Demand Response Load Following of Source and Load Systems

Hu, Jianqiang; Cao, Jinde; Yong, Taiyou; Guerrero, Josep M.; Z. Q. Chen, Michael; Li, Yaping

Published in:
I E E E Transactions on Control Systems Technology

DOI (link to publication from Publisher):
10.1109/TCST.2016.2615087

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

Demand Response Load Following of Source and Load Systems

Jianqiang Hu, Jinde Cao, *Fellow, IEEE*, Taiyou Yong, *Senior Member, IEEE*, Josep M. Guerrero, *Fellow, IEEE*, Michael Z. Q. Chen, *Senior Member, IEEE*, Yaping Li, *Member, IEEE*

***Abstract*—This paper presents a demand response load following strategy for an interconnected source and load system, in which we utilize traditional units and population of cooling thermostatically controlled loads (TCLs) to follow the mismatched power caused by the load activities and the renewable power injection in real time. In the demand side of power systems, these TCLs are often affiliated to a bus load agent and can be aggregated to multiple TCL aggregators. Firstly, aggregate evaluation of the TCL aggregator is carried out based on a bilinear aggregate model so as to derive the available regulation capacities and regulation rates of aggregators. Based on the evaluation results, the dispatch center optimizes the real time load following trajectories for the generating units and the flexible load agents via look-ahead optimization by considering the injection of renewable power. Furthermore, we mainly focused on the distributed control of multiple TCL aggregators. By proposing a distributed pinning control strategy and designing a spare communication network among the aggregators, the reference power tracking of the load agent can be shared by all aggregators inside it in a distributed way. Finally, simulation results on a modified IEEE-9 bus system are provided to demonstrate the effectiveness of the proposed load following strategy.**

***Index Terms*—Demand response, Load following, Distributed pinning control, Thermostatically controlled loads (TCLs).**

NOMENCLATURE

A. Notations for a single TCL

N_L	Number of TCLs in the load aggregator
$\theta(t)$	Internal temperature [°C]
$s(t)$	Operation state [on/off]
θ_a	Ambient temperature [°C]
δ_{db}	Temperature deadband
θ_{set}	Temperature setpoint
θ^+	Upper bound of the temperature deadband
θ^-	Lower bound of the temperature deadband
T_c	Steady state cooling time
T_h	Steady state heating time
θ_{set}^{des}	Preferred setpoint temperature
$\Delta\theta_{set}^{max}$	Upper bound of the comfort interval
$\Delta\theta_{set}$	Universal broadcast control signal

B. Notations for a single TCL aggregator

N_j	Number of TCL aggregators in jth load agent
$x(t)$	State variable for the bilinear aggregate model
$u(t)$	Control signal for the bilinear aggregate model
$y(t)$	Approximate aggregated power from the bilinear aggregate model
Q	Dimension of the state variable $x(t)$
$P_T(t)$	Direct aggregate power by Monte Carlo for the load aggregator
$\mu_1(\mu_2)$	Regulation gains for the virtual leader
β	Coupling strength for the distributed protocol

C. Notations for the joint real-time economic dispatch

T	Total number of periods in the scheduled time horizon
$G(L)$	Set of generating units (load agents)
$P_{G,i}^{(0)}$	Day-ahead generating scheduling for ith generating unit
$P_{L,j}^{(0)}$	Day-ahead estimated utilization for jth load agent
$\Delta P_{G,i}$	Real-time regulation value for ith generating unit
$\Delta P_{L,j}$	Real-time regulation value for jth load agent
$C_{g,i}$	Dispatch cost function for ith generating unit
$C_{l,j}$	Regulation cost function for jth load agent
P_D	Real-time load forecasting for the system
P_{pv}	Real-time photovoltaic power injection to the system
P_{wi}	Real-time wind power injection to the system
$P_{G,i}^{max}$	Upper bound for ith generating unit
$P_{G,i}^{min}$	Lower bound for ith generating unit
$R_{G,i}^{up}$	Upward ramping rate for ith generating unit
$R_{G,i}^{dn}$	Downward ramping rate for ith generating unit
$P_{L,j}^{max}$	Upper regulation limit for jth load agent
$P_{L,j}^{min}$	Lower regulation limit for jth load agent
$R_{L,j}^{up}$	Increase rate for jth load agent
$R_{L,j}^{dn}$	Decrease rate for jth load agent

D. Abbreviation

$\otimes$	Kronecker product
TCL	Thermostatically controlled load
DR	Demand response
EDP	Economic dispatch problem

This work was jointly supported in part by the State Grid Corporation of China under Project DZ71-13-045, in part by the Research Grants Council, Hong Kong, through the General Research Fund under Grant 17205414, the HKU CRCG Seed Funding Programme for Basic Research 201411159037, in part by the National Natural Science Foundation of China under Grant nos. 61374053, 61573096 and 61272530, and the 333 Engineering Foundation of Jiangsu Province of China under Grant no. BRA2015286.

J. Hu and J. Cao are with Research Center for Complex Systems and Network Sciences, and Department of Mathematics, Southeast University, Nanjing 210096, China. E-mail: jqhuseu@gmail.com, jdcao@seu.edu.cn.

T. Yong and Y. Li are with China Electric Power Research Institute, Nanjing 210003, China. E-mail: yongtaiyou@epri.sgcc.com.cn.

J. M. Guerrero is with the Department of Energy Technology, Aalborg University, Aalborg DK-9220, Denmark. E-mail: joz@et.aau.dk.

M. Z. Q. Chen is with the Department of Mechanical Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China. E-mail: mzqchen@hku.hk.

I. INTRODUCTION

DEMAND response (DR) in smart grids [1] has been proposed and investigated for several decades [2], [3], [4], which can be served as a utility manager to redistribute the electricity demand for a certain period of time, e.g., time-of-day, day-of-week. Several control techniques have been proposed for DR management, such as, direct load control [5], dynamic demand control [6], real-time pricing control [7], [8]. However, the basic problem for DR is how to make efficient energy consumption schedules for massive amounts of smart loads. Traditionally, DR is focused on the curtailment of some electrical loads during peaking periods so as to alleviate the demand for peaking generation sources. On the other hand, consumers are more willing to participate in DR programs and try to shift some of their high-load household appliances to off-peak hours to reduce their energy expenses.

Nowadays, with the development of fast communication and advanced control technologies, more and more electrical loads in the demand side can be equipped with communication and control modules, which are more active and can be deployed all the time instead of just the peaking periods. Therefore, demand-side resources can provide various ancillary services, such as regulation, load following, frequency responsive spinning reserve, and supplemental reserve. And good candidates for these ancillary services including washers, dryers, water heaters, *heating, ventilating, and air-conditioning* (HVAC) systems with thermal storage, refrigerator, *plug-in hybrid electric vehicles* (PHEV) etc., among which aggregate populations of *thermostatically controlled loads* (TCLs), such as air conditioners, refrigerators, and electric water heaters can respond the power demand or be dispatched when subject to some control signals, such as pricing or thermostat set point signals.

In fact, electrical loads in the demand side can be grouped into two categories: rigid loads and flexible loads [9]. And flexible loads are composed of controllable loads and responsive loads, where controllable loads mean that the consumption behavior of the loads can be controlled to the expected objective designedly and responsive loads denote that the electrical loads can respond to the electricity price or system frequency by their own decisions, which may be uncertain and not controllable. In this paper, we only consider the demand response case of controllable loads and mainly focus on the cooling TCLs in the demand side, such as air-conditioning in the summer. Specifically, we utilize the traditional generating units and controllable load agents to follow the mismatched power caused by the load activities and the renewable power injection in real time.

