

District Cooling System Control for Providing Operating Reserve based on Safe Deep Reinforcement Learning

Peipei Yu, *Student Member, IEEE*, Hongxun Hui, *Member, IEEE*, Hongcai Zhang, *Member, IEEE*,
Ge Chen, *Student Member, IEEE*, and Yonghua Song, *Fellow, IEEE*

Abstract—Heating, ventilation, and air conditioning (HVAC) systems are well proved to be capable to provide operating reserve for power systems. As a type of large-capacity and energy-efficient HVAC system (up to 100 MW), district cooling system (DCS) is emerging in modern cities and has huge potential to be regulated as a flexible load. However, strategically controlling a DCS to provide flexibility is challenging, because one DCS services multiple buildings with complex thermal dynamics and uncertain cooling demands. Improper control may lead to significant thermal discomfort and even deteriorate the power system's operation security. To address the above issues, we propose a model-free control strategy based on the deep reinforcement learning (DRL) without the requirement of accurate system model and uncertainty distribution. To avoid damaging “trial & error” actions that may violate the system's operation security during the training process, we further propose a safe layer combined to the DRL to guarantee the satisfaction of critical constraints, forming a safe-DRL scheme. Moreover, after providing operating reserve, DCS increases power and tries to recover all the buildings' temperature back to set values, which may probably cause an instantaneous peak-power rebound and bring a secondary impact on power systems. Therefore, we design a self-adaption reward function within the proposed safe-DRL scheme to constrain the peak-power effectively. Numerical studies based on a realistic DCS demonstrate the effectiveness of the proposed methods.

Index Terms—District cooling system, operating reserve, model-free control, safe deep reinforcement learning.

I. INTRODUCTION

A. Background

THE increasing intermittent renewable energy resources bring more uncertainties to the generation-side, and scale up the demands for operating reserve services in modern power systems [1]. Traditionally, the service is majorly provided by thermal or gas generating units, which are carbon-intensive and being phased out [2]. With the development of Internet of Things technologies, active control of demand-side resources has emerged as an alternative solution to provide operating reserve by curtailing or transferring power consumption [3].

This paper is funded in part by the Science and Technology Development Fund, Macau SAR (File no. SKL-IOTSC(UM)-2021-2023, 0137/2019/A3, and 0003/2020/AKP), and in part by the National Natural Science Foundation of China under Grant 52007200. (Corresponding author: *Hongcai Zhang*.)

P. Yu, H. Hui, H. Zhang, G. Chen, and Y. Song are with the State Key Laboratory of Internet of Things for Smart City and Department of Electrical and Computer Engineering, University of Macau, Macao, 999078 China; H. Zhang is also with the Smart City Research Center, Zhuhai UM Science & Technology Research Institute, Zhuhai, 519031 China (email: hc Zhang@um.edu.mo).

The heating, ventilation, and air conditioning (HVAC) system is an ideal resource, because it can shift its power consumption flexibly while assuring the comfortable indoor temperature by utilizing the building's inherent thermal inertia [4]. Besides, HVAC has large regulation capacity as it accounts for over 40% of power consumption in modern cities [5].

Compared with a common household HVAC system, the district cooling system (DCS) is one type of HVAC with larger capacity and higher efficiency, and thus DCS is emerging and being developed in many cities [6]. As shown in Fig. 1, DCS is composed of one energy station and some pipelines to produce chilled water for multiple buildings [7]. Generally, one DCS's capacity can be up to 100 MW, which is more than 10,000 times of a household HVAC [8]. Therefore, DCS has huge regulation potential to provide operating reserve, which, however, is only studied by few published papers. For example, Lo et al. [9] use least squares regression to optimize the day-ahead power dispatch for a large cooling system to perform demand response. Cox et al. [7] and Chen et al. [10] design day-ahead power scheduling strategies for DCS to minimize electricity costs with time-of-use pricing. Tang et al. [11] propose a direct load control strategy for a centralized AC system for requests of smart grids. The chillers are assumed to be operated in the on-off mode. All of the aforementioned studies focus on the day-ahead or hour-ahead control while ignore the real-time uncertainties in cooling demands. Besides, the operation mode of chillers is usually continuous so that assuming it to be on-off mode may not fully utilize DCS's regulation capacity. To fill this research gap, this study focuses on the real-time control of a DCS to provide operating reserve subject to the comfortable temperature constraint in each building. In most electricity markets, the start time for resources (i.e. DCS) to provide operating reserve is uncertain, while the time interval for operating reserve is fixed (e.g., 10 minutes in PJM [12], 15~30 minutes in China [13]). As illustrated by the load curve in Fig. 1, there are two control stages for a DCS to provide operating reserve:

- 1) *In the power reduction stage*, the controller cuts down the DCS operating power following the instruction from the power system operator. In the meantime, it also tries to fulfill the temperature requests of heterogeneous buildings, when the cooling supply from DCS gets decreased as a result of power reduction.
- 2) *In the power recovery stage*, the DCS stops providing reserve and begins to restore all the buildings' indoor

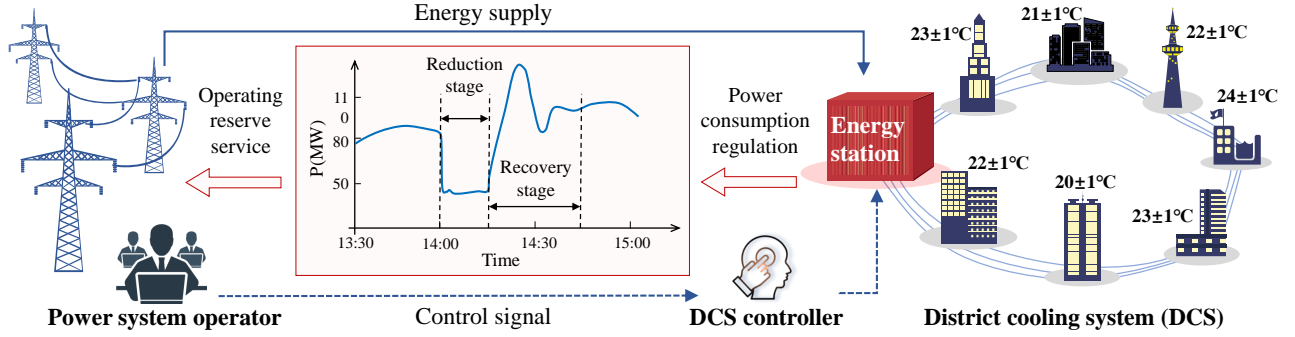


Fig. 1: The supply methods and scale of the DCS.

temperature back to set values by increasing its cooling supply. During this stage, the DCS shall recover its power consumption smoothly to avoid the peak-power rebound that may cause a secondary impact on the power system, which has just returned to the stable state.

The above two-stage control of DCS is quite challenging because of both the system complexity and cooling demand uncertainty, detailed as follows:

Complexity: The power consumption of a DCS is usually adjusted automatically according to its operation state, so that we cannot control the power directly to provide operating reserve like other demand-side resources [14]. Instead, the mass flow in pipelines is regulated to adjust the DCS's operating power indirectly [15], which can achieve faster-response effects than regulating the set temperature [16]. However, describing the relationship between the mass flow and operating power needs an accurate thermal dynamic model. This is challenging because the thermal dynamics of a DCS, including cooling power generation, transportation and consumption, is usually quite complex [6]. Furthermore, a DCS is a large networked system whose dynamic parameters are hard and expensive to measure. Obtaining an accurate model that is completely in sync with the real system is nontrivial in practice [17]. Therefore, traditional model-based control methods for HVAC systems are hard to be used in DCS [18].

Uncertainty: The DCS's power consumption and buildings' indoor temperatures are related to the ambient temperature and indoor human behaviors [19]. Higher ambient temperature and more indoor human activities call for more cooling supply and higher power consumption. However, the indoor human behaviors are stochastic and hard to accurately predict [5]. The ambient temperature can bring different influence to heterogeneous buildings. As a result, it is nontrivial to control a DCS subject to heterogeneous indoor temperature constraints in multiple buildings, especially when the DCS power consumption is cut down to provide operating reserve.

B. Literature Reviews

In recent years, some researchers adopt model predictive control (MPC) method for regulating DCS to achieve the cost reduction in energy systems [20]. However, MPC usually requires a reliable dynamic model of the system, which is often unavailable in practice [21]. Furthermore, when the system model is too complex, MPC can be quite computationally

expensive and may fail to work in real-time control scenarios. Mixed-integer linear programming (MILP) has also been used to the complex system scheduling problems [22], [23]. Unfortunately, at each step, optimization methods need to recalculate from the beginning, resulting in too large calculation cost to be applied in real-time control. Moreover, the execution time of MILP increases exponentially according to the problem dimensions and cannot solve complex issues. Another commonly used control method is heuristic algorithm, including genetic algorithm (GA), particle swarm optimization (PSO), ants colony optimization (ACO), etc. Stoppato et al. [24] combine heuristic algorithms to obtain system's optimal operation, while the convergence of heuristic algorithms cannot be proved mathematically and is less robust.

Compared with the aforementioned control methods, deep reinforcement learning (DRL) has become increasingly popular to handle model-free and high dimensional decision-making problems [25]. DRL has been proved to be more robust with stable convergence results to effectively handle uncertainties of systems through the prediction in neural networks. Some researchers have adopted DRL to control traditional HVACs. For instance, Du et al. [26] use DRL to control residential HVACs as to respond dynamic electricity prices. Xu et al. [27] adopt DRL to schedule home energy consumption considering uncertain PV generation. Liang et al. [28] present a DRL-based control strategy to minimize both the HVAC's energy consumption and user's thermal discomfort. Ruelens et al. [29] propose a DRL-based direct control method of HVAC to provide ancillary services. Zhang et al. [30] utilize DRL for cost-effective control of a HVAC in commercial buildings. However, to the best of our knowledge, published papers have not studied DRL-based control for DCS to provide operating reserve for power systems.

Generally, a DRL-based controller has to be trained through lots of "trail-and-errors" before being intelligent [31]. It means some "bad" decisions may be made during the training process, part of which may cause constraint violations. However, in power systems, some critical constraint violations can cause damaging results [32]. For example, if a DCS fails to provide sufficient operating reserve as it promised to the system operator, the power system may face the stability problem. To address this challenge from the constraint safety in power systems, the safe-DRL framework is needed to ensure the satisfaction of critical constraints during the training process. Considering that safe-DRL is an emerging concept in the

application in power systems, there are little research to combine HVAC or DCS with safe-DRL. Some published studies propose safe-DRL frameworks for voltage control problems [33]–[36], emergency load-shedding control problems [37]–[39] and demand-side resource scheduling problems (e.g., EV, building's equipment) [40]–[42]. However, DCS is different from EVs and household HVACs due to its complex thermal dynamic process and uncertainties. Thus, the above methods can not be adopted directly in this paper.

