Fast Frequency Regulation Utilizing Non-aggregate Thermostatically Controlled Loads Based on Edge Intelligent Terminals

Taoyi Qi, Chengjin Ye, *Member, IEEE,* Hongxun Hui, *Member, IEEE,* and Yuming Zhao

***Abstract*--Thermostatically controlled loads (TCLs) have promising regulation potential and are regarded as valuable flexible resources. Due to the relatively lower operating power, TCLs are always aggregated to participate in various regulation programs. However, the intrinsic response delays of aggregation always result in relatively slower speed and limit further implementation in the fast frequency regulation (FFR) service, which is urgently needed by renewable power systems without sufficient inertias. Therefore, a non-aggregate method based on edge intelligent terminal (EIT) is proposed to utilize TCLs to provide FFR without unacceptable rebound effects. Firstly, the FFR infrastructure, consisting of the distributed frequency response and centralized recovery optimization, is developed based on functions of the EIT, such as local measurement, edge computation, and flexible control. Then, the equivalent mirror recovery strategy is proposed to mitigate power fluctuations and reduce occupations of other regulation resources. On this basis, the optimal recovery problem of TCLs is formulated to achieve flexible power control without violations of FFR requirements. To overcome the curse of dimensionality, the dynamic clustering method is proposed to divide non-aggregate TCLs into tens of groups, thereby the optimization problem is solved by EITs separately and the solution speed is independent of system scales. Finally, the proposed FFR is validated based on several simulation cases, and results show that TCLs significantly contribute to enhancing frequency resilience for low-inertia power systems.**

***Index Terms*--Fast frequency regulation, thermostatically controlled load, intelligent terminal, cloud-edge collaboration.**

I. Introduction

FREQUENCY security is one of the most important power system criteria, whose variations indicate real-time supply-demand mismatches [1]. As massive conventional units are replaced by growing renewable energy resources (RES) [2], not only decreases the accessible frequency regulation capacity but also the inertia of power systems reduced dramatically [3]. The frequency resilience, related to the system inertia and frequency response provision speed, is becoming gradually insufficient to withstand the frequency drop caused by sudden infeed loss and recover its initial frequency at the given time [4].

Worse still, kinds of extreme weather have occurred more frequently worldwide in the past decade, always along with dual effects caused by the demand burgeon of TCLs and the supply failure of generators, which threatens the secure and stable operation of power systems [5]. For instance, the intense heat swept across China, resulting in significant cooling load growth in 2022. In Sichuan province, many hydro generators are forced to stop due to the lack of reservoir capacity. Even though other resources are regulated with maximum output, the severe supply-demand imbalance still leads to widespread electricity restriction accidents [6]. Under such circumstances, full-load or overload operations with higher failure rates and less available regulation capacity are likely to result in severe frequency accidents, raising many more concerns about frequency security [7].

To enhance resilience against various extreme weather events, the frequency security problems cannot be handled by dispatching generators alone. Therefore, fast frequency regulation (FFR) services provided by other resources, such as wind turbines (WTs), energy storage systems (ESSs), and flexible loads, are implemented to complement the virtual inertia. FFR aims to limit the maximum rate of change of frequency (RoCoF) in the low-inertia systems, via responding to the frequency deviations within a very short period, e.g., 0.5 seconds in Texas and 1 second in the UK [8].

As one of the typical flexible loads, TCLs attract much more research interests and have been investigated to provide frequency regulation services for power systems recently [9], [10]. Existing researches have leveraged the potential of TCLs for providing both primary frequency regulation (PFR) and secondary frequency regulation (SFR) services [11], [12]. Benefiting from the significant shares of buildings' total power consumption and inherent thermal inertia, TCLs can also provide promising FFR capacity without notable temperature discomforts during a relatively short frequency regulation period [13]. The demands of FFR are increasing steadily with the development of RESs, while the ESSs and WTs are facing the problems of higher costs and shorter duration, respectively. However, several obstacles should be overcome before using TCLs to provide FFR due to more strict response requirements on response speed, and rebound effects are put forward.

To guarantee such a short response period, the centralized control method, such as direct load control (DLC), is widely used in industrial or commercial loads with significant power. Nevertheless, centralized control is not cost-effective to be applied considering the vast number of TCLs [14]. With the progress of technologies, several advancing equipment, such as smart meters, have been investigated to realize the fast frequency response. Besides, the control method should also be suitable for TCLs without violating their operating requirements. Hence, the distributed control method is technically feasible to achieve large-scale participation of residential TCLs and has been adopted in many previous research studies [9], [15], [16]. Among them, the aggregators always play an essential role in PFR or SFR regulation infrastructures. Compared with the DLC method, the involvement of aggregators brings in extra communicational layers and control steps between the system operators and

This work was supported by the key science and technology project of China Southern Power Grid Corporation (090000KK52220020). (Corresponding author: Chengjin Ye.)

T. Qi is with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, and the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau 999078, China (eeqty@zju.edu.cn).
C. Ye is with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China (yechenjing@zju.edu.cn).
H. Hui is with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau 999078, China (hongxunhui@um.edu.mo).
Y. Zhao is with the Shenzhen Power Supply Bureau Co., Ltd. Shenzhen, China (zhaoym97@sina.com).

TCLs, accompanied by more information interactions as well as longer time delays, making it more challenging to satisfy the fast frequency response. Consequently, the non-aggregate scheme should be further exploited to facilitate the FFR participation of TCLs.

Apart from speed requirements, some indices on the recovery process are also proposed to prevent the rebound effects after the deployment of FFR [17], e.g., the maximum ramp rate and the maximum power rebound rate. To mitigate the rebound effects, researchers have proposed multiple recovery strategies in previous PFR and SFR programs through optimizing and controlling the response time instant of TCLs in the response period, so as to restrain the synchronous power engagement relying on mutual coordination. Ref [18] proposed an optimal sequential dispatch strategy to mitigate the power rebound. In [19], the TCLs and other loads are coordinated to respond sequentially. Ref [20] mitigated the rebound effects by dispatching heat pumps with various response delays. However, these methods are diverging from the purpose of FFR due to the sacrifice of regulation potential, because not all accessible TCLs can be dispatched in the shortest period.

Considering the contradictions between response speed and regulation capacity, coordination of TCLs is transferred from the response period to the recovery period, so as to break the synchronizations of *on-off* cycling. In [21], a random recovery delay method is used to minimize long-term frequency oscillations. Ref [22] proposed resetting the states of TCLs to eliminate the power oscillations during the restoration period. Similarly, a distributed recovery method by modifying the temperature dead-bands of TCLs is presented in [9] to restrain the power rebound effects. However, the losses of thermal comforts are ignored without full consideration of regulation duration. Among these strategies, acceptable temperature ranges are adopted to evaluate thermal comforts, which are satisfied only if the temperature does not violate the preset ranges, e.g., the regulation periods of 15 minutes and 5 minutes are regarded as the same discomforts. Obviously, the inappropriate assumption will lead to severe discomforts for part of residents due to the longer duration time.

To participate in the FFR program, both the ability to fast response and the controllability of rebound power should be guaranteed simultaneously, and the thermal comforts of residents cannot also be ignored. To address these challenges, the FFR infrastructure based on cloud-edge collaboration is developed based on edge intelligent terminals (EIT), so as to dispatch massive non-aggregate TCLs to provide promising FFR with a controllable recovery process. The significant contributions are three-fold:

1) A novel FFR infrastructure with distributed frequency response and centralized recovery optimization of massive non-aggregate TCLs based on EITs is developed, aiming to guarantee the FFR requirements on fast response and controllable rebound effects simultaneously. Firstly, TCLs can be dispatched directly by EITs to realize fast response in 1 second according to local frequency measurements. Moreover, the computing advantages of EITs are also utilized to identify the time-varying operating parameters accurately which are challenging to obtain in practice.

2) The equivalent mirror recovery strategy (EMRS) on the edge side is proposed to restore all TCLs to their initial states at the given time instant, including operating states and temperatures. Not only can it mitigate the intrinsic power fluctuations effectively after the recovery process, but the computational burdens of recovery optimization are also shared by EITs. In addition, several criteria on rebound effects and thermal comforts are also proposed to quantify the tradeoff between power systems and residents during the recovery process.

3) Following the solving results of EMRS, the recovery optimization problem on the cloud side is established to determine the operating states of TCLs. Benefiting from the proposed dynamic clustering methods, large-scale and non-aggregate TCLs can be reduced to limited clusters, whose number is specific and only depends on the recovery duration and timesteps. Therefore, the curse of dimensionality is addressed effectively, and the optimization problem can be solved by the cloud control center within a short period.

The remainder of this paper is organized as follows. Section II introduces the modeling of TCLs and the specific FFR infrastructure based on EITs. The multistage processes of FFR are presented in section III to illustrate the detailed coordination between the cloud control center and EITs. The proposed cloud-edge collaboration for recovery optimization is given in Section IV. Section V introduces the simulation results of case studies. Finally, the main conclusions are summarized in Section VI.

II. FFR INFRASTRUCTURE OF TCLS BASED ON EITS

As one of the most significant flexible loads, TCLs possess the remarkable regulation potential to provide FFR for power systems. Fortunately, users are also motivated to participate in the FFR program by proactively sacrificing temporary electricity demands to prevent indistinctive and long-term load shedding when the frequency drops significantly.

A. The operating model of the individual TCL

When extreme weather occurs and there are severe electricity shortages, part of industry and commerce is required to suspend production so as to provide a sufficient supply of electricity for residents to withstand the intense heat in China. Under such circumstances, residential TCLs mainly are the accessible regulation capacity. Consequently, the fixed-frequency air conditioners (ACs) of residents are chosen to develop the typical operating model considering their extensive stock. Denoting $\mathbf{\Gamma}$ as the set of N_{AC} ACs. For the *i*-th AC, when it is operating with the set temperature $T_{set,i}$, the operating states are cyclic between *on* and *standby*, along with the indoor temperature varying in the range $[T_{in,i}^{-}, T_{in,i}^{+}]$ [23]. Taking the cooling mode as the example, the specific state of the *i*-th AC $s_i(t)$ can be determined by equation (1):

$$s_i(t)=\begin{cases}1 & T_{in,i}(t)>T_{in,i}^{+}\\ 0 & T_{in,i}(t)<T_{in,i}^{-}\\ s_i(t-\Delta t) & T_{in,i}^{-}\le T_{in,i}(t)\le T_{in,i}^{+}\end{cases} \tag{1}$$

where $s_i(t)$ is the binary variable, 0 and 1 indicate *standby* and *on* states, respectively. $T_{in,i}^{-}$ and $T_{in,i}^{+}$ represent the upper and lower temperature bands, respectively. $T_{in,i}(t)$ is the real-time indoor temperature of the room.