In order to control these interactive units, it is urgent to clearly understand the mathematical models of them. The physical model of the generating units can be seen widely in the literature; while for the controllable loads, the load modelling is varied for different kinds of application scenarios. We are concerned with the model of a single TCL and the aggregate model of a population of TCLs. Compared with limited number of generating units, the number of these terminal loads is pretty large. Since there are numerous small terminal TCLs, it is difficult to control them from the dispatch center in a centralized way. One solution is to consider the load aggregator, which is aggregated by a certain number of TCLs. And these aggregators can be controlled in different kinds of control strategies.

Next, we review the basic control model of a single TCL and the aggregate model of multiple TCLs. Mathematical modeling is critical in controlling and regulating the output power of the aggregate TCLs. And these aggregate models (homogeneous/heterogeneous) can always be divided into two categories: on/off models and temperature setpoint models. Research along this line can be traced back to [10], in which the authors had proposed a physical model to capture the hybrid dynamics of each thermostat in the population. Furthermore, the authors in [11] derived a discrete-time stochastic computer model, which showed good agreements with basic physical principles. In fact, the aggregate TCL modeling was originally motivated by the "cold load pickup" phenomenon after a service interruption [12]. Malhame and Chong [13] investigated the aggregate dynamics of all TCLs and proposed a coupled Fokker-Planck equation based on the statistical probability density of a hybrid state Markov system. By dividing the cycling period of TCLs, the authors of [14], [15] proposed a state-queueing model to analyze the price response of aggregate loads consisting of multiple thermostatically controlled appliances. The authors of [16] derived a transfer function model related to the aggregate response of a population of homogeneous TCLs, which could be realized by a second-order linear state-space model with the control input being the shift in the deadband of all TCLs.

Beside the above mentioned aggregate on/off control models, which were mainly utilized for direct load control that interrupting the consumption power directly to all loads. Recent works are concerned with the temperature setpoint models used for indirect load control. Callaway [17] studied the effectiveness of utilizing setpoint temperature control signals to regulate the aggregate demand of a population of homogeneous TCLs based on the Fokker-Planck model. Inspired by such a control mechanism and the existing diffusion-based load modelling, Bashash and Fathy [18], [19] set up a bilinear state-space model for aggregate homogeneous TCLs with the setpoint temperature as the control input signal and the aggregate power as the output signal and showed good power tracking performance between the aggregate power output and the reference power curve. It has been shown that the aggregate flexibility of a population of TCLs can be succinctly modeled as a stochastic battery with dissipation [20].

On the other hand, the corresponding on/off and temperature setpoint models for heterogeneous TCLs can be found in literature as well. The authors of [21], [22], [35] considered the modeling and control problems for aggregate heterogeneous TCLs by proposing an approximate linear system model, in which the control input is the on-off rate for the TCLs in the temperature deadband. By considering the indoor air temperature and inner mass temperature, the authors of [23], [24] utilized a second-order equivalent thermal parameter model to characterize a single TCL, which could better capture both the transient and steady state aggregate response for a population of heterogeneous TCLs. By adding some simple imbedded instructions and memory to the temperature controller of each TCL, Sinitsyn *et. al.* [25] showed that short-term pulses of het-

erogeneous populations of TCLs can be created, which could avoid temporary synchronization of the TCLs' dynamics.

For the aggregate population of TCLs, there are always three kinds of control strategies for different DR ancillary services: centralized control, decentralized control and distributed control. The authors in [26] utilized a centralized model predictive control (MPC) method [27], [28] to regulate the aggregate output power of heterogeneous second-order TCLs to provide regulation service; MPC had been utilized for climate control in swiss office buildings by authors in [29] as well, which are centralized optimization and control solutions for DR services. Compared with centralized control, decentralized and distributed control receive much attention as well for lower communication costs. Decentralized control was utilized for DR flexible loads [30] to provide primary frequency control service. A stochastic decentralized control strategy [31] was proposed for a heterogeneous set of TCLs in DR services. Distributed optimization or distributed control have been widely used in Microgrids for power allocation or voltage/current control. As for distributed DR, it has been shown in [32], [33] that distributed framework can shift the burden of load leveling from the grid to end users.

Different control strategies have different advantages: Centralized control is high-efficient in optimization and control with global constraint while poor in scalability and robustness. Decentralized control is good at handling optimization or control problems with local constraints while may be helpless in global objective. Distributed control is a coordinate strategy for multiple interactive units which can exhibit the emergence behavior, such as synchronization or consensus, see [34] for more details, which has good scalability and strong robustness. In this paper, we consider a scenario of how to regulate multiple heterogeneous TCL aggregators to perform load following service, i.e., aggregator of aggregator. This problem has a global constraint, i.e., all aggregators achieve an expected power trajectory. Meanwhile, aggregators may be plugged in or plugged out flexibly according to their own enabling conditions.

Specifically, we utilize the bilinear aggregate model [19] for demand response load following of multiple heterogeneous TCL aggregators. The homogeneous TCLs in an intelligent residential district are modelled as a load aggregator. Thus, there may be several heterogeneous load aggregators in each district and numerous districts are managed by a bus load agent. Generally, each bus load agent needs to collect all information of districts globally and process the huge amount of data, and then issues the control signal to each aggregator simultaneously. These centralized solutions [36] are costly to be implemented especially when the number of the participants is large, and it is often susceptible to single-point failures. We propose an distributed pinning control strategy to coordinate the operation of multiple generating units and multiple load aggregators such that these two controllable resources can provide the load following service.

Distributed pinning control [37], [38] is flexible, reliable and less expensive to be implemented, which has a good scalability and can survive single-point failures. Pinning control means the global information is only shared with a small fraction of participants, and other ones communicate each other through a spare communication network, which can handle the optimization and control problems with global constrains. Here, the load agent just needs to send the global objective information to the adjacent districts and other districts communicate with their neighbors. This is a combined centralized-distributed control strategy, which is also a partially-distributed one.

On the other hand, the active power trajectory is analyzed and optimized by the dispatch center. The load agent must evaluate its regulation capacities and ramping rates exactly so as to obtain a feasible reference power trajectory. The dispatch center needs to collect all the private information of each aggregator to perform the optimal power calculation. Based on the bilinear aggregate mode, we first derive the regulation capacities and ramping rates of TCL aggregator analytically, which is one of the main contributions in this paper. On the other hand, we introduce a novel distributed pinning control solution for the coordination of multiple heterogeneous TCL aggregators in a load agent, which can provide load following ancillary service with traditional generating units.

Such a control strategy can guarantee the convergence of aggregate power of the load agents with any size of aggregators; meanwhile the distributed control can be implemented using a simple communication network, such as Wi-Fi connections. To realize the reference load following, the generating units can respond to the generating commands by regulating the speed of the units exactly by feedback control. While, for TCL load aggregators, the temperature setpoint is regulated in a distributed way so as to change the aggregate power of all TCLs in the DR program.

The rest of this paper is organized as follows. In Section II, problem formulation and the basic physical model and the approximated aggregate model are provided. Furthermore, aggregate evaluation of the TCL aggregator is carried out so as to provide the available capacity constraints and increasing/decreasing rate constraints. Based on the evaluation results, Section III presents a real time look-ahead economic dispatch for both generating units and bus load agents, which can provide the load following power trajectories for them. Section IV describes the detailed distributed pinning control implementation for numerous TCL aggregators in a load agent. Section V illustrates a simulation case with two load agents on a modified IEEE 9-bus system. Finally, conclusion and future work are presented in Section VI.