C. Contributions

In this paper, we propose a safe-DRL control strategy for DCS to provide operating reserve while satisfying major critical constraints. This paper advances the relevant published literature in the following aspects:

- 1) The DCS control problem is developed as a Markov Decision Process (MDP) mathematically to provide operating reserve. The designed reward function of DRL aims to achieve minimum impacts on buildings' thermal comfort when providing the required operating reserve. Besides, the proposed iteration algorithm does not need the accurate system model of DCS nor the distribution of uncertainties, which can address challenges from both the system complexity and uncertainty.
- 2) A novel safe-DRL framework is proposed for constraint assurance, where a safe layer is designed on the top of traditional DRL algorithm. The proposed safe layer avoids critical power constraint violations to protect power systems from undesirable "trial-and-errors", through fine tuning possible unsafe control signals into safe ones during the training process.
- 3) A self-adaption target method is proposed and designed as the reward function in the safe-DRL framework during the power recovery stage. The proposed method can effectively achieve the smooth power recovery and avoid peak-power rebound that probably brings secondary impacts to power systems.

Besides, numerical studies verify the effectiveness of our proposed strategy, based on a real-world DCS. The analysis shows DCSs are qualified to provide operating reserve with mild impacts on buildings' indoor thermal comforts, subject to the critical power constraints.

The rest is organized as follows. Section II introduces the physical architecture and control logic of DCS. Section III proposes the safe-DRL framework. Numerical studies are carried out in Section IV. Section V concludes this paper.

II. MODELLING OF THE DCS

This section establishes the DCS model as the simulated environment to interact with the proposed DRL agent. Noted that the only information received by the agent is the feedback from the established environment, while not the details about the accurate DCS model.

A. DCS Framework

The schematic diagram of a DCS is shown in Fig. 2, in which blue lines represent the chilled water (or cooling wind)

to supply thermal energies for buildings; red lines are the returned warm water (or warm wind). Its heat transmission process includes three isolated loops:

In the first water loop, chillers produce chilled water with a set temperature $T^{\text{ch},s}$, which is transported through pipelines to distributed buildings to supply cooling demands. The total mass flow m_t^{ch} is separated to different buildings by their independent two-port valves, which determine each building's own mass flow rate $m_{i,t}^{\text{I}}$ [43]. After the heat exchange process, the chilled water in pipelines becomes warm with temperature $T_t^{\text{ch},r}$ and then is pumped back to chillers. The decoupler between the supply and return water balances pressure when the mass flow rate changes.

In the second water loop (i.e., water cycle in buildings), the water temperature $T_{i,t}^{\text{II},s}$ in buildings is cooled down by the chilled water in the first water loop through heat exchangers. Then the cool water transfers its thermal energy to the air in Air Handle Units (AHUs) to form cooling winds. The temperature of return water $T_{i,t}^{\text{II},r}$ reflects fluctuating cooling demands in buildings and further influences chillers' power consumption automatically.

In the air loop, AHUs blow cooling winds with the temperature $T_{i,t}^{\text{w}}$ into each room, which can further influence the indoor temperature $T_{i,t}^{\text{A}}$ and refresh the indoor air.

Note that the aforementioned three loops are mutually independent while interactional. Specifically, the total power consumption of a DCS majorly comes from chillers in the first water loop, whose operations are automatically and indirectly adjusted based on the buildings' cooling demands in the third loop. Therefore, it is significant for the DCS control to find the relationship between these loops.

B. Modelling of Key Components

1) *Chillers*: Chillers consume most electricity in DCS. Their power consumption can be calculated based on the energy and mass balance, as follows:

$$P_t^{\text{ch}} = \frac{Q_t^{\text{ch}}}{\text{COP}}, \quad \forall t, \quad (1)$$

where P_t^{ch} is chillers' electrical power at time t , in kW; Q_t^{ch} is the cooling power, in kW; COP denotes chiller's coefficient of performance. Generally, Q_t^{ch} is determined by chillers' varying return water temperature $T_t^{\text{ch},r}$, in $^{\circ}\text{C}$, and instantaneous mass flow rate m_t^{ch} , in kg/s, as follows:

$$Q_t^{\text{ch}} = m_t^{\text{ch}} c^{\text{w}} (T_t^{\text{ch},r} - T^{\text{ch},s}), \quad \forall t, \quad (2)$$

where c^{w} is the specific heat capacity of water, in $\text{kJ}/(\text{kg}\cdot^{\circ}\text{C})$. The set temperature of supply chilled water is represented by $T^{\text{ch},s}$, which is usually a designed constant [44]. Therefore, controlling the mass flow rate m_t^{ch} can influence the electrical power P_t^{ch} effectively. Further, we can rewrite $T_t^{\text{ch},r}$ and m_t^{ch} according to the mass balance as:

$$m_t^{\text{ch}} = m_t^{\text{dec}} + \sum_{i \in \mathcal{I}} m_{i,t}^{\text{I}}, \quad \forall t, \quad (3)$$

$$T_t^{\text{ch},r} = \frac{m_t^{\text{dec}} c^{\text{w}} T_t^{\text{dec}} + \sum_{i \in \mathcal{I}} m_{i,t}^{\text{I}} c^{\text{w}} T_{i,t}^{\text{II},r}}{m_t^{\text{ch}} c^{\text{w}}}, \quad \forall t, \quad (4)$$

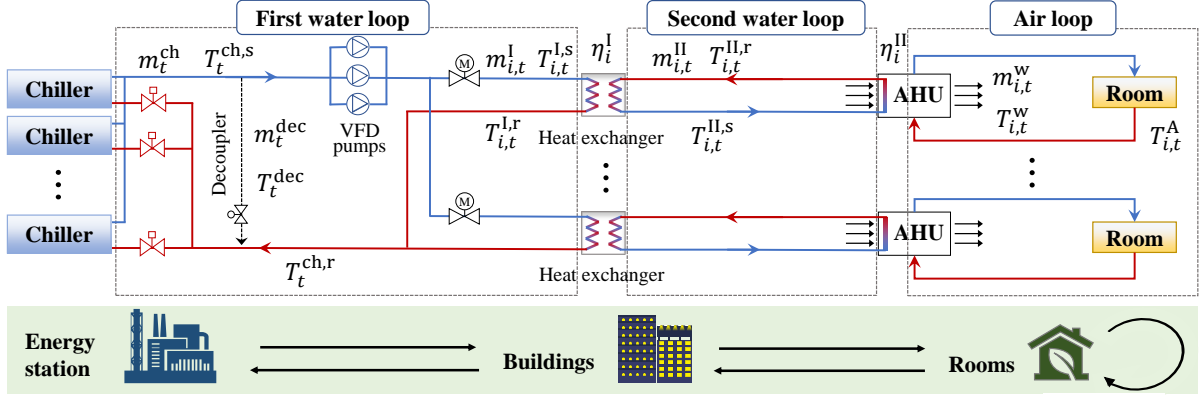


Fig. 2: Schematic diagram of a DCS.

where set \mathcal{I} denotes the set of terminal buildings; $m_{i,t}^I$ and $T_{i,t}^{I,r}$ are each buildings' mass flow rate and return water temperature in first water loop, respectively; m_t^{dec} and T_t^{dec} are the mass flow rate and return water temperature of the decoupler, respectively. Eqs. (3)-(4) show the mass flow and energy balances between chillers and buildings.

2) *Heat exchangers*: Heat exchangers transfer cooling supply from the first water loop to the second water loop. Considering the heat loss in pipelines, each building's actual supply water temperature can be calculated by [45]:

$$T_{i,t}^{I,s} = T_t^{\text{out}} + \eta_i^{\text{pipe}} (T_t^{\text{ch,s}} - T_t^{\text{out}}), \quad \forall i \in \mathcal{I}, \forall t, \quad (5)$$

where η_i^{pipe} is the heat transfer coefficient of supply pipelines; T_t^{out} is the ambient temperature; $T_{i,t}^{I,s}$ is the supply chilled water temperature for building i . Further, the corresponding exchanging heat in building i , $Q_{i,t}^{\text{HE}}$ in kW, can be given by:

$$\begin{aligned} Q_{i,t}^{\text{HE}} &= m_{i,t}^{\text{II}} c^w (T_{i,t}^{\text{II,r}} - T_{i,t}^{\text{II,s}}) \\ &= \eta_i^{\text{I}} m_{i,t}^{\text{I}} c^w (T_{i,t}^{\text{I,r}} - T_{i,t}^{\text{I,s}}), \quad \forall i \in \mathcal{I}, \forall t, \end{aligned} \quad (6)$$

where η_i^{I} indicates the transfer efficiency of first water loop to second water loop; $m_{i,t}^{\text{I}}$ and $m_{i,t}^{\text{II}}$ are the mass flow rate of two sides, respectively. Similarly, $T_{i,t}^{\text{I,r}}$, $T_{i,t}^{\text{I,s}}$ and $T_{i,t}^{\text{II,r}}$, $T_{i,t}^{\text{II,s}}$ are the return and supply water temperature of each side, respectively. In addition, $Q_{i,t}^{\text{HE}}$ is determined by the performance of the heat exchanger, which can be calculated by [46]:

$$\frac{Q_{i,t}^{\text{HE}}}{k_i^{\text{HE}}} = \int_0^{F_i^{\text{HE}}} \Delta T_{i,t} dF_i \approx F_i^{\text{HE}} \Delta T_{i,t}^{\text{mean}}, \quad \forall i \in \mathcal{I}, \forall t, \quad (7)$$

where k_i^{HE} is heat exchangers' transfer coefficient, in $\text{kW}/(\text{m}^2 \cdot ^\circ\text{C})$; F_i^{HE} is the surface area, in m^2 ; Symbol $\Delta T_{i,t}^{\text{mean}}$ is defined as the mean difference between the water's temperature of two sides, which is a function formulated as:

$$\Delta T_{i,t}^{\text{mean}} = \frac{(T_{i,t}^{\text{II,r}} - T_{i,t}^{\text{I,s}}) - (T_{i,t}^{\text{II,s}} - T_{i,t}^{\text{I,r}})}{\ln((T_{i,t}^{\text{II,r}} - T_{i,t}^{\text{I,s}})/(T_{i,t}^{\text{II,s}} - T_{i,t}^{\text{I,r}}))}, \quad \forall i \in \mathcal{I}, \forall t. \quad (8)$$

The above Eqs. (6)-(8) determine the dynamic exchanging heat in each building i between the first and second water loops.