To evaluate the accessible regulation time and thermal comfort, it is necessary to calculate the detailed temperature variations during the FFR deployment. The first-order equivalent thermal parameter (ETP) model is widely used to describe temperature variations. The thermal dynamic transition process can be expressed as follows [24]:

$$C_i\frac{dT_{in,i}(t)}{dt}=\frac{T_{out}-T_{in,i}(t)}{R_i}-P_{AC,i}^{rated}\cdot EER_i \tag{2}$$

where $T_{in,i}(t)$ denotes the indoor temperature at the instant t. C_i and R_i are the equivalent heat capacity and heat resistance of the room, respectively. $P_{AC,i}^{rated}$ is the rated power of the AC. EER_i represents the energy efficiency ratio of the AC, within such a short period like the FFR process, EER_i can be regarded as a constant value.

B. The response infrastructure of FFR based on EITs

As illustrated in the Introduction, the aggregators make FFR more complicated in the view of both communications and coordination. Consequently, to improve the performance of FFR, a non-aggregate response infrastructure based on EITs is developed. As shown in Fig. 1, all ACs are equipped with EITs, whose demo has been designed and developed. EITs possess various functions, including local measurement, edge computation, flexible control of ACs, and bidirectional communication with the control center [25]. It should be noted that all data is solved and stored locally to protect privacy.

1) Local measurement

During the operating period, the EIT continues measuring the operating power and recording the duration time of *on* state $\tau_{on,i}^{ope}$ and *standby* state $\tau_{off,i}^{ope}$, respectively. Relative data are saved and analyzed within the EIT, thereby the privacy of residents can be protected effectively. Moreover, the communication burdens between the EIT and the control center can be alleviated dramatically, with the advantage of decreasing the risk caused by communication congestion and data losses.

Meanwhile, the EIT is metering the local frequency of power systems with a high resolution. Once a contingency occurs and the severe RoCoF is monitored, the FFR services provided by ACs will be activated instantaneously, instead of waiting for the response signal from aggregators or the control center like traditional response infrastructure.

2) Edge computation

There are such a large amount of ACs and buildings with heterogenous parameters, some physical parameters are very difficult to be obtained precisely, such as C_i and R_i. Benefit from the computational advantages, the specific parameters can be identified precisely by solving the equation (2) and updated timely with regard to the measured $\tau_{on,i}^{ope}$ and $\tau_{off,i}^{ope}$.

In addition, edge computation also contributes to breaking the obstacle of coordination between large-scale non-aggregate ACs. Firstly, the kinds of data generated with their operations and corresponding analysis can be solved locally, avoiding computational as well as communicational burdens caused by centralized processing. Besides, upon the FFR being activated, the EIT needs to be involved in the cloud-edge collaboration through finishing the EMRS solution. The specific duration time of *on* state $\tau_{on,i}^{REC}$ and *standby* state $\tau_{off,i}^{REC}$ will be determined, respectively, and be submitted to the cloud control center for further centralized optimization.

3) Flexible Control

The EIT is able to control the affiliated AC through several methods, including direct load control (DLC), state switches, and temperature adjustment. When the FFR is activated, the DLC method is adopted to guarantee the AC can be turned off within 1 second, so as to deliver FFR to power systems and avoid severe frequency drop.

During the FFR process, the EIT needs to keep the *off* state and provide continuous regulation capacity, earning more time for dispatching other regulation resources at a slower speed. In the recovery period, the EIT should switch the operating states of the AC and recover its initial state according to the centralized optimization results rather than its operating philosophy.

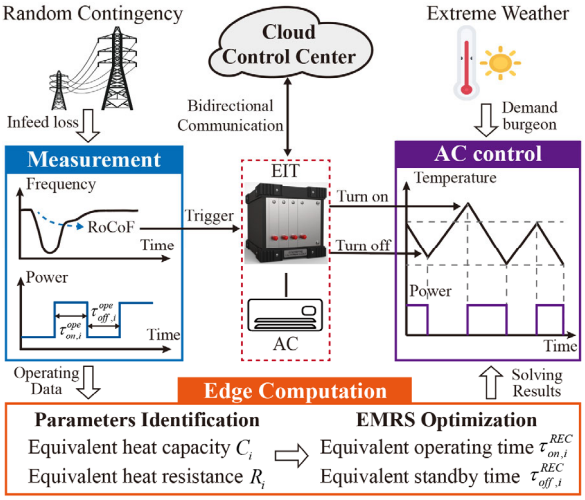


Fig. 1 The infrastructure of FFR provide by ACs based on EITs

Through the proposed response infrastructure, not only can non-aggregate ACs respond to the frequency variations quickly based on the distributed measurement and control, but also the rebound effects of the recovery process can be solved effectively by the cloud-edge collaboration.

III. The Specific Frequency Regulation Process with ACs and Generators

For receiving-end power systems, contingencies such as the transmission line failure will cause severe frequency drops. Once the RoCoF is monitored out-of-limit, the FFR process provided by ACs will be triggered to enhance the system inertia and compensate for the power loss, along with the PFR and SFR provided by conventional units, the detailed frequency regulation process and the cloud-edge collaboration are introduced in this section.

A. FFR process

The frequency deviation of the power system during the whole frequency regulation process is usually calculated by the first-order differential equation as follows [26]:

$$2H\frac{d\Delta f(t)}{dt}+D\cdot P_D\cdot\Delta f(t)=FR(t)-P_{loss} \tag{3}$$

where H denotes the inertia of the power system. $\Delta f(t)$ is the frequency deviation. D represents the load-damping rate. P_D is the sum of load demand, which can be considered as a fixed value during frequency regulation. P_{loss} is the sudden power loss. $FR(t)$ is the frequency regulation capacity provided by both ACs and generators at time t.

Assuming the contingency occurs at t_0 and the nominal operating power of ACs is P_{AC}^{ref}. When the frequency drop is monitored, it is complicated to figure out the specific power loss and dispatch response capacity within 1s. To take full use of the fast response advantages of EITs, all accessible ACs need to be involved in maximizing the provision of FFR capacity. On this basis, the RoCoF is chosen as the condition to activate the FFR to guarantee the fast and accurate response of ACs, the main reasons are as follows: 1) RoCoF is the earliest index that can be measured to reveal the severity of contingencies at the beginning, using RoCoF as the index can prevent misoperation from slight frequency fluctuations [27]. 2) Faster involvement of FFR can achieve smaller frequency deviations, so FFR services of ACs should be provided as fast as possible to maximize their contributions, [28]. 3) Other primary regulation resources such as generators are able to

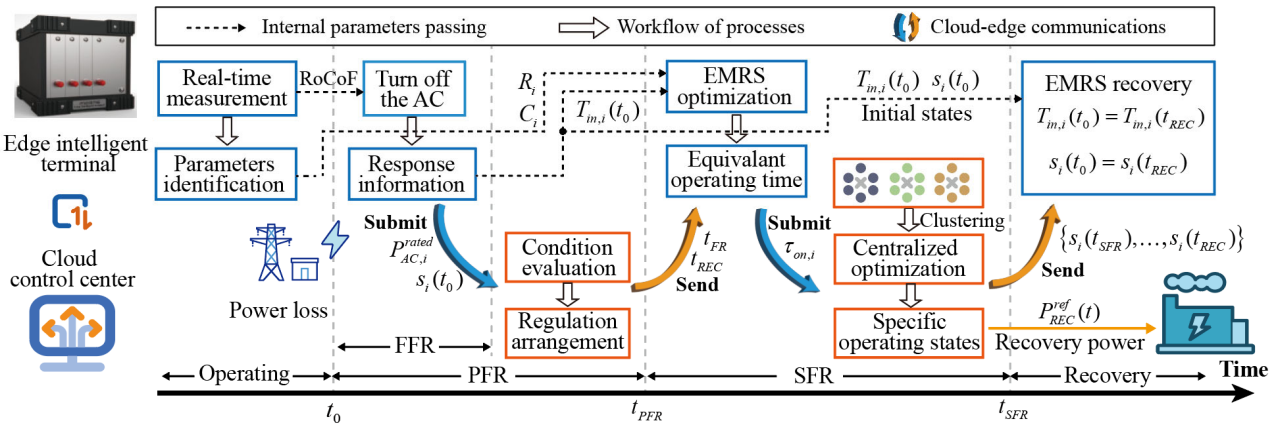


Fig. 2 The specific FFR processes and cloud-edge collaborations.

adjust their frequency regulation capacities with regard to frequency variations after FFR.

Once the measured RoCoF is beyond the σ_{RoCoF}, the EIT will use the DLC method to turn off the affiliated AC immediately.

$$\frac{d\Delta f(t)}{dt} > \sigma_{RoCoF} \tag{4}$$

Considering the different delays of power response, it is assumed that the frequency regulation capacity provided by ACs $FR_{AC}(t)$ is increasing linearly with time, and the complete response time is set as t_{res}^{AC}, which can be expressed as follows:

$$FR_{AC}(t) = \begin{cases} P_{AC}^{ref} \cdot \dfrac{t-t_0}{t_{res}^{AC}} & t \in (t_0, t_{res}^{AC}] \\ P_{AC}^{ref} & t \in (t_{res}^{AC}, t_{SFR}] \end{cases} \tag{5}$$

Apart from the control of the AC, the EIT also needs to record the current states, including the temperature $T_{in,i}(t_0)$ and the operating mode $s_i(t_0)$. Then, the specific response capacity $P_{AC,i}^{rated} s_i(t_0)$ should be submitted to the control center to help it determine the specific FFR duration time in the PFR process.