II. Problem Formulation and Aggregate Evaluation of TCLs

A. Problem formulation

In smart grids, renewable energy power generations (such as wind and photovoltaic) are developing rapidly. Connecting renewable energy sources to the power grid may result in some uncertainties or instability problems to the whole system. Therefore, more operation reserves are needed to provide ancillary services so as to balance the supply and demand [39]. And ancillary services are traditionally provided by regulating generators, that is generating units follow the load activities. Some alternative techniques, e.g., flywheels,

distributed generations, electric vehicles, distributed energy storages and demand response resources, can also provide the same ancillary services, among which ancillary services contributed from the demand side is a promising technique since flexible loads can provide faster regulation or load following services without environmental contaminants [20]. Therefore, how to make efficient load management strategies is critical in DR control. And we are concerned with how to control numerous TCL load aggregators to provide the load following service.

In this paper, we intend to solve the load following problem provided by both generating units and TCL aggregators. Firstly, the dispatch center optimizes the optimal generating schedules for regulating units and planned utilizations for TCL aggregators. Furthermore, the generating units can be controlled in a centralized way to follow the generating command and TCL aggregators can be controlled in a distributed way to follow the reference power trajectory. The load following structure of interconnected source and load systems is given in Fig. 1. That is, we utilize the traditional units and demand response TCL aggregators to follow the load activities caused by rigid and responsive loads and the injection of renewable energy power.

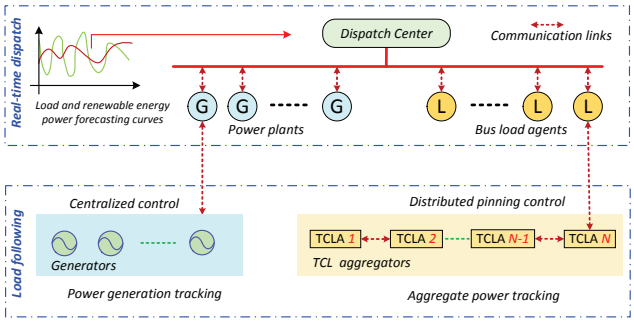

Fig. 1. Demand response load following structure of source and load systems.

In the proposed DR load following framework, the direction of the power flow is still from the generating units to loads. While, the difference lies in that the actual load curve is reduced by incorporating the dynamic response of TCL aggregators and the power flow is reduced slightly as well. The current research is focused on the active power distribution among generating units and flexible load agents and the load following service of multiple TCL aggregators in each load agent. Therefore, the active/reactive power flow constraints are not considered in the optimization dispatch layer.

B. Basic model of a single TCL and an aggregator

A TCL aggregator is composed of a group of individual terminal TCLs and the aggregate power of all these devices is critical for analysis and synthesis. It has been shown in [11] that this kind of TCL can be modeled as a thermal capacitance C [kWh/°C] in series with a thermal resistance R [°C/kW], the dynamic behavior of which can be modeled by two state variables, the internal temperature of the conditioned mass $\theta(t)$ and the discrete operation state $s(t)$ [0/1]. Suppose there are N_L TCLs in the load aggregator and the basic model of ith TCL is characterized by the following hybrid first-order ordinary differential equation [11]:

$$\frac{\mathrm{d}\theta_i(t)}{\mathrm{d}t} = \frac{1}{C_i R_i}\big(\theta_{a,i} - \theta_i(t) - s_i(t) R_i P_{r,i}\big), \tag{1}$$

for $i = 1, \ldots, N_L$. $P_{r,i}$ [kW] is the rate of energy transfer to or from the thermal mass, which is positive for cooling TCLs and negative for heating TCLs. The operation state $s_i(t)$ is a discrete switching sequence governed by a thermostatic switching law with predetermined temperature deadband. For the cooling TCLs, one has the following switching law of the operation state:

$$s_i(t) = \begin{cases} 0 & \text{if } s_i(t-\epsilon) = 1 \ \&\ \theta_i(t) < \theta_i^- \\ 1 & \text{if } s_i(t-\epsilon) = 0 \ \&\ \theta_i(t) > \theta_i^+ \\ s_i(t-\epsilon) & \text{otherwise} \end{cases} \tag{2}$$

where ϵ is the sampling period for the switching sequence, and the boundaries of the temperature deadband are given as:

$$\theta_i^- = \theta_{set,i} - \frac{\delta_{db,i}}{2}; \quad \theta_i^+ = \theta_{set,i} + \frac{\delta_{db,i}}{2}. \tag{3}$$

The on/off states of the TCL in the temperature deadband is not changed until the internal temperature hits the boundaries of the deadband, thus the operation period of the TCL can be divided into a cooling period and a heating time. The steady state cooling time $T_{c,i}$ and the heating time $T_{h,i}$ for ith thermostatic load can be calculated according to the solution of Equations (2) and (3) [16]:

$$T_{c,i} = C_i R_i \ln\left(\frac{P_{r,i}R_i + \theta_i^+ - \theta_{a,i}}{P_{r,i}R_i + \theta_i^- - \theta_{a,i}}\right),$$

$$T_{h,i} = C_i R_i \ln\left(\frac{\theta_{a,i} - \theta_i^-}{\theta_{a,i} - \theta_i^+}\right).$$

To illustrate the dynamic evolutions of a TCL, we simulate the hybrid dynamic equation with the parameters $C = 1.8$, $R = 1/0.3$, $\theta_a = 38$, $P_r = 16$, $\theta(0) = 23$, $\theta_{set} = 24$ and $\delta_{db} = 1$. Therefore, it is easy to calculate the cooling period $T_c = 9.153$ min and the heating period $T_h = 25.725$ min. The internal temperature and the operation state of a TCL is given in Fig. 2.

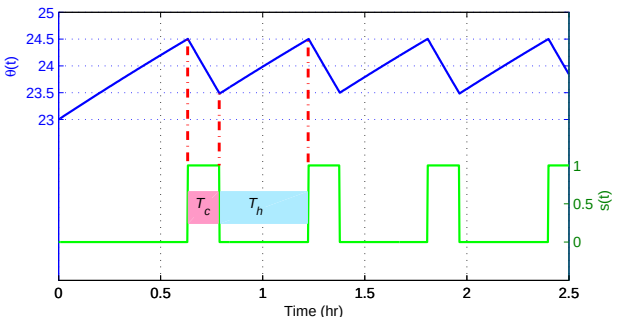

Fig. 2. Dynamic evolution for the temperature and on/off state of a thermostatic load.

Furthermore, the aggregate power output of the aggregator with N_L TCLs can be calculated by:

$$P_T(t) = \sum_{i=1}^{N_L} \frac{1}{\eta_i} s_i(t) P_{r,i}, \tag{4}$$

where η_i is the power energy's transmission efficiency.

According to the hybrid model, one can conclude that the setpoint temperature $\theta_{set,i}$ is in fact the control signal for each TCL. Meanwhile, different customers have different comfort levels under control. That is, the control input must satisfy the customer's constraint condition,

$$\theta^-_{set,i} \le \theta_{set,i} \le \theta^+_{set,i}, \tag{5}$$

where $\theta^-_{set,i}$ and $\theta^+_{set,i}$ are the lower and upper limits of the customers' acceptable regulation thresholds.

