Evaluation and Sequential Dispatch of Operating Reserve Provided by Air Conditioners Considering Lead–Lag Rebound Effect

Wenqi Cui, Yi Ding, *Member, IEEE*, Hongxun Hui, *Student Member, IEEE*, Zhenzhi Lin, Pengwei Du, *Senior Member, IEEE*, Yonghua Song, *Fellow, IEEE*, and Changzheng Shao, *Student Member, IEEE*

***Abstract*—Air conditioners (ACs) are widely considered as good candidates to provide operating reserve. Demand response rebound, i.e., the rebound peak of aggregate power, may exist when ACs are controlled by changing the set point temperature. The rebound peak during the recovery period, named lag rebound, may cause significantly higher demand than that prior to the reserve deployment event. The rebound peak during the reserve deployment period, named lead rebound, is rarely considered in previous researches but will constrain the duration time to a short period (e.g., 10 min), which greatly limits the utilization of ACs. This paper proposes an optimal sequential dispatch strategy of ACs to mitigate the lead–lag rebound, and thus realize flexible control of the duration time from minutes to several hours. To quantify the effects of lead–lag rebound, a capacity-time evaluation framework of the operating reserve is developed. On this basis, ACs are grouped to be dispatched in sequence to mitigate the lead–lag rebound. The co-optimization of sequential dispatch on the capacity dimension and time dimension forms a mixed-integer nonlinear bilevel programming problem, in which the consumers' thermal comfort is also guaranteed. Case studies are conducted to validate the proposed strategy.**

***Index Terms*—Air conditioning, load management, dispatching, demand response, operating reserve.**

Nomenclature

i	Index of an individual AC.
g	Index of an AC group.
k	Index of an AC group to be dispatched.
q	Index of an AC group to be recovered.
d	Index of the reserve deployment period.
r	Index of the recovery period.
t	Index of time (h).
$\theta_i(t)$	Room temperature corresponding to the i-th AC at the time t (°C).
$\theta_a(t)$	Ambient temperature (°C).
$T_{set,i}$	Set point temperature of the i-th AC (°C).
θ_i^+/θ_i^-	The upper/lower temperature hysteresis band of the i-th AC (°C).
$\boldsymbol{\Gamma}$	The set of all the ACs under an aggregator.
$N^{\max}$	The number of ACs in $\boldsymbol{\Gamma}$.
τ_g^d/τ_g^r	The reserve deployment/recovery time instant of group g (h).
RC_g^d/RC_g^r	Reserve capacity of group g during the reserve deployment/recovery period (MW).
BC_g^d/BC_g^r	Rebound capacity of group g during the reserve deployment/recovery period (MW).
PD_g	The difference of the aggregate power before and after the changes of the set point temperature (MW).
$\underset{t_1\to t_2}{SD}(x(t))$	The standard deviation of the variable $x(t)$ during the time period $[t_1, t_2]$.
PV_g	Power volatility of aggregate power of group g after the end of rebound process (MW).
DT_g	Deployment duration of operating reserve provided by ACs in group g (h).
RT_g	Ramp time of operating reserve provided by ACs in group g (min).
RR_g	Ramp rate of operating reserve provided by ACs in group g (MW/min).
RC_g^*	Required reserve capacity instructed by the system operator (MW).
DT_g^*	Required deployment duration instructed by the system operator (h).

Manuscript received November 23, 2017; revised April 17, 2018; accepted May 25, 2018. Date of publication June 11, 2018; date of current version October 18, 2018. This work was supported in part by the National Science Foundation of China under Grants 51577167 and 51537010, and in part by the State Grid Corporation of China (52110416001T). Paper no. TPWRS-01767-2017. *(Corresponding author: Yi Ding.)*

W. Cui, Y. Ding, H. Hui, Z. Lin, and C. Shao are with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: cuiwenqi@zju.edu.cn; yiding@zju.edu.cn; huihongxun@zju.edu.cn; linzhenzhi@zju.edu.cn; shaochangzheng@zju.edu.cn).

P. Du is with the Electric Reliability Council of Texas, Taylor, TX 76574 USA (e-mail: Pengwei.du@ercot.com).

Y. Song is with the Department of Electrical and Computer Engineering, University of Macau, Macau 999078, China, and also with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: yhsongcn@zju.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2018.2846270

I. Introduction

The growing penetration of renewable energy sources into the electric power system calls for a huge amount of balancing services at multiple timescales [1], [2]. Air conditioners (ACs) offer an alternative of traditional generation units for balancing the system by actively reducing or increasing electricity consumption [3]. Statistical data have shown that ACs account for approximately 35%, 33% and 40% of the electricity consumption during the peak hours in many cities in China [4], Spain [5] and India [6], respectively. With the development of smart grid technologies and real-time telemetry [7], [8], it is technically feasible for ACs to respond to instructions within a short period and provide operating reserve at various time scales [9]. However, demand response rebound is one possible obstacle of using ACs for the provision of operating reserve [10]. This phenomenon is the rebound peak that arises when a large amount of loads are re-connected to the grid at approximately the same time [11]. The existence of the demand response rebound may cause significantly higher demand than that prior to the demand response event. In extreme cases, the increased load current derived from the rebound peak may even lead to the melting of overhead lines, which harms system security considerably [12].

The demand response rebound at the end of a demand response event, which is when ACs are recovered to the initial states, is also referred to as load payback effect [13], [14], load recovery effect [15], [16] or cold load pickup [12], [17] in the existing literatures. This phenomenon has been observed in many pilot projects, including the Californian pilot study of time-of-use and critical peak pricing [18] and the Norwegian project of direct control of residential water heaters in 475 households [19]. Several studies have addressed the impacts of demand response rebound on the market scheme [20], [21] and system scheduling [10], [22]. Most researches reduce the level of rebound by increasing the diversity of loads [19] or randomizing the reconnection of appliances over time [13]. Apart from that, reference [23] copes with demand response rebound by leveraging chilled-water capacity through a least enthalpy estimation based thermal comfort control. The above approaches can reduce the demand response rebound to some extent but cannot mitigate the rebound entirely. The concept of dispatching groups of devices in sequence is initially presented to prolong the deployment duration of operating reserve without increasing the interruption duration of individual consumers [24], [25]. Reference [26] further indicates that restoring group of devices to the initial state in sequence can also reduce the demand response rebound. This is examined by the recovery of water heaters in [17] and [27], which illustrate that both the magnitude and mean value of demand response rebound can be largely reduced. However, the devices are evenly divided into several groups and are recovered at a regular time interval (e.g., 10 min) in [17], [24]–[27]. As a result, the aggregate dynamics of devices are not considered, which cannot guarantee that the rebound peak is reduced to the minimum value. Reference [13] presents the concept to achieve better control on the demand response rebound by the coordination among different groups of devices, but there lacks mathematical model to determine the coordination process among the groups. Moreover, these research studies only consider the rebound load during the recovery period, while neglecting the rebound load during the reserve deployment period.

For clarity, rebound peak of aggregate power during the reserve deployment period is named as the lead rebound effect in this paper. By contrast, rebound peak of aggregate power during the recovery period is named as the lag rebound effect. The lead rebound effect is neglected in previous research studies because the ACs are converted to the *off* state when providing operating reserve. Hence, ACs can reliably provide load reduction with different duration time, as long as they remain in the *off* state [10], [22]. However, shutting the units off directly may cause short-cycling of ACs, which can reduce their lifetime, increase maintenance, and potentially damage them [28], [29]. Therefore, recent research studies have focused on controlling ACs through changing the set point temperature instead [11], [30]. In this case, according to the Law of Energy Conservation [23], demand response rebound will also occur during the reserve deployment period. For example, ACs are in the cooling mode in summer. Upon receiving the load-reduction instruction, ACs will migrate to the new upper temperature hysteresis band and stay longer in the *standby* state. More internal heat will accumulate with longer *standby* period of ACs. Hence, higher demand for cooling occurs when ACs have reached the new upper temperature hysteresis band, resulting in the increase of electricity consumption, i.e., lead rebound effect.

The lead rebound effect poses a new challenge to the provision of the operating reserve by ACs because it limits the duration time before the rebound occurs. If the rebound occurs within the period of reserve deployment, then ACs cannot sustain the required reserve capacity in the required duration time. Requirements on the duration time of the operating reserve are crucial because it ensures the system operator has adequate time to correct the imbalance between load and generation [31]. Typically, the required duration time of spinning/non-spinning reserve can be 30 minutes or even 60 minutes [32]. Considering that the lead rebound may occur approximately 10 minutes after the control of ACs, ACs encounter difficulty in fulfilling the requirements on the duration time. To make the most of the ACs' potential to provide operating reserve, the duration time of the load reduction or load increase should be flexibly controlled. This entails a need to mitigate the lead rebound entirely during the required reserve deployment period, which can be potentially realized by dispatching different groups of ACs in sequence. However, as illustrated above, there lacks the consideration on the dynamics of ACs, and so does the co-optimization among the reserve capacity and dispatch time instant of different AC groups in existing literatures, making it difficult to guarantee the entire mitigation of the lead-lag rebound effect. Moreover, the control of the rebound time, which is crucial for the determination of the duration time, is not involved. Consequently, the time of rebound remains uncontrollable and the utilization of operating reserve provided by ACs is still limited by the rebound period.

This paper proposes an optimal sequential dispatch strategy of ACs to mitigate the lead-lag rebound entirely and thus realize the flexible provision of various types of operating reserve. Because of the constraint of the duration time imposed by the lead rebound, traditional methods that quantify the duration time between the reserve deployment and the recovery are not suitable [32]. Therefore, the first step is to develop an evaluation framework of operating reserves to quantify the effect of the rebound load on the capacity dimension and time dimension. Then, ACs dispatched at the same time instant are defined as an AC group, and different AC groups are dispatched in sequence. In order to guarantee that the rebound load of each group is mitigated entirely by the reserve capacity of the latter group, the dispatch time instant of each group is optimized to minimize the deviation between the actual load changing and the required value, while the selection of ACs in each group are optimized to make full use of ACs' available reserve capacity and guarantee consumers' comfort. Co-optimization of the above problems on the time dimension and capacity dimension forms a mixed integer nonlinear bi-level programming problem, which is then solved by genetic algorithm. In addition, a three-layer structure is designed to integrate the proposed strategy with the interactions between aggregators and consumers. Case studies are conducted to verify the proposed strategy for providing operating reserve with multiple duration time. The major contributions of this paper are as follows:

1) The lead rebound effect, which results in the special obstacle of controlling the duration time, is considered for the provision of operating reserve with ACs. To the best of the authors' knowledge, it is the first time that the lead rebound is modeled and analyzed.
2) A capacity-time evaluation framework of the operating reserve provided by ACs is developed to quantify the impacts of the lead-lag rebound effect. Compared with the existing evaluation method for traditional generating units [32], the proposed evaluation framework can better characterize the dynamics of demand-side operating reserve.
3) An optimal sequential dispatch strategy, which can entirely mitigate both the lead rebound and lag rebound, is proposed to realize the flexible control of duration time from minutes to several hours.

The remainder of the paper is organized as follows. Section II analyzes the lead-lag rebound effect from the aggregate response of ACs controlled by the changes of the set point temperature. Section III quantifies the impacts of the lead-lag rebound by a proposed capacity-time evaluation framework of operating reserve. On this basis, Section IV proposes an optimal sequential dispatch strategy of ACs for the entire mitigation of the lead-lag rebound and the provision of operating reserve with multiple duration time. Numerical evidence for the effectiveness of the proposed strategy is provided in Section V. Conclusions are given in Section VI.