B. PFR process

The PFR process has the same start time instant as the FFR, where the conventional units are the main resources to provide frequency regulation. Assuming that the PFR provided by units is also increasing linearly, and the rising time is set as t_{res}^{g} [29]:

$$FR_{g}^{PFR}(t) = \begin{cases} P_{g}^{PFR} \cdot \dfrac{t-t_0}{t_{res}^{g}} & t \in (t_0, t_{res}^{g}] \\ P_{g}^{PFR} & t \in (t_{res}^{g}, t_{SFR}] \end{cases} \tag{6}$$

where $FR_{g}^{PFR}(t)$ is the PFR provided by generators at time t. P_{g}^{PFR} is the total PFR capacity. Following the engineering experience, the t_{res}^{AC} and t_{res}^{g} are set as 1s and 10s in this paper, respectively [30], [4].

Due to the load-shedding during the response period, the higher power demand and extra electricity consumption after deployment are unavoidable, which are named the power rebound and electricity rebound in this paper, respectively. If any recovery strategy is employed, severe power rebounds and fluctuations will appear and violate the FFR rules, even resulting in cascading failures.

To prevent the harmful effects of rebound, FFR rules require that the rebound power should be constrained as:

$$\Delta P_{AC}(t) \le P_{AC}^{ref} \times \alpha\%, t \in (t_{SFR}, t_{REC}) \tag{7}$$

$$\Delta P_{AC}(t) = P_{AC}(t) - P_{AC}^{ref} \tag{8}$$

where $\alpha\%$ is the acceptable rate of power rebound. t_{SFR} is the end time instant of FFR provided by ACs. t_{REC} represents the end time instant of the recovery process, where the operating power and states of ACs are restored to their nominal values.

Let $\tau_{FR} = t_{FR} - t_0$ and $\tau_{REC} = t_{REC} - t_{FR}$ denote the duration time of FFR and recovery process, respectively. τ_{FR} can be determined by the control center with regard to the power loss, available regulation capacity, operating statuses, and so on. The rebound phenomenon is related to the τ_{FR} and τ_{REC}, hence, the following problem is to determine the appropriate τ_{REC} without violence of FFR constraints.

When all ACs are operating without disturbances, they can be regarded as an equivalent AC which has stable power $P_{AC}(t)$ with indoor temperature $T_{in}^{agg}(t)$:

$$P_{AC}(t) = \sum_{i=1}^{N_{AC}} P_{AC,i}(t) \tag{9}$$

$$T_{in}^{agg}(t) = \frac{\sum_{i=1}^{N_{AC}} [T_{in,i}(t) P_{AC,i}^{rated}]}{\sum_{i=1}^{N_{AC}} P_{AC,i}^{rated}} \tag{10}$$

Similarly, the equivalent heat capacity C_{AC}^{agg}, the equivalent heat resistance R_{AC}^{agg}, and the equivalent energy efficiency ratio EER_{AC}^{agg} can be obtained through the similar weighted average. Based on the developed equivalent model and equation (2), the recovery duration time τ_{REC} can be expressed as follows.

$$\tau_{REC} \ge -R_{AC}^{agg} C_{AC}^{agg} \ln\left\{ \frac{T_{out} - (1+\alpha\%) P_{AC}^{ref} R_{AC}^{agg} EER_{AC}^{agg} - T_{in}^{agg}(t_0)}{T_{out} - (1+\alpha\%) P_{AC}^{ref} R_{AC}^{agg} EER_{AC}^{agg} - T_{in,agg}^{\max}} \right\} \tag{11}$$

$$T_{in,agg}^{\max} = T_{out} - [T_{out} - T_{in}^{agg}(t_0)] \exp\left(-\frac{\tau_{FR}}{R_{AC}^{agg} C_{AC}^{agg}}\right) \tag{12}$$

where $T_{in,agg}^{\max}$ is the equivalent maximum temperature at the end of the frequency regulation process.

It should be noted that equation (11) only determines the minimum duration of the recovery process with constant operating power. In practice, both the operating power and the rate of power rebound can be altered according to different situations. In this paper, the minimum value of τ_{REC} is chosen to minimize the thermal discomforts on residents. Besides, these works should be finished and τ_{REC} needs to be sent to EITs before the SFR process as shown in Fig. 2. Obviously, the original data with privacy is transformed into

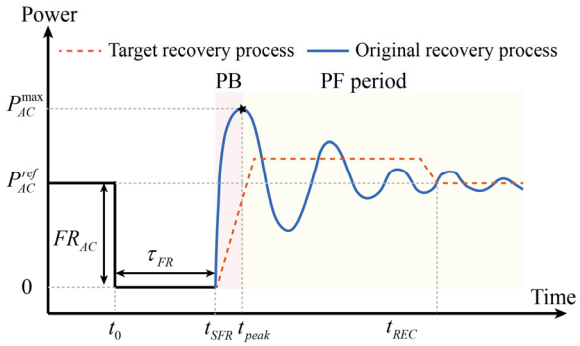


Fig. 3 The rebound effects of ACs with different recovery process

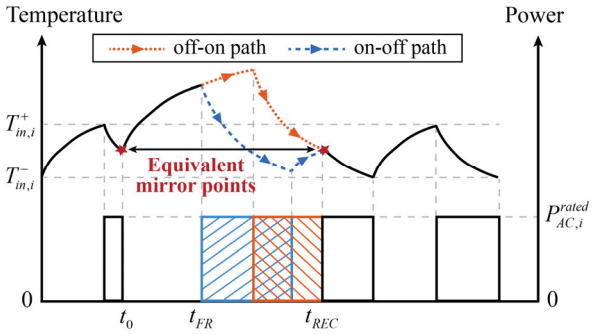


Fig. 4 The illustration of EMRS

indirect data for communication, and the data transmission only occurs during the FFR process, which are beneficial for protecting residents' privacy.

C. SFR process

After the frequency reaches the quasi-steady state at the time instant t_{SFR}, the generators start to provide SFR to eliminate frequency deviations. The SFR provided by generators is governed by auto generator control (AGC) according to the area control error (ACE) signal [31]:

$$ACE(t) = P_g^{PFR} + P_{AC}^{ref} + FR_g^{SFR}(t) - P_{loss} - 10 \cdot \gamma \cdot \Delta f(t) \quad (13)$$

where γ is the frequency bias of the power system. $FR_g^{SFR}(t)$ is the SFR provided by AGC at time t, which can be calculated by the following equation:

$$FR_g^{SFR}(t) = -\delta_P ACE(t) - \frac{1}{\delta_I} \int_{t_{SFR}}^{t} ACE(\tau) d\tau \quad (14)$$

where δ_P and δ_I denote the proportional and integral parameters of AGC, respectively.

According to (14), we can find that the $FR_g^{SFR}(t)$ is related to real-time ACE and the integration of ACE. The integration part aims to realize the zero-error frequency regulation and eliminate the effect caused by small fluctuations. Therefore, relying on the increasing SFR, the deviations of frequency can be eliminated at the end of SFR.

Meanwhile, ACs are also keeping to provide frequency regulation capacity during SFR. On the one hand, the frequency can be recovered more quickly and there is sufficient time to activate other regulation resources to manage the following recovery of ACs. On the other hand, both the cloud control center and EITs need to take SFR time to optimize coordinately the specific recovery processes of ACs, which will be illustrated in the following section.

IV. Evaluation and Optimization of the Recovery Process for Non-Aggregate ACs

In this section, the specific criteria from dimensions of power systems and residents are proposed to quantify multiple rebound effects of ACs recovery. On this basis, the detailed recovery optimization problem of non-aggregate ACs is solved efficiently based on cloud-edge collaborations.

A. The criteria of rebound effects caused by ACs

After providing FFR, the ACs need to be recovered to the nominal operations as soon as possible for residents' thirsty demands for comfort under extreme weather. Nevertheless, the intrinsic rebound effects of ACs also cannot be ignored. According to the different features, the recovery process of ACs can be divided into a power rebound (PR) period and a power fluctuations (PF) period, which are represented by light pink and yellow areas in Fig. 3, respectively.

The PR period starts at t_{SFR} when the ACs are permitted to restore then the total power profiles are increasing sharply due to the intense operating. After reaching the summit, the cumulated power starts fluctuating in the following PF period due to the synchronous *on-off* switches, as shown by the blue solid curve in Fig. 3. Additionally, the thermal discomforts of residents caused by regulation should also be evaluated. Therefore, several criteria are proposed as follows to evaluate the rebound effects of the recovery process.

1) Power ramp rate (PRR)

PRR aims to describe the growth rate of operating power at the beginning of recovery. From the view of power systems, *PRR* is expected to be limited to the feasible response speed of AGC, avoiding the second frequency drop, which may result in the off-line of RES and cause a more severe failure.

PRR is defined as the quotient between maximum rebound power and the rising time:

$$\underset{t_{FR} \to t_{peak}}{PRR} = \frac{P_{AC}(t_{peak})}{P_{AC}^{ref}(t_{peak} - t_{SFR})} \times 100\%, t \in (t_{SFR}, t_{REC}) \quad (15)$$

where t_{peak} is the time instant that the operating power of ACs reaches the peak value.

2) Maximum power rebound ratio (MPRR)

Apart from the *PRR* index, the *MPRR* is also a key criterion to mirror the rebound effects. Considering there are few available regulation capacities to utilize after the power loss under extreme weather within such a short period, significant power rebound will pose an immense challenge to the SFR provided by generators. To describe the severity of power rebound, the *MPRR* is developed as follows:

$$\underset{t_{peak}}{MPRR} = \frac{P_{AC}(t_{peak}) - P_{AC}^{ref}}{P_{AC}^{ref}} \times 100\% \quad (16)$$

3) Power fluctuation intensity (PFI)

The uniform starting destroys the original randomness of ACs' operations, thereby there exist cyclic fluctuations caused by the on-off switches of ACs. To accommodate the power fluctuations, the AGC resources have to alter and respond to the power deviations quickly and repeatedly. For the power system which is recovering from a severe contingency, regulation resources will be over-occupied persistently, causing the loss of regulation ability to withstand the possible cascading failures.