3) *Buildings*: AHU transfers the heat from the second water loop to the third air loop by blowing cooling wind whose energy balance is give as:

$$m_{i,t}^w c^A (T_{i,t}^A - T_{i,t}^w) = \eta_i^{\text{II}} m_{i,t}^{\text{II}} c^w (T_{i,t}^{\text{II,r}} - T_{i,t}^{\text{II,s}}), \quad \forall i \in \mathcal{I}, \forall t, \quad (9)$$

$$T_{i,t}^w = \frac{1}{2} (1 - \alpha_i) (T_{i,t}^{\text{II,s}} + T_{i,t}^{\text{II,r}}) + \alpha_i T_t^{\text{out}}, \quad \forall i \in \mathcal{I}, \forall t, \quad (10)$$

where c^A , $m_{i,t}^w$ are air's specific heat capacity and wind's mass flow rate; η_i^{II} is the exchanging heat efficiency of second water loop to AHU; $T_{i,t}^A$, T_t^{out} are the indoor and ambient temperature, respectively; $T_{i,t}^w$ represents the temperature of the cooling air out from AHU, mixing the outdoor fresh air with proportion α_i . Then the indoor thermal dynamic is described as [47]:

$$c^A \rho^A V_i \frac{dT_{i,t}^A}{dt} = Q_{i,t}^{\text{loss}} - Q_{i,t}^{\text{DCS}}, \quad \forall i \in \mathcal{I}, \forall t, \quad (11)$$

where ρ^A is the density of the air, in kg/m^3 ; V_i is the space volume of the i th building, in m^3 ; $Q_{i,t}^{\text{loss}}$ is the i th building's heat loss because of its heat exchange with the ambient environment and $Q_{i,t}^{\text{DCS}}$ is its cooling gain from DCS, which are given as:

$$Q_{i,t}^{\text{DCS}} = m_{i,t}^w c^A (T_{i,t}^A - T_{i,t}^w), \quad \forall i \in \mathcal{I}, \forall t, \quad (12)$$

$$Q_{i,t}^{\text{loss}} = U_i^{\text{O-A}} A_i^S (T_t^{\text{out}} - T_{i,t}^A) + \zeta_{i,t}, \quad \forall i \in \mathcal{I}, \forall t, \quad (13)$$

where $U_i^{\text{O-A}}$ is the heat transfer coefficient, in $\text{kW}/(\text{m}^2 \cdot ^\circ\text{C})$; A_i^S is the surface area of the i th building, in m^2 ; $\zeta_{i,t}$ is the heat load from indoor sources (e.g., stochastic human behaviors and electric equipment), in kW. Eqs. (11)-(13) give the temperature dynamic in buildings. If $\Delta T_{i,t}^A$ is used to represent the temperature fluctuation in a stable operating state (T_{i,t_0}^A , m_{i,t_0}^w , T_{i,t_0}^w), then the temperature dynamic's first-order Taylor series expression can be expressed as:

$$\begin{aligned} c^A \rho^A V_i \frac{d\Delta T_{i,t}^A}{dt} &= -(m_{i,t_0}^w c^A + U_i^{\text{O-A}} A_i^S) \Delta T_{i,t}^A + m_{i,t_0}^w c^A \Delta T_{i,t}^w \\ &\quad + c^A (T_{i,t_0}^w - T_{i,t_0}^A) \Delta m_{i,t}^w + U_i^{\text{O-A}} A_i^S \Delta T_t^{\text{out}} \\ &\quad + \Delta \zeta_{i,t}, \quad \forall i \in \mathcal{I}, \forall t, \end{aligned} \quad (14)$$

where $\Delta T_{i,t}^w$, $\Delta m_{i,t}^w$, ΔT_t^{out} and $\Delta \zeta_{i,t}$ are the changes in the corresponding four variables. Then, the Laplace transformation of Eq. (14) can be obtained as following:

$$\begin{aligned} &(K_1 s + K_2 + K_4) T_i^A(s) \\ &= K_2 T_i^w(s) + K_3 m_i^w(s) + K_4 T^{\text{out}}(s) + \zeta_i(s), \quad \forall i \in \mathcal{I}, \end{aligned} \quad (15)$$

where $K_1 = c^A \rho^A V_i$, $K_2 = m_{i,t_0}^w c^A$, $K_3 = c^A (T_{i,t_0}^w - T_{i,t_0}^A)$ and $K_4 = U_i^{\text{O-A}} A_i^S$. Therefore, the temperature dynamic in each building is an inertial process with inertia time constant

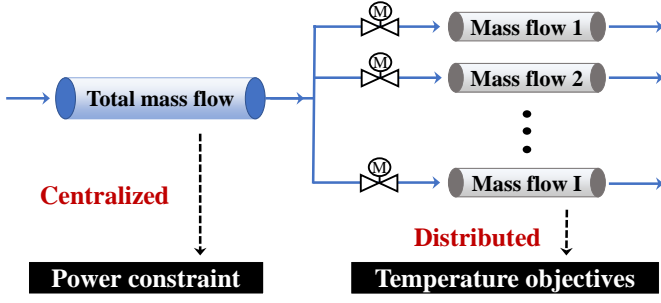


Fig. 3: Objectives and constraints during DCS control process.

$K_1/(K_2+K_4)$, which is mainly determined by each building's inherent characteristics.

The above models from Eq. (1) to Eq. (15) describe the whole thermal dynamics in a DCS. In summary, a DCS provides cooling supply to multiple buildings through two water loops and one air loop to transmit thermal energies.

Remark 1. The chillers' cooling power is not only determined by the mass flow rate m_i^{ch} but also the uncertain return water temperature $T_i^{ch,r}$. The later is further influenced by stochastic ambient temperature T^{out} and heat load ζ_i of buildings in Eq. (13). Besides, the accurate thermal model parameters in three loops are unknown and difficult to obtain in practice, which makes the conventional model-based control strategy infeasible for a DCS. To deal with these challenges, a model-free DRL method is proposed in the following Section III.

III. DCS CONTROL BASED ON SAFE DEEP REINFORCEMENT LEARNING

As shown in Fig. 3, when the DCS receives signals to provide operating reserve, its total mass flow, i.e., the sum of all the buildings' mass flows, should be regulated to satisfy the power constraint. Given the total mass flow, the second problem is to allocate the mass flow among different buildings. An ideal control strategy shall properly allocate these mass flows to make heterogeneous buildings have similar temperature deviations, so that all the buildings' comforts can be guaranteed to the most extent and their thermal inertia can be fully utilized. Therefore, the power requirement from the power system and the temperature comfort requirements from buildings should be both considered in the DCS control problem when providing operating reserve.

A. Formulation of the DCS Control Problem

The DCS control problem is a typical sequential decision-making problem that can be described as a MDP [48], which aims to minimize temperature impacts on buildings by controlling mass flows continuously. Considering there is a critical power constraint during the control process, the studied problem needs to be formulated as a Constrained Markov Decision Process (CMDP) [49].

In the CMDP framework, a centralized smart controller, called *agent*, is designed to send each building signals to control its mass flow rate $m_{i,t}^l$. When a DCS provides operating reserve during the period $\mathcal{T} = [t_0, t_1]$, the DCS is

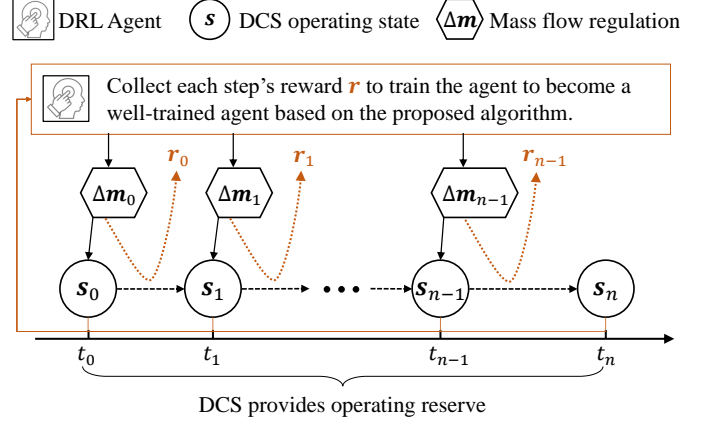


Fig. 4: Interactions between the agent and the DCS.

regarded as an *environment* whose real-time operation state s_t at time slot $t \in \mathcal{T}$ is observed by the agent. Then according to the information in s_t , the agent makes one decision for DCS to execute action a_t , which means there is a complete trajectory $\tau = \{s_{t_0}, a_{t_0+1}, s_{t_0+1}, \dots, a_{t_1}, s_{t_1}\}$ to describe the control process. Here, both the state s_t and action a_t are multi-dimensional vectors rather than a scalar. The probability from the state s_t to s_{t+1} after taking action a_t is defined by a transition function $P(s_{t+1}|s_t, a_t)$, which is not necessary (assumed unknown) in model-free methods.

In DCS control process, its power consumption and buildings' indoor temperature are main considerations. The temperature deviation $\Delta T_{i,t} = T_{i,t}^A - T_{i,t}^{\text{set}}, \forall i \in \mathcal{I}, t \in \mathcal{T}$, is defined as the temperature comfort indicator, in which $T_{i,t}^{\text{set}}$ is the set temperature. Thus, the state and the action are defined by:

$$s_t = [\Delta P_t, m_{i,t}^l, T_{i,t}^{\text{r}}, \Delta T_{i,t} | i \in \mathcal{I}]^T \in \mathcal{S}, \quad \forall t \in \mathcal{T}, \quad (16)$$

$$a_t = [\Delta m_{1,t}^l, \Delta m_{2,t}^l, \dots, \Delta m_{|\mathcal{I}|,t}^l]^T \in \mathcal{A}, \quad \forall t \in \mathcal{T}, \quad (17)$$

where ΔP_t equals to the gap between the actual power P_t^{ch} and required power cap P^{cap} of power systems. The scale of the state space \mathcal{S} and action space \mathcal{A} are $|\mathcal{S}| = 3|\mathcal{I}| + 1$ and $|\mathcal{A}| = |\mathcal{I}|$, respectively. As shown in Fig. 4, the DRL agent gives its control signal Δm at each time slot t during operating reserve. According to the past experiences, the agent updates its policy with the proposed algorithm to become more intelligent. Because the mass flow can be regulated by valves continuously, the action space is a continuous space and $\Delta m_{i,t}^l$ is a continuous variable. The positive (or negative) $\Delta m_{i,t}^l$ means to increase (or decrease) the mass flow rate, in which there are upper and lower limits $\underline{m}_i^l, \bar{m}_i^l$ in a real DCS. Thus, the maximum value of action is constrained by $|\Delta m_{i,t}| \leq \bar{m}_i^l$.

An arbitrary mapping from the state space to the action space $\pi : \mathcal{S} \rightarrow \mathcal{A}$ is called a *policy*. Essentially, the agent's task is to find an *optimal policy* that will be used as a guide for future online controlling. In order to evaluate a policy's performance, r_{t+1} is defined as *reward* for the action a_t in one step, which is formulated as:

$$r_{t+1} = -\theta^r \mathbb{E}_{i \in \mathcal{I}}[|\Delta T_{i,t+1}|] - \sigma_{i \in \mathcal{I}}^2[\Delta T_{i,t+1}], \quad \forall t \in \mathcal{T}. \quad (18)$$

Eq. (18) includes two parts: the average and variance of all the buildings' temperature deviation at next time $t+1$. The

former item $\mathbb{E}_{i \in \mathcal{I}}[\Delta T_{i,t+1}]$ is the average temperature of all the buildings' indoor temperature deviations from their corresponding set values. A smaller average value means a less temperature influences to buildings. The later item $\sigma_{i \in \mathcal{I}}^2[\Delta T_{i,t+1}]$ is the variance of all the buildings' temperature deviations, where a smaller variance means less difference of the influences among different buildings. Parameter θ^r is the weight factor to determine the importance of the two parts.