II. Analysis of the Lead-Lag Rebound Effect

A. Model of an Individual AC

The operation process of an individual AC is described by the general state model for the thermostatically-controlled-loads (TCLs) [33]:

$$\frac{d\theta_i(t)}{dt} = -\frac{1}{C_i R_i}[\theta_i(t) - \theta_a(t) + m_i(t) R_i Q_i] \tag{1}$$

where $\theta_i(t)$ is the room temperature corresponding to the *i*-th AC at time *t*, $\theta_a(t)$ is the ambient temperature. C_i and R_i are the thermal capacity and thermal resistance corresponding to the room of the *i*-th AC, respectively. $m_i(t)$ represents the *on* or *standby* state of the *i*-th AC. Q_i is the energy transfer rate of the *i*-th AC, which is equal to the product of the input power p_i and the coefficient of performance COP_i of the *i*-th AC.

The *i*-th AC operates cyclically around its set point temperature $T_{set,i}$ with a dead band of ΔT_i. For example, if the AC is in the cooling mode in summer, it will switch to the *on* state when the room temperature reaches the upper band ($\theta_i^+ = T_{set,i} + 0.5 \times \Delta T_i$). Similarly, when the room temperature reaches the lower band ($\theta_i^- = T_{set,i} - 0.5 \times \Delta T_i$), it will switch to the *standby* state. The temperature range between θ_i^- and θ_i^+ is defined as the hysteresis band $[\theta_i^-, \theta_i^+]$ [33]:

$$m_i(t) = \begin{cases} 1, & \theta_i(t) > \theta_i^+ \\ 0, & \theta_i(t) < \theta_i^- \\ m_i(t-1), & otherwise \end{cases} \tag{2}$$

B. Aggregate Response of ACs

The AC load model (1)–(2) reveals that the state of ACs can be quickly controlled by changing the set point temperature. However, if ACs are controlled by consumers manually, the response may not be sufficiently fast to meet the requirement of ramp time [31]. Therefore, it is assumed that the ACs are installed with smart meters. By signing contracts with consumers, aggregators can control ACs at the permitted periods [34]. When the duration time has reached the required value, ACs are recovered to the initial states [13].

Denoting $\mathbf{\Gamma}$ as the set of all the ACs under an aggregator, the number of ACs in $\mathbf{\Gamma}$ is $N^{\max}$. The *i*-th AC is permitted to be controlled by the aggregator during the period $[t_i^s, t_i^e]$. The reserve deployment period instructed by the system operator is$[t_{ins}, t_{end}]$. The available AC is defined as the unit that is permitted to be controlled by the aggregator during the required duration. The availability of ACs is labeled by the vector $\mathbf{V} \in \mathbb{R}^{N^{\max} \times 1}$, in which the *i*-th element is:

$$v_i = \begin{cases} 1, & [t_{ins}, t_{end,}] \in [t_i^s, t_i^e] \\ 0, & otherwise \end{cases}, \quad \forall i \in \mathbf{\Gamma} \tag{3}$$

The ACs dispatched at the same time instant τ_g are defined as an AC group *g*. $\mathbf{S}_g \in \mathbb{R}^{N^{\max} \times 1}$ is the vector that represents the ACs belonging to group *g*. The *i*-th element in $\mathbf{S}_g$ is:

$$s_{g,i} = \begin{cases} 1, & i \in group\ g \\ 0, & otherwise \end{cases}, \quad \forall i \in \mathbf{\Gamma} \tag{4}$$

The physical parameters of the *i*-th AC, including C_i and R_i, can be usually assumed as a constant value. The input power $p_i$is usually set by the AC manufacturer and cannot be changed by consumers. Considering that there may exist tens of thousands of units under an aggregator, it may be difficult to obtain the parameters of all the ACs. In this case, aggregators can randomly select N^e ACs to measure their parameters. Kernel density estimation, which is one of the non-parametric probability density estimation methods, can be utilized to obtain the probability density distribution of parameters from the known data points corresponding to N^e selected ACs [35]. For example, $(R_1^e, R_2^e, ..., R_{N^e}^e)$ denotes the thermal resistance of the randomly selected N^e ACs. The estimated probability density

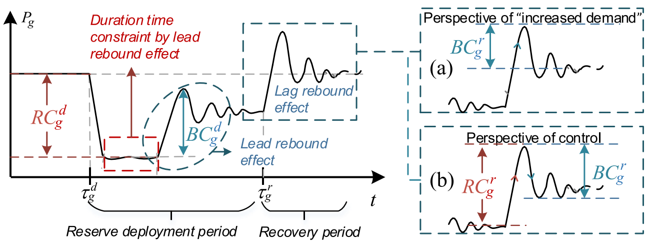


Fig. 1. Typical curve of the aggregate response of ACs.

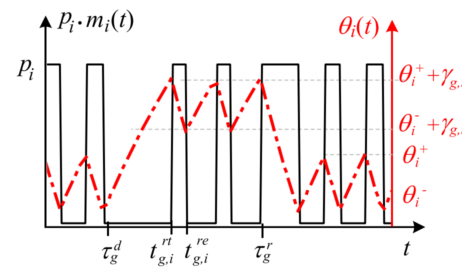


Fig. 2. Consumed power and room temperature of the i-th AC when the temperature hysteresis band is increased by $\gamma_{g,i}$ at τ_g^d and decreased by $\gamma_{g,i}$ at τ_g^r.

$\hat{f}_{R^e}$ of the thermal resistance R^e can be expressed as:

$$\hat{f}_{R^e}(R^e) = \frac{1}{N^e h_{R^e}} \sum_{i_e=1}^{N^e} K\left(\frac{R^e - R^e_{i_e}}{h_{R^e}}\right) \tag{5}$$

where $K(\cdot)$ is the normal kernel function. h_{R^e} is the bandwidth of the kernel function and can be determined by the rule-of-thumb bandwidth estimator method [35]. Similarly, the probability density of other parameters can also be estimated with the kernel density estimation. Then, the parameters of the ACs can be randomly set according to the probability density distribution of these parameters.

The temperature hysteresis band of individual AC in $\mathbf{\Gamma}$ is assembled in the vector $\boldsymbol{\theta}^{-(+)} = [\theta_1^{-(+)}, \theta_2^{-(+)} \cdots \theta_{N^{\max}}^{-(+)}]^T$. Changes of the set point temperature during the reserve deployment/recovery of group g are assembled in the vector $\boldsymbol{\gamma}_g^{d/r} = [\gamma_{g,1}^{d/r}, \gamma_{g,2}^{d/r} \cdots \gamma_{g,N^{\max}}^{d/r}]^T$. Similarly, $\bar{\boldsymbol{\gamma}}^{d/r}$ and $\underline{\boldsymbol{\gamma}}^{d/r}$ are the highest increase and decrease of the set point temperature determined by consumers, respectively. Apart from τ_g and the time t, the aggregate power is primarily influenced by parameters $\boldsymbol{\theta}^{-(+)}$, $\boldsymbol{\gamma}_g^{d/r}$ and $\mathbf{S}_g$, which are assembled in the array $\boldsymbol{u}_g = [\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_g^{d/r}, \mathbf{S}_g] \in \mathbb{R}^{N^{\max} \times 1}$. Consequently, the aggregate power P_g of group g is a function of the mentioned parameters:

$$P_g(\mathbf{u}_g, \tau_g, t) = \sum_{i \in \mathbf{\Gamma}} p_i \cdot m_i(t) \cdot s_{g,i} \tag{6}$$

C. Lead Rebound Effect and Lag Rebound Effect

A typical curve of the aggregate response of an AC group is shown in Fig. 1, in which the ACs change the set point temperature for load reduction at τ_g^d and recover the set point temperature to the initial value at τ_g^r. Two rebound peaks of aggregate power exist during the reserve deployment period and the recovery period, respectively. In this paper, the former is named as the lead rebound effect, and the latter is named as the lag rebound effect.

During the reserve deployment period, the decrease of aggregate power is corresponding to the provision of reserve capacity RC_g^d, while the increase of aggregate power is corresponding to the rebound capacity BC_g^d of the lead rebound. The lead rebound effect is rarely considered in existing research studies because it only occurs when ACs are controlled by changing the set point temperature. This property can be explained by the migration of ACs' room temperature through the Law of Energy Conservation [23]. For example, ACs are in cooling mode in summer. The variation of room temperature and the corresponding consumed power of the i-th AC are illustrated in Fig. 2. Upon receiving the load-reduction instruction at τ_g^d, the i-th AC increases the set point temperature and remains in the *standby* state until the room temperature reaches the new upper temperature hysteresis band at $t_{g,i}^{rt}$, after which the AC will switch to the *on* state. The internal heat is not transferred outdoors and thus accumulates during the longer *standby* period shown by the segment from τ_g^d to $t_{g,i}^{rt}$. Hence, higher load demand occurs in the subsequent *on* state. Consequently, the lead rebound is the nature increase of aggregate power resulted from the cyclic operation characteristics of ACs. It is different from the opt-out behavior of consumers, which is the consumers' initiative behavior to choose not to participate in the demand response event when their comfort levels cannot be maintained [37], [38].

By contrast, the lag rebound effect is the result of consumers' recovery behavior and a common phenomenon after a demand response event [12], [19]. As shown in Fig. 2, the AC decreases the set point temperature after receiving the load-reduction instruction at τ_g^r and remains in the *on* state for longer time. Hence, aggregate power of ACs increases instantly when the loads are recovered to the initial states at τ_g^r. The lag rebound capacity BC_g^r is considered as the increase of aggregate power larger than the initial value in the existing research studies [12], [19], as illustrated in Fig. 1(a). From the control perspective in this paper, the recovery process can be regarded as the reserve deployment process with the required reserve capacity provided by the load-increase operation. In this case, the increase of aggregate power is corresponding to the provision of reserve capacity RC_g^r, while the decrease of aggregate power is corresponding to the rebound capacity BC_g^r, as illustrated in Fig. 1(b).

As depicted in Fig. 1, the duration time is constrained within the period before the lead rebound occurs. It is required to notice that sudden changes of the set point temperature will cause temporary synchronization of the ACs [33]. In other words, after the ACs have reached the new temperature hysteresis band, it will switch between the *on* state and *standby* state at the approximately same time, which will increase the level of rebound and also lead to large power fluctuations. However, few methods are available to quantify the effect of lead-lag rebound and the associated power fluctuations, making it difficult to mathematically describe the objectives of AC load control. Therefore, Section III proposes several indices to evaluate operating reserves considering the effect of lead-lag rebound and power fluctuations.

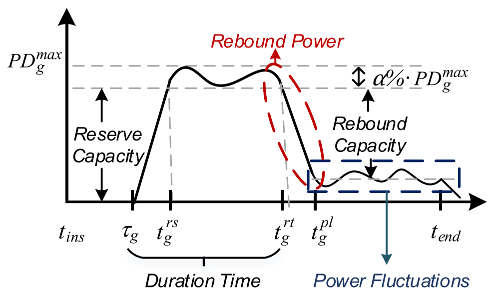


Fig. 3. Power difference of the aggregate power before and after the changes of the set point temperature.

III. Capacity-Time Evaluation of the Operating Reserve Considering Lead-Lag Rebound Effect

Operating reserve providers are required to respond to different types of events over different time frames [31]. Therefore, two fundamental dimensions, *capacity* and *time* [32], are required to be considered for evaluating the operating reserve. The capacity dimension entails assigned amount of load reduction/increase during the reserve deployment period. The time dimension involves duration time and ramp rate. The effect of the lead-lag rebound and the associated power fluctuations are also quantified on these two dimensions.