To evaluate the intensity of power fluctuations, the criterion *PFI* is established to describe the accumulated effects of power deviations on power systems:

$$\underset{t_{peak} \to t_{REC}}{PFI} = \sum_{t=t_{peak}}^{t_{REC}} \left| P_{AC}(t + \Delta t) - P_{AC}(t) \right| \quad (17)$$

4) Thermal discomforts (TD)

Residents are more sensitive to temperature variations under extreme weather, which are regarded as the same value

only if the indoor temperature is within the maximum acceptable ranges in previous researches. The thermal discomforts are related to both the temperature deviations and time duration. For the *i*-th AC, the RD_i can be calculated by:

$$RD_i \underset{t_0 \to t_{RES}}{=} \sum_{t=t_0}^{t_{RES}} [T_{in}^i(t) - T_{set,i}^+]\Delta t \quad if\ T_{in}^i(t) > T_{set,i}^+ \tag{18}$$

where $T_{in}^i(t)$ is the real-time indoor temperature of *i*-th AC. $T_{set,i}^+$ represents the upper boundary of the temperature dead-band. Δt indicates the timestep.

B. Edge-side optimization based on EMRS

Considering the optimization problem of massive ACs with heterogenous parameters is difficult to solve by the control center alone, the computational abilities of EITs should be appropriately utilized. On the other hand, the edge optimization of ACs is to the benefit of protecting the privacy of residents as well as reducing the occupation of valuable communication resources. Consequently, the EMRS is proposed to break the synchronous cycling switches and mitigate the power fluctuations of ACs as shown in Fig. 4, which aims to recover all ACs to their initial states at the given recovery period τ_{REC}.

$$\tau_{REC} = \tau_{on,i}^{REC} + \tau_{off,i}^{REC} \tag{19}$$

$$T_{in,i}(t_0) = T_{in,i}(t_{REC}) \tag{20}$$

where $\tau_{on,i}$ and $\tau_{off,i}$ represent the equivalent duration time of *on* state and *off* state of the *i*-th AC, respectively.

In order to guarantee all ACs can be recovered to their initial states at the given time instant, specific $\tau_{on,i}^{REC}$ and $\tau_{off,i}^{REC}$ needs to be solved by EITs. As shown in Fig. 4, there are two recovery paths respectively named *off-on* path and *on-off* path. The *off-on* path keeps *off* state in the beginning, thereby more thermal discomforts are generated with less electricity demand. On the contrary, the *on-off* path turns on ACs instantly with more energy consumption.

Taking a typical AC as an example, the corresponding parameters and specific results can be found in Section V and TABLE I, respectively. The initial indoor temperature is 25.50 ℃, then increases to 27.77 ℃ after 300s regulation. Obviously, the average indoor temperature of *off-on* path is higher than that of *on-off* path. According to equation (2), the cooling performance of ACs is inversely proportional to the deviations between outdoor and indoor temperatures. Therefore, the *on-off* path has longer operating time of about 759s, which is 22.8% longer than that of *off-on* path. Meanwhile, the thermal comforts of users in *on-off* path are better due to the more electricity consumption.

TABLE I
The numerical comparisons between two paths

	$T_{in}(t_0)$	$T_{in}(t_{FR})$	τ_{on}^{REC}	τ_{off}^{REC}	$T_{in}^{\max}$
off-on	25.50℃	27.77℃	618s	372s	30.07℃
on-off	25.50℃	27.77℃	759s	241s	27.77℃

In fact, there are several state switches during the recovery period, which will formulate a more balanced path between the *off-on* path and the *on-off* path. Therefore, the average value is chosen to approximately depict the equivalent duration time of *on* state. According to the ETP model in (2), the equation (19) and (20) can be transferred as follows:

$$\tau_{on,i}^{REC} = (\tau_{on,i}^{REC(+)} + \tau_{on,i}^{REC(-)})/2 \tag{21}$$

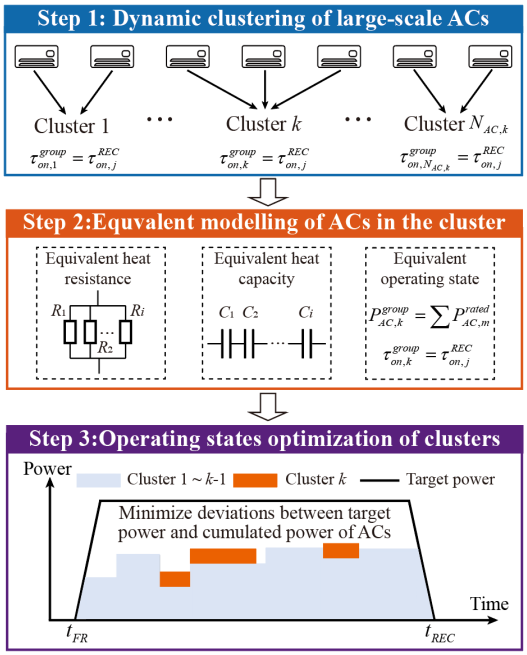


Fig. 5 Processes of recovery optimization at cloud side

$$\begin{cases} T_{in,i}^{\max} = T_{out} - [T_{out} - T_{in,i}(t_0)]\exp(-\dfrac{\tau_{FR} + \tau_{off,i}^{REC(+)}}{R_i C_i}) \\ T_{in,i}(t_0) = T_{out} - R_i P_{AC,i}^{rated} EER_i \\ \qquad -[T_{out} - R_i P_{AC,i}^{rated} EER_i - T_{in,i}^{\max}]\exp(-\dfrac{\tau_{on,i}^{REC(+)}}{R_i C_i}) \end{cases} \tag{22}$$

$$\begin{cases} T_{in,i}(t_0) = T_{out} - [T_{out} - T_{in,i}^{\min}]\exp(-\dfrac{\tau_{off,i}^{REC(-)}}{R_i C_i}) \\ T_{in,i}^{\min} = T_{out} - R_i P_{AC,i}^{rated} EER_i \\ \qquad -[T_{out} - R_i P_{AC,i}^{rated} EER_i - T_{in,i}(t_{SFR})]\exp(-\dfrac{\tau_{on,i}^{REC(-)}}{R_i C_i}) \end{cases} \tag{23}$$

where $\tau_{on,i}^{REC(+)}$ and $\tau_{off,i}^{REC(+)}$ are the equivalent duration time of *on* state and *off* state for *off-on* path, respectively. $\tau_{on,i}^{REC(-)}$ and $\tau_{off,i}^{REC(-)}$ are the equivalent duration time of *on* state and *off* state for the *on-off* path, respectively. $T_{in,i}^{\max}$ and $T_{in,i}^{\min}$ represent the maximum temperature and minimum temperature of the two paths at the turning point, respectively.

In addition, the EIT should guarantee the state at t_{REC} is equal to its initial state.

$$s_i(t_0) = s_i(t_{REC}) \tag{24}$$

Through solving the above problem, the EIT only needs to submit the $\tau_{on,i}^{REC}$ to the cloud control center for further centralized optimization.

C. Cloud-side optimization based on dynamic clustering

Through the optimization of EITs, power fluctuations can be mitigated entirely, due to all ACs are recovered to the initial states where they are operating steadily. As shown in Fig. 5, the remaining problem that the control center needs to solve is optimizing their detailed operating states in the recovery process, with the aim of limiting the power rebound within the acceptable range.

Firstly, the control center should determine the maximum acceptable values of $\overline{PRR}$ and $\overline{MPRR}$ with regard to accessible regulation resources in the PFR process. In this paper, to minimize the thermal discomforts, the $\overline{PRR}$ and

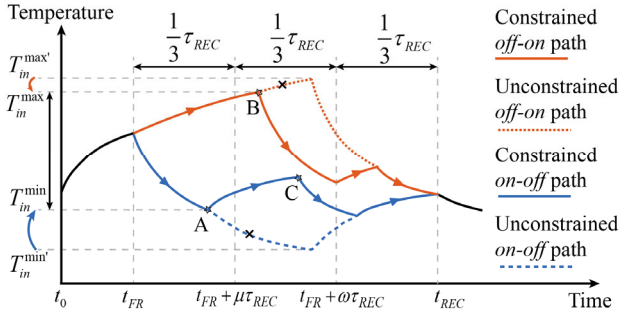


Fig. 6 Effects on recovery process caused by constraints of equivalent on/off time

$\overline{MPRR}$ are set as the maximum value and the duration of the recovery process is set as the minimum value according to FFR rules. Assuming the power of ACs is decreasing to the initial value with the same rate of change at the end of the recovery period, as shown by the dashed orange line in Fig. 3.

Consequently, the optimization problem aims to minimize the cumulated deviations between the actual power profiles and the optimal power profiles, through determining the specific operating states with the timestep Δt of all ACs during the whole recovery process. For the *i*-th AC, denoting the set of the optimization variables that indicate its operating states is expressed as $\mathbf{S_i} = \{s_i(t_{FR}), s_i(t_{FR}+\Delta t), \ldots, s_i(t_{REC})\}$, the number of total variables will be $N_{AC} \times (\tau_{REC} / \Delta t)$.

It should be noted that the value of Δt will determine the number of ACs groups, which will influence the solving efficiency and accuracy. Too small value of timestep will increase the communication burdens dramatically, while too large Δt will lead to unacceptable performance. Therefore, considering the practical response process of ACs, the state switches will take 15-60 seconds [32], the Δt should be capable of their dynamic process.

Obviously, although EITs are implemented to reduce the computational burden, the curse of dimensionality also cannot be avoided as the number of ACs increases. Therefore, a dynamic clustering method based on the $\tau_{on,i}^{REC}$ is proposed to overcome the obstacle. Compared with other clustering methods, the dynamic clustering method has two aspects of advantages. First, the clustering is only focusing on all online ACs according to the latest $\tau_{on,i}^{REC}$, the effects on clustering results caused by the offline of ACs can be prevented effectively. More importantly, the number of clusters is strictly restrained within the total timesteps $\tau_{REC} / \Delta t$, no matter how many ACs are involved and no matter what the distributions of parameters are.