For a TCL aggregator, Eqs. (1)–(5) denote a hybrid physical aggregate model with $\theta_{set,i}$ and P_T being the input and output variables, which is a multiple-input single-output (MISO) system, which is difficult for application in practice. Reference [19] provides a single-input single-output (SISO) bilinear system to approximate the aggregate response of (1)–(5) for a group of homogeneous TCLs,

$$\begin{cases} \dot{x}(t) = Ax(t) + Bx(t)u(t), \\ y(t) = \tilde{C}x(t), \end{cases} \tag{6}$$

where $x(t) = (x_1(t), \ldots, x_Q(t))^T \in \mathbb{R}^Q$ is the state bin representing the average number of off-state $(x_1, \ldots, x_N)$ or on state $(x_{N+1}, \ldots, x_Q)$ TCLs in each temperature interval, see Fig. 10 in Appendix for more details. The control input $u(t)$ is set to be $u(t) = \Delta\dot{\theta}_{set} \in [\bar{\alpha}_{on}, \bar{\alpha}_{off}]$ (where $\bar{\alpha}_{on}$ and $\bar{\alpha}_{off}$ are given in Appendix as well) and $y(t)$ is the approximate aggregate power of the population of TCLs. Furthermore, the temperature setpoint $\theta_{set,i}$ of each terminal TCL is generated by

$$\theta_{set,i} = \theta^{des}_{set,i} + \Delta\theta_{set}(t) \text{ and } |\Delta\theta_{set}(t)| \le \Delta\theta^{max}_{set},$$

where the maximal bound of the comfort interval $\Delta\theta^{max}_{set} = \max\limits_{1\le i\le N_L} \{|\theta^+_{set,i} - \theta^{des}_{set,i}|, |\theta^-_{set,i} - \theta^{des}_{set,i}|\}$ and $\Delta\theta_{set}(t)$ is the universal broadcast control signal for the load aggregator which makes the temperature setpoints of all the TCLs in this load aggregator move by the same amplitude. The coefficient matrices $A \in \mathbb{R}^{Q\times Q}$, $B \in \mathbb{R}^{Q\times Q}$, and $\tilde{C} \in \mathbb{R}^Q$ are given in Appendix.

C. Aggregate evaluation of TCL aggregator

The dispatch center need to analyze the dynamical response of TCLs under different ambient temperatures and evaluate the available maximal and minimal regulation capacities and the increasing/decreasing rates (ramping rates) for load agents so as to execute the optimal economic dispatch.

According to the aggregate power equation (4), it can be derived that the aggregate power of a homogeneous TCL aggregator is bounded by $0 \le P_T(t) \le \frac{P}{\eta}N_L$. However, the regulation objective of the load agent must ensure users' comfort levels, i.e., the temperature setpoint θ_{set} must lie in the interval $[\theta^-_{set}, \theta^+_{set}]$. For a homogeneous group of TCLs, the steady-state power consumption P_{Ts} without control can be calculated as

$$\begin{aligned} P_{Ts} =& \frac{P_r}{\eta} \cdot \frac{T_c}{T_c + T_h} \cdot N_L \approx \frac{P_r}{\eta} \cdot \frac{\theta_a - \theta_{set}}{P_r R} \cdot N_L \\ \approx& \frac{N_L}{\eta R}(\theta_a - \theta_{set}). \end{aligned}$$

Therefore, the available stable lower and upper bounds of the TCL aggregator can be obtained approximately based on the steay-state aggregate power as follows:

$$P^{min}_{Ts} = \frac{N_L}{\eta R}(\theta_a - \theta^+_{set}), \quad P^{max}_{Ts} = \frac{N_L}{\eta R}(\theta_a - \theta^-_{set}), \tag{7}$$

and we set the mininal and maximal capacities of the TCL aggregator to be $P^{min}_{Agg} = 0.5P^{min}_{Ts}$ and $P^{max}_{Agg} = 1.5P^{max}_{Ts}$ on the basis of simulation experiences.

On the other hand, the ramping rate $\dot{y}(t)$ can be calculated as

$$\begin{aligned} \dot{y}(t) =& \tilde{C}(A + B \otimes u(t))x(t) \\ =& \begin{bmatrix} 0 & \tilde{C}_2 \end{bmatrix} \begin{bmatrix} A_{11}+B_{11}\otimes u(t) & A_{12}+B_{12}\otimes u(t) \\ A_{21}+B_{21}\otimes u(t) & A_{22}+B_{22}\otimes u(t) \end{bmatrix} x(t) \\ =& \begin{bmatrix} \tilde{C}_2(A_{21}+B_{21}\otimes u(t)) & \tilde{C}_2(A_{22}+B_{22}\otimes u(t)) \end{bmatrix} x(t) \\ =& \frac{P_r}{\eta}[\bar{\alpha}_{off} - u(t)]x_N + \frac{P_r}{\eta}[\bar{\alpha}_{on} - u(t)]x_{N+1}. \end{aligned}$$

By the constraint $\bar{\alpha}_{on} \le u(t) \le \bar{\alpha}_{off}$, one can further derive that

$$\frac{P_r}{\eta}(\bar{\alpha}_{on} - \bar{\alpha}_{off})x_{N+1} \le \dot{y}(t) \le \frac{P_r}{\eta}(\bar{\alpha}_{off} - \bar{\alpha}_{on})x_N.$$

Suppose the distribution of off-state and on-state TCLs in all temperature subintervals follows a uniform distribution, i.e.,

$$x_N + x_{N+1} = \frac{N_L \Delta T}{\delta},$$

and the fact that

$$\bar{\alpha}_{off} - \bar{\alpha}_{on} = \frac{RP_r}{CR\Delta T} \quad \text{and} \quad x_{N+1}\frac{\delta}{\Delta T}\frac{P_r}{\eta} = y,$$

where ΔT is the length of the temperature subinterval (see Fig. 10 in the Appendix). Furthermore, one has

$$-\frac{P_r y}{C\delta} \le \dot{y}(t) \le \frac{P_r(N_L P_r - \eta y)}{\eta C\delta},$$

which concludes that the increasing and decreasing rates of TCL aggregator are, respectively,

$$R^{dn}_{Agg} = -\frac{P_r y}{C\delta} \quad \text{and} \quad R^{up}_{Agg} = \frac{P_r(N_L P_r - \eta y)}{\eta C\delta}. \tag{8}$$

Since there are multiple TCL aggregators under a bus load agent, each DR aggregator has to report its evaluation information to the load agent. Based on the information received, the load agent can calculate the total capacity and ramping rate values, which are summarized in Algorithm 1.

III. Look-ahead Economic Dispatch to Provide Load Following Trajectories

Optimal economic dispatch problem (EDP) is one of the key problems in power system operation, which aims at minimizing the total economic cost of the stable operation for power systems subject to the supply-demand balance [40]. Traditionally, solving EDP can provide a solution for unit commitment and power output distributions for all generators. However, with the extensive application of renewable energy generation and the development of microgrid techniques, power resources and loads are becoming diversified. TCLs,

Algorithm 1 Aggregate Evaluation of a load agent

Input: Ambient temperature θ_a;
The number of TCLs in each TCL aggregator N_L;
TCL aggregator's private parameters P_r, C, R, η, δ, θ_{set}.
Output: Aggregate evaluation of load agent P_L^{min}, P_L^{max}, R_L^{dn}, R_L^{up}.
1: Calculate the available lower bound P_{Ts}^{min} and upper bound P_{Ts}^{max} of the TCL aggregator in steady-state by Eq. (7);
2: Set the mininal and maximal capacities of the TCL aggregator P_{Agg}^{min} and P_{Agg}^{max};
3: Calculate the ramping rates R_{Agg}^{dn} and R_{Agg}^{up} by Eq. (8) based on step 2;
4: Suppose there are N_j aggregators in jth load agent, then calculate the load agent's parameters based on $P_{L,j}^{min} = \sum_{i=1}^{N_j} P_{Agg,i}^{min}$ and so on;
5: **return** Aggregate evaluation values.

as one kind of flexible adjustable loads, are accounting for a large proportion of prosumers which can provide services of regulation or load following for active power allocation. These aggregate TCLs, which are managed by the load agents, can participate in the load following service together with traditional generating units.

