional value-at-risk (CVaR)

- can consider constraint variance
- trade-off between optimality & risk

Risk-aware soft actor-critic

Training objective: $\max_{\pi} \min_{\kappa \ge 0} J(\pi) - \kappa(\Gamma_\pi(s,a,\alpha) - d)$

Actor-critic structure:

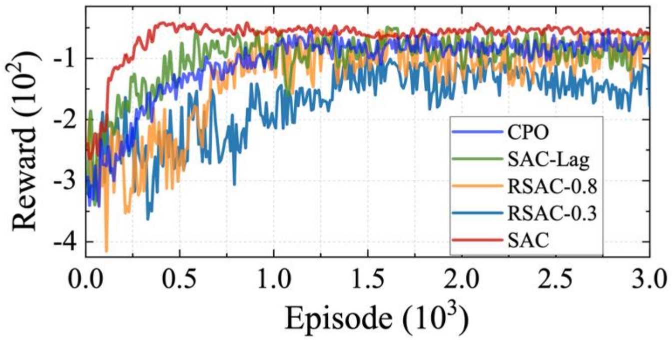
Introduce a Safety-critic neutral network to guide probabilistic constraints satisfaction

- Wasserstein metric for evaluating distribution distance:

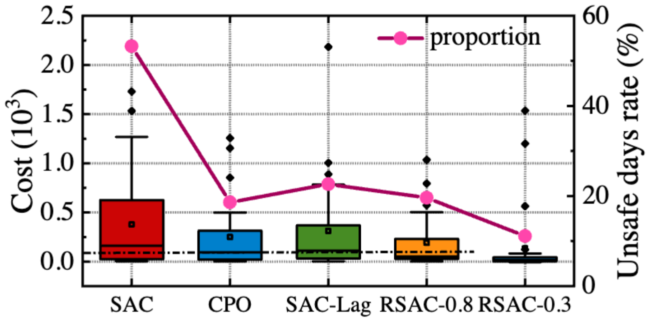
$$W_p(a,b) \doteq \left(\int_0^1 |F_a^{-1}(s) - F_b^{-1}(s)|^p ds \right)^{1/p}$$

Scenario 3: CVaR-based RL (RSAC) control for problem without formula of critical constraints

- **Results:** By selecting different risk levels, the proposed RSAC **can self-adaptively achieve the trade-off** between policy optimality and constraint safety