Consequently, all ACs will be clustered $N_{AC,k}$ groups with regard to their equivalent operating time, let the group k be expressed as Ω_k with the same $\tau_{on,k}^{group}$:

$$\tau_{on,k}^{group} = \tau_{on,j}^{REC}, \forall j \in \Omega \tag{25}$$

$$P_{AC,k}^{group} = \sum P_{AC,m}^{rated}, \forall m \in \Omega_k \tag{26}$$

Benefiting from the clustering methods, the cloud control center only needs to solve the 0-1 integer programming problem as follows. The objective aims to minimize the bias between real power profiles and optimal power profiles.

$$Min \sum_{t=t_{FR}}^{t_{REC}} \left| \left[\sum_{k=1}^{N_{AC,k}} s_k(t) P_{AC,k}^{group} \right] - P_{REC}^{ref}(t) \right| \tag{27}$$

$$s.t. \sum_{t=t_{FR}}^{t_{REC}} s_k^{group}(t) = \tau_{on,k}^{group} \tag{28}$$

$$1-\beta \le \frac{\sum_{i=1}^{N} s_k^{group}(t) P_{AC,k}^{group}}{P_{REC}^{ref}(t)} \le 1+\beta, \forall t \in (t_{SFR}, t_{REC}) \tag{29}$$

$$\sum_{t}^{\min[t+\tau_{on}^{\min}, t_{REC}]} s_k(t) \ge \min[\tau_{on}^{\min}, (t_{REC}-t)][s_i(t) - s_i(t-\Delta t)] \tag{30}$$

$$\sum_{t}^{\min[t+\tau_{off}^{\min}, t_{REC}]} s_k(t) \le \min[\tau_{off}^{\min}, (t_{REC}-t)][1 + s_i(t) - s_i(t-\Delta t)] \tag{31}$$

$$\sum_{t=t_{FR}}^{t_{FR}+\mu\tau_{REC}} s_k(t) \le \tau_{on,k}^{group} \lambda\% \tag{32}$$

$$\sum_{t=t_{FR}}^{t_{FR}+\omega\tau_{REC}} s_k(t) \ge \tau_{on,k}^{group} \eta\% \tag{33}$$

Equation (28) aims to guarantee that the total duration times of *on* states are equal to the EMRS results submitted by EITs. Although the operating mode is different from that assumed in edge optimization, the equivalent cumulated operating time can prevent significant deviations. Equation (29) restrains the bias between the real power and the optimal power, β represents the allowable rate of power deviations. Equation (30) requests that the continuous operating time should be longer than the minimum value $\tau_{on}^{\min}$. Equation (31) requests that the continuous standby time should be longer than the minimum value $\tau_{off}^{\min}$.

Equation (32) and (33) are utilized to balance thermal comforts of residents. In terms of the *on-off* path, the AC is operating continuously to decrease the indoor temperature as mentioned before; not only is more energy consumed, but also other ACs' recovery processes are influenced. Considering the cumulated operating power is limited to relieve the harmful effects caused by the recovery process, the long-time operation of some ACs will make other ACs be kept off compulsively. To this end, Equation (32) is involved to ensure that the cumulated operating time during the period $[t_{FR}, t_{FR}+\mu\tau_{REC}]$ is less than $\tau_{on,k}^{group}\lambda\%$. As shown in Fig. 1, the **A** point is where the constraint works. Compared with the unconstrained path, the minimum temperature grows from $T_{in}^{\min'}$ to $T_{in}^{\min}$, which also increases the thermal comforts if the $T_{in}^{\min'}$ is beyond the comfortable temperature intervals.

For the *off-on* path, the AC is kept off state, so the temperature increases steadily and reaches $T_{in}^{\max'}$ finally. Obviously, this recovery path will cause severe thermal discomforts to residents. To prevent these bad effects, equation (33) is involved to require that the cumulated operating time during the period $[t_{SFR}, t_{SFR}+\omega\tau_{REC}]$ should be more than $\tau_{on,k}^{group}\eta\%$. As shown in Fig. 1, the AC turns to on state to decrease the indoor temperature at point **A**, aiming to satisfy the requirements of the shortest operating time. Therefore, the maximum temperature decreases from $T_{in}^{\max'}$ to $T_{in}^{\max}$. Besides, the constraint also plays a role in point **C** of *on-off* path.

V. Case Study

In this section, simulations on a receiving-end system in a typical developed Southern China city are conducted to validate the effectiveness of the proposed FFR infrastructure for enhancing frequency resilience.

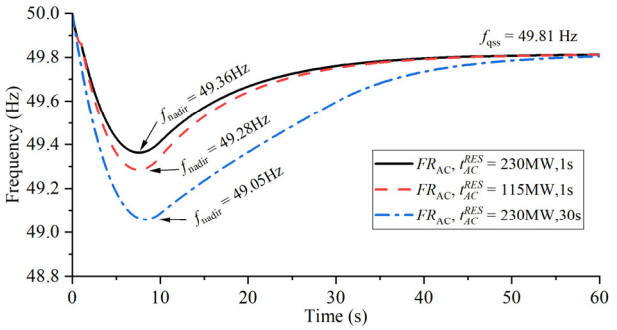


Fig. 7 Frequency response performances with different FFR

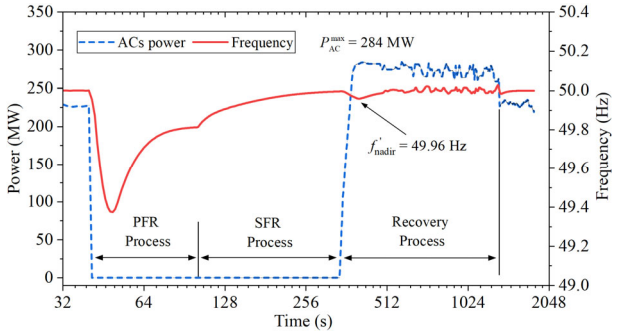


Fig. 8 The FFR process of ACs and corresponding frequency variations

A. Parameter setting

The total demand is set as 19000 MW on a summer day at 38 ℃. The nominal frequency is taken as 50 Hz. The inertia is set as 5 s [33]. The proportional coefficient δ_p and the integral coefficient δ_I of AGC are set as 0.05 and 200, respectively. γ is set as 1% of the peak load[1]. The duration of the PFR process and SFR process are set as 60 s and 240 s [1], respectively. The trigger threshold σ_{RoCoF} is set as 0.125 Hz/s. The response time of ACs t_{res}^{AC} and generators t_{res}^{g} are set as 1 s and 10 s, respectively [30], [4].

The parameters of ACs are shown in TABLE II. Considering that ACs account for about 40% of total demand and 14% of electricity is consumed by residents, there are a total number of 200,000 residential ACs for approximately providing up to 230 MW FFR capacity.

TABLE II
Physical parameters of ACs

Parameters	Descriptions	Values	Units
$A_{room,i}$	Square of room	$N(30,10^2)$ [18]	m^2
$P_{AC,i}^{rated}$	Rated power	$U(50,80)\times A_{room}$ [18]	kW
R_i	Heat resistance	$100 / A_{room}$ [9]	℃/kW
C_i	Heat capacity	$15A_{room}$ [9]	kJ/℃
$T_{set,i}$	Set temperature	$U(24,27)$ [34]	℃
EER_i	Energy efficiency ratio	$U(3.1,3.5)$ [34]	\

Normal distribution with the mean value μ and the standard deviation of σ is abbreviated to $N(\mu,\sigma^2)$; Uniform distribution with the minimum and maximum values of a and b, respectively, is abbreviated to $U(a,b)$.

During the recovery process, the maximum power ramp rate $\overline{PRR}$ and the maximum power rebound rate $\overline{MPRR}$ are set as 2%/s and 20%, respectively. Accordingly, the recovery period τ_{REC} can be solved with the value of 1000 s, the timestep Δt is set as 10 s. Besides, the allowable rate of power deviations is set as 5%, and the minimum operating time and standby time are set as 180 s uniformly. The corresponding simulations have been carried out using Python 3.7.9 and Cplex 12.1 on an Intel® i7 13700kf, 32 GB RAM desktop.

B. Analysis of contributions to the PFR performance with various FFR provided by ACs

Simulations with different response parameters are conducted to illustrate the contributions of the FFR provided by ACs, as shown in TABLE III. The sudden infeed loss and total PFR are set as 1000 MW and 930 MW, respectively [9]. The PFR is provided by generators, the FFR is provided by ACs and the corresponding response capacity as well as the complete response time are alterable. Considering that part of the PFR capacity is replaced by FFR, the remaining capacity will be reserved for the following recovery process of ACs.

As shown in Fig. 7, ACs in Case A have provided 230 MW FFR within 1 s, thereby the frequency security under sudden power loss is increased a lot, with a frequency nadir at 49.36 Hz. On this basis, Case B and Case C are conducted to evaluate the contributions of response capacity and response speed, respectively. In Case B, the FR_{AC} is set as half of that in Case A and the t_{res}^{AC} keeps the same. Due to the lack of sufficient FFR, the frequency nadir reaches 49.28 Hz. In Case C, the FR_{AC} is as same as that of Case A while the t_{AC} is set as 30 s, which can also reflect the current situations with higher penetration of RES and lower inertia. It can be seen from Fig. 7 that frequency resilience decreases significantly, the frequency nadir has reached 49.05 Hz, which may cause a more severe cascading failure and trigger UFLS (under frequency load shedding).

TABLE III
The results of PFR with different ACs parameters

	FR_{AC}	t_{res}^{AC}	FR_g^{PFR}	Frequency nadir
Case A	230 MW	1 s	700 MW	49.36 Hz
Case B	115 MW	1 s	815 MW	49.28 Hz
Case C	230 MW	30 s	700 MW	49.05 Hz

Obviously, the deviations of frequency facing the sudden loss are related to both the response capacity and the response speed. Compared with conventional units, FFR provided by ACs can contribute to enhancing the frequency resilience under extreme weather. Case B represents the aforementioned strategies, which divide ACs into several groups to eliminate rebounds via mutual coordination, such as sequential dispatch [18]. However, the sacrifice of FFR capacity also results in the reduction of frequency regulation performance. Case C is similar to methods that have longer response delay, e.g., centralized control. These regulation capacities can be regarded as supplementation for PFR or SFR, while the improvements for frequency resilience and virtual inertia are limited. Therefore, to maintain the frequency within the threshold, it is of great importance to arrange sufficient FFR capacity to defend the extreme weather.