Further, compared with the immediate reward r_t , the *return* G_t is defined as the accumulated reward in the future, which considers not only the immediate reward but also the expected influence to future rewards caused by the current action. The total discounted reward at time slot t is expressed as:

$$G_t = r_{t+1} + \gamma r_{t+2} + \dots = \sum_{\tau=0}^{t-1} \gamma^\tau r_{t+\tau+1}, \quad \forall t \in \mathcal{T}, \quad (19)$$

where $\gamma \in [0, 1]$ is a discount factor to represent the weight of the influence to future rewards [50]. For instance, when $\gamma = 1$, the agent considers the immediate and future rewards with the same importance. By contrary, when $\gamma = 0$, the agent only considers the current reward and $G_t = r_t$. Then, an *action-value function* $Q^\pi(s_t, a_t)$ is defined as the expected return from state s_t , taking action a_t and following policy π :

$$Q^\pi(s_t, a_t) = \mathbb{E}_\pi[G_t | s_t, a_t], \quad \forall t \in \mathcal{T}, \quad (20)$$

where the *optimal action-value function* $Q^*(s_t, a_t)$ means the maximum action-value over all policies $\max_\pi Q^\pi(s_t, a_t)$. According to the theorem in MDP [31], optimal policy π^* is defined to satisfy $Q^{\pi^*}(s_t, a_t) = Q^*(s_t, a_t), \forall t$. Therefore, the agent's objective is to maximize the expected return J^π :

$$\max_{\pi} J^\pi = \mathbb{E}_{s_t \sim \mathcal{S}, a_t \sim \pi} [G_t] = \mathbb{E}_{s_t \sim \mathcal{S}} [Q^\pi(s_t, \pi(s_t))]. \quad (21)$$

However, different with the conventional policy optimization problem, there is a critical power constraint for a DCS during the power reduction stage and formulated as:

$$P_t^{\text{ch}} \leq P^{\text{cap}}, \quad \forall t \in \mathcal{T}, \quad (22)$$

where P^{cap} is the required power cap from the power system operator to constrain DCS's operating power¹. If it is violated, the DCS may be heavily penalized by the power system operator. Thus, Eq. (22) turns the DCS control problem from a traditional MDP into a CMDP.

B. Policy Gradient Algorithm

To solve the optimal policy π^* in Eq. (21), a safe-DRL algorithm is proposed as shown in Fig. 5, which combines the actor-critic framework and deep Q-learning. Two neural networks are adopted to represent the action-value function Q and policy π , with parameters θ^Q, θ^π , respectively. The network to approximate Q value is called *critic network*, and another one that outputs actions is called *actor network*. In Fig. 5, the agent firstly interacts with the DCS environment to obtain transitions (s_t, a_t, r_t, s_{t+1}) , and collects all transitions into an *experience reply buffer* R . Secondly, the agent randomly sample a mini-batch data from R to update two networks. Finally, the DCS

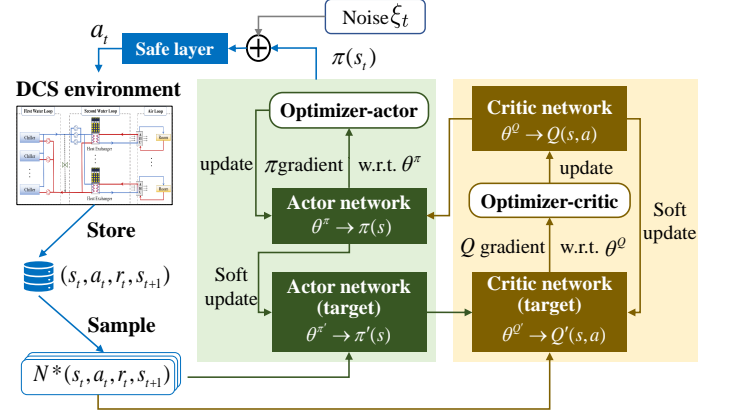


Fig. 5: Scheme of the safe-DRL algorithm.

receives an action that is produced by the actor network $\pi(s_t)$ and further fine tuned by the safe layer. The convergence of the DRL algorithm has been proved mathematically [52].

Using the experience reply buffer R , randomly sampled data keeps weak correlationship with each other, which effectively avoids over-fitting of the two networks. The update rule for actor network θ^π is given in its gradient direction as [31]:

$$\nabla_{\theta^\pi} J^\pi = \mathbb{E}[\nabla_a Q(s, \pi(s)) \nabla_{\theta^\pi} \pi(s)], \quad (23)$$

where the gradient of Q needs to be estimated through the critic network. Moreover, because the sampled transitions from R are all guided by policy π , the Monte-Carlo approach is adopted to give an un-biased estimate of Eq. (23) as:

$$\nabla_{\theta^\pi} J^\pi \approx \frac{1}{K} \sum_{k=1}^K \nabla_a Q(s_k, \pi(s_k)) \nabla_{\theta^\pi} \pi(s_k), \quad (24)$$

where k is the index of samples; K is the size of the sampled mini-batch data set \mathcal{K} . For the critic network θ^Q , the mean squared error (MSE) is used as the loss function:

$$L = \frac{1}{K} \sum_{k=1}^K [y_k - Q(s_k, a_k)]^2, \quad (25)$$

where y_k is the target value of $Q(s_k, a_k)$ and needs to be estimated. To stabilize the training process and guarantee the convergence, the target y_k should not change frequently. According to the Bellman Expectation Equation of Eq. (20), two target networks (Q', π') , copies of ordinary networks (Q, π) , are designed to calculate y_k as:

$$Q^\pi(s_t, a_t) = \mathbb{E}[r_t + \gamma Q^\pi(s_{t+1}, a_{t+1})], \quad \forall t \in \mathcal{T}, \quad (26)$$

$$y_k = r_k + Q'(s_{k+1}, \pi'(s_{k+1})), \quad \forall k \in \mathcal{K}. \quad (27)$$

To be more stabilized, the target networks are updated following the running average method, which are given by:

$$\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'}, \quad (28)$$

$$\theta^{\pi'} \leftarrow \tau \theta^\pi + (1 - \tau) \theta^{\pi'}, \quad (29)$$

where τ is the smooth factor, $0 \leq \tau \ll 1$. Finally, to improve the efficiency of the exploration, an independent noise ξ_t is added to each action subject to the Gaussian distribution $\xi \sim N(0, \sigma^2)$. The proposed algorithm is summarized in Table I, where the safe layer showed as row 06 will be described in detail in the next subsection.

¹Power cap is determined by the regulation capacity that the DCS offered to the electrical market ahead of one day [51].

TABLE I: Safe-DRL algorithm

01	Initialize the random process ξ , the experience reply buffer R and the actor, critic networks $Q(s, a), \pi(s)$ with weights θ^Q, θ^π , respectively. Initialize corresponding two target networks Q', π' with weights $\theta^{Q'} \leftarrow \theta^Q, \theta^{\pi'} \leftarrow \theta^\pi$.
02	For episode = 1 : 1 : M do
03	Receive initial observation state s_{t_0} .
04	For $t = 1 : 1 : T$ do
05	Select DCSs' control action $a_{t_0+t} = \pi(s_{t_0+t}) + \xi_{t_0+t}$.
06	fine tune a_{t_0+t} by the safe layer.
07	Execute the action a_{t_0+t} , then obtain the reward r_{t_0+t} and the next state s_{t_0+t+1} .
08	Collect the transition $(s_{t_0+t}, a_{t_0+t}, r_{t_0+t}, s_{t_0+t+1})$ to R , and randomly sample a mini-batch data from R .
09	Update the actor and critic networks by (27) and (25).
10	Update the two target networks by (28)-(29).
11	Endfor
12	Endfor

C. Constrained Policy by the Safe Layer

As shown in step 06 of Table I, the action a_t needs to be fine tuned by the safe layer before being executed, which aims to guarantee the critical constraint of the operating power in Eq. (22). The proposed safe layer achieves the required power cap to assure the high-quality performance in the operating reserve, which effectively makes the DRL agent's control results more reliable in practice.

In each time step $t \in \mathcal{T}$, according to the output action $a_t = \Delta m_t^I$, the next mass flow rate of buildings m_{t+1}^I can be obtained as:

$$m_{t+1}^I = m_t^I + \Delta m_t^I, \quad \forall t \in \mathcal{T}. \quad (30)$$

Thus, based on energy balance Eqs. (1)-(2), the power consumption at the next state is calculated as:

$$P_{t+1}^{\text{ch}} = \sum_{i \in \mathcal{I}} m_{i,t+1}^I \Theta_t, \quad \forall t \in \mathcal{T}, \quad (31)$$

where $\Theta_t = \frac{1}{\text{COP}} [c^w (T_t^{\text{ch},r} - T^{\text{ch},s})]$ is the known parameter related with the return water temperature. Then, if the power consumption satisfies the constraint $P_{t+1}^{\text{ch}} \leq P^{\text{cap}}$, the action Δm_t^I will be executed directly; Otherwise, the action Δm_t^I should be optimized to decrease the operating power.

To address this issue, we propose the following linear mapping rule to adjust Δm_t^I as:

$$\Delta \tilde{m}_t^I = \Delta m_t^I + \mu_t \Delta m_t^I + v_t m_t^I, \quad \forall t \in \mathcal{T}, \quad (32)$$

where μ_t and v_t are the correction coefficients for adjusting the action Δm_t^I , and $\mu_t, v_t \leq 0$; $\Delta \tilde{m}_t^I$ is the updated action from the original agent's output and will finally be executed in DCS. When μ_t and v_t are close to 0, the last two correction terms in Eq. (32) will take small function, i.e., the original agent's action Δm_t^I will not be adjusted too much by the safe layer. By contrast, when μ_t and v_t are negative and far from 0, the original agent's action Δm_t^I will be adjusted significantly. In other words, the safe layer is not only a simple saturation function, but also needs to train the agent to converge. If the decision from the agent Δm_t^I is changed quite a lot by the safe layer, which probably decreases the agent's training efficiency and even leads to the failure of its convergence. Therefore, the coefficients μ_t and v_t are expected to be large and close to 0.

TABLE II: Safe layer method

01	Obtain the next mass flow rate m_{t+1}^I and operating power P_{t+1}^{ch} by (30), (31).
02	If $P_{t+1}^{\text{ch}} \leq P^{\text{cap}}$ then : execute Δm_t^I directly;
03	Else
04	Solve the optimal coefficients μ_t and v_t by (33)-(36);
05	Optimize the next mass flow rate Δm_t^I using (32);
06	Execute the fine tuned mass flow rate $\Delta \tilde{m}_t^I$.
06	End

On this basis, the two coefficients μ_t and v_t can be optimized by following linear programming:

$$\max_{\mu_t, v_t} \mu_t + v_t, \quad (33)$$

$$\text{s.t.: } \sum_{i \in \mathcal{I}} (\mu_t \Delta m_{i,t}^I + v_t m_{i,t}^I) \Theta_t \leq P^{\text{cap}}, \quad \forall t \in \mathcal{T}, \quad (34)$$

$$\underline{m}_i^I \leq \mu_t \Delta m_{i,t}^I + v_t m_{i,t}^I \leq \bar{m}_i^I, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (35)$$

$$\mu_t, v_t, \leq 0, \quad \forall t \in \mathcal{T}, \quad (36)$$

where the objective in Eq. (33) represents the minimum changes on the original agent's action Δm_t^I . The constraint in Eq. (34) is to satisfy the required power cap from power systems. Inequalities (35)-(36) define the domain of parameters $\mu_t, v_t \leq 0$ and $\underline{m}_i^I \leq m_{i,t+1}^I \leq \bar{m}_i^I$. The calculation process of the safe layer is illustrated in Table II to achieve the fine tuning of the "unsafe" action.