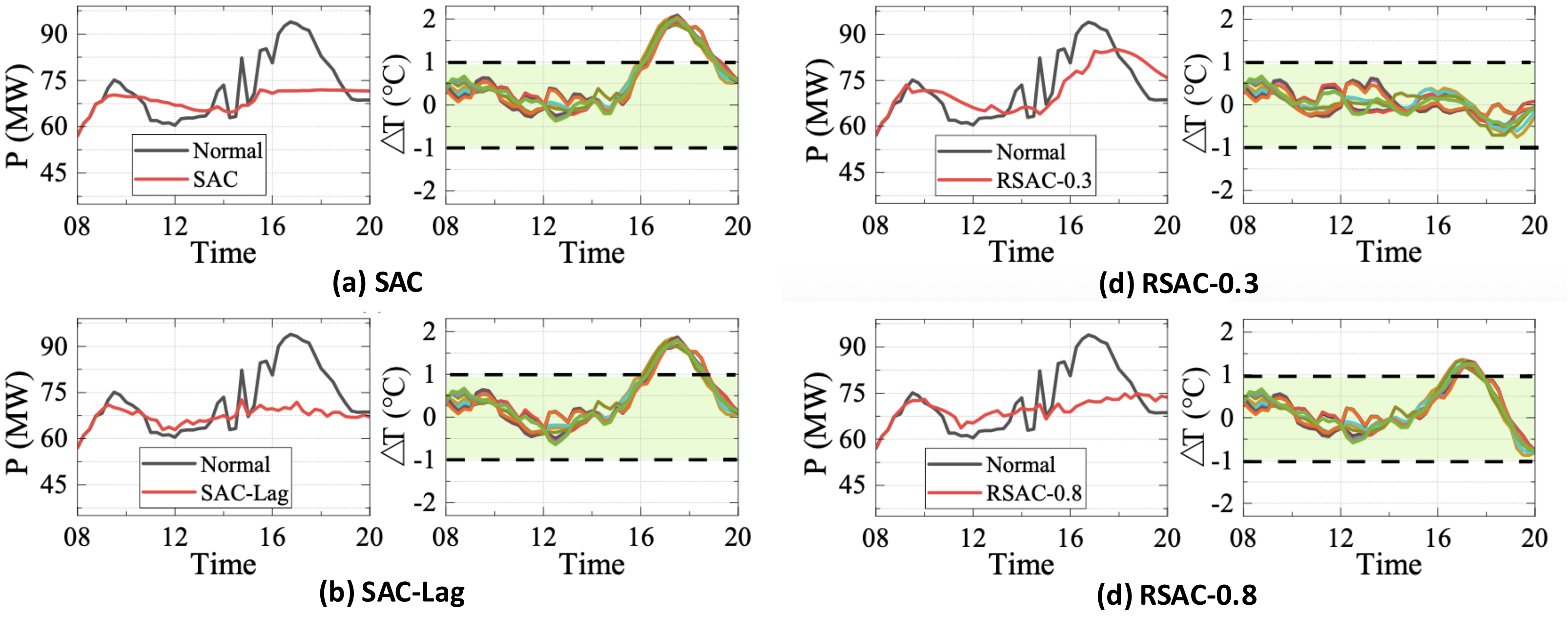


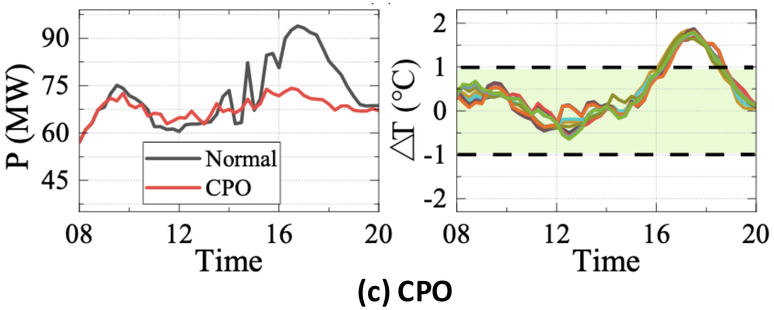
(a)



(b)

Training process & costs analysis
(a) Reward (tie-line smoothing), (b) Cost (temperature violations)

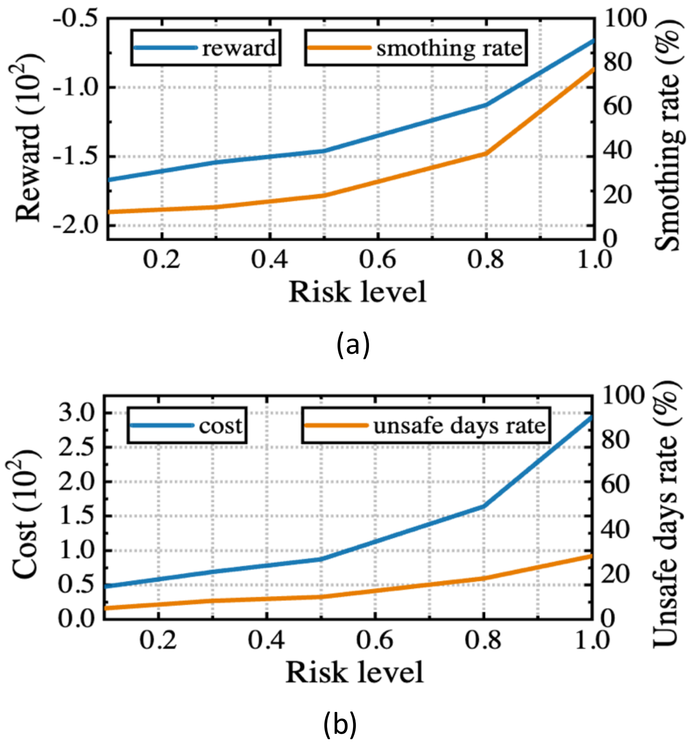




Daily (8:00 to 20:00) regulation performance of different methods (left: tie-line power, right: indoor temperature)