C. FFR process provided by non-aggregate ACs

The complete FFR process of massive non-aggregate ACs is simulated based on the proposed infrastructure. As shown in Fig. 8, all ACs are operating normally with the power of about 230 MW in the beginning. When the power system is suffering from an infeed loss at $t_0 = 40\,s$, ACs are triggered to provide 230 MW FFR capacity. Due to the fast provisions of FFR, the nadir of frequency is limited to 49.36 Hz. Then, the frequency is increasing gradually with the progress of PFR provided by conventional units. At the end of the PFR process, the frequency reaches the quasi-state at 49.81 Hz.

When the ACs are providing FFR, recovery optimization is being carried out at the same time. Through the cloud-edge collaboration, the specific operating states of all ACs are determined by the control center within the SFR process, on

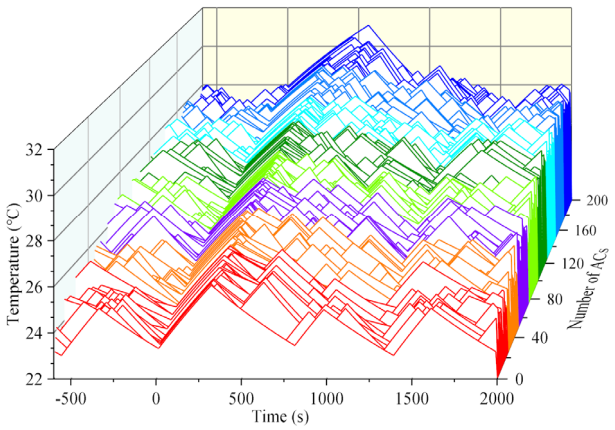


Fig. 9 The temperature variations of 200 ACs for the provision of FFR.

the basis of EMRS implemented in EITs and the dynamic clustering method adopted in the cloud control center.

After providing the FFR for 300 s, ACs are guided to recover to their nominal states with regard to the optimization results. As shown in Fig. 8, the cumulated power of ACs starts to grow to the summit $P_{AC}^{\max} = 284\,\text{MW}$ within the maximum power ramp rated $\overline{PRR} = 2\%/\text{s}$. Meanwhile, the power gaps caused by the withdrawal of ACs are filled by the generators which should have been utilized in PFR. Benefiting from the sufficient dispatching time and accurate power profiles of the recovery process, generators can track the power variations effectively, so only minor frequency deviations are caused with a second frequency nadir at 49.96 Hz.

With the progress of recovery, ACs are operating with relatively stable power within the allowable range. Fig. 9 demonstrates the detailed temperature variations of 200 ACs. It can be seen that almost all maximum temperature rises are limited within 3 ℃, benefiting from the involved proactively constraints (32) and (33). Besides, constraints (30) and (31) also worked effectively so frequent *on-off* switches are avoided. After the recovery process, all ACs are recovered to their initial states through the proposed EMRS, where they are operating without significant oscillations.

Firstly, the performance of the proposed EMRS is analyzed through randomly chosen 2000 ACs. As shown in Fig. 10, the red solid line represents the initial temperatures of ACs after being sorted, and the surrounding pink areas are the temperature errors of EMRS. Generally, most errors are restrained within ± 0.3 ℃ and possess an average value of 0.025 ℃. Besides, the temperature errors distribute relatively uniformly, which means that ACs are recovered to another stable status near the initial status, thereby no fluctuations occur after the recovery process.

The light blue line represents the corresponding maximum temperature, and the dark blue dashed line is the linear fit of the maximum temperature. It can be seen that the maximum temperature rise is decreasing with the growth of the initial temperature, e.g., 3.25 ℃ rise for the lowest initial temperature while 2.01 ℃ for the highest initial temperature. Therefore, the thermal comforts are more balanced in the proposed method.

On this basis, the relationships between maximum temperature rise and residents' discomforts are further analyzed as shown in Fig. 11. With the growth of discomforts, the colors of points are changing from dark blue to dark red. Residents' discomforts were mainly distributed at the relatively lower levels, with a mean value of 9.91 (℃·min) and a medium value of 8.39 (℃·min), respectively. More

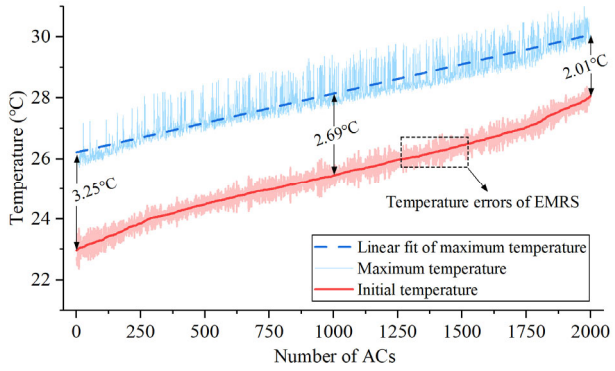


Fig. 10 Performance of the proposed EMRS

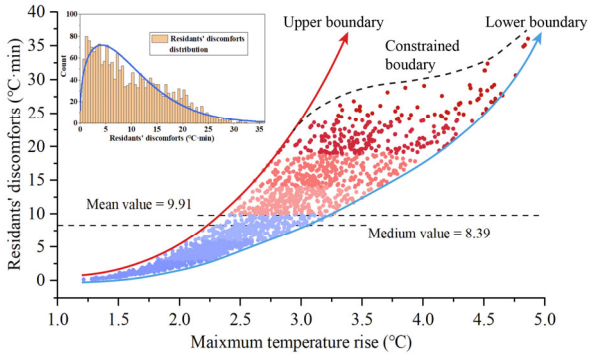


Fig. 11 Relationships between maximum temperature rise and thermal discomforts

specifically, the detailed distribution of *TD* is also shown at the top left corner, which indicates that most residents possess relatively comfortable experiences.

Considering that the temperature rise leads to discomforts in both up & down temperature variation processes, the *TD* is approximately proportional to the square of the maximum temperature rise in accordance with the definition of *TD*. Due to the variety of thermal parameters of ACs and rooms, the increasing trends are slightly different, but all points are constrained within the upper boundary and lower boundary, which are represented by the red line and blue line, respectively. It can be seen that the varying intervals continue to expand with the maximum temperature rise. When the maximum temperature rise reaches 3 ℃, it is worth noting that the thermal discomforts are restrained and the varying intervals are narrowed steadily in the direction of the lower boundary. Therefore, these dots are mainly located in the lower areas, which proves the effectiveness of the constraints (32) and (33). Moreover, these constraints have more noticeable effects on the residents with higher temperature rise, since the points are nearly restricted at the lower boundary after the temperature rise is beyond 4.5 ℃. Even though the differences in ACs and rooms will lead to various thermal discomforts during the same process, the proposed method contributes to protecting these residents with poor envelopes from severe thermal discomforts.

D. Performance comparisons between different recovery strategies

The recovery performances are evaluated among the free recovery method, sequential delay recovery method, and the proposed recovery method. The specific recovery processes are illustrated in Fig. 12, and the criteria for both the power system and residents are shown in TABLE IV and TABLE V, respectively. $T_{rise}^{\max}$, $T_{rise}^{\min}$, and $\overline{T_{rise}}$ indicate the maximum, minimum, and average temperature rises, respectively. $TD^{\max}$,

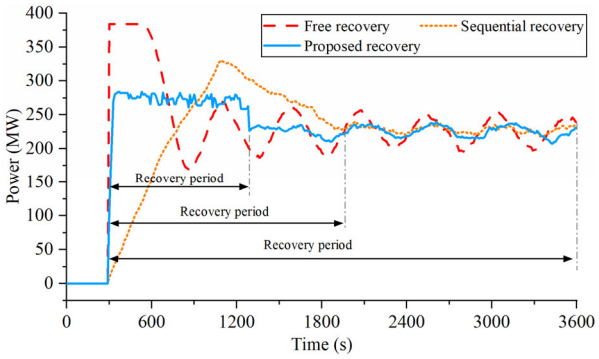


Fig. 12 Specific recovery processes of different methods

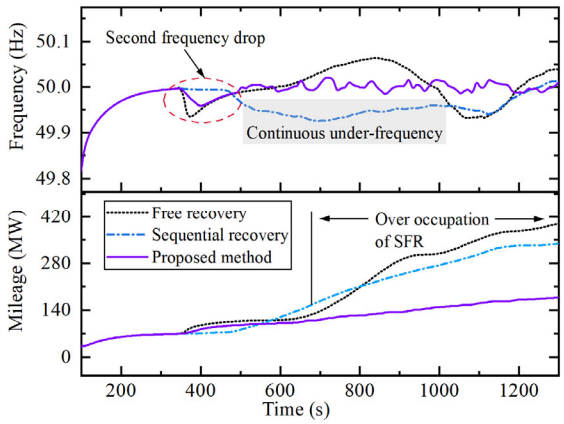


Fig. 13 Comparisons of frequency variations and SFR mileages with different recovery method

$TD^{\min}$, and $\overline{TD}$ are the maximum, minimum, and average thermal discomforts, respectively.

TABLE IV
The criteria on power system dimension of different recovery methods

	PRR	*MPRR*	*PFI*
Free recovery	17.08 %/s	70.79 %	1087.2 MW
Sequential recovery	0.17 %/s	42.68 %	157.3 MW
Proposed recovery	1.78 %/s	24.82 %	412.2 MW

TABLE V
The criteria on residents' dimension of different recovery methods

	$T_{rise}^{\max}$	$T_{rise}^{\min}$	$\overline{T_{rise}}$	$TD^{\max}$	$TD^{\min}$	$\overline{TD}$
Unit		℃			(℃·min)	
Free recovery	3.52	1.11	2.27	18.97	0.03	4.62
Sequential recovery	8.36	1.27	4.52	86.79	0.17	32.99
Proposed recovery	4.86	1.21	2.66	36.03	0.18	9.91

Free recovery means that any method is not taken to eliminate the rebound effects, consequently, the power is increasing sharply with 17.08 % *PRR* and causes a severe power rebound with 70.79 % *MPRR*. There is no doubt that the free recovery is likely to lead to a second frequency drop and cause a cascading failure, especially since the power system suffered a severe contingency several minutes ago. Besides, there are also striking power fluctuations. The *PFI* can be regarded as the utilization of AGC resources, according to its definition which is similar to the mileage. On the contrary, the remarkable sacrifice of power system stability earns lower temperature rises and thermal discomforts for residents. It is worth noting that the maximum temperature rise reaches 3.52 ℃ even though all ACs are recovered instantly. Therefore, the differences in physical parameters can never be ignored, especially for those buildings with poor heat capacities and heat resistances.