Load agents and generating units can obtain their regulation capacities from bidding in a regulated market or from system operator in a deregulated system. Here, we consider the joint EDP for both flexible generating units and flexible load agents. Based on the load and renewable energy power forecasting curves, the look-ahead EDP can be solved for both participant units, i.e., the reference power trajectories for flexible generating units and flexible load agents can be derived respectively. Since generating units can carry out the generating schedules exactly based on the traditional centralized generating control of units, we mainly focus on the demand response load following control and implementation problem of TCL aggregators in each load agent.

The objective of the energy schedule is to minimize the total regulation cost for all the generating units and load agents:

$$\min \mathbb{F} = \sum_{t=1}^{T} \Big(\sum_{i\in G} C_{g,i}\big(\Delta P_{G,i}(t)\big) + \sum_{j\in L} C_{l,j}\big(\Delta P_{L,j}(t)\big) \Big), \tag{9}$$

where $\Delta P_{G,i}(t) = P_{G,i}(t) - P_{G,i}^{(0)}(t)$ and $\Delta P_{L,j}(t) = P_{L,j}(t) - P_{L,j}^{(0)}(t)$; and $C_{g,i}(\cdot)$ and $C_{l,j}(\cdot)$ are often described by quadratic functions, such as, $C_{g,i}(P_i) = a_i P_i^2 + b_i P_i + c_i$ and $C_{l,j}(P_j) = \tilde{a}_j P_j^2 + \tilde{b}_j P_j + \tilde{c}_j$, in which $a_i(\tilde{a}_i)$, $b_i(\tilde{c}_i)$ and $c_i(\tilde{c}_i)$ are predetermined constants reported by the generating units and the load agents, and T is the number of the look-ahead periods.

The energy mismatch can be derived as $\Delta P_{total}(t) = P_D(t) - P_{pv}(t) - P_{wi}(t) - \sum_{i\in G} P_{G,i}^{(0)}(t)$ based on the prediction of the load and new energy power, and such a mismatched power will be counterbalanced by both generating units and load agents together in real time, i.e., $\sum_{i\in G} \Delta P_{G,i}(t) + \sum_{j\in L} \Delta P_{L,j}(t) = \Delta P_{total}(t)$. By denoting Δt as the optimization period (such as $\Delta t = 5$ min), then the power balance constraint and the inequality constraints for such an optimization problem can be summarized as follows:

$$\begin{cases} \sum_{i\in G} P_{G,i}(t) + P_{in}(t) = P_D(t) - \sum_{j\in L} \Delta P_{L,j}(t), \\ P_{G,i}^{min} \le P_{G,i}(t) \le P_{G,i}^{max}, \\ -R_{G,i}^{dn} \le \dfrac{1}{\Delta t}\big(P_{G,i}(t+1) - P_{G,i}(t)\big) \le R_{G,i}^{up}, \\ P_{L,j}^{min} \le P_{L,j}(t) \le P_{L,j}^{max}, \\ -R_{L,j}^{dn} \le \dfrac{1}{\Delta t}\big(P_{L,j}(t+1) - P_{L,j}(t)\big) \le R_{L,j}^{up}, \end{cases} \tag{10}$$

for $\forall i \in G$, $j \in L$ and $t = 1, \ldots, T$, where $P_{in}(t) = P_{pv}(t) + P_{wi}(t)$. In this optimization problem, the dispatch cost $\mathbb{F}$ is the total regulation cost with respect to the active power regulation variable $\Delta\mathcal{P}(t) = [\Delta P_1(t), \Delta P_2(t), \ldots, \Delta P_{|G|+|L|}(t)]^T$ (the vector of generation scheduling for generating units and consumption scheduling for load agents), which is a quadratic function. And the regulation capacities and ramping rates for generating units are reported to the dispatch center; the regulation capacities and ramping rates for load agents are estimated by Algo. 1 and then reported to the dispatch center as well.

Such a look-ahead optimization problem (9) with the constraints (10) is a convex optimization and can be solved by the interior point method easily [41]. Therefore, one can obtain the energy scheduling $\mathcal{P}(t) = \mathcal{P}^{(0)}(t) + \Delta\mathcal{P}(t)$ for generating units and load agents, where $\mathcal{P}^{(0)}(t) = [P_1^{(0)}(t), P_2^{(0)}(t), \ldots, P_{|G|+|L|}^{(0)}(t)]^T$.

In the following, we are concerned with the DR load following control and implementation problem of multiple TCL aggregators in each load agent. Based on the temperature setpoint control signal, multiple TCL aggregators can regulate their aggregate power consumption behavior to follow the reference power trajectory from the dispatch center in a distributed way.

IV. Distributed Pinning Control of Multiple Aggregated TCLs

TCLs have already accounted for a significant proportion of the residential load and aggregate power of a large population of TCLs can vary drastically under small changes of the temperature setpoint, which instead do not influence the comfort levels of customers. It is more suitable for them to provide minute-minute ancillary services due to the temperature field characteristics of this kind of loads. In this section, we consider the DR load following control problem of multiple heterogeneous TCL aggregators managed by a flexible load agent, where each aggregator is composed of numerous homogeneous TCLs.

Here, we just need to control all TCL aggregators in each load agent to perform the active power tracking. The EDP has already provided a reference power trajectory for each load agent, and we will design a distributed pinning control algorithm to coordinate the operation of all aggregators inside

the load agent such that the aggregate power response could track the reference power trajectory.

Suppose there are N_j TCL aggregators in the jth load agent. For symbolic simplification, we consider the scenario that the reference power value $P_{ref}(t)$ is shared by N TCL aggregators with the ith aggregator modeled by the following bilinear model,

$$\begin{cases} \dot{x}_i(t) = A_i x_i(t) + B_i x_i(t) u_i(t), \\ y_i(t) = \tilde{C}_i x_i(t), \quad i = 1, 2 \ldots, N, \end{cases} \tag{11}$$

where $u_i(t) = \Delta\dot{\theta}_{set,i}(t) \in [\bar{\alpha}_{on,i}, \bar{\alpha}_{off,i}]$ is the control input for ith aggregator and $y_i(t)$ is the approximated aggregate power of the aggregator. Since the reference power values are provided every optimization period, which is a slow time-scale power trajectory. And it is transformed to be a real time trajectory for control by the spline interpolation method.

Thus, the total amount of the aggregate power under the load agent is $\sum_{i=1}^{N} y_i(t)$, which can be measured by the transformer terminal unit (TTU) installed at the terminal of the load agent. For a given reference power trajectory $P_{ref}(t)$, we need to design an efficient control strategy such that the aggregate response of all TCL aggregators under this load agent can track the reference power curve, i.e., $\sum_{i=1}^{N} y_i(t) - P_{ref}(t) \to 0$.