Remark 2. The mass flow rate is fine tuned by a mapping rule in Eq. (32) to satisfy the power constraint, which maintains all the buildings' relative relation. In this way, the influence to the agent's training process is linear and feasible to learn, which guarantees the convergence of the agent's policy iteration.

D. Self-adaptive Target Method

After providing operating reserve, DCS stops following power systems' regulation signals and enters the power recovery stage. Thus the power cap constraint in Eq. (22) is relaxed and the DCS tends to recover buildings' comfort temperature as soon as possible. However, a too rapid recovery of the temperature may cause an instantaneous increase in the power consumption, called "power rebound". It may lead to a new power peak and cause stability problems for power systems. In some extreme cases, the large load current brought by the power rebound could cause the melting of overhead lines, which damages power system security considerably [53]. To avoid the "unsafe" power rebound, we further propose a self-adaption target method combined to the proposed safe-DRL scheme to achieve a smooth recovery, as follows:

$$r_t = -\mathbb{E}_{i \sim \mathcal{I}} [|\Delta T_{i,t+1} - \varphi_{i,t+1}|], \quad \forall t \in [t_1, t_2], \quad (37)$$

where r_t is the reward of the indoor temperature in the recovery stage; $\varphi_{i,t}$ is the self-adaptive factor; t_1 is the end time of the power reduction stage and also the beginning time of the power recovery stage; t_2 is the required time for recovering the indoor temperature to the set value.

The reward r_t in Eq. (37) is different from the definition during the reduction stage in Eq. (18). Because Eq. (37) considers not only the buildings' set values, but also the self-adaptive factor $\varphi_{i,t}$ to design a expected temperature-

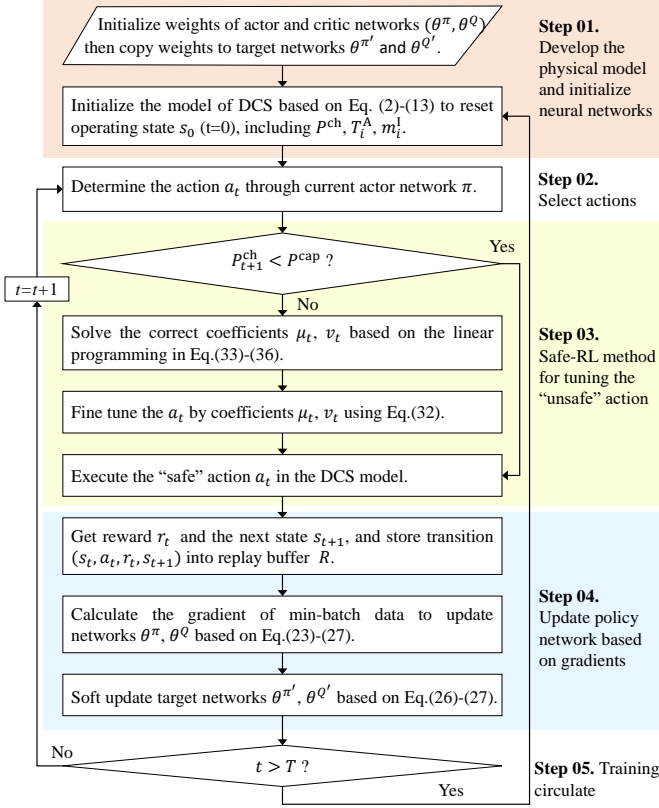


Fig. 6: Step-by-step calculation process.

decreasing trend. In this way, the sharp increase of the DCS's operating power can be alleviated. We propose the following configuration method for the self-adaptive factor²:

$$\varphi_{i,t} = \frac{\Delta T_{i,t_1}}{1 + e^{\lambda[\frac{t-t_1}{t_2-t_1} - \frac{1}{2}]}}, \quad \forall i \in \mathcal{I}, \forall t \in [t_1, t_2], \quad (38)$$

where λ is determined according to the required recovery extent of the indoor temperature at time t_2 . For example, when λ is set as 6, the recovery extent of the indoor temperature can reach 95% of $\Delta T_{i,t_1}$ at time t_2 . Therefore, we can set the values of λ and t_2 to obtain the self-adaptive factor $\varphi_{i,t}$.

Moreover, in order to constrain the increased operating power during the recovery stage strictly, we also design a safe layer for the agent, similar with that during the reduction stage in Eq. (27)-(33). The difference is that the P^{cap} in Eq. (31) is replaced by the power consumption $P_{t_0}^{ch}$ at time t_0 , given by:

$$P_t^{ch} \leq \bar{P}^{ch} = P_{t_0}^{ch}, \quad \forall t \in [t_1, t_2], \quad (39)$$

where \bar{P}^{ch} is the upper limit of the operating power during the recovery stage. The training process of the safe-DRL agent is organized in Fig. 6 to show the step-by-step calculation process. It includes 5 steps, where the first step is to develop the environment based on Eqs. (2)-(13) and initialize the policy; the second step is to select an action according to the policy; the third step is to tune the "unsafe" action by the proposed safe layer as Table II; the fourth step is to update the policy according to samples.

²The configuration principle is to make the indoor temperature recover to 50% of $\Delta T_{i,t_1}$ when the time goes halfway, i.e., $t = t_1 + \frac{1}{2}(t_2 - t_1)$.

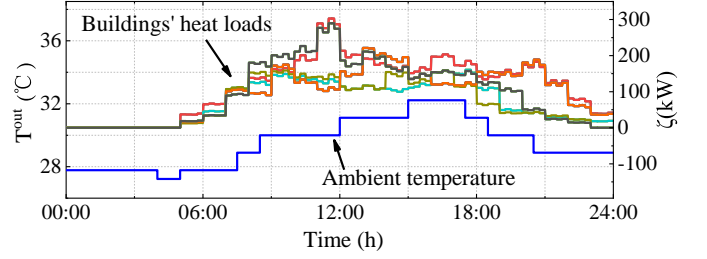


Fig. 7: The ambient temperature and buildings' heat loads.

Remark 3. The proposed self-adaptive target method in Eq. (37) can regulate the DCS operating power to avoid the power rebound in the power recovery stage and minimize the buildings' comfort impacts.

IV. CASE STUDIES

A. Test System

The test system is modelled based on a realistic DCS in Hengqin, China, following its technical guidelines (the 4th Edition) [54]. The total installed cooling capacity in the energy station is 41,000 RT (≈ 144 MW) with COP=5.5. The designed supply and return water temperature in two loops at time $t=0$ is $T^{ch,s} = 3$ °C, $T_{i,0}^{l,r} = 12$ °C, $T_{i,0}^{ll,s} = 13$ °C, $T_{i,0}^{ll,r} = 18$ °C, respectively. In addition, based on the national standard in China (JGJ 134-2010, GB 12021.3-2010, GB 31349-2014), the following parameters are designed as $k_i^{HE} = 4.5$ kW/(m²·°C), $U^{O-A} = 0.0036$ kW/(m²·°C), $c^w = 4.2$ kJ/(kg·°C), $c^A = 1.005$ kJ/(kg·°C) and $\rho^A = 1.205$ kg/m³. The efficiency coefficients of heat exchanging process between different loops are set as $\eta_i^I = 0.9$, $\eta_i^{II} = 0.9$, respectively. The heat transfer coefficient of supply water η^{pipe} is 0.95. The air mixing proportion is set as $\alpha_i = 0.1$.

The DCS in Hengqin provides cooling services for 12 buildings. The maximum value of the mass flow rate \bar{m}_i^1 ranges from 600 kg/s to 1,200 kg/s in different buildings, and the corresponding minimum value \underline{m}_i^1 is 3% of \bar{m}_i^1 . Each building's floor area A_i^s and its set temperature $T_{i,t}^{set}$ are distributed in 100,000~300,000 m², and 20~23 °C, respectively. The maximum deviation of the required comfortable indoor temperature is ± 1 °C. Moreover, the ambient temperature T_t^{out} and each building's heat load $\zeta_{i,t}$ adopt the realistic data in Hengqin, from June 1, 2020 to August 31, 2020 (one typical day's profiles are shown in Fig. 7).

The control objective of DCS is to provide operating reserve from 14:00pm to 14:15pm, as shown in Fig. 8.³ The black curve is the original power consumption, and regarded as the power baseline before regulation. The red shadow area is the required decrease of energy consumption, and the operating power should be lower than the power cap $P^{cap}=60$ MW during this period. In the recovery stage, the new power cap is set as the peak power of the baseline, i.e., $\bar{P}^{ch} = 96$ MW.

³Note that these experimental settings are for illustrative purpose. In practice, the service duration, the operating reserve period and the power cap P^{cap} are determined by the system operator. The effectiveness of the proposed methodology is not affected by these parameter settings.

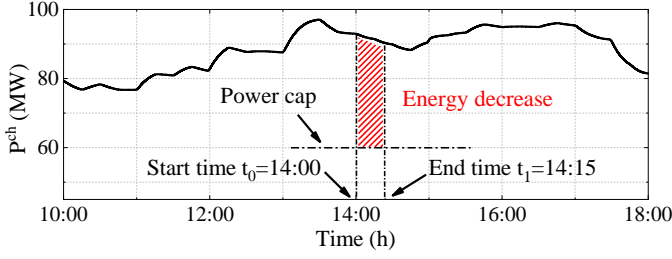


Fig. 8: The original power consumption of DCS.

B. Benchmarks

To validate the superiority of the proposed safe-DRL scheme, we implement another two centralized control methods as our benchmarks: the traditional proportional-integral (PI) controller [55] and the DRL controller [25]. Here, the superiority includes providing a higher-quality service and preventing the power rebound with the minor impacts on buildings' temperature comforts.

In the PI method, the control signal is determined by the feedback of both the power cap violation and indoor temperature comfort⁴. The buildings' mass flow regulation at each time step t can be expressed as:

$$\Delta m_t^{\text{ch}} = P^{\text{ch}}(P_t^{\text{ch}} - P_{t-1}^{\text{ch}}) + I^{\text{ch}}(P_t^{\text{ch}} - P^{\text{cap}}), \quad (40)$$

$$\Delta m_{i,t}^{\text{I}} = [P_i(T_{i,t}^{\text{A}} - T_{i,t-1}^{\text{A}}) + I_i(T_{i,t}^{\text{A}} - T_{i,t}^{\text{set}})] + m_{i,t-1}^{\text{I}} \Delta m_t^{\text{ch}} / \sum_{i \in \mathcal{I}} m_{i,t-1}^{\text{I}}, \quad (41)$$

where $P^{\text{ch}}, I^{\text{ch}}$ are parameters of the PI controller in pipelines to follow power caps; P_i, I_i are the parameters of PI controllers in buildings to follow set indoor temperatures. Eq. (41) means the regulation of the total mass flow is achieved by adjusting each building proportionally. During the power reduction stage, parameters are set as $P^{\text{ch}}=0.2, I^{\text{ch}}=0.02$. During the power recovery stage, the power constraint is relaxed and $P^{\text{ch}}=I^{\text{ch}}=0^5$.

In the traditional DRL method, it does not have the safe layer so that the power constraint P^{cap} is considered as a penalty item in its reward function, which is formulated as:

$$r_{t+1} = -\theta^{\text{r}} \mathbb{E}_{i \in \mathcal{I}} [\Delta T_{i,t+1}] - \sigma_{i \in \mathcal{I}}^2 [\Delta T_{i,t+1}] - \theta^{\text{p}} |P_{t+1}^{\text{ch}} - P^{\text{cap}}|, \quad \forall t \in \mathcal{T}, \quad (42)$$

where θ^{p} is the weight factor of the penalty item.