A. Universal Expression of the Load Reduction/Increase

The difference of the aggregate power before and after the changes of set point temperature represents the effects of AC load control. The aggregate power changes in an adverse direction in load-reduction and load-increase operation. Therefore, the power difference PD_g is expressed as (7), so that the reserve capacity is a positive number and thus evaluation indices can be represented as a universal form.

$$PD_g\left(\mathbf{u}_g, \tau_g, t\right) = \begin{cases} P_g^0\left(\mathbf{u}_g, \tau_g, t\right) - P_g\left(\mathbf{u}_g, \tau_g, t\right), & load - reduction \\ P_g\left(\mathbf{u}_g, \tau_g, t\right) - P_g^0\left(\mathbf{u}_g, \tau_g, t\right), & load - increase \end{cases} \tag{7}$$

where $P_g^0(\mathbf{u}_g, \tau_g, t)$ and $P_g(\mathbf{u}_g, \tau_g, t)$ are the aggregated power of group g before and after the changes of the set point temperature, respectively. Equation (7) is not expressed in the form of absolute value because the rebound load may exceed the aggregate power at the initial state. Hence, evaluation indices represented by the PD_g can be applied to both the load-reduction operation during the reserve deployment period and the load-increase operation during the recovery period, respectively. A general evaluation framework for the control of the lead-lag rebound is developed based on the dynamics of power difference after the changes of set point temperature, which is shown by Fig. 3.

B. Evaluation of the Operating Reserve Provided by ACs on the Capacity Dimension

1) Reserve Capacity (RC): The aggregate power fluctuates in nature due to the cyclic operation characteristics of ACs. Let $PD_g^{\max}$ denotes the maximum power difference of group g during reserve deployment period, the valid reserve deployment is then defined as the threshold between $(1-\alpha\%)\cdot PD_g^{\max}$ and $PD_g^{\max}$. t_g^{rt} and t_g^{rs} are the time instant corresponding to two endpoints of the defined threshold. Hence, the aggregate reserve capacity (*RC*) is:

$$RC_g(\mathbf{u}_g, \tau_g) = PD_g^{\max}(\mathbf{u}_g, \tau_g) - PD_g^{\max}(\mathbf{u}_g, \tau_g) \times \alpha\% \tag{8}$$

2) Rebound Capacity (BC): As is illustrated in Fig. 2, the power difference declines because of the rebound. The decline process stops when the aggregate power of the ACs enter the steady state. Denote t_g^{pl} as the end of the rebound process, which is defined as the time instant when power difference stops declining. The rebound capacity (*BC*) is the difference between the reserve capacity and power difference at t_g^{pl}:

$$BC_g(\mathbf{u}_g, \tau_g) = RC_g(\mathbf{u}_g, \tau_g) - PD_g(\mathbf{u}_g, \tau_g, t_g^{pl}) \tag{9}$$

3) Power Volatility (PV): Standard deviation (*SD*) is adopted to represent the fluctuations of aggregate power. The standard deviation of the variable *x*(*t*) during the time period $[t_1, t_2]$ is defined as:

$$\underset{t_1 \to t_2}{SD}(x(t)) = \sqrt{\left[\int_{t_1}^{t_2} (x(t) - \bar{x}(t))^2 \Delta t\right] / (t_2 - t_1)} \tag{10}$$

where $\bar{x}(t)$ is the mean value of the variable $x(t)$ from t_1 to t_2.

The standard deviation of the power difference after the power spike caused by the rebound peak represents the power fluctuations caused by the AC load control. Power volatility (*PV*) is defined as the ratio of the standard deviation to the initial aggregate power at the time instant of reserve deployment:

$$PV_g = \underset{t_g^{pl} \to t_{end}}{SD}(PD_g(\mathbf{u}_g, \tau_g, t)) / P_g(\mathbf{u}_g, \tau_g, \tau_g) \tag{11}$$

C. Evaluation of the Operating Reserve Provided by ACs on the Time Dimension

1) Duration Time (DT): Duration time (*DT*) is the period that reserve service providers maintain the required reserve capacity RC^*. Traditionally, *DT* is calculated as the period between the reserve deployment and the recovery [32]. However, ACs cannot maintain RC^* once the lead rebound occurs. In this case, *DT* is the period within the defined reserve capacity threshold in (8). Therefore, *DT* is quantified according to the rebound time instant:

$$DT_g = \begin{cases} t_g^{rt} - \tau_g, & t_g^{rt} < \tau_g + DT^* \\ t_{end} - \tau_g, & otherwise \end{cases} \tag{12}$$

2) Ramp Rate (RR): Ramp time (*RT*) is the period that the reserve service providers control their output to the required reserve capacity:

$$RT_g = t_g^{rs} - \tau_g \tag{13}$$

Ramp rate (*RR*) is the speed that the reserve service providers control their output to the required value [32]. *RR* is defined as the ratio of the reserve capacity to the ramp time:

$$RR_g(\mathbf{u}_g, \tau_g) = RC_g(\mathbf{u}_g, \tau_g) / RT_g \tag{14}$$

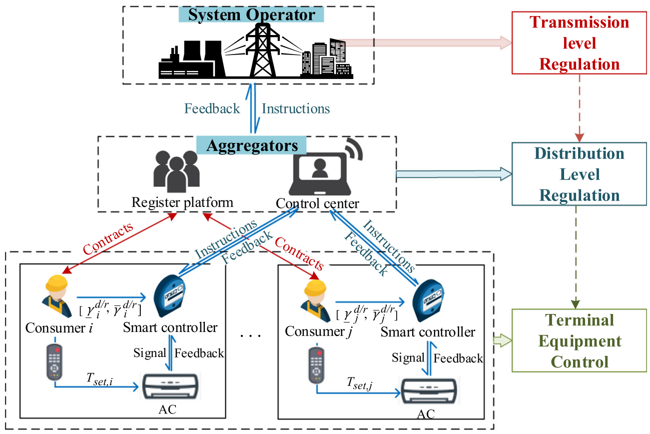


Fig. 4. The interactions among the system operator, aggregators and consumers.

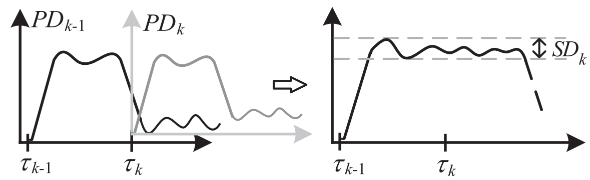


Fig. 5. Principle of the sequential dispatch strategy.

IV. Sequential Dispatch Strategy of ACs for Providing Operating Reserve With Multiple Duration Time

A. The Interactions Among the System Operator, Aggregators and Consumers

Three types of entities involve in the provision of operating reserve by ACs, i.e., the system operator, the aggregators and the consumers [39]. The interactions among the entities are shown in Fig. 4.

The role of system operator (e.g., the Independent System Operator (ISO) in the United States, Transmission System Operator (TSO) in the European Commission or the grid company in China) is to operate the transmission system [39], [40]. The system operator will run multiple DR programs to motivate consumers for benefiting transmission system operations by actively reducing or increasing electricity consumption [41]. Typically, ancillary service market programs allow consumers to act as reserve service providers for providing operating reserve on equal terms with the generators [41]. During the operating hours, in order to correct the imbalance between generation and demand, the system operator will instruct reserve service providers about when and how much the operating reserves to be deployed and recovered [34].

Small consumers, including the owners of ACs, are grouped by aggregators to bid in the market because the limited capacity of individual consumer cannot fulfill the requirement on the minimum amount of bids in the spot market [39]. Depending on the specific design and structure of a DR program, the aggregators can be distribution system operators, load-serving entities, or DR providers [41]. By signing contracts with consumers, aggregators can control ACs at the permitted periods [34]. If the bids from aggregators for the provision of operating reserve are accepted in the market, the aggregators will respond to the instructions from the system operator by regulating the controllable ACs with the proposed sequential dispatch strategy during the operating hours. On this basis, aggregators will send instruction signals to the smart controllers of the ACs to be dispatched [38].

It is assumed that the ACs are installed with smart controllers, which share the functions of communication, sensor and control [42]. The smart controllers enable consumers to easily set the parameters, such as the temperature ranges, controllable periods and control modes [9], [43]. After receiving the instructions from aggregators, the smart controllers will then control ACs with local embedded control strategy according to the parameters set by consumers.

B. Sequential Dispatch Strategy of ACs to Mitigate the Lead-Lag Rebound Effect

In order to mitigate the rebound load and thus enable the flexible control of duration time, a sequential dispatch strategy of ACs is proposed. The principle of sequential dispatch strategy is shown in Fig. 5. The rebound capacity can be considered to be an updated required reserve capacity. If ACs are divided into several groups and if the reserve capacity of each group is equal to the rebound capacity caused by the previous group, then the rebound load of the previous group is mitigated entirely.

The first group of ACs is dispatched when receiving reserve deployment instruction at t_{ins}. The required reserve capacity of the first group is equal to the value instructed by the system operator:

$$RC_1^* \left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_1^{d/r}, \mathbf{S}_1, t_{ins} \right) = RC^* \tag{15}$$

The latter groups of ACs are dispatched continuously to mitigate the rebound load caused by the previous group.

$$\begin{aligned} & RC_k^* \left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_k^{d/r}, \mathbf{S}_k, \tau_k \right) \\ & \quad = BC_{k-1} \left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_k^{d/r}, \mathbf{S}_{k-1}, \tau_{k-1} \right) \end{aligned} \tag{16}$$

The process of sequential dispatch terminates when the rebound capacity is sufficiently small. Note that the stabilized aggregate power also fluctuates within a range and the fluctuations will increase with larger number of ACs. Therefore, the termination condition of the sequential-dispatching process is set as a value proportional to the aggregate power at t_{ins}. β denotes the proportional coefficient of the termination condition, and its value can be set according to the power volatility PV in steady state. Suppose the rebound load has been mitigated entirely when:

$$BC_k \leq \beta\% \cdot P_k(\mathbf{u}_k, \tau_k, t_{ins}) \tag{17}$$

As illustrated in Figs. 1 and 2, the rebound load and the fluctuations of aggregate power are highly relevant to the temporary synchronization of the ACs. Existing methods on avoiding the synchronization of ACs can be classified into two categories. The first is to track power profiles by subtle changes of the set point temperature in real time based on two-way communication between control center and ACs [33], [44], [45]. In [33]

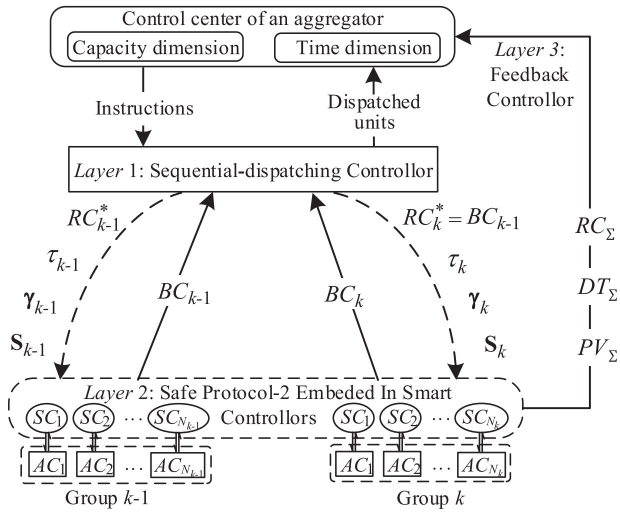


Fig. 6. Three-layer structure for the implementation of sequential dispatch strategy.

and [44], synchronization of homogenous ACs is governed by broadcasting subtle temperature set point changes as the output signal of a feedback-controller. Reference [45] includes AC parameter heterogeneity by controlling the on/off states of ACs based on the state bin transition model. However, this method requires careful tuning of parameters for specific scenarios and is difficult to be applied to the control of ACs with changes of set point temperature. Therefore, this approach is more suitable for the provision of load following or regulation services, which usually accommodate ACs with fast communication equipment and control ACs through subtle changes of set point temperature frequently (e.g., minute-to-minute changes of set point temperature smaller than 0.1° C [33]).