To achieve a fair participation of all aggregators, we choose the relative amplitude of the temperature setpoint as a consensus variable. By utilizing the distributed pinning control, all the control inputs of the aggregators will be synchronized to a common reference temperature setpoint dynamics, which can drive the aggregate power output to the reference power trajectory. So, the main objective is to develop a reference temperature setpoint updating equation and a distributed control algorithm such that the aggregate output tracking can be achieved just by communications with the neighbors.

Firstly, we consider how to generate a reference temperature setpoint signal. For the load agent, the aggregate response deviation can be measured, based on which we can design a centralized pinner, which is utilized as a virtual leader for distributed pinning control, and the group of the aggregators are called followers. Pinning means that the centralized pinner is only connected to a small fraction of the TCL aggregators and the remaining aggregators communicate with their neighbors based on the spare communication network.

For the TCL aggregators in the load agent, we define a reference temperature setpoint variable $\Delta\theta_{set}^{ref}$, which is regulated by the response variation between the total aggregate power and the reference power $\sum_{i=1}^{N} y_i(t) - P_{ref}(t)$. The changing rate of the reference temperature setpoint is closely related to the response variation. The following first-order regulation differential equation is used to model the reference temperature setpoint's dynamics:

$$\begin{aligned} \Delta\dot{\theta}_{set}^{ref}(t) =& \mu_1 \Big(\sum_{i=1}^{N} y_i(t) - P_{ref}(t) \Big) \\ &+ \mu_2 \int_{t_0}^{t} \Big(\sum_{i=1}^{N} y_i(t) - P_{ref}(t) \Big) \mathrm{d}t, \end{aligned} \tag{12}$$

where the initial value $\Delta\theta_{set}^{ref}(0) = 0$ and $|\Delta\theta_{set}^{ref}(t)|$ is bounded by the $\Delta\theta_{set}^{max}$ for the optimal comfort interval of the costumers and the rate is bounded by $\max_{1\le i\le N} \bar{\alpha}_{on,i} \le \Delta\dot{\theta}_{set}^{ref}(t) \le \min_{1\le i\le N} \bar{\alpha}_{off,i}$, the coefficients μ_1 and μ_2 are approximate positive regulation gains.

So far, the reference temperature setpoint variable is derived, which can guide the group of the aggregators to move towards the direction of the reference power trajectory and track the reference power. This reference signal can be issued out to all load aggregators in a centralized way. However, such a control strategy becomes both time-consuming and inefficient especially when the number of the aggregators is very large and the distribution of these load aggregators is decentralized in different control areas.

We consider the distributed pinning control for all TCL aggregators in the load agent. Distributed control has many advantages, such as easy implementation, low complexity, high robustness, and good scalability, which enable the participants to be plugged-in or plugged-out flexibly. The temperature variation $\Delta\theta_{set,i}(t)$ of ith TCL aggregator is regulated by the following equation,

$$\begin{aligned} \Delta\dot{\theta}_{set,i}(t) =& \beta \sum_{j=1}^{N} a_{ij} \big(\Delta\theta_{set,j}(t) - \Delta\theta_{set,i}(t) \big) \\ &- \beta d_i \big(\Delta\theta_{set,i}(t) - \Delta\theta_{set}^{ref}(t) \big), \end{aligned} \tag{13}$$

where the initial value $\Delta\theta_{set,i}(0) = \theta_{set,i}(0) - \theta_{set,i}^{des}$; a_{ij} is the element of the adjacency matrix A of the communication topology among the TCL aggregators; and $d_i = 1$ if the ith TCL aggregator is pinned by the load agent (centralized pinner), otherwise $d_i = 0$. Generally, the communication topology needs to have a directed spanning tree for a directed communication structure or to be connected for an undirected communication structure.

The reason why we utilize a distributed pinning strategy is at least threefolds: (1) Compared with the limited number of generating units, the number of TCL aggregators is large and the distribution of these aggregators is decentralized in a wide range of areas, so the traditional centralized control strategy is time-consuming and inefficient. (2) TCL aggregators have more flexibility, as they can choose to participate the DR or not by their own enabling condition, a distributed strategy can provide a more robust solution to handle this scenario and realize the plug-in and plug-out of TCL aggregators. (3) Different from the classical distributed average consensus strategy, there is a global coordination objective to fulfill the reference active power tracking in this problem. Therefore, the distributed pinning strategy is an efficient allocation strategy for coordinating multiple load aggregators.

According to the distributed pinning communication protocol (13) and the bilinear model (11), it is easy to derive the control signal $u_i(t)$ for ith TCL aggregator, i.e.,

$$\begin{aligned} u_i(t) =& \beta \sum_{j=1}^{N} a_{ij} \big(\Delta\theta_{set,j}(t) - \Delta\theta_{set,i}(t) \big) \\ &- \beta d_i \big(\Delta\theta_{set,i}(t) - \Delta\theta_{set}(t) \big). \end{aligned} \tag{14}$$

In the implementation of the distributed protocol, the interaction among TCL aggregators and centralized pinner occurs at discrete time steps. Therefore, continuous-time differential equations (12) and (13) need to be transformed to discrete-time difference equations. Here, we utilize the Euler method to derive the discrete-time close-loop system:

$$\begin{cases} \Delta\theta_{set}^{(k+1)} = \Delta\theta_{set}^{(k)} + h\mu_1\Big(\sum_{i=1}^{N} y_i(k) - P_{ref}(k)\Big) \\ \qquad\qquad + h\mu_2 \sum_{l=0}^{k}\Big(\sum_{i=1}^{N} y_i(l) - P_{ref}(l)\Big), \\ \Delta\theta_{set,i}^{(k+1)} = \Delta\theta_{set,i}^{(k)} + h\beta\sum_{j=1}^{N} a_{ij}\big(\Delta\theta_{set,j}^{(k)} - \Delta\theta_{set,i}^{(k)}\big) \\ \qquad\qquad - h\beta d_i\big(\Delta\theta_{set,i}^{(k)} - \Delta\theta_{set}^{ref}(k)\big), \\ u_i(k) = \big(\Delta\theta_{set,i}^{(k+1)} - \Delta\theta_{set,i}^{(k)}\big)/h, \end{cases} \tag{15}$$

where the output power $y_i(k)$ is updated from the discrete-time aggregate model:

$$\begin{cases} x_i(k+1) = x_i(k) + h\big(A_i x_i(k) + B_i x_i(k) u_i(k)\big), \\ y_i(k) = \tilde{C}_i x_i(k), \quad i = 1, 2 \ldots, N, \end{cases} \tag{16}$$

where h is the discretization step, i.e., the sampling period for the practical operation system.

Remark 1. *TCL aggregator has the flexibility to switch on/off the DR program by setting/removing the communication links with its neighbors. New participants can be involved in the DR program by just setting up a communication link with the existing participants. The important and complex computational task of the load agent is greatly reduced and now it just needs to receive the reference signal from the dispatch center and send out the control commands to the pinned TCL aggregators.*

Remark 2. *The plugging-in and the plugging-out of the TCL aggregator will change the structure of the communication network. In fact, the consensus convergence speed is determined by the real part of the minimal nonzero eigenvalue of Laplacian matrix of the communication network. And, the larger the real part of the minimal nonzero eigenvalue, the larger convergence rate of the convergence speed. On the other hand, the required condition for the communication network is that the communication network has a directed spanning tree for a directed communication network and connected for an undirected communication network.*

We summarize the steps for the load following of multiple TCL aggregators as an algorithm presented as follows.