C. Training Process of the Safe-DRL Agent

The parameters of the proposed safe-DRL are designed as Table III. The key hyper-parameters are designed based on the experience concluded in the existing literature [56], including the discount factor, learning rates, replay buffer capacity, etc.. The actor and critic networks are composed by one input

⁴A more complex version of PI controller is PID controller, in which a D parameter is added to reflect the differential control process. We have also tested PID controller's performance for DCS control. However, our experiments show that PID does not obviously outperform PI (but with a more complex structure). Furthermore, when adopting a PID controller with a large D parameter, there may exist a strong noise in the controlling results. Hence, we only include PI controller in this paper to save space.

⁵The tuning rule is to satisfy the daily operating and maintain buildings' comfortable set temperatures.

TABLE III: Parameters for safe-DRL and DRL methods.

Symbols	Definitions	Values
τ	Target smooth factor	0.005
γ	Discount factor	0.9
$ R $	Replay buffer capacity	10000
ξ	Exploration noise	0.3
M	Max episodes	2500
T	Max step	15
K	Mini batch size	200
δ_{θ^Q}	Learning rate of critic network Q	0.001
δ_{θ^π}	Learning rate of actor network π	0.0001
θ^{r}	Weight factor of temp deviations	0.01
θ^{p}	Weight factor of power violations	0.05

layer, two hidden layers and one output layer, respectively⁶. The neurons number in each hidden layer is set as 128. The Rectified Linear Unit is used as the activation function.

The parameters in DRL (benchmark) adopt the same experimental settings with safe-DRL. The simulation is implemented by Windows system, using PyTorch in Python with an Intel core i7 CPU @3.0 GHz and 16GB memory.

The training process is shown in Fig. 9, and the number of training episode is 2500. Fig. 9(a) presents the reward value for appraising the agent's decision in each episode. It can be seen that the rewards in safe-DRL and DRL have oscillations at first because of the unknown knowledge about the training environment (i.e., the DCS). With the increase of training episodes, the rewards converge to their respective stable values, called convergence reward. Then, both of the two agents obtain the optimal policy in Eq. (21).

The comparison of training efficiency between the proposed safe-DRL and the traditional DRL method is shown in Table IV, where the efficiency indicator includes sample efficiency, convergence time and convergence reward. Sample efficiency is the estimated minimum number of samples to converge as illustrated in Fig. 9(a). It can be seen that the proposed safe-DRL needs less sampled episodes to converge, so it has higher sample efficiency and shorter convergence time (2.9 mins). Besides, the convergence reward of safe-DRL is larger (-30) than that of DRL (-45), which means the safe-DRL agent can achieve the temperature objective better than DRL.

Fig. 9(b) shows the constraint violation during the agents' training processes, where $\Delta P = P^{\text{ch}} - P^{\text{cap}}$ is the power gap to the required power cap. It can be seen that the power constraint violation is conspicuous and even reaches to over 40MW in DRL, which may harm the stable operation of the power system. However, the operating power can satisfy the power cap strictly in safe-DRL, which proves the effectiveness of the proposed safe-layer. Besides, the operating power is quite close to the power cap, because the agent wants to make full use of the allowable power to decrease the indoor temperature deviations. Thus the well-trained agent can be applied to the online control of DCS for providing operating reserve.

D. Online Control of DCS for Providing Operating Reserve

For a random case, it is assumed that the power system has the regulation demand at 14:00pm, and sends the regulation

⁶Layer number adopts existing literature models that performs well [57].

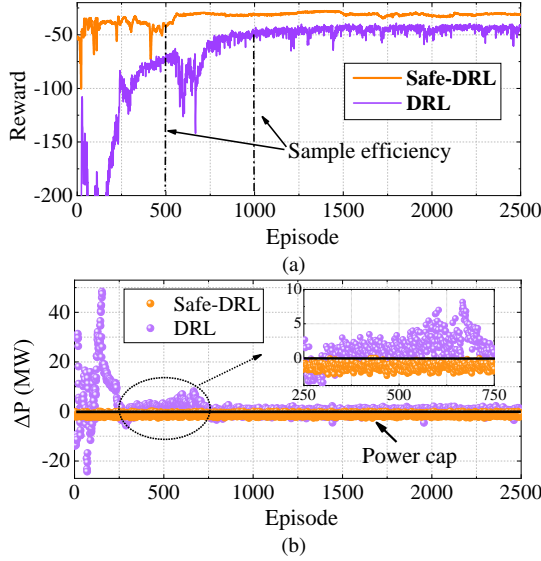


Fig. 9: The training process of the safe-DRL agent. (a) The reward value; (b) The constraint value of operating power.

TABLE IV: The training efficiency results for 2500 episodes.

Methods	Sample efficiency	Convergence time	Convergence reward
Safe-DRL	~500	2.9 mins	-30
DRL	~1000	4.3 mins	-45

signal to the agent to cut down the DCS's operating power to be lower than 60 MW in this dispatch period (15 mins). The control power results of DCS for providing operating reserve is shown in Fig. 10, which applies three different controllers (i.e., PI, the traditional DRL and the proposed safe-DRL) to the system in Fig. 10(a), Fig. 10(b) and Fig. 10(c), respectively.

It can be seen from Fig. 10(a) that DCS operating power is cut down and satisfies the required power after 5 mins. Because the PI controller is designed based on the feedback, it cannot respond to the changing environment immediately and results in some time delay. In Fig. 10(b) and Fig. 10(c), the DRL and safe-DRL controllers can decrease the operating power more quickly compared with PI controller, where the power reduction is achieved only within 1 min. Moreover, during the whole dispatch period, the operating power in Fig. 10(a) cannot be maintained below 60 MW and exceeds the required power cap at 14:11 due to the dynamic cooling demand in buildings (e.g., variational heat loads caused by people flows). By contrast, the operating power can be controlled under the power cap during all the dispatch period in Fig. 10(c), which validates the effectiveness of the proposed safe-DRL agent to satisfy power system's critical constraint strictly. In Fig. 10(b), the traditional DRL method can also achieve the required power cap after training, however its training process can not satisfy the constraint.

After the power reduction stage, three controllers in Fig. 10 increases DCS's operating power to restore buildings' indoor temperatures. However, a new peak power 114 MW and 104 MW appears in the recovery stage in Fig. 10(a) and Fig. 10(b), respectively. They are even much higher than the original daily maximum operating power (96 MW). This phenomenon may cause a secondary impact on the power system that has just

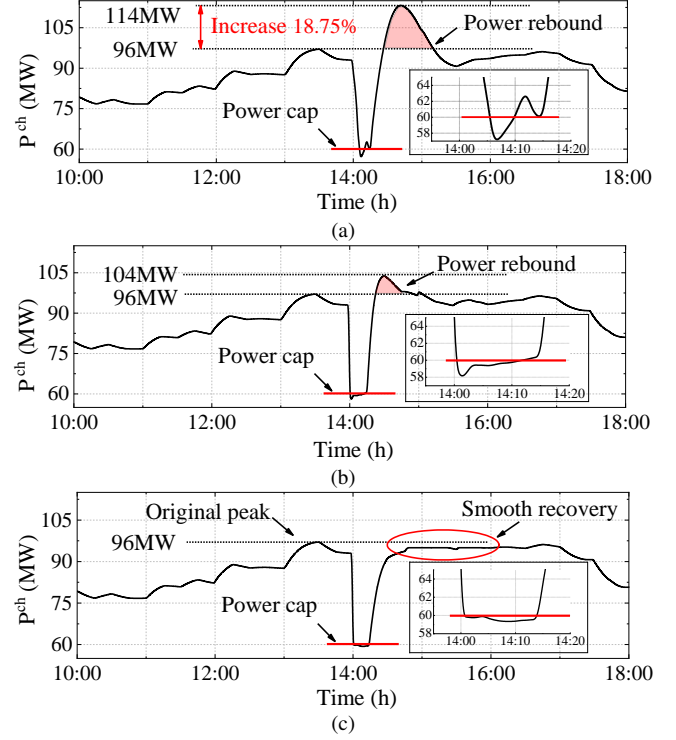


Fig. 10: The control power results of DCS based on (a) PI controller; (b) Traditional DRL method; (c) Safe-DRL method.

TABLE V: The statistical indicator of the temperature influence to buildings

Methods	Max deviation	Uncomfortable number	Average deviation
PI	1.57 °C	6	0.75 °C
DRL	1.18 °C	2	0.80 °C
Safe-DRL	0.93 °C	0	0.85 °C

returned to the stable state. By utilizing the proposed safe-DRL method in Fig. 10(c), the safe-layer limits the peak value during the recovery stage and guarantees the smooth recovery of the operating power without a new peak power rebound.

Moreover, when DCS is controlled to provide operating reserve, building's indoor temperature will get influenced and deviates from its set value, as shown in Fig. 11. The blue area shows the comfortable temperature range in buildings, and ΔT denotes each building's temperature deviation. In the power reduction stage, all the buildings' indoor temperatures increase due to the reduction of cooling power supplies. In Fig. 11(a), more than half of the buildings' indoor temperatures deviate larger than 1 °C and enter the uncomfortable area. It means that some buildings get seriously impacted during the regulation process while some others do not. In Fig. 11(b), the DRL method can maintain the temperature comfort better than PI, while some buildings' indoor temperature still exceed the comfortable range. By contrast, in Fig. 11(c), the temperature deviations in different buildings are close and maintained within 1 °C by using the safe-DRL controller. Although all buildings have different floor areas, heights and heat loads, the temperature influence to each building is always similar. Thus, the safe-DRL method is insensitive to buildings' different models. As shown in Table V, it can be seen the max deviation of buildings in safe-DRL is the smallest, which makes sure

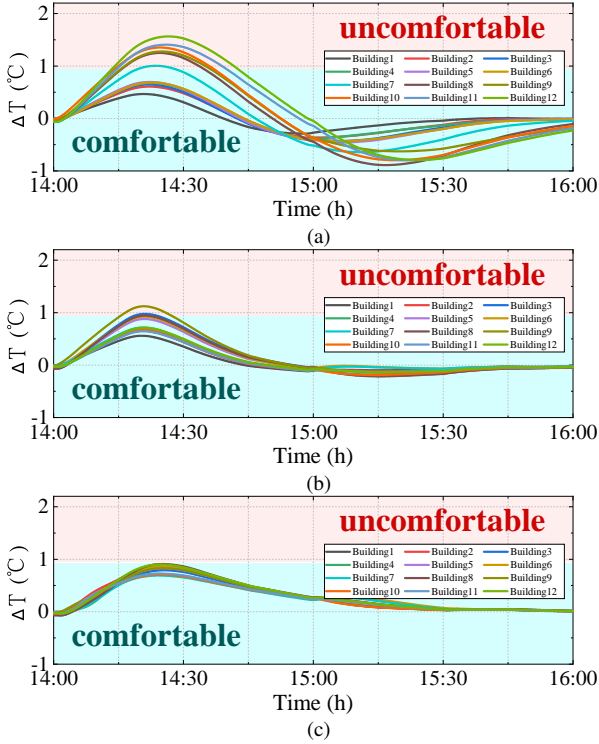


Fig. 11: The temperature deviation results based on (a) PI controller; (b) Traditional DRL method; (c) Safe-DRL method.

more buildings comfortable. However, the average temperature deviation of all the buildings is a little higher (around 0.05°C) than the other two methods, which is small and can be neglected. This validates the advantage of the proposed method to regulate each building's mass flow rate dynamically and guarantee their temperature requirements.