The second is to implement the shift in the set point temperature according to safe protocols embedded in the smart controller of each individual AC [36], [46]–[48]. References [36] and [46] propose several safe protocols to generate different power pulse shapes and avoid the sudden changes of ACs' state, which have been proved to effectively avoid the synchronization for both heterogeneous and homogeneous ACs with different changes of set point temperature [47], [48]. Therefore, such approach based on safe protocols is more suitable for providing operating reserve in this paper, which controls heterogeneous ACs with larger changes of set point temperature for only two times (one at the reserve deployment time instant and the other at the recovery time instant). Among all the safe protocols proposed in [36] and [46], the safe protocol-2 (SP-2) can reduce power fluctuations to the lowest level and is adopted in this paper.

Implementation of the sequential dispatch strategy coordinated with the SP-2 is illustrated in Fig. 6. The smart controller of the *i*-th AC is abbreviated to SC_i. Firstly, according to the requirements of operating reserve on the capacity dimension and the time dimension, the sequential-dispatching controller optimizes the use of all the controllable ACs under an aggregator with the proposed sequential dispatch strategy. The sequential-dispatching controller will then send signals to the smart controllers of ACs in group *k* about the reserve deployment time instant and the changes of set point temperature. Secondly, after receiving the instructions from the sequential-dispatching controller, the control of ACs will follow SP-2, which is embedded in the smart controller of an individual AC. In this way, the fluctuations of aggregate power can be largely reduced without adding additional computational burden between aggregators and consumers. Thirdly, the detailed information of the reserve deployment is feedback to aggregators.

C. Capacity-Time Co-Optimization of Sequential Dispatch Process During the Reserve Deployment Period

The dispatch time instant and available reserve capacity of each AC group are co-optimized to make full use of ACs' potential for the provision of operating reserve. For clarity, the subscript *d* and *r* denote the parameters of reserve deployment period and the recovery period, respectively. Fig. 5 shows that the dispatch time instant τ_k^d of group *k* greatly influences the performance of dispatching group *k* to mitigate the rebound load caused by group *k*-1. For example, if group *k* is dispatched too early, then the rebound load of group *k* and group *k*-1 will accumulate, resulting in higher rebound load. By contrast, if group *k* is dispatched too late, then the aggregate power will still rebound until group *k* is dispatched. Therefore, the deployment time instant τ_k^d of group *k* is optimized to minimize the deviation between the actual power difference and the required value RC^*, that is,

$$\min_{\tau_k^d} \quad \underset{t_{ins} \to t_k^{rt}}{SD} \left\{ \left(\sum_{j=1}^{k} PD_j \left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}^d, \mathbf{S}_j^d, \tau_j^d, t \right) \right) - RC^* \right\} \tag{18}$$

The largest available reserve capacity of an AC group can be calculated when the set point temperature of all the ACs are changed to the bound set by consumers. Aggregators tend to make the most of the available reserve capacity to earn more benefits. Therefore, the selection of ACs $\mathbf{S}_k^d$ in group *k* and the corresponding changes of set point temperature $\boldsymbol{\gamma}_k^d$ are optimized to follow the requirements of reserve capacity with minimum available reserve capacity:

$$\min_{\mathbf{S}_k^d, \boldsymbol{\gamma}_k^d} \quad RC_k^d \left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \widehat{\boldsymbol{\gamma}}^d, \mathbf{S}_k^d, \tau_k^d \right) \tag{19}$$

where $\widehat{\boldsymbol{\gamma}}^d$ is the maximum changes of the set point temperature within the bound set by consumers. In other words, $\widehat{\boldsymbol{\gamma}}^d = \bar{\boldsymbol{\gamma}}^d$ for load-reduction in summer or load-increase in winter, $\widehat{\boldsymbol{\gamma}}^d = \underline{\boldsymbol{\gamma}}^d$ for load-increase in summer or load-reduction in winter.

Equations (18) and (19) form a bi-level optimization:

$$\min_{\tau_k^d} \quad \underset{t_{ins} \to t_k^{rt}}{SD} \left\{ \left(\sum_{j=1}^{k} PD_j (\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_j^d, \mathbf{S}_j^d, \tau_j^d, t) \right) - RC^* \right\} \tag{20}$$

$$\text{s.t.} \quad \tau_k^d > \tau_{k-1}^d \tag{21}$$

$$\tau_k^d < t_{ins} + DT^* \tag{22}$$

$$\left[\mathbf{S}_k^d,\ \boldsymbol{\gamma}_k^d\right] = \arg\min RC_k^d\left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \widehat{\gamma}^d, \mathbf{S}_k^d, \tau_k^d\right) \tag{23}$$

$$\text{s.t.}\quad RC_k^{d,*}\left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_k^d, \mathbf{S}_k^d, \tau_k^d\right) = BC_{k-1}^d\left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_{k-1}^d, \mathbf{S}_{k-1}^d, \tau_{k-1}^d\right) \tag{24}$$

$$\underline{\gamma}^d \le \boldsymbol{\gamma}_k^d \le \overline{\gamma}^d \tag{25}$$

$$\sum_{j=1}^{k} \sum_{i=1}^{N^{\max}} s_{j,i}^d \times v_i \le \sum_{i=1}^{N^{\max}} v_i \tag{26}$$

The high-level problem (20)–(22) optimize the deployment time instant of group k to minimize the deviation between the actual power difference and the required value. Equation (21) ensures that the deployment time instant of the current group is later than the previous group, while (22) limits the dispatch operation within the required duration time DT^*. The low-level problem (23)–(26) optimize the selection of ACs in group k and the corresponding changes of set point temperature to follow the requirements of reserve capacity with minimum available reserve capacity of ACs. Equation (24) constraints the reserve capacity of group k according to the rebound capacity of the previous group. Equation (25) limits the changes of the set point temperature within the range set by the consumers. Equation (26) constraints the total number of dispatched ACs within maximum number of available ACs. The bi-level optimization formed by (20)–(26) is a mixed integer nonlinear bi-level programming problem. Genetic algorithm (GA) provides a flexible modeling framework that allows considering the nonlinearities and non-convexities associated with the mixed integer nonlinear bi-level programming problem [49]. Therefore, GA is applied to solve the bi-level optimization formed by (20)–(26) in this paper.

The optimization of the low-level problem (23)–(26) cannot continue when (24) and (26) cannot be satisfied at the same time. In other words, the remaining ACs are not adequate to mitigate the rebound load caused by the previous group entirely. K denotes the number of all the dispatched AC groups during the reserve deployment period. In this case, the units belonging to the K-th group are selected as the remaining ACs:

$$\mathbf{S}_K^d = \mathbf{V} - \sum_{j=1}^{K-1} \mathbf{S}_j^d \tag{27}$$

After all the available ACs have been dispatched, the sequential dispatch process is terminated. Hence, the duration time is determined by the rebound time instant of group k and is represented by:

$$DT = t_K^{rt} - t_{ins} \tag{28}$$

D. Capacity-Time Co-Optimization of Sequential Dispatch Process During the Recovery Period

ACs will be recovered to the initial states when the duration time reaches the required value. Sequential dispatch strategy can also be utilized to mitigate the lag rebound. Since the ACs to be recovered are the same as those dispatched during the reserve deployment period, there is no need to conduct the optimization revealed in (23)–(26). Equation (25) has ensured that the changes of set point temperature are within the range set by consumers and therefore the consumers' basic comfort levels can be guaranteed. However, consumers' thermal comfort levels will still decrease with longer deployment duration or larger changes of the set point temperature [42]. In order to avoid the further dissatisfaction of consumers, ACs with the lowest comfort levels should be recovered earlier. $DT_i^{\max}$ denotes the maximum allowable control duration of the i-th AC according to the contract with the aggregator. The objective function to determine ACs in the q-th group is then represented by (29), so that the ACs with the lowest thermal comfort levels are selected to be recovered.

$$\min_{\mathbf{S}_q^r} \quad \sum_{i=1}^{N^{\max}} \sum_{m=1}^{K} \left(1 - \frac{\tau_q^r - \tau_m^d}{DT_i^{\max}} \cdot \frac{\gamma_{m,i}^d}{\widehat{\gamma}_i^d}\right) \cdot s_{m,i}^d \cdot s_{q,i}^r \tag{29}$$

Similar with the reserve deployment process, the optimization of the recovery time instant τ_q^r and the selection of ACs $\mathbf{S}_q^r$ of group q form bi-level optimization:

$$\min_{\tau_q^r} \quad \underset{t_{end}\to t_q^{rt}}{SD}\left\{\left(\sum_{j=1}^{q} PD_j(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_j^r, \mathbf{S}_j^r, \tau_j^r, t)\right) - RC^*\right\} \tag{30}$$

$$\text{s.t.}\quad \tau_q^r > \tau_{q-1}^r \tag{31}$$

$$\tau_q^r < t_{end} + \mu \tag{32}$$

$$\mathbf{S}_q^r = \arg\min \quad \sum_{i=1}^{N^{\max}} \sum_{m=1}^{K} \left(1 - \frac{\tau_q^r - \tau_m^d}{DT_i^{\max}} \cdot \frac{\gamma_{m,i}^d}{\widehat{\gamma}_i^d}\right) \cdot s_{m,i}^d \cdot s_{q,i}^r \tag{33}$$

$$\text{s.t.}\quad \boldsymbol{\gamma}_q^r = -\boldsymbol{\gamma}_K^d \tag{34}$$

$$\left|RC_q^{r,*}\left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_q^r, \mathbf{S}_q^r, \tau_q^r\right) - BC_{q-1}^r\left(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\gamma}_{q-1}^r, \mathbf{S}_{q-1}^r, \tau_{q-1}^r\right)\right| < \underline{p} \tag{35}$$

$$\underline{p} = \min\left(p_i \middle| \sum_{m=1}^{K} s_{m,i}^d \ne \sum_{l=1}^{q-1} s_{l,i}^r, i = 1, 2, \ldots, N^{\max}\right) \tag{36}$$

$$\mathbf{S}_q^r \le \sum_{m=1}^{K} \mathbf{S}_m^d - \sum_{l=1}^{q-1} \mathbf{S}_l^r \tag{37}$$

The high-level problem (30)–(32) optimize the recovery time instant of group q, which is similar to (20)–(22). μ denotes the duration in which the recovery process has to be finished. Therefore, the high-level optimization is limited within the period $[t_{end}, t_{end} + \mu]$. The low-level problem (33)–(37) optimize the selection of ACs in group q so that the ACs with the lowest thermal comfort levels are recovered earlier. Equation (34) resets the set point temperature to its original value. In other words, the aggregate power of ACs cannot be flexibly controlled with changing set point temperature. Therefore, (35) represents that the rebound load of group q–1 is mitigated entirely by the reserve capacity of group q when their difference is smaller than

TABLE I
AC PHYSICAL PARAMETERS

Parameters	Descriptions	Values	Units
A_i	Room area	$N(20,25)$ [55]	m^2
C_i	Thermal capacity	$0.015 \cdot A_i$ [33]	kWh/°C
R_i	Thermal resistance	$100 \cdot A_i^{-1}$ [33]	°C/kW
COP_i	Coefficient of performance	$-0.0384\lvert\theta_a(t)-\theta_i(t)\rvert$ $+3.9051$ [56]	/
p_i	Input power	$U(40 \cdot A_i\ ,\ 70 \cdot A_i)$ [55]	W
T_{set}	Set point temperature	$U(23,28)$	°C

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

$\underline{p}$, which denotes the minimum power of the remaining ACs and is calculated as (36). Equation (37) constraints ACs in group q within dispatched ACs that have not been recovered.