Algorithm 2 Distributed Load Following of TCL Aggregators

1: Solving the look-ahead optimization (9) and (10) to provide the active power reference trajectories $\mathcal{P}(t)$ every 5 min;
2: Using the cubic spline interpolation to derive the reference power trajectories for load agents in each 5-min time interval;
3: Measuring the aggregate response of the load agent and generate the reference temperature setpoint $\Delta\theta_{set}^{ref}$ by Eqs. (15) and (16) every 10 s;
4: Distributed calculation the private relative amplitude $\Delta\theta_{set,i}$ based on (15) to generate control input u_i of the TCL aggregator every 10 s;
5: Estimating the aggregate response based on the bilinear model (16), then back to Step 3 and continue the loop.

V. Simulation and Results

In this section, we validate the performance of the proposed load following strategy in a regional power system with three generating units and two demand response load agents. The units and load agents are placed on a modified IEEE-9 bus system with one wind power plant and one photovoltaic power plant. The physical connections for the units and load agents and communication links for the TCL aggregators are given in Fig. 3.

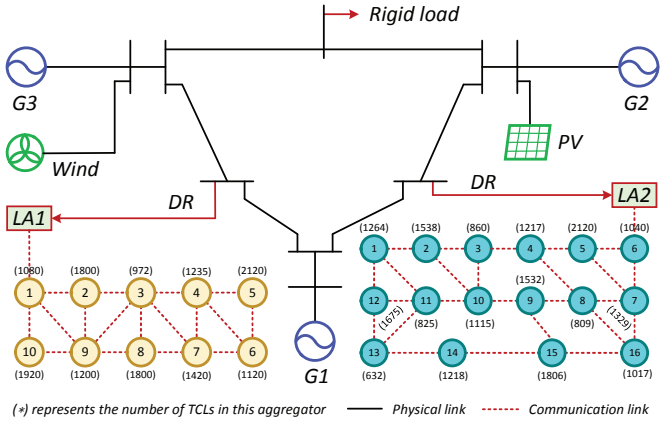

Fig. 3. IEEE 9 bus system with demand response load agents and renewable energy power injection.

The demand response load agents are utilized to balance the power generation and power utilization, especially the large power mismatch resulted from the intermittent injection of the renewable energy power. Suppose there are 26 TCL aggregators with 34664 cooling air conditioners, which are distributed under two demand response load agents and will participate in real time load following service with generating units. The reference active power trajectories for the load agents can be received from the dispatch center every 5 min.

In the following, we illustrate the load following service in detail. We consider a short-term load following service, such as peak load period in the afternoon. Taking a summer day in Nanjing, China as an example. Suppose the average ambient temperature is $38^{\circ}C$ from 13:00 $\sim$ 15:00 and all the demand response air-conditioners will be controlled in real time.

A. Aggregate evaluation of the aggregator

Suppose all the cooling air-conditioners in one district are homogeneous TCLs and they are aggregated as an aggregator.

The aggregator can be an energy management system [42] installed in this district.

As for the demand response load agent $LA1$, there are 10 TCL aggregators with a total number of 14667 cooling air-conditioners. The width of all temperature deadbands is $\delta_{db} = 1^\circ$ C; the thermal resistance R of each load aggregator is chosen from $[4.5, 5.5]$ uniformly; the thermal capacitance C of each load aggregator is chosen from $[8, 12]$ uniformly; the output cooling energy P_{ra} of each load aggregator is chosen from $[16, 20]$ uniformly; the coefficient of performance η of each load aggregator is chosen from $[2.6, 3]$ uniformly. The preferred setpoint for each aggregator is given with $\theta^{des}_{set,i} = [23.15, \ldots, 24.45]$. The initial proportion of off TCLs in each load aggregator is $pro_{\text{off}} = [0.472, \ldots, 0.516]$, and the initial temperature for each air conditioning in each load aggregator is chosen uniformly in the first temperature deadband. By Algorithm 1, one can derive the private parameters for $LA1$: $P^{min}_{L,1} = 14$ MW, $P^{max}_{L,1} = 67$ MW, $R^{dn}_{L,1} = -76$ MW/min, and $R^{up}_{L,1} = 125$ MW/min.

On the other hand, for the demand response load agent $LA2$, there are 16 TCL aggregators with a total number of 19997 cooling air-conditioners. Suppose the width of the temperature deadband $\delta_{db} = 1^\circ C$; and $R \in [1.5, 2.5]$, $C \in [8, 12]$, $P_{ra} \in [12, 16]$, $\eta \in [2.6, 3]$ uniformly. The preferred setpoint for each aggregator is given with $\theta^{des}_{set,i} = [23.15, \ldots, 24.45]$. The initial proportion of off-state TCLs in each load aggregator is $pro_{\text{off}} = [0.172, \ldots, 0.116]$, and the initial temperature for each air conditioning in each load aggregator is chosen uniformly in the first temperature deadband. By Algorithm 1, one can derive the private parameters for $LA2$: $P^{min}_{L,2} = 19$ MW, $P^{max}_{L,2} = 92$ MW, $R^{dn}_{L,2} = -82$ MW/min, and $R^{up}_{L,2} = 95$ MW/min.

B. Reference power trajectories solving

Based on the reported aggregate evaluation results of the load agents and the dispatch cost functions of units and load agents, the dispatch center can perform the look-ahead optimization easily. The detailed private coefficients for units and load agents are given in Tab. I. Meanwhile, the day-ahead generating plans (estimated flexible utilization) for generating units (load agents), and the real time new energy power injection P_{wi}, P_{pv} and the real time load forecasting P_D for a $2h+$ period are given in Tab. II. Based on the centralized look-ahead optimization (look-ahead period $T = 4$) carried out by the dispatch center, the real time reference power trajectories are issued to all participants every 5 min.

TABLE I
GENERATION AND DEMAND UNITS' PRIVATE PARAMETERS.

	Generation Unit Parameters(unit: MW)					
G	a_i	b_i	$R^{dn}_{G,i}$	$R^{up}_{G,i}$	$P^{min}_{G,i}$	$P^{max}_{G,i}$
G1	0.1151	5.2034	20	25	10	250
G2	0.0856	1.2104	25	30	10	300
G3	0.1225	1.1518	22	27	10	270
	Demand Unit Parameters(unit: MW)					
L	$\tilde{a}_j$	$\tilde{b}_j$	$R^{dn}_{L,j}$	$R^{up}_{L,j}$	$P^{min}_{L,j}$	$P^{max}_{L,j}$
L1	0.1021	6.2015	76	125	14	67
L2	0.1265	6.3472	82	95	19	92

TABLE II
THE DAY-AHEAD SCHEDULED (ESTIMATED) VALUES FOR GENERATING UNITS (LOAD AGENTS) AND THE PREDICTED VALUES FOR P_{wi}, P_{pv} AND P_D.