It can be seen from Figs. 10-11 that the power rebound is more obvious with the PI method, because the indoor temperatures can recover to their set values quickly after the power reduction period and cause the temperature overshoot. Compared with the PI method, the proposed safe-DRL method can control the indoor temperature to recover the set values smoothly. That is to say, the safe-DRL method addresses the power rebound at the cost of a longer recovery time.

E. Physical Operating States of DCS During Control Process

Fig. 12 shows the control results of all the buildings' water mass flows \dot{m}_t^l in the first water loop. Fig. 12(a) and Fig. 12(b) are obtained based on the PI controller and safe-DRL method, respectively. In the power reduction stage, all the buildings' mass flows are decreased slowly based on the feedback with the same proportion in Fig. 12(a), which is regardless of the differences among buildings. However, mass flows are decreased quickly at first, and then adjusted in different directions for satisfying different cooling demands in heterogeneous buildings in Fig. 12(b). Therefore, the proposed safe-DRL method can consider different buildings' thermal inertia characteristics and guarantee their comforts. In the power recovery stage, water mass flows in Fig. 12(a) recover with a faster speed than that in Fig. 12(b), which makes it easier for the PI controller to cause power rebound.

Buildings' wind mass flows \dot{m}_t^w are dependent variables during the control process. If buildings' indoor temperature is

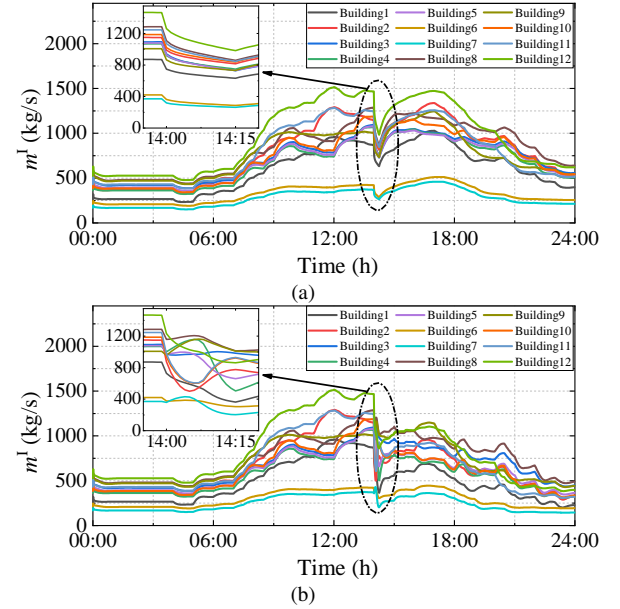


Fig. 12: The water mass flow in buildings based on (a) PI controller; (b) Safe-DRL method.

higher than the set value, wind mass flows \dot{m}_t^w will increase to cool down the buildings. The dynamic processes of the wind mass flow in different buildings are illustrated in Fig. 13, where Fig. 13(a) and Fig. 13(b) are the control results of the PI controller and safe-DRL, respectively. In the power reduction stage, wind mass flows in two figures both increase automatically due to the decrease of chilled water. The increase speed in Fig. 13(b) is faster than that in Fig. 13(a), because a faster power decrease in the safe-DRL method makes a sharper increase of buildings' indoor temperatures. Buildings want to increase their cooling winds to cool down the indoor temperature.

In the power recovery stage, wind mass flows in Fig. 13(b) decrease more slowly than that in Fig. 13(a). It means buildings' indoor temperatures are recovered more slowly using the safe-DRL method for preventing the power rebound. However, in the PI controller, the water mass flows are increased quickly in the recovery stage, which results in the quick decrease of wind mass flows in Fig. 13(a). To sum up, the proposed safe-DRL method can slow down the adjustment of wind mass flows to avoid the power rebound.

F. Sensitivity of the Key Parameters of DCS for Providing Operating Reserve

Considering the real-time demands from power systems, the service requirements of the power cap and duration are probably various. Thus, the sensitive analysis is carried out to validate the effectiveness of the proposed method in different scenarios. Fig. 14(a) and Fig. 14(b) show the analysis results of different duration periods and power caps, respectively. There are two observations: each building's maximum temperature deviation ΔT^{\max} , and the maximum power consumption P^{\max} . The ΔT^{\max} is for representing the indoor temperature comfort in buildings, and P^{\max} is for quantifying power cap violations.

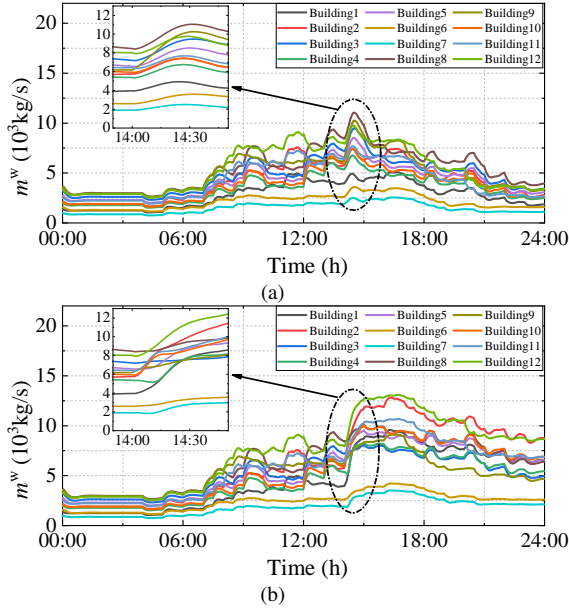


Fig. 13: The wind mass flow in buildings based on (a) PI controller; (b) Safe-DRL method.

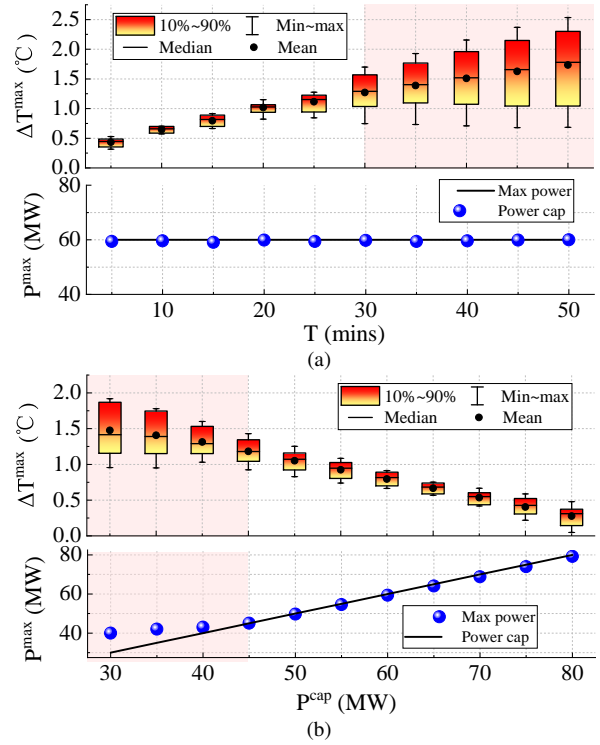


Fig. 14: The sensitivity results of (a) Duration; (b) Power cap.

Fig.14(a) shows the control results based on different scenarios of duration time, which ranges from 5 minutes to 50 minutes, where the power cap is set as 60MW. It can be seen that the power cap can be always satisfied during the regulation period with different duration time. However, the impacts on buildings' indoor temperature are more significant with the increase of duration time. When the duration time is longer than 30 minutes, some buildings' temperature deviations can be up to 1.5°C. By contrast, when the duration time is less than 30 minutes, all the buildings' temperature deviations are comparatively marginal with small variances.

Fig.14(b) shows the control results based on different required power caps, which range from 30MW to 80MW with the same duration time 15 minutes. It can be seen that the DCS is capable to provide high-quality operating reserve when the power cap is higher than 45 MW. However, the power cap can not be satisfied when it is lower than 45 MW, because DCS has the minimum physical limits in water mass flows. Besides, a lower power cap can result in larger temperature impacts on buildings because of less cooling supplies. It can be seen from Fig.14(b) that the average temperature deviation is out of the comfortable range and the corresponding variances are large, when the power cap is lower than 45 MW.

In summary, the proposed safe-DRL can generally perform well with different service duration and power cap scenarios. However, long service duration time and low power cap may deteriorate its performance. A reasonable power cap and service duration time are significant for the final control results of DCS, which are generally determined by the capacity offer of DCS in electrical market [51]. The strategy for the DCS operator to provide its operating reserve capacity is beyond the scope of this paper, but will be our future work.

V. CONCLUSION

This paper proposes a model-free safe-DRL scheme for DCS control problem to provide operating reserve. A safe layer is proposed to effectively guarantee the critical power constraint in the power reduction stage. A self-adaptive target method is further adopted to tackle the power rebound in the power recovery stage. Meanwhile, it minimizes the impacts on buildings' indoor temperature to keep all the buildings as comfortable as possible. Numerical studies show that the DCS's operating power is always below the power cap during training, which ensures the "safety" for providing operating reserve. Besides, the DCS's operating power can recover smoothly and avoid an undesirable peak power rebound. All the buildings' temperature deviations can be guaranteed within the required range $\pm 1^\circ\text{C}$ to stay comfortable.

The well-trained agent in the proposed safe-DRL method can usually take effects on the similar DCS. If the physical system is totally different, the agent probably should be trained again using the new system's historical data. To be more efficient, the agent can transfer the old system's knowledge to the new system by less training episodes, which also will be our future work. For the ancillary service that requires a long duration (e.g., more than one hour), it is probably more efficient for DCS to regulate both the chilled water's mass flow and supply temperature, which will be studied in our future work as well.

REFERENCES

- [1] S. Imprim, S. V. Nese, and B. Oral, "Challenges of renewable energy penetration on power system flexibility: A survey," *Energy Strategy Rev.*, vol. 31, p. 100539, Sep. 2020.
- [2] H. Nosair and F. Bouffard, "Reconstructing operating reserve: Flexibility for sustainable power systems," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1624–1637, Oct. 2015.