Considering that the operation process of an individual AC is described by the general state model for TCLs in (1), the proposed method may be applied to other TCLs, such as refrigerators, heat pump space heaters and electric water heaters [50]. Statistical data have shown that TCLs account for 48%, 35%, 40% and 51% of residential electricity consumption in the U.S. [51], the UK [52], Australia [53] and China [54], respectively. Therefore, the proposed method could be applied to different regions considering the widespread of TCLs. Among all the common TCLs, the cycle time of AC is relatively short, leading to a short deployment duration constrained by the lead rebound. Therefore, AC is taken as a typical type of TCL to show the effect of the lead-lag rebound and the effectiveness of the proposed strategy.

V. CASE STUDIES AND SIMULATION RESULTS

Case studies are conducted to validate the effectiveness of the sequential dispatch strategy for providing operating reserve with various duration time. First, the potential of ACs for the provision of operating reserve is evaluated by the sequential dispatch of ACs without the recovery program. Second, the performance of the sequential dispatch and recovery strategy is verified by the reserve deployment of ACs with various required reserve capacity and duration time. Furthermore, operating reserve provided by ACs are used to relieve congestion resulted from system peak load in IEEE-30-bus test system. On this basis, different dispatch strategies of ACs in existing research studies are compared to verify the necessity of mitigating lead-lag rebound with the proposed sequential dispatch strategy.

An aggregator with controllable ACs in a residential area in summer, which is when all of the ACs operate in cooling mode, is modeled. The simulation parameters are illustrated in Table I. The operation parameters of ACs are generated from a pilot study which obtained spinning reserve from responsive air conditioning loads at a motel over a year [55]. The coefficient of performance COP_i of the i-th AC is set according to [56], which generates this parameter from AC operating data

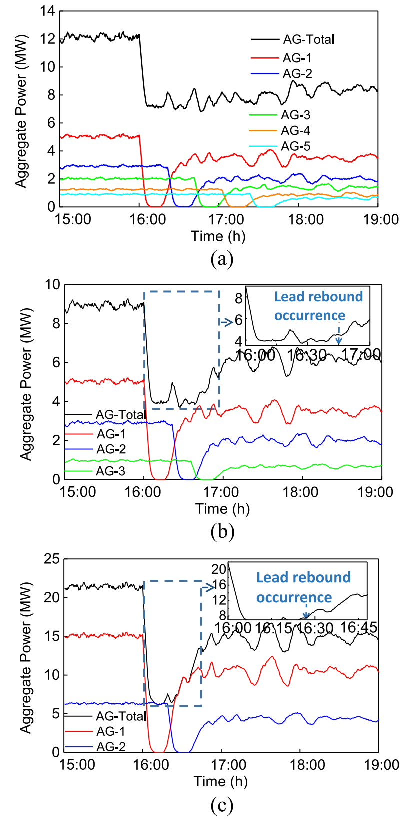


Fig. 7. Operating reserve provided by ACs when receiving reserve deployment signal at 16:00. (a) $RC^* = 5$ MW $N^{\max} = 60{,}000$. (b) $RC^* = 5$ MW $N^{\max} = 25{,}000$. (c) $RC^* = 15$ MW $N^{\max} = 60{,}000$.

published by Bosch Termoteknik. Thermal parameters of rooms are set according to [33], which lists the bulk thermal properties of buildings from measurement data in experimental studies. It is assumed that aggregators are permitted to control the ACs for at least one hour by contract. The ambient temperature is set as 32° C. The temperature dead band and the maximum changes of set point temperature are set to 1° C and 2° C, respectively. The proportional coefficient $\beta\%$ in (17) is set as 10%.

A. Evaluation of ACs' Potential for the Provision of Operating Reserve

This case simulates the reserve deployment of ACs without the recovery program, through which the maximum duration time corresponding to the required reserve capacity RC^* can be observed. ACs are dispatched to provide load-reduction service at the time 16:00. The dynamics of the sequential dispatch process of $N^{\max}$ ACs are demonstrated in Fig. 7. The aggregate power of all dispatched ACs is labeled as AG-Total, below which the aggregate power of the g-th group is labeled as AG-g. The number of ACs and the dispatch time instant of the g-th group τ_g are shown in Table II.

The curve of AG-Total in Fig. 7(a) shows that aggregate power of all the dispatched ACs decreases from 12 MW to 7 MW after receiving the reserve deployment signal and maintains at 7 MW since then. Hence, the lead rebound is eliminated entirely after the sequential dispatch of five groups of ACs (shown by AG-1

TABLE II
DISPATCH RESULTS OF ACS DURING THE RESERVE DEPLOYMENT PERIOD

RC^*/ N^{max}	AG-No.	Number	τ_g	AG-No.	Number	τ_g
5MW/60,000	1	14,143	16:00	2	8,143	16:22
	3	5,663	16:41	4	3,518	17:03
	5	2,524	17:24	-	-	-
5MW/25,000	1	14,143	16:00	2	8,143	16:22
	3	2,714	16:36	-	-	-
15MW/60,000	1	42,279	16:00	2	17,721	16:19

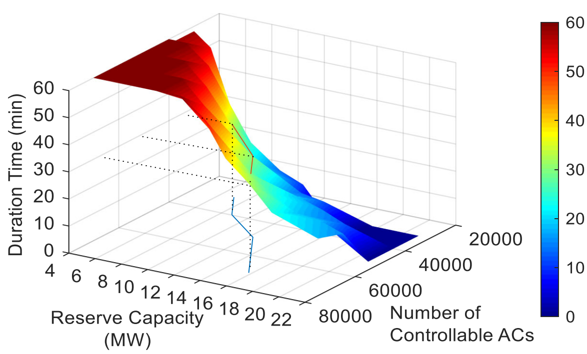


Fig. 8. Maximum duration time corresponding to different numbers of controllable ACs and different required reserve capacity.

to AG-5) when $N^{\max}$ is 60,000. By contrast, RC^* in Fig. 7(a) is the same with that in Fig. 7(b), while $N^{\max}$ of the latter is 35,000 smaller than the former. Table II shows that the dispatch results of AG-1 and AG-2 are the same. However, all the remaining ACs in Fig. 7(b) are utilized in AG-3 and the curve of AG-Total in Fig. 7(b) shows that the aggregate power of dispatched ACs rebounds at around 16:45. Hence, 25,000 controllable ACs are not sufficient to mitigate the rebound load entirely and the maximum duration time is only about 45 minutes. On the other hand, $N^{\max}$ in Fig. 7(a) and Fig. 7(c) is the same and equals to 60,000, while RC^* of the latter is 10 MW higher than the former. The curve of AG-Total in Fig. 7(c) shows that the aggregate power of dispatched ACs rebounds at around 16:25. Fig. 7(c) and Table II show that after the dispatch of two groups (AG-1 and AG-2), all the controllable ACs are utilized. Therefore, 60,000 controllable ACs are not sufficient to mitigate the rebound load entirely when RC^* is 15 MW and the maximum duration time is only about 25 minutes. Consequently, there exists constraints between the feasible reserve capacity and feasible deployment duration, both of which are also limited by total number of controllable ACs.

In order to evaluate ACs' potential for the provision of operating reserve, the maximum duration time $DT^{\max}$ corresponding to different numbers of controllable ACs and RC^* is simulated, as demonstrated in Fig. 8. In this way, the feasibility ranges of operating reserve are the areas below the surface in Fig. 8. The simulation cases are conducted on a PC with Intel 2.3 GHz 2-core processor (4MB L3 cache), 8 GB memory. The computational time of the sequential dispatch and recovery process when $N^{\max}$ controllable ACs are required to provide reserve capacity of RC^* for $DT^{\max}$ is shown in Fig. 9. Since all the ACs are assumed to be controllable for at least one hour, the

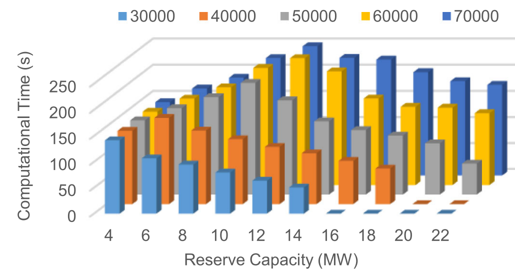


Fig. 9. Computational time corresponding to different numbers of controllable ACs and different required reserve capacity.

longest deployment duration is set as 60 min, after which the sequential dispatch procedure will stop. For a given number of ACs, if the maximum duration time in Fig. 8 equals to 60 min, the computational time in Fig. 9 increases with RC^* because more ACs are dispatched with larger RC^*. On the other hand, if the maximum duration time in Fig. 8 is smaller than 60 min, all the controllable ACs are dispatched to fulfill the requirement of RC^*. The computational time in Fig. 9 decreases with RC^* because less groups of ACs will be dispatched/recovered with larger RC^*. Hence, the computational time reaches the maximum value when DT^* has just reached 60 min. The maximum computational time in Fig. 9 is 248 s, which is corresponding to the provision of 12 MW reserve capacity for 60 min with 70,000 ACs.

For a given number of ACs, the maximum duration time decreases with the increase of RC^*. Consequently, reserve capacity in Fig. 8 reaches the maximum value when the maximum duration time equals to the required value DT^*. Assume that the required duration time is 30 min [31], the maximum reserve capacity corresponding to the number of controllable ACs is plotted on the bottom of Fig. 8, which shows that the maximum reserve capacity is 14.09 MW when there are 60,000 ACs. Hence, when RC^* is lower than 14.09 MW, ACs can fulfill the requirements of duration time, as illustrated in Fig. 7(a). By contrast, when RC^* is 15 MW, the duration time is not enough, as illustrated in Fig. 7(c). Therefore, the maximum reserve capacity can be effectively evaluated according to the expected duration time and total number of controllable ACs. Such evaluation is helpful for the selection of RC^* and DT^* in the following cases.

B. Provision of Operating Reserve With Various Duration Time and Reserve Capacity

In this case, the ACs are required to provide operating reserve with a specified reserve capacity RC^* and a specified duration time DT^*. Hence, ACs should ensure that the lead rebound is mitigated entirely during DT^*, after which the recovery program will be triggered. The maximum number of controllable ACs is set as 60,000. The dynamics of reserve deployment with the recovery process are shown in Fig. 10, in which the dispatch results during the reserve deployment period and the recovery period are separated by the dotted line. The room temperature profile corresponding to the load control in Fig. 10(a)–(e) are shown in Fig. 11(a)–(e), respectively. The indices for the simulated case are presented in Table III. Standard deviation *SD* and

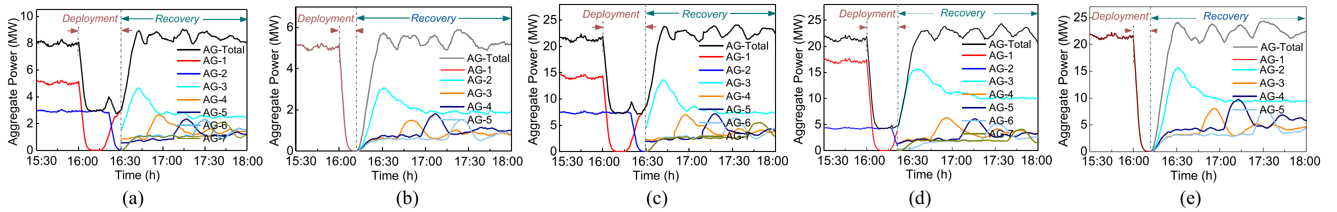


Fig. 10. Sequential dispatch and recovery of ACs for the provision of operating reserve with various RC^* and DT^*. (a) $RC^* = 5\,\text{MW}$, $DT^* = 0.5\,\text{h}$. (b) $RC^* = 5\,\text{MW}$, $DT^* = 0.2\,\text{h}$. (c) $RC^* = 14\,\text{MW}$, $DT^* = 0.5\,\text{h}$. (d) $RC^* = 17\,\text{MW}$, $DT^* = 0.36\,\text{h}$. (e) $RC^* = 21\,\text{MW}$, $DT^* = 0.2\,\text{h}$.