The sequential delay recovery limits the power rebounds and fluctuations via turning on the ACs sequentially, which represents the interests of power. The *PRR* has promising performance due to the uniformly distributed recovery processes of ACs. However, there are also non-negligible shortcomings. Firstly, the *MPRR* is difficult to evaluate and control, which still reaches 42.68 % with such a slower *PRR*. For clarity, the reduction of *MPRR* is realized by prolonging the ranges of sequential delay, which will further exacerbate the thermal discomforts of residents. It can be seen that the maximum temperature and average maximum temperature reach 8.36 ℃ and 4.52 ℃, respectively. There is no doubt that the significant temperature increase will affect the FFR participation of users.

Compared with free recovery and sequential delay recovery, the proposed recovery method aims to balance the demands of power systems and residents. Firstly, the *PRR* and *MPRR* are involved in the optimization model, thus these criteria can be satisfied strictly and altered according to sufficient situations. Especially for the critical criterion *MPRR*, the power deficit is likely to cause more severe accidents under the situation lacking sufficient regulation capacity. On the other hand, thermal discomfort is also taken into consideration. As a result, the $T_{rise}^{\max}$ and $TD^{\max}$ are limited to lower values compared with sequential delay recovery, which contributes to preventing bad experiences for some residents. Besides, the proposed method can control the cumulated power, thereby the criterion of power ramp rate can be easily optimized, referring to that of sequential recovery only if the power system has relevant demand.

Moreover, the specific effects on frequency variations and SFR mileages are analyzed in Fig. 13. At the beginning of the SFR process, the frequency is restored gradually to its nominal value with the growth of SFR capacity. At 350 s, the frequency reaches 49.99 Hz and the ACs are permitted to recover. In terms of the free recovery method, the frequency suffers a severe drop of about 0.062 Hz due to the limitations of SFR response speed. For the sequential recovery strategy, there are very slight and slow frequency drops, because the ramp rate of aggregated power of ACs is much lower than that of the free method, which can also be proved by the increasing speed of the SFR mileage. The proposed method aims to restrain the power rebound effects without losing too much thermal comforts, hence, the ramp rate of ACs' power is relatively smaller and the frequency drop is limited to 0.039 Hz.

During the time between from 500s to 1000s, there are significant frequency differences between the 3 recovery methods. After suffering from the sharp power rebound, the aggregated operating power of the free method turns to decrease due to the disadvantages named periodical fluctuations, thereby the over-frequency problem is appearing. On the contrary, the sequential strategy leads to continuous under-frequency problems during the period. Although the ramp rate is better than the other methods, the steady growth of ACs' operating power poses huge challenges to SFR performance. It is worth noting that the maximum SFR capacity is not limited in the case study, otherwise more severe frequency drop will emerge if the maximum SFR mismatches the rebound power under extreme weather conditions. Fortunately, the proposed method keeps the frequency more stable with some fluctuations around the nominal value, which can also be regarded as normal oscillations when power systems are operating.

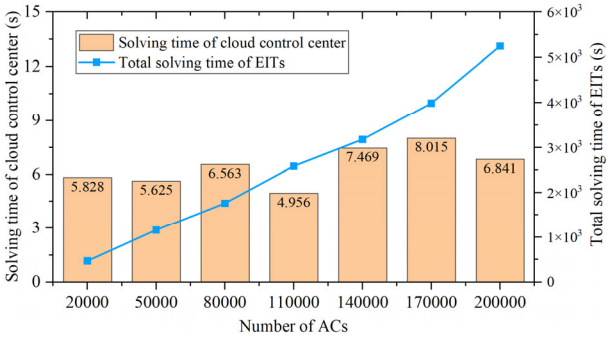


Fig. 14 Computational performance of cloud control center and EITs

Besides, the proposed method achieves much better performance on SFR mileage, which reveals the usage of SFR because both the up and down regulations contribute to the mileage cumulations. The SFR mileage of the proposed method is 196.5 MW, while those of the free method and sequential recovery are 415.8 MW and 365.9 MW, respectively. Apart from the costs increase caused by the large usage of SFR, the over-occupation of SFR is likely to result in severe accidents without sufficient SFR capacity, considering the dual effects of extreme weather and lack of regulation capacity.

Therefore, the proposed method realizes much better balance on both power systems and residents, the proposed method also performs better on the critical criterion *MPRR* and makes it constrained strictly within the permitted range.

E. Computational and communication performance of EITs

To prove the feasibility of real-time optimization, the specific computational and communication performance are analyzed in this subsection. As shown in Fig. 14, the blue points and orange columns are the total solving time of EITs and the solving time of the cloud control center, respectively. It should be noted that the solving time of EITs is the sum of all ACs due to the simulations are finished sequentially. In fact, the EMRS of all ACs are processed by EITs simultaneously, and the average solving time of EMRS is around 0.022 seconds. Then, benefiting from the proposed dynamic clustering method, large-scale ACs can be divided into tens of groups. Hence, the solving time of the cloud control center is relatively stable with an average value of 6.471 seconds, no matter how many ACs are involved. Both of the computational performances of EITs and the cloud control center prove that the proposed FFR infrastructure can be implemented widely and possess extraordinary expansibility.

Moreover, considering the time costs of several times communications, even if the time delay of communication is set as 10 seconds, which is twice as long as that used in [35], the real-time optimization can also be guaranteed. For clarity, in the 60 s PFR process, twice communications between EITs and the control center only account for 1/3 time, while the determinations of FFR and recovery durations can be finished within 1 second. In the 240 s SFR process, the time becomes more abundant, twice communications and solving time of EITs as well as the control center take no more than 30 s. Consequently, the developed FFR infrastructure and proposed recovery optimization not only can satisfy the real-time requirements, but also reserve abundant time to handle kinds of possible uncertainties.

Apart from simulation analysis, the initial demo of the EIT has been designed and developed to validate the possibility of application, which is shown in Fig. 15. The EIT mainly possesses function modules, named computation module,

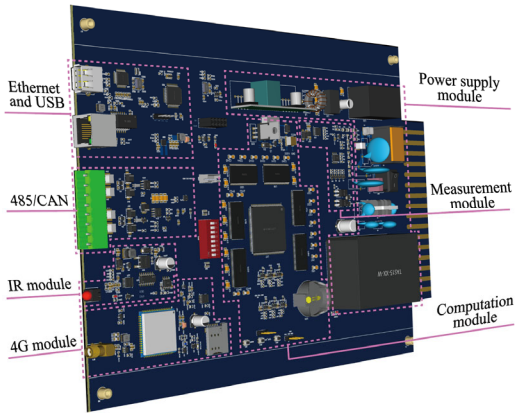


Fig. 15 The circuit diagram of the EIT

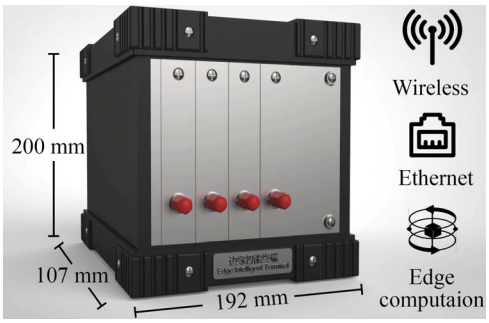


Fig. 16 Specific parameters of EIT's overall dimensions

infrared ray (IR) module, 485/CAN module, 4G module, measurement module, and power supply module.

Computation module is the core to achieve edge computation, Rockchip RK3399 is chosen as the CPU considering the balance of performance and economy. Even though its computational performance is much lower than the desktop CPU, the average solving time of 0.46 seconds still can satisfy the solving speed at the edge side.

In order to realize remote communications with the control center, the 4G module, and Ethernet are equipped with the EIT to provide more flexible networking modes, which can be determined according to practical configuration conditions. Besides, the communication performance is analyzed practically at a distance of 9.3 km based on 4G networks, results indicate that the single communication delay is about 0.15 seconds. It should be noted that the 5G module can also be upgraded to satisfy better communication performance.

The 485/CAN module and IR module are used to communicate or control with on-site flexible loads, which are compatible with most equipment, such as TCLs, EVs, batteries, and so on. Consequently, the EITs are designed to play the role of home energy management center instead of a dedicated device for controlling TCLs. The specific parameters of overall dimensions are 192*107*200 mm, as shown in Fig. 16.

Consequently, the proposed FFR scheme can satisfy both the computational and communication requirements.

VI. CONCLUSION

This paper proposed a novel regulating infrastructure and the corresponding recovery strategy, to utilize large-scale ACs to provide high-quality FFR services for power systems. Based on the distributed frequency response of EITS, all ACs can be regulated within 1 second. Furthermore, recovery optimization utilizing cloud-edge collaboration is adopted to satisfy the requirements of the recovery process with regard to FFR rules. Specifically, an EMRS is implemented in the

EIT to mitigate power fluctuations by restoring all ACs to their initial states at the same time. The cloud control center finishes the operating optimization with the aim of realizing the flexible controllability of rebound power, and the computational burdens caused by massive ACs are solved effectively by the proposed dynamic clustering method.

In the near future, the proposed functions of EITs can be integrated into smart meters, so as to utilize existing ACs to increase the virtual inertia and enhance the frequency resilience of power systems. Regarding the future works and challenges: 1) the integration of lightweight artificial intelligence into EITs should be further explored to enhance intelligent operation and automatic response capabilities; 2) the security and reliability of EITs need to be improved to protect against potential cyber-attacks.

REFERENCES

[1] Y. Yang, J. C.-H. Peng, C. Ye, Z.-S. Ye, and Y. Ding, "A Criterion and Stochastic Unit Commitment Towards Frequency Resilience of Power Systems," *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 640–652, Jan. 2022, doi: 10.1109/TPWRS.2021.3095180.

[2] Y. Cheng, N. Zhang, D. S. Kirschen, W. Huang, and C. Kang, "Planning multiple energy systems for low-carbon districts with high penetration of renewable energy: An empirical study in China," *Appl. Energy*, vol. 261, p. 114390, Mar. 2020, doi: 10.1016/j.apenergy.2019.114390.