- [3] P. Siano, "Demand response and smart grids survey," *Renew. Sust. Energ. Rev.*, vol. 30, pp. 461–478, Feb. 2014.
- [4] J. Wang, S. Huang, D. Wu, and N. Lu, "Operating a commercial building hvac load as a virtual battery through airflow control," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 158–168, Jan. 2021.
- [5] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Appl. Energy*, vol. 236, pp. 1078–1088, Feb. 2019.
- [6] S. Werner, "International review of district heating and cooling," *Energy*, vol. 137, pp. 617–631, Oct. 2017.
- [7] S. J. Cox, D. Kim, H. Cho, and P. Mago, "Real time optimal control of district cooling system with thermal energy storage using neural networks," *Appl. Energy*, vol. 238, pp. 466–480, Mar. 2019.
- [8] X. Chen, Q. Hu, Q. Shi, X. Quan, Z. Wu, and F. Li, "Residential hvac aggregation based on risk-averse multi-armed bandit learning for secondary frequency regulation," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 6, pp. 1160–1167, Nov. 2020.
- [9] C.-C. Lo, S.-H. Tsai, and B.-S. Lin, "Ice storage air-conditioning system simulation with dynamic electricity pricing: A demand response study," *Energies*, vol. 9, no. 2, p. 113, 2016.
- [10] G. Chen, B. Yan, H. Zhang, D. Zhang, and Y. Song, "Time-efficient strategic power dispatch for district cooling systems considering the spatial-temporal evolution of cooling load uncertainties, to appear in CSEE J. Power Energy Syst., 2021. DOI: 10.17775/CSEEJPES.2020.06800.
- [11] R. Tang, S. Wang, K. Shan, and H. Cheung, "Optimal control strategy of central air-conditioning systems of buildings at morning start period for enhanced energy efficiency and peak demand limiting," *Energy*, vol. 151, pp. 771–781, 2018.
- [12] PJMINT.L.L.C., "PJM manual 11: Energy&ancillary services market operations," Revision: 115, pp. 83–98, Jun. 01, 2021. [Online]. Available: <https://www.pjm.com/-/media/documents/manuals/m11.ashx>.
- [13] C. Cheng, B. Yang, and F. Wang, "Research on the application of large scale air conditioning load in demand response," *Power Demand Side Management*, vol. 19, no. 3, pp. 57–59, Apr. 2017.
- [14] S. Buffa, M. H. Fouladfar, G. Franchini, I. Lozano Gabarre, and M. Andrés Chicote, "Advanced control and fault detection strategies for district heating and cooling systems a review," *Appl. Sci.*, vol. 11, no. 1, p. 455, Jan. 2021.
- [15] W. Zheng and D. J. Hill, "Incentive-based coordination mechanism for distributed operation of integrated electricity and heat systems," *Appl. Energy*, vol. 285, p. 116373, Mar. 2021.
- [16] Y. Lu, *Practical heating and air conditioning design manual*, Beijing, China, 2008.
- [17] Y. Li, Y. Rezugui, and H. Zhu, "District heating and cooling optimization and enhancement towards integration of renewables, storage and smart grid," *Renew. Sust. Energ. Rev.*, vol. 72, pp. 281–294, May 2017.
- [18] Y. Ding, W. Cui, S. Zhang, H. Hui, Y. Qiu, and Y. Song, "Multi-state operating reserve model of aggregate thermostatically-controlled-loads for power system short-term reliability evaluation," *Appl. Energy*, vol. 241, pp. 46–58, May 2019.
- [19] R. Yin, E. C. Kara, Y. Li, N. DeForest, K. Wang, T. Yong, and M. Stadler, "Quantifying flexibility of commercial and residential loads for demand response using setpoint changes," *Appl. Energy*, vol. 177, pp. 149–164, Sep. 2016.
- [20] M. D. Knudsen and S. Petersen, "Model predictive control for demand response of domestic hot water preparation in ultra-low temperature district heating systems," *Energy Build.*, vol. 146, pp. 55–64, Jul. 2017.
- [21] C. Saletti, A. Gambarotta, and M. Morini, "Development, analysis and application of a predictive controller to a small-scale district heating system," *Appl. Therm. Eng.*, vol. 165, p. 114558, Jan. 2020.
- [22] D. M. Alghool, T. Y. Elmekawy, M. Haouari, and A. Elomri, "Optimization of design and operation of solar assisted district cooling systems," *Energy Convers. Manag.*, vol. 6, p. 100028, Apr. 2020.
- [23] M. Wirtz, L. Neumaier, P. Remmen, and D. Müller, "Temperature control in 5th generation district heating and cooling networks: An milp-based operation optimization," *Appl. Energy*, vol. 288, p. 116608, Apr. 2021.
- [24] A. Stoppato, A. Benato, N. Destro, and A. Mirandola, "A model for the optimal design and management of a cogeneration system with energy storage," *Energy Build.*, vol. 124, pp. 241–247, Jul. 2016.
- [25] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
- [26] Y. Du, H. Zandi, O. Kotevska, K. Kurte, J. Munk, K. Amasyali, E. Mckee, and F. Li, "Intelligent multi-zone residential hvac control strategy based on deep reinforcement learning," *Appl. Energy*, vol. 281, p. 116117, Jan. 2021.
- [27] X. Xu, Y. Jia, Y. Xu, Z. Xu, and C. S. Lai, "A multi-agent reinforcement learning-based data-driven method for home energy management," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3201–3211, Jul. 2020.
- [28] L. Yu, Y. Sun, Z. Xu, C. Shen, D. Yue, T. Jiang, and X. Guan, "Multi-agent deep reinforcement learning for hvac control in commercial buildings," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 407–419, Jan. 2021.
- [29] F. Ruelens, B. J. Claessens, P. Vrancx, F. Spiessens, and G. Deconinck, "Direct load control of thermostatically controlled loads based on sparse observations using deep reinforcement learning," *CSEE J. Power Energy Syst.*, vol. 5, no. 4, pp. 423–432, Dec. 2019.
- [30] X. Zhang, D. Biagioni, M. Cai, P. Graf, and S. Rahman, "An edge-cloud integrated solution for buildings demand response using reinforcement learning," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 420–431, Jan. 2021.
- [31] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, Sep. 2016.
- [32] R. Dobbe, P. Hidalgo-Gonzalez, S. Karagiannopoulos, R. Henriquez-Auba, G. Hug, D. S. Callaway, and C. J. Tomlin, "Learning to control in power systems: Design and analysis guidelines for concrete safety problems," *Electr. Power Syst. Res.*, vol. 189, p. 106615, Dec. 2020.
- [33] P. Kou, D. Liang, C. Wang, Z. Wu, and L. Gao, "Safe deep reinforcement learning-based constrained optimal control scheme for active distribution networks," *Appl. Energy*, vol. 264, p. 114772, Apr. 2020.
- [34] Y. Gao, W. Wang, J. Shi, and N. Yu, "Batch-constrained reinforcement learning for dynamic distribution network reconfiguration," *IEEE Trans. on Smart Grid*, vol. 11, no. 6, pp. 5357–5369, Nov. 2020.
- [35] W. Wang, N. Yu, Y. Gao, and J. Shi, "Safe off-policy deep reinforcement learning algorithm for volt-var control in power distribution systems," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3008–3018, Jul. 2019.
- [36] H. Liu and W. Wu, "Two-stage deep reinforcement learning for inverter-based volt-var control in active distribution networks," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2037–2047, May 2020.
- [37] T. L. Vu, S. Mukherjee, R. Huang, and Q. Hung, "Barrier function-based safe reinforcement learning for emergency control of power systems," *arXiv preprint arXiv:2103.14186*, May 2021.
- [38] T. L. Vu, S. Mukherjee, T. Yin, R. Huang, Q. Huang et al., "Safe reinforcement learning for emergency loadshedding of power systems," *arXiv preprint arXiv:2011.09664*, Nov. 2020.
- [39] C. Zhang, Z. Li, J. Zheng, and Q. Wu, "Power system emergency control to improve short-term voltage stability using deep reinforcement learning algorithm," in *2019 IEEE 3rd International Electrical and Energy Conference (CIEEC)*. IEEE, Sep. 2019, pp. 1872–1877.
- [40] Z. Liang, C. Huang, W. Su, N. Duan, V. Donde, B. Wang, and X. Zhao, "Safe reinforcement learning-based resilient proactive scheduling for a commercial building considering correlated demand response," *IEEE Open Access Journal of Power and Energy (OAJPE)*, vol. 8, pp. 85–96, Mar. 2021.
- [41] H. Li, Z. Wan, and H. He, "Constrained ev charging scheduling based on safe deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2427–2439, 2019.
- [42] C. Zhang, S. R. Kuppannagari, R. Kannan, and V. K. Prasanna, "Building hvac scheduling using reinforcement learning via neural network based model approximation," in *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation*, Nov. 2019, pp. 287–296.
- [43] W. Zheng and D. Hill, "Distributed real-time dispatch of integrated electricity and heat systems with guaranteed feasibility," *IEEE Trans. Industr. Inform.*, May 2021.
- [44] S. Xiaochun, "Technical analysis of district cooling design for west zone of zhongguancun," *Heating Ventilating & Air Conditioning*, Oct. 2004.
- [45] W. Zheng, Y. Hou, and Z. Li, "A dynamic equivalent model for district heating networks: formulation, existence and application in distributed electricity-heat operation," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2685–2695, Jan. 2021.
- [46] N. Marshall, "Heat exchange design handbook," *Int. J. Heat Fluid Flow*, vol. 4, no. 2, p. 77, 1983.
- [47] H. Hui, Y. Ding, W. Liu, Y. Lin, and Y. Song, "Operating reserve evaluation of aggregated air conditioners," *Appl. Energy*, vol. 196, pp. 218–228, Jun. 2017.
- [48] O. Alagoz, H. Hsu, A. J. Schaefer, and M. S. Roberts, "Markov decision processes: a tool for sequential decision making under uncertainty," *Med. Decis. Making*, vol. 30, no. 4, pp. 474–483, Dec. 2010.
- [49] E. Altman, *Constrained Markov decision processes*. CRC Press, 1999, vol. 7.

- [50] O. Alagoz, H. Hsu, A. J. Schaefer, and M. S. Roberts, "Markov decision processes: a tool for sequential decision making under uncertainty," *Med. Decis. Making*, vol. 30, no. 4, pp. 474–483, 2010.
- [51] Z. Shi, C. Wang, X. Lei, X. Ye, W. Yuan, and Z. Cao, "Research on the technical economy and market mechanism of electric heat storage system participating in auxiliary service," in *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, Oct. 2018, pp. 1–6.
- [52] S. Singh, T. Jaakkola, M. L. Littman, and C. Szepesvári, "Convergence results for single-step on-policy reinforcement-learning algorithms," *Mach. Learn.*, vol. 38, no. 3, pp. 287–308, Mar. 2000.
- [53] K. P. Schneider, E. Sortomme, S. S. Venkata, M. T. Miller, and L. Ponder, "Evaluating the magnitude and duration of cold load pick-up on residential distribution using multi-state load models," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3765–3774, Nov. 2016.
- [54] New District Hengqin: District Cooling and Heating System Technical Guidelines (the 4th Edition), Zhuhai, China, Mar. 2016.
- [55] Y. Chen, H. Yan, Y. Luo, and H. Yang, "A proportionalintegral (pi) law based variable speed technology for temperature control in indirect evaporative cooling system," *Appl. Energy*, vol. 251, p. 113390, Oct. 2019.
- [56] S. S. Mousavi, M. Schukat, and E. Howley, "Deep reinforcement learning: an overview," in *Proceedings of SAI Intelligent Systems Conference*. Springer, 2016, pp. 426–440.
- [57] A. J. Thomas, M. Petridis, S. D. Walters, S. M. Gheytassi, and R. E. Morgan, "Two hidden layers are usually better than one," in *International conference on engineering applications of neural networks*. Springer, Aug. 2017, pp. 279–290.