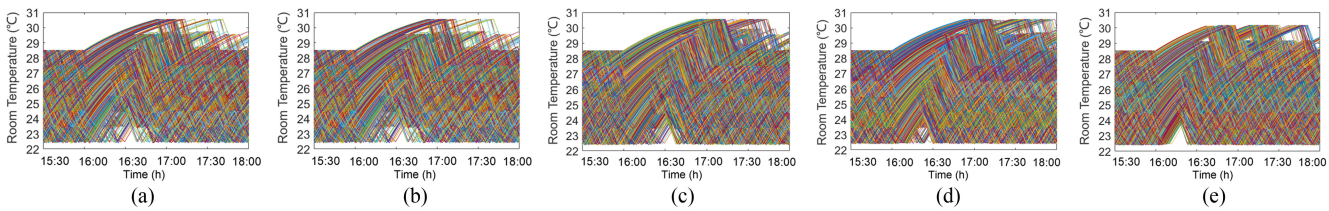


Fig. 11. The room temperature profile of ACs for the provision of operating reserve with various RC^* and DT^*. (a) $RC^* = 5\,\text{MW}$, $DT^* = 0.5\,\text{h}$. (b) $RC^* = 5\,\text{MW}$, $DT^* = 0.2\,\text{h}$. (c) $RC^* = 14\,\text{MW}$, $DT^* = 0.5\,\text{h}$. (d) $RC^* = 17\,\text{MW}$, $DT^* = 0.36\,\text{h}$. (e) $RC^* = 21\,\text{MW}$, $DT^* = 0.2\,\text{h}$.

TABLE III
INDICES OF THE OPERATING RESERVE WITH VARIOUS RC^* AND DT^*

Instruction (RC^*/DT^*)	SD (MW)	PV (%)	RT_d (min)	RT_r (min)	RR_d (MW/min)	RR_r (MW/min)
5MW/0.5h	0.30	3.75	5.03	8.99	0.99	0.56
5MW/0.2h	0.29	5.86	4.92	12.08	1.02	0.41
14MW/0.5h	0.83	3.96	5.98	8.52	2.81	1.64
17MW/0.36	0.88	4.19	5.40	9.67	3.15	1.76
21MW/0.2h	1.20	5.71	5.11	12.19	4.11	1.74

TABLE IV
DISPATCH RESULTS OF ACS DURING THE RESERVE DEPLOYMENT PERIOD AND THE RECOVERY PERIOD

RC^*/DT^*	Period	AG-No.	Number	τ_g	AG-No.	Number	τ_g
5MW/ 0.5h	Deployment	1	14,143	16:00	2	8,143	16:21
	Recovery	3	7,329	16:30	4	4,031	16:49
		5	3,435	17:08	6	3,804	17:21
		7	3,687	17:34	-	-	-
5MW/ 0.2h	Deployment	1	14,143	16:00	-	-	-
	Recovery	2	6,013	16:12	3	2,605	16:40
		4	3,050	17:12	5	2,475	17:14
14MW/ 0.5h	Deployment	1	39,431	16:00	2	20,569	16:22
	Recovery	3	20,673	16:30	4	10,099	16:50
		5	10,104	17:09	6	9,868	17:22
		7	9,256	17:39	-	-	-
17MW/ 0.36h	Deployment	1	48,014	16:00	2	11,986	16:18
	Recovery	3	28,166	16:22	4	9,299	16:48
		5	8,601	17:08	6	7,446	17:23
		7	6,498	17:42	-	-	-
21MW/ 0.2h	Deployment	1	60,000	16:00	-	-	-
	Recovery	2	24,354	16:11	3	10,942	16:45
		4	14,092	17:02	5	10,612	17:22

power volatility PV are the values for the load recovery process. The threshold of valid reserve capacity calculated by (8) is set between $0.9 \cdot PD_g^{\max}$ and $PD_g^{\max}$. The number of ACs and the dispatch/recovery time instant of the g-th group τ_g are shown in Table IV.

Aggregate power of all dispatched ACs maintains at the reduced value during the reserve deployment period and returns the initial value steadily after receiving the recovery signal, as is illustrated by the curve of AG-Total in Fig. 10. This means that both the lead rebound and lag rebound are mitigated entirely by dispatching different AC groups in sequence. Dynamics of reserve deployment are various with different DT^* and RC^*. On the one hand, RC^* in Fig. 10(a) and (b) is the same (5 MW), while DT^* of the latter is 0.3 h lower than the former. Table IV shows that only 14,143 ACs dispatched at 16:00 are enough to realize the deployment duration of 0.2 h in Fig. 10(b), while another 8,143 ACs are dispatched at 16:21 in Fig. 10(a) to extend the deployment duration to 0.5 h. Hence, ACs are divided into more groups to be recovered in Fig. 10(a), leading to the decrease of PV by $2.11\% \ (= 5.86\% - 3.75\%)$ compared to that in Fig. 10(b), as shown in Table III. Fig. 11(a) and (b) show that the reduced deployment duration in Fig. 10(b) leads to the decreased changes of set point temperature by approximately 0.5° C compared to that in Fig. 10(a). However, the ramp time of Fig. 10(b) during the recovery period is $3.09\,\text{min} \ (= 12.08\text{min} - 8.99\text{min})$ longer than that in Fig. 10(a). This is because ACs in Fig. 10(b) have just reached the new temperature hysteresis band when they are recovered to the initial states and therefore it takes longer to migrate to the initial temperature hysteresis band according to the rule of SP-2 [34].

On the other hand, DT^* in Figs. 10(a) and (c) is the same (0.5 h), while RC^* of the latter is 9 MW higher than the former. It can be seen from Table IV that ACs are divided into two groups to be dispatched and five group to be recovered both in Fig. 10(a) and (c). Table III shows that such increase of RC^* leads to the increase of SD from 0.30 MW to 0.83 MW, while PV is similar. All of the controllable ACs are utilized for the provision of operating reserve in Fig. 10(c)–(e), where the percentage of power reduction are 63.6%, 80% and 100%, respectively. DT^* in Fig. 10(c)–(e) is set as the maximum deployment duration obtained from Fig. 8. Similar to Fig. 10(b), ACs are divided into less group

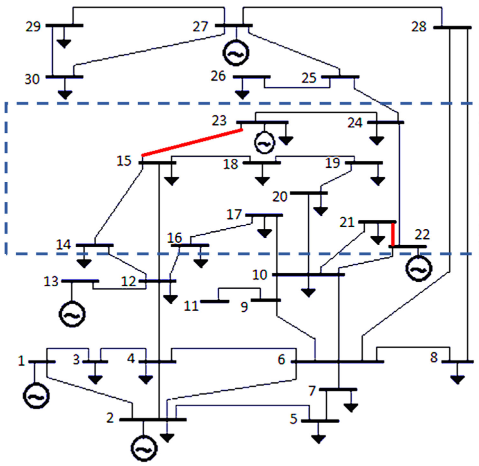


Fig. 12. Diagram of IEEE-30 bus system.

to be recovered during the recovery period in Fig. 10(e), leading to the increase of *PV* by 1.75% (= 5.71% − 3.96%) than that in Fig. 10(c) and 1.52% (= 5.71% − 4.19%) than that in Fig. 10(d). Hence, shorter duration time and larger reserve capacity will lead to larger fluctuations of aggregate power. As a result, the possible reserve capacity and duration time should be constrained by the maximum power fluctuations, which is quantified by *SD* and *PV*. Moreover, Table III shows that the ramp time of reserve deployment is within 5.98 min, which can fulfill the requirement on ramp time of 10-min spinning reserve, 30-min spinning reserve, etc. [31]. The ramp time during the recovery period is within 12.19 min, which is also shorter than the maximum limit (between 15 min to 90 min) for different types of operating reserve [57]. Therefore, the sequential dispatch strategy of ACs can mitigate both the lead rebound and lag rebound entirely, which enables flexible control of the duration time and reserve capacity to fulfill the requirements of different types of operating reserve.

C. Comparison of Different Dispatch Strategy of ACs for the Provision of Operating Reserve

In order to validate the necessity of mitigating both the lead rebound and lag rebound, the operating reserve provided by ACs is utilized to relieve congestion resulted from the system peak load in IEEE-30-bus test system [58], diagram of which is shown in Fig. 12. To obtain the load patterns for the summer day, 1.23% of the historical hourly load in the COAST weather zone in Electric Reliability Council of Texas (ERCOT) in 24th August 2017 is utilized to generate the total system load [59]. In this way, the peak of total system load equals to approximately 120% of the load demand in standard IEEE-30-bus test system [58]. The profile of total system load [59] and the corresponding ambient temperature [60] are shown in Fig. 13.

Figure 13 shows that the system peak load exists at around 15:00. At 16:00, ACs located from bus 14 to 24 provide operating reserve of 14 MW as illustrated by Fig. 10(c), which accounts for 15% of the electricity consumption in these buses. The operating reserve provided by ACs is replaced by 30-min operating reserve at 16:30, after which ACs are recovered to

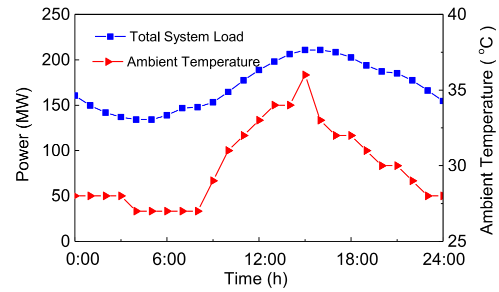


Fig. 13. Total system load and ambient temperature at each time instant.

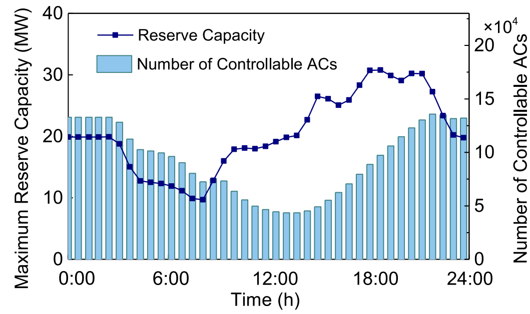


Fig. 14. Number of controllable ACs and the potential reserve capacity at each time instant.

the initial states. The total number of controllable ACs located from bus 14 to bus 24 is generated from the consumer travel habit data collected by National Household Travel Survey [61] and is shown by the bars in Fig. 14. The total aggregate power corresponding to the controllable ACs is shown by the square-scattered lines in Fig. 14. The distribution of controllable ACs located from bus 14 to bus 24 is proportional to the base load in these buses [58]. Locational marginal price (LMP) at each bus is evaluated by optimal power flow (OPF).

The proposed sequential dispatch strategy (SDS) is compared with the four other methods, which includes: 1) GDS [17], [27]: The concept of grouping devices to reduce the lag rebound in existing literatures, which divide ACs into several groups and recover them at a regular time interval. In this case, ACs are divided into three groups to be dispatched every 10 min and divided into five groups to be recovered every 10 min. In addition, the control of ACs in each group also follows safe protocol-2, which is the same as the proposed SDS; 2) RDS [13]: Randomizing the deployment/recovery of ACs over time, which is the most common way to mitigate the demand response rebound in existing researches. In this case, the deployment and recovery of ACs are randomized between 0 and 10 min; 3) SP-2 [36]: The safe protocol-2 to avoid synchronization of ACs, which reduces the power fluctuations and the level of demand response rebound; 4) CDS [62]: Traditional centralized dispatch strategy in which all the ACs are deployed/recovered when receiving reserve deployment/recovery signal instantly.