Scenario 3: CVaR-based RL (RSAC) control for problem without formula of critical constraints

- **Results:** By selecting different risk levels, the proposed RSAC **can self-adaptively achieve the trade-off** between policy optimality and constraint safety



Sensitivity analysis on risk levels (a) Reward (tie-line smoothing), (b) Cost (temperature violations)

Methods	Reward	Cost	Smoothing rate (%)	Unsafe days rate(%)
RSAC-0.1	-167.0 ± 13.1	47.4 ± 4.9	12.4	5.0
RSAC-0.3	-154.3 ± 9.5	69.0 ± 8.4	14.6	8.3
RSAC-0.5	-146.0 ± 8.3	87.2 ± 7.6	19.8	10.0
RSAC-0.8	-112.6 ± 7.8	164.2 ± 15.7	38.9	18.3
RSAC-1.0	-101.5 ± 7.4	294.0 ± 27.3	77.1	28.3
SAC-Lag	-98.2 ± 8.1	306.5 ± 28.7	74.3	30.0
CPO	-104.6 ± 6.9	242.1 ± 32.5	79.2	25.0
SAC	-65.8 ± 4.3	478.6 ± 16.9	90.1	53.3

Risk Level

Small

Large

Relevant publications

1. **H. Zhang**, Y. Song, G. Chen, and P. Yu, "Reliable Non-Parametric Techniques for Energy System Operation and Control: Fundamentals and Applications of Constraint Learning and Safe Reinforcement Learning Methods," ***Elsevier***, 2025.
2. P. Yu, **H. Zhang** and Y. Song, "Equivalent System Model of District Cooling System in Frequency Domain to Provide Primary Frequency Regulation," ***CSEE Journal of Power and Energy Systems***, Early Access, 2023.
3. P. Yu, **H. Zhang**, Y. Song, et. al., "District Cooling System Control for Providing Operating Reserve Based on Safe Deep Reinforcement Learning," ***IEEE Transactions on Power Systems***, vol. 39, pp. 40-52, 2023.
4. P. Yu, **H. Zhang** and Y. Song, "District Cooling System Control for Providing Regulation Services based on Safe Reinforcement Learning with Barrier Functions," ***Appllied Energy***, vol. 347, pp. 121396, 2023.
5. P. Yu, **H. Zhang**, Y. Song, et. al., "Frequency Regulation Capacity Offering of District Cooling System: An Intrinsic-motivated Reinforcement Learning Method," ***IEEE Transactions on Smart Grid***, vol. 14, no. 4, pp. 2762-2773, 2023.
6. P. Yu, **H. Zhang** and Y. Song, "Adaptive Tie-Line Power Smoothing of District Cooling System with Renewable Generation based on Risk-aware Reinforcement Learning," ***IEEE Transations on Power Systems***, vol. 39, no. 6, pp. 6819-6832, 2024.
7. P. Yu, **H. Zhang**, Z. Hu, and Y. Song, "Voltage control of distribution grid with district cooling systems based on scenario-classified reinforcement learning," ***Applied Energy***, vol. 377, Part B, No. January, p. 124415, 2025.
8. P. Yu, **H. Zhang**, Y. Song, et. al., "Safe Reinforcement Learning for Power System Control: A Review," ***Renewable and Sustainable Energy Reviews***, vol. 223, p. 116022, 2025.

Thank you!

Hongcai Zhang
University of Macau
hczhang@um.edu.mo