[3] G. Lalor, A. Mullane, and M. O'Malley, "Frequency control and wind turbine technologies," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1905–1913, Nov. 2005, doi: 10.1109/TPWRS.2005.857393.

[4] V. Trovato, I. M. Sanz, B. Chaudhuri, and G. Strbac, "Advanced Control of Thermostatic Loads for Rapid Frequency Response in Great Britain," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2106–2117, May 2017, doi: 10.1109/TPWRS.2016.2604044.

[5] Y. Wang, C. Chen, J. Wang, and R. Baldick, "Research on Resilience of Power Systems Under Natural Disasters—A Review," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1604–1613, Mar. 2016, doi: 10.1109/TPWRS.2015.2429656.

[6] H. Mortaji, S. H. Ow, M. Moghavvemi, and H. A. F. Almurib, "Load Shedding and Smart-Direct Load Control Using Internet of Things in Smart Grid Demand Response Management," *IEEE Trans. Ind. Appl.*, vol. 53, no. 6, pp. 5155–5163, Nov. 2017, doi: 10.1109/TIA.2017.2740832.

[7] M. Aunedi, P.-A. Kountouriotis, J. E. O. Calderon, D. Angeli, and G. Strbac, "Economic and Environmental Benefits of Dynamic Demand in Providing Frequency Regulation," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 2036–2048, 2013, doi: 10.1109/TSG.2013.2258047.

[8] L. Meng *et al.*, "Fast Frequency Response From Energy Storage Systems—A Review of Grid Standards, Projects and Technical Issues," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1566–1581, Mar. 2020, doi: 10.1109/TSG.2019.2940173.

[9] T. Jiang, P. Ju, C. Wang, H. Li, and J. Liu, "Coordinated Control of Air-Conditioning Loads for System Frequency Regulation," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 548–560, Jan. 2021, doi: 10.1109/TSG.2020.3022010.

[10] H. Hui, Y. Chen, S. Yang, H. Zhang, and T. Jiang, "Coordination control of distributed generators and load resources for frequency restoration in isolated urban microgrids," *Appl. Energy*, vol. 327, p. 120116, Dec. 2022, doi: 10.1016/j.apenergy.2022.120116.

[11] Y. Lin, P. Barooah, S. Meyn, and T. Middelkoop, "Experimental Evaluation of Frequency Regulation From Commercial Building HVAC Systems," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 776–783, Mar. 2015, doi: 10.1109/TSG.2014.2381596.

[12] H. Hui, Y. Ding, and M. Zheng, "Equivalent Modeling of Inverter Air Conditioners for Providing Frequency Regulation Service," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1413–1423, Feb. 2019, doi: 10.1109/TIE.2018.2831192.

[13] M. Waseem, Z. Lin, Y. Ding, F. Wen, S. Liu, and I. Palu, "Technologies and Practical Implementations of Air-conditioner Based Demand Response," *J. Mod. Power Syst. Clean Energy*, vol. 9, no. 6, pp. 1395–1413, Nov. 2021, doi: 10.35833/MPCE.2019.000449.

[14] G. De Carne, G. Buticchi, M. Liserre, and C. Vournas, "Real-Time Primary Frequency Regulation Using Load Power Control by Smart Transformers," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5630–5639, Sep. 2019, doi: 10.1109/TSG.2018.2888687.

[15] X. Wu, J. He, Y. Xu, J. Lu, N. Lu, and X. Wang, "Hierarchical Control of Residential HVAC Units for Primary Frequency Regulation," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3844–3856, Jul. 2018, doi: 10.1109/TSG.2017.2766880.

[16] D. Zhang *et al.*, "Multi-Objective Control of Residential HVAC Loads for Balancing the User's Comfort With the Frequency Regulation Performance," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3546–3557, Sep. 2022, doi: 10.1109/TSG.2022.3171847.

[17] ERCOT, "Overview of Renewables in the ERCOT System." [Online]. Available: https://integrationworkshops.org/winddublin/wp-content/uploads/sites/18/2018/11/1_3_ERCOT_presentation_Julia_Matevosyan.pdf

[18] W. Cui *et al.*, "Evaluation and Sequential Dispatch of Operating Reserve Provided by Air Conditioners Considering Lead–Lag Rebound Effect," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6935–6950, Nov. 2018, doi: 10.1109/TPWRS.2018.2846270.

[19] X. Zhou *et al.*, "Exploiting Integrated Demand Response for Operating Reserve Provision Considering Rebound Effects," *IEEE Access*, vol. 10, pp. 15151–15162, 2022, doi: 10.1109/ACCESS.2022.3148398.

[20] E. Georges, B. Cornélusse, D. Ernst, V. Lemort, and S. Mathieu, "Residential heat pump as flexible load for direct control service with parametrized duration and rebound effect," *Appl. Energy*, vol. 187, pp. 140–153, Feb. 2017, doi: 10.1016/j.apenergy.2016.11.012.

[21] K. Dehghanpour and S. Afsharnia, "Designing a novel demand side regulation algorithm to participate in frequency control using iterated mappings," *IET Gener. Transm. Distrib.*, vol. 8, no. 10, pp. 1687–1699, 2014, doi: 10.1049/iet-gtd.2013.0701.

[22] X. Shi, J. Yu, J. Cao, and T. Yong, "Decentralised frequency-based control of a population of heterogeneous ACs without power oscillations," *IET Gener. Transm. Distrib.*, vol. 12, no. 10, pp. 2471–2479, May 2018, doi: 10.1049/iet-gtd.2017.1213.

[23] W. Zhang, J. Lian, C.-Y. Chang, and K. Kalsi, "Aggregated Modeling and Control of Air Conditioning Loads for Demand Response," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4655–4664, 2013, doi: 10.1109/TPWRS.2013.2266121.

[24] M. Song, C. Gao, H. Yan, and J. Yang, "Thermal Battery Modeling of Inverter Air Conditioning for Demand Response," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5522–5534, 2018, doi: 10.1109/TSG.2017.2689820.

[25] A. Delavari and I. Kamwa, "Sparse and Resilient Hierarchical Direct Load Control for Primary Frequency Response Improvement and Inter-Area Oscillations Damping," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5309–5318, Sep. 2018, doi: 10.1109/TPWRS.2018.2795462.

[26] L. L. Grigsby, *Power System Stability and Control*. CRC Press, 2017.

[27] E. Ela, V. Gevorgian, A. Tuohy, B. Kirby, M. Milligan, and M. O'Malley, "Market Designs for the Primary Frequency Response Ancillary Service—Part I: Motivation and Design," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 421–431, Jan. 2014, doi: 10.1109/TPWRS.2013.2264942.

[28] Statnett, "future-system-inertia." [Online]. Available: https://www.statnett.no/globalassets/for-aktorer-i-kraftsystemet/utvikling-av-kraftsystemet/nordisk-frekvensstabilitet/future-system-inertia-phase-1.pdf

[29] L. Badesa, F. Teng, and G. Strbac, "Optimal Portfolio of Distinct Frequency Response Services in Low-Inertia Systems," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4459–4469, 2020, doi: 10.1109/TPWRS.2020.2997194.

[30] F. Teng, V. Trovato, and G. Strbac, "Stochastic Scheduling With Inertia-Dependent Fast Frequency Response Requirements," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1557–1566, Mar. 2016, doi: 10.1109/TPWRS.2015.2434837.

[31] N. Jaleeli, L. S. VanSlyck, D. N. Ewart, L. H. Fink, and A. G. Hoffmann, "Understanding automatic generation control," *IEEE Trans. Power Syst.*, vol. 7, no. 3, pp. 1106–1122, Aug. 1992, doi: 10.1109/59.207324.

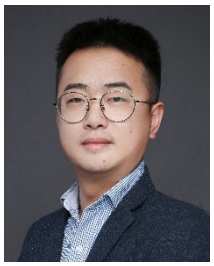
[32] H. Hui, Y. Ding, and M. Zheng, "Equivalent Modeling of Inverter Air Conditioners for Providing Frequency Regulation Service," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1413–1423, Feb. 2019, doi: 10.1109/TIE.2018.2831192.

[33] L. Badesa, F. Teng, and G. Strbac, "Simultaneous Scheduling of Multiple Frequency Services in Stochastic Unit Commitment," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 3858–3868, Sep. 2019, doi: 10.1109/TPWRS.2019.2905037.

[34] 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, doi: 10.1016/j.apenergy.2016.12.004.

[35] Ş. Sönmez, S. Ayasun, and C. O. Nwankpa, "An Exact Method for Computing Delay Margin for Stability of Load Frequency Control Systems With Constant Communication Delays," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 370–377, Jan. 2016, doi: 10.1109/TPWRS.2015.2403865.

Taoyi Qi received the B.E. and M.E. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2020 and 2023, respectively. He is currently pursuing the Ph.D degree with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao, China. His research interests mainly focus on digital twins, demand response, intelligent buildings, and deep reinforcement learning.



Chengjin Ye (Member, IEEE) received the B.E. and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2010 and 2015, respectively. From 2015 to 2017, he served as a Distribution System Engineer with the Economics Institute, State Grid Zhejiang Electric Power Company Ltd., Hangzhou, China. From 2017 to 2019, he was an Assistant Research Fellow with the College of Electrical Engineering, Zhejiang University. Since 2020, he has been a Tenure-track Professor there. His research interests mainly include resilience enhancement of power grids and integrated energy systems, as well as market mechanism and control strategy towards the integration of demand resources into power system operation.

Yuming Zhao received the B.S. and Ph.D. degrees from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 2001 and 2006, respectively. He is currently a Senior Engineer (Professor Level) with Shenzhen Power Supply Company Ltd., Shenzhen, China. His main research interests include the DC distribution power grid.

Hongxun Hui (Member, IEEE) received the B.E. and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2015 and 2020, respectively. From 2018 to 2019, he was a Visiting Scholar with the Advanced Research Institute, Virginia Tech, Blacksburg, VA, USA, and CURENT Center, University of Tennessee, Knoxville, TN, USA. He is currently an Assistant Professor with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao, China. His research interests include optimization and control of power system, demand response, and Internet of Things technologies for smart energy.