The total system load profile after the reserve deployment of ACs controlled by different dispatch strategies at 16:00 is shown in Fig. 15. The indices for the simulated cases are presented in Table V. Branch *m*-*n* denotes the branch between bus *m* and bus *n*. The branch loading index (BLI) in branch 21-22 and branch

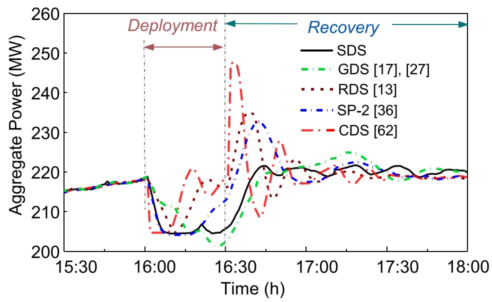


Fig. 15. Total system load after the reserve deployment of ACs controlled by different dispatch strategies at 16:00.

TABLE V
INDICES OF THE OPERATING RESERVE PROVIDED BY ACS WITH DIFFERENT DISPATCH STRATEGIES

Dispatch strategies	DT (min)	BC_d (MW)	BC_r (MW)	RT_d (min)	RT_r (min)
Proposed SDS	30.00	0	0	5.98	8.52
GDS [17], [27]	30.00	0	5.22	23.40	8.68
RDS [13]	15.11	12.02	16.71	9.73	7.03
SP-2 [36]	19.27	6.13	17.89	4.87	8.95
CDS [62]	8.93	15.63	30.45	0.62	0.55

15-23 are shown by curves in Fig. 16. LMP in 5-minute intervals in bus 21 is shown by the bars in Fig. 16.

The profile of BLI in Fig. 16 shows that congestion exists in branch 21-22 and branch 15-23 since 15:00 because of the system peak load, leading to the increase of LMP from 44$/MW to 72$/MW. The deployment of operating reserve provided by ACs at 16:00 relieves the congestion and reduces the LMP from 72$/MW to 44$/MW. Table V and the deployment segment from 16:00 to 16:30 in Fig. 15 show that the lead rebound is mitigated entirely in SDS. Accordingly, LMP in Fig. 16(a) remains at the level of 44$/MW after 16:00.

Similar to SDS, GDS can also mitigate the lead rebound entirely. However, the ramp time prolongs to 23.40 min, which cannot fulfill the requirement of many types of operating reserves (e.g., 10-min spinning reserve). By contrast, RDS, SP-2 and CDS cannot mitigate the lead rebound entirely and the value reaches 12.02 MW, 6.13 MW and 15.63 MW, respectively. Compared to CDS and RDS, almost all the power fluctuations are removed by SP-2. However, SP-2 cannot entirely mitigate the lead rebound either and the aggregate power still rebound within the required duration time. Consequently, the actual duration time is only 19.27 min, which is shorter than DT^* (30 min). Congestion still exists during the reserve deployment period and the LMP increases to the level of 72$/MW again, as shown in Fig. 16(c)–(e). Therefore, it is crucial to mitigate the lead rebound entirely so that the duration time can be flexibly controlled.

On the other hand, Table V and the recovery segment from 16:30 to 18:00 in Fig. 15 show that the lag rebound is also mitigated entirely by SDS, but cannot be mitigated entirely by RDS, SP-2 and CDS, whose lag rebounds reach 16.71 MW, 17.89 MW and 30.45 MW, respectively. As a result, the aggregate power attained by RDS, SP-2 and CDS are very large,

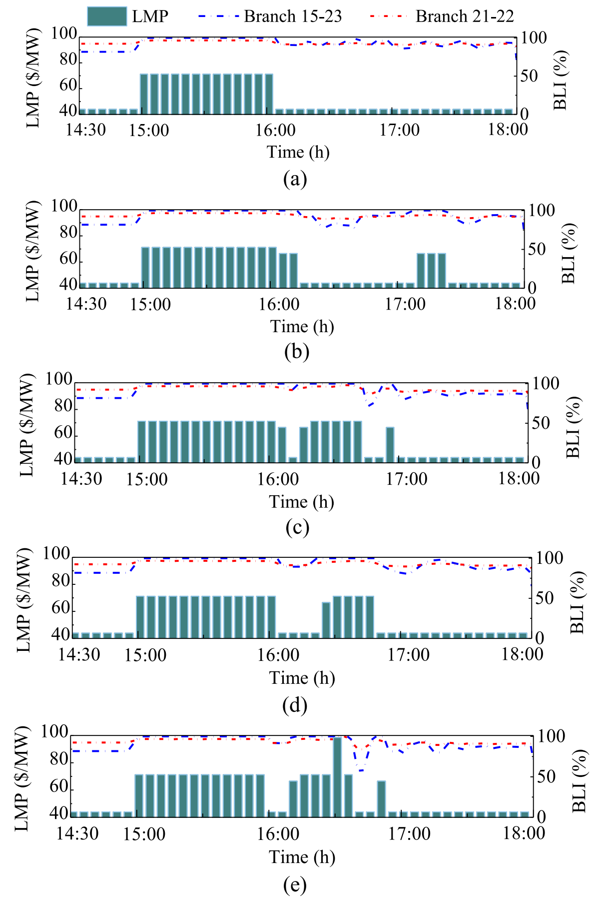


Fig. 16. Locational marginal price in 5-minute intervals in bus 21 and branch loading index of the congestion branches corresponding to different control strategies. (a) SDS. (b) GDS. (c) RDS. (d) SP-2. (e) CDS.

resulting in the increase of LMP to over 72$/MW. Because of lacking the co-optimization among the reserve capacity and dispatch time instant of different AC groups in GDS, a rebound peak of 5.22 MW still exists at around 17:10, leading to the increase of LMP to around 67$/MW. It is required to mention that although the ramp time in SDS and SP-2 is a little longer than that in CDS and RDS, it can fulfill the requirements of most types of operating reserve. Therefore, the proposed SDS is better than the other dispatch strategies on the entire mitigation of the lead-lag rebound attained by the capacity-time co-optimization during the reserve deployment/recovery process.

VI. CONCLUSION

This paper presents a novel sequential dispatch strategy of ACs for the provision of the operating reserve. The impacts of the lead-lag rebound on the capacity dimension and the time dimension are quantified by a proposed evaluation framework. Illustrative results demonstrate that the sequential dispatch strategy and recovery algorithm enable ACs to provide operating reserve with multiple duration time. The maximum reserve capacity and the corresponding feasible duration time range are constrained by parameters including the total number of ACs and the power volatility limit. Aggregators should carefully balance these constraints to determine the reserve capacity and duration time. Moreover, the comparison of the proposed strategy with other methods illustrates that the proposed capacity-time co-optimization among different AC groups enable the entire mitigation of the lead-lag rebound. In this way, ACs can be utilized to relieve congestion or reduce peak load without adding additional burden to the power system.

Acknowledgment

The authors would like to thank all reviewers and editors who provided valuable comments for improving the quality of this paper. The suggestions for the technique descriptions of the proposed strategy and the formulation of case studies are greatly appreciated.

References

[1] 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.

[2] N. Zhang, C. Kang, Q. Xia, and J. Liang, "Modeling conditional forecast error for wind power in generation scheduling," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1316–1324, May 2014.

[3] H. Zhong, Q. Xia, C. Kang, M. Ding, J. Yao, and S. Yang, "An efficient decomposition method for the integrated dispatch of generation and load," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 2923–2933, Nov. 2015.

[4] C. Gao, Q. Li, and Y. Li, "Bi-level optimal dispatch and control strategy for air-conditioning load based on direct load control," *Proc. Chin. Soc. Elect. Eng.*, vol. 34, pp. 1546–1555, 2014.

[5] M. Izquierdo, A. Moreno-Rodríguez, A. González-Gil, and N. García-Hernando, "Air conditioning in the region of Madrid, Spain: An approach to electricity consumption, economics and CO_2 emissions," *Energy*, vol. 36, no. 3, pp. 1630–1639, Mar. 2011.

[6] S. Cox, "Cooling a warming planet: A global air conditioning surge," 2012. [Online]. Available: https://e360.yale.edu/features/cooling_a_warming_planet_a_global_air_conditioning_surge

[7] P. Siano, C. Cecati, H. Yu, and J. Kolbusz, "Real time operation of smart grids via FCN networks and optimal power flow," *IEEE Trans. Ind. Inf.*, vol. 8, no. 4, pp. 944–952, Nov. 2012.

[8] C. Ju, P. Wang, L. Goel, and Y. Xu, "A two-layer energy management system for microgrids with hybrid energy storage considering degradation costs," *IEEE Trans. Smart Grid*, to be published. doi: 10.1109/TSG.2017.2703126

[9] T. Jin, C. Kang, and H. Chen, "Integrating consumer advance demand data in smart grid energy supply chain," in *Smart Grids: Clouds, Communications, Open Source, and Automation*. Boca Raton, FL, USA: CRC Press, 2014, pp. 251–274.

[10] V. Trovato, F. Teng, and G. Strbac, "Role and benefits of flexible thermostatically controlled loads in future low-carbon systems," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5067–5079, Sep. 2018.

[11] C. H. Wai, M. Beaudin, H. Zareipour, A. Schellenberg, and N. Lu, "Cooling devices in demand response: A comparison of control methods," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 249–260, Jan. 2015.

[12] 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, Sep. 2016.

[13] A. Abiri-Jahromi and F. Bouffard, "Contingency-type reserve leveraged through aggregated thermostatically-controlled loads–Part I: Characterization and control," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1972–1980, May 2016.

[14] N. Lu and S. Katipamula, "Control strategies of thermostatically controlled appliances in a competitive electricity market," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, San Francisco, CA, USA, 2005, pp. 202–207.

[15] N. G. Paterakis, M. Gibescu, A. G. Bakirtzis, and J. P. S. Catalão, "A multi-objective optimization approach to risk-constrained energy and reserve procurement using demand response," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 3940–3954, Jul. 2018.

[16] D. T. Nguyen, M. Negnevitsky, and M. de Groot, "Modeling load recovery impact for demand response applications," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1216–1225, May 2013.

[17] N. Saker, M. Petit, and J. L. Coullon, "Demand side management of electrical water heaters and evaluation of the cold load pick-up characteristics (CLPU)," in *Proc. IEEE Trondheim PowerTech*, 2011, pp. 1–8.

[18] Y. W. Law, T. Alpcan, V. C. S. Lee, A. Lo, S. Marusic, and M. Palaniswami, "Demand response architectures and load management algorithms for energy-efficient power grids: A survey," in *Proc. 7th Int. Conf. Knowl. Inf. Creativity Support Syst.,* Melbourne, VIC, Australia, Nov. 2012, pp. 134–141.

[19] T. Ericson, "Direct load control of residential water heaters," *Energy Policy*, vol. 37, no. 9, pp. 3502–3512, Sep. 2009.

[20] P. Siano and D. Sarno, "Assessing the benefits of residential demand response in a real time distribution energy market," *Appl. Energy*, vol. 161, pp. 533–551, Jan. 2016.

[21] Q. Wu, P. Wang, and L. Goel, "Direct load control (DLC) considering nodal interrupted energy assessment rate (NIEAR) in restructured power systems," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1449–1456, Aug. 2010.

[22] L. Goel, Q. Wu, and P. Wang, "Fuzzy logic-based direct load control of air conditioning loads considering nodal reliability characteristics in restructured power systems," *Elect. Power Syst. Res.*, vol. 80, no. 1, pp. 98–107, Jan. 2010.

[23] C. M. Chu and T. L. Jong, "A novel direct air-conditioning load control method," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1356–1363, Aug. 2008.

[24] B. J. Kirby, "Load response fundamentally matches power system reliability requirements," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, 2007, pp. 1–6.

[25] P. Grünewald and J. Torriti, "Demand response from the non-domestic sector: Early UK experiences and future opportunities," *Energy Policy*, vol. 61, pp. 423–429, Oct. 2013.

[26] N. Motegi, M. A. Piette, D. S. Watson, S. Kiliccote, and P. Xu, "Introduction to commercial building control strategies and techniques for demand response—Appendices," Lawrence Berkeley Nat. Lab., Berkeley, CA, USA, Rep. LBNL-59975, May 2007.

[27] H. Johal, K. Anaparthi, and J. Black, "Demand response as a strategy to support grid operation in different time scales," in *Proc. IEEE Energy Convers. Congr. Expo.*, 2012, pp. 1461–1467.

[28] B. M. Sanandaji, T. L. Vincent, and K. Poolla, "Ramping rate flexibility of residential HVAC loads," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 865–874, Apr. 2016.

[29] X. Wu, J. He, Y. Xu, J. Lu, N. Lu, and X. Wang, "Hierarchical control of residential HVAC units for primary frequency control," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3844–3856, Jul. 2018.

[30] N. Lu, "An evaluation of the HVAC load potential for providing load balancing service," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1263–1270, Sep. 2012.

[31] J. Wang, M. Shahidehpour, and Z. Li, "Contingency-constrained reserve requirements in joint energy and ancillary services auction," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1457–1468, Aug. 2009.

[32] J. F. Ellison, L. S. Tesfatsion, V. W. Loose, and R. H. Byrne, "Project report: A survey of operating reserve markets in U.S. ISO/RTO-managed electric energy regions," Sandia Nat. Lab., Livermore, CA, USA, Rep. SAND2012-1000, 2012.

[33] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Convers. Manage.*, vol. 50, no. 5, pp. 1389–1400, May 2009.

[34] C. Shao, Y. Ding, J. Wang, and Y. Song, "Modeling and integration of flexible demand in heat and electricity integrated energy system," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 361–370, Jan. 2018.

[35] D. Xie, H. Hui, Y. Ding, and Z. Lin, "Operating reserve capacity evaluation of aggregated heterogeneous TCLs with price signals," *Appl. Energy*, vol. 216, pp. 338–347, Apr. 2018.

[36] N. A. Sinitsyn, S. Kundu, and S. Backhaus, "Safe protocols for generating power pulses with heterogeneous populations of thermostatically controlled loads," *Energy Convers. Manage.*, vol. 67, pp. 297–308, Mar. 2013.

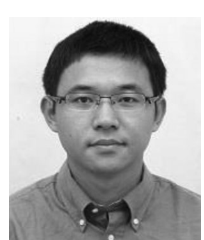
[37] A. Zipperer *et al.*, "Electric energy management in the smart home: Perspectives on enabling technologies and consumer behavior," *Proc. IEEE*, vol. 101, no. 11, pp. 2397–2408, Nov. 2013.

[38] Q. Hu, F. Li, X. Fang, and L. Bai, "A framework of residential demand aggregation with financial incentives," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 497–505, Jan. 2018.
[39] Q. Wang, C. Zhang, Y. Ding, G. Xydis, J. Wang, and J. Østergaard, "Review of real-time electricity markets for integrating distributed energy resources and demand response," *Appl. Energy*, vol. 138, pp. 695–706, Jan. 2015.
[40] J. Wang, C. N. Bloyd, Z. Hu, and Z. Tan, "Demand response in China," *Energy*, vol. 35, no. 4, pp. 1592–1597, Apr. 2010.
[41] M. Parvania and M. Fotuhi-Firuzabad, "Integrating load reduction into wholesale energy market with application to wind power integration," *IEEE Syst. J.*, vol. 6, no. 1, pp. 35–45, Mar. 2012.
[42] M. Shafie-khah and P. Siano, "A stochastic home energy management system considering satisfaction cost and response fatigue," *IEEE Trans. Ind. Inf.*, vol. 14, no. 2, pp. 629–638, Feb. 2018.
[43] Q. Hu and F. Li, "Hardware design of smart home energy management system with dynamic price response," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1878–1887, Dec. 2013.
[44] C. Perfumo, E. Kofman, J. H. Braslavsky, and J. K. Ward, "Load management: Model-based control of aggregate power for populations of thermostatically controlled loads," *Energy Convers. Manage.*, vol. 55, pp. 36–48, Mar. 2012.
[45] J. L. Mathieu, S. Koch, and D. S. Callaway, "State estimation and control of electric loads to manage real-time energy imbalance," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 430–440, Feb. 2013.
[46] N. Mehta, N. A. Sinitsyn, S. Backhaus, and B. C. Lesieutre, "Safe control of thermostatically controlled loads with installed timers for demand side management," *Energy Convers. Manage.*, vol. 86, pp. 784–791, Oct. 2014.
[47] B. Zhao, J. Liu, G. Zhang, and J. Su, "Recovery strategy and probability model for emergency control of thermostatically controlled load groups," in *Proc. IEEE Transp. Electrific. Conf. Expo, Asia-Pac.*, 2017, pp. 1–5.
[48] J. Bendtsen and S. Sridharan, "Efficient desynchronization of thermostatically controlled loads," *IFAC Proc. Vol.*, vol. 46, no. 11, pp. 245–250, Jan. 2013.
[49] E. G. Talbi, *Metaheuristics for Bi-Level Optimization*. Berlin, Germany: Springer-Verlag, 2013.
[50] J. L. Mathieu, M. Dyson, and D. S. Callaway, "Using residential electric loads for fast demand response: The potential resource and revenues, the costs, and policy recommendations," in *Proc. ACEEE Summer Study Build.*, 2012, pp. 189–203.
[51] U.S Energy Information Administration, "Residential energy consumption survey (RECS)," 2012. 2047. [Online]. Available: https://www.eia.gov/todayinenergy
[52] Z. Jean-Paul, E. Matt, G. Jonathan, and K. Nicola, "Household electricity survey," Dept. Environ. Food Rural Affairs, London, U.K., May 2012.
[53] Government of South Australia, "Home energy use." 2017. [Online]. Available: https://www.sa.gov.au/topics/energy-and-environment/using-saving-energy/home-energy-use
[54] X. Zheng *et al.*, "Characteristics of residential energy consumption in China: Findings from a household survey," *Energy Policy*, vol. 75, pp. 126–135, Dec. 2014.
[55] B. J. Kirby, "Spinning reserves from controllable packaged through the wall air conditioner (PTAC) units," Oak Ridge Nat. Lab., Oak Ridge, TN, USA, Rep. ORNL/TM-2002/286, Apr. 2003.
[56] 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, 2017.
[57] E. Ela, M. Milligan, and B. Kirby, "Operating reserves and variable generation," Nat. Renewable Energy Lab., Golden, CO, USA, Tech. Rep. NREL/TP-5500-51978, Aug. 2011.
[58] M. Shahidehpour and Y. Wang, "Appendix C: IEEE-30 bus system data," in *Communication and Control in Electric Power Systems: Applications of Parallel and Distributed Processing*. Hoboken, NJ, USA: Wiley, 2003.
[59] Electric Reliability Council of Texas, "Hourly load data archives," 2017. [Online]. Available: http://www.ercot.com/gridinfo/load/load_hist
[60] Time and Date AS, "Weather in August 2017 in Houston, Texas, USA." 2017. [Online]. Available: https://www.timeanddate.com/weather/usa/hous-ton/historic?month=8&year=2017
[61] U.S. Department of Transportation and Federal Highway Administration, "National household travel survey." 2009. [Online]. Available: http://nhts.ornl.gov.
[62] C. Perfumo, J. Braslavsky, and J. K. Ward, "A sensitivity analysis of the dynamics of a population of thermostatically-controlled loads," in *Proc. Australasian Univ. Power Eng. Conf.*, 2013, pp. 1–6.

Wenqi Cui received the B.Eng. degree in electrical engineering from Southeast University, Nanjing, China, in 2016. She is currently working toward the M.S. degree in electrical engineering from Zhejiang University, Hangzhou, China. Her research interests include smart grid and modeling and optimization of demand-side resources and the power market.

Yi Ding (M'12) received the bachelor's and Ph.D. degrees in electrical engineering from Shanghai Jiaotong University, Shanghai, China, and Nanyang Technological University, Singapore, in 2000 and 2007, respectively. He is currently a Professor with the College of Electrical Engineering, Zhejiang University, Hangzhou, China. His research interests include power systems reliability/performance analysis incorporating renewable energy resources, smart grid performance analysis, and engineering systems reliability modeling and optimization.

Hongxun Hui (S'17) received the bachelor's degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2015, where he is currently working toward the Ph.D. degree in electrical engineering. His research interests include smart grid, modeling and optimization of demand-side resources and the power market.



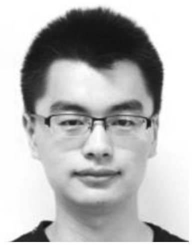
Zhenzhi Lin received the Ph.D. degree in electrical engineering from South China University of Technology, Guangzhou, China, in 2008. From 2007 to 2008, he was a Research Assistant with the Department of Electrical Engineering, The Hong Kong Polytechnic University. From 2010 to 2011, he was a Research Scholar with Min Kao Department of Electrical Engineering and Computer Science, The University of Tennessee. From 2013 to 2014, he was a Research Associate with the School of Engineering and Computing Sciences, Durham University. He is currently an Associate Professor with the School of Electrical Engineering, Zhejiang University, Hangzhou, China. His research interests include power system wide-area monitoring and control, controlled islanding, and power system restoration.

Pengwei Du (M'06–SM'12) received the Ph.D. degree in electric power engineering from Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently with the Electric Reliability Council of Texas, Taylor, TX, USA. Prior to this, he was a Senior Research Engineer with the Department of Energy, Pacific Northwest National Laboratory. He is an Editor for the IET Generation, Transmission & Distribution and the IEEE Transactions on Power Systems. He was the recipient of the IEEE PES Power System Dynamic Performance Committee Best Paper Prize in 2016.

Yonghua Song (M'90–SM'94–F'08) received the B.Eng. degree from Chengdu University of Science and Technology, Chengdu, China, in 1984, and the Ph.D. degree from China Electric Power Research Institute, Beijing, China, in 1989.

He is currently the President of the University of Macau, Macau, China, and an Adjunct Professor with the College of Electrical Engineering, Zhejiang University, Hangzhou, China, and the Department of Electrical Engineering, Tsinghua University, Beijing, China. From 1989 to 1991, he was a Postdoctoral Fellow with Tsinghua University. He then held various positions with Bristol University, Bristol, U.K., Bath University, Bath, U.K., and Liverpool John Moores University, Liverpool, U.K., from 1991 to 1996. In 1997, he was a Professor in power systems with Brunel University, Uxbridge, U.K., where he has been the Pro-Vice Chancellor of Graduate Studies since 2004. In 2007, he took up the Pro-Vice Chancellorship and Professorship of electrical engineering with the University of Liverpool, Liverpool, U.K. He was a Professor with the Department of Electrical Engineering, Tsinghua University, where he was an Assistant President and the Deputy Director with the Laboratory of Low Carbon Energy in 2009. From 2012 to 2017, he was the Executive Vice-President of Zhejiang University. His research interests include smart grid, electricity economics, and operation and control of power systems.

Prof. Song was elected the Vice-President of the Chinese Society for Electrical Engineering (CSEE) and appointed the Chairman of the International Affairs Committee of the CSEE in 2009. In 2004, he was elected a Fellow of the Royal Academy of Engineering, U.K. He was the recipient of the D.Sc. Award from Brunel University in 2002 for his original achievements in power systems research.



Changzheng Shao (S'17) received the B.S. degree in electric engineering from Shandong University. He is currently working toward the Ph.D. degree in electric engineering with Zhejiang University, Hangzhou, China. His research interests include operation and optimization of the integrated energy system and the power market.