Time(k)	$P^{(0)}_{G,1}$	$P^{(0)}_{G,2}$	$P^{(0)}_{G,3}$	$P^{(0)}_{L,1}$	$P^{(0)}_{L,2}$	P_{pv}	P_{wi}	P_D
13 : 00	200	260	219	37	52	30.2	45.3	763
13 : 05	208	262	223	35	51	30.1	69.8	779
13 : 10	214	263	227	34	49	29.8	52.4	784
13 : 15	212	266	225	32	47	29.2	48.2	781
13 : 20	210	262	236	35	51	28.8	60.8	787
13 : 25	216	258	230	32	47	28.3	65.2	788
13 : 30	210	263	231	36	52	28.6	58.4	794
13 : 35	214	258	225	35	51	28.1	75.5	791
13 : 40	210	266	232	38	49	27.9	72.3	793
13 : 45	216	258	228	32	53	27.6	68.4	802
13 : 50	217	263	229	38	55	27.5	72.8	808
13 : 55	214	258	231	35	49	27.7	57.3	801
14 : 00	210	266	223	32	49	27.4	52.1	801
14 : 05	212	263	221	36	53	27.2	62.4	809
14 : 10	214	261	227	34	49	27.3	58.2	804
14 : 15	210	260	225	32	50	27.1	51.9	811
14 : 20	206	259	231	34	51	26.9	58.2	817
14 : 25	216	266	223	36	55	26.8	72.2	809
14 : 30	214	263	231	33	53	26.3	53.5	824
14 : 35	212	258	225	38	49	26.7	52.1	821
14 : 40	214	266	223	36	51	26.6	48.9	813
14 : 45	208	263	222	32	53	26.2	55.3	822
14 : 50	216	264	229	36	55	26.5	65.2	828
14 : 55	210	266	231	35	49	26.4	45.8	821
15 : 00	214	260	222	38	53	26.3	52.9	813
15 : 05	216	265	224	34	49	25.9	67.7	820
15 : 10	212	262	230	36	51	25.7	50.5	825
15 : 15	217	267	227	32	49	25.5	60.3	828
15 : 20	213	263	222	38	53	25.6	60.4	824

Based on the above parameters' setting, the reference power trajectories for the generating units and load agents can be calculated by the optimization problem (9) and (10), which are shown in Figs. 4 and 5 vs the day-ahead scheduled/estimated power trajectories.

Previously, we have discussed the real time look-ahead optimal dispatch for generating units and load agents. Since the generating schedules are optimized by considering the physical constraints, the units can fulfill the generating commands by the existing centralized feedback control strategy. Next, we will illustrate the detailed load following problems for load agents based on distributed pinning control of multiple aggregate TCLs.

C. Demand response load following control of TCLs

Suppose the comfortable levels of customers are $\pm 2^\circ$ C around the setpoint temperature and all the approximate aggregate models have the same dimensions ($Q = 100$ for models in $LA1$ and $LA2$) in the simulation. By calculations, one can derive the maximal and minimal bounds of the control inputs for models in $LA1$ and $LA2$, which are given as $-22.0334 \leq u_i^{LA1} \leq 14.2868$ and $-9.0951 \leq u_i^{LA2} \leq 14.8180$.

In the following, the distributed pinning control strategy (15) is utilized to coordinate the operation of all TCL aggregators in $LA1$ and $LA2$. The communication matrix (a_{ij}) and the pinning links (d_i) can be easily derived from Fig. 3 and the reference power values are provided by the dispatch center every 5 min and we utilize the method of cubic spline interpolation to derive a more detailed reference power trajectory during

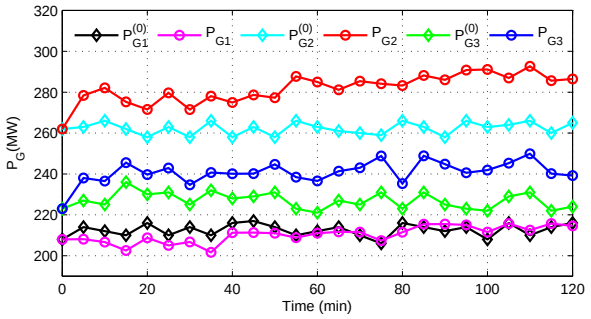

Fig. 4. The optimized power trajectories P_{Gi} and the day-ahead scheduled power trajectories $P_{Gi}^{(0)}$ for generating units.

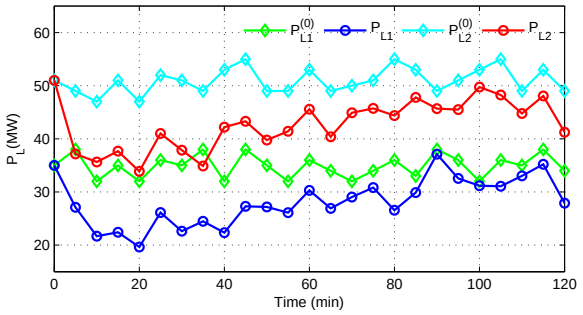

Fig. 5. The optimized power trajectories P_{Lj} and the day-ahead estimated power trajectories $P_{L}^{(0)}$ for load agents.

each optimization period. For the real-time control of the TCL aggregators, the sampling period and the control commands are 10 s for a cycle.

By setting the control gains $\mu_1 = \mu_2 = 1.2$ and the coupling strength $\beta = 8$ and running the simulation, one can obtain the relative incremental temperature $\Delta\theta_{set,i}$ and control input u_i for each TCL aggregator in load agent $LA1$, which are given in Fig. 6. Furthermore, the power tracking curves followed by the approximate aggregate model and the Monte-Carlo method are given in Fig. 7 as well.

For the 16 TCL aggregators in $LA2$, by setting the control gains $\mu_1 = \mu_2 = 1.8$ and the coupling strength $\beta = 6$ and performing the simulation, one can obtain the relative incremental temperature $\Delta\theta_{set,i}$ and control input u_i for each TCL aggregator in load agent $LA2$, which are given in Fig. 8. Furthermore, the power tracking curves followed by the aggregate model and the Monte-Carlo method and the reference power trajectory are given in Fig. 9.

Based on the load forecasting and the cost quotation of generating units and load agents, EDP aims at providing the reference power trajectories for generating units and load agents in real time (e.g. every 5 min in the simulation). Furthermore, the generating units can execute the generating schedules via the existing centralized control strategy, not covered in this paper; the load aggregators respond to the load following control in a distributed way. From the simulation results of the Monte-Carlo method, the load following strategy proposed in this paper can guide the aggregate power of the load agent in an orderly manner.

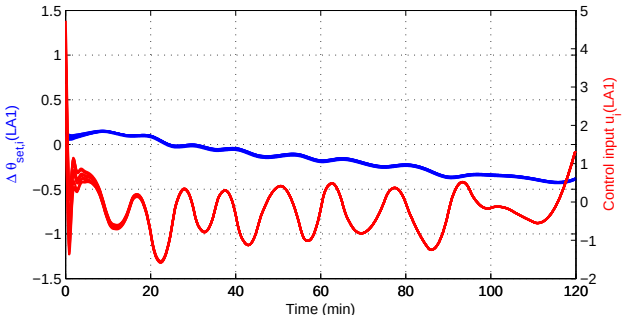

Fig. 6. The relative incremental temperature $\Delta\theta_{set,i}$ and the control input u_i in $LA1$.

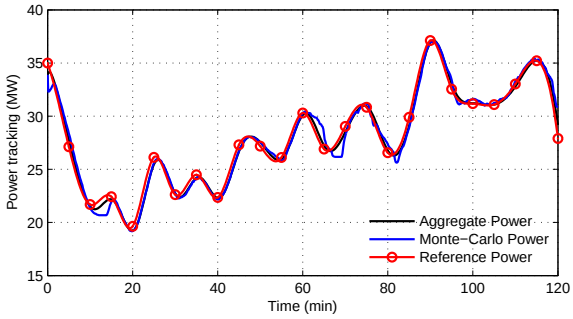

Fig. 7. The power tracking curves for $LA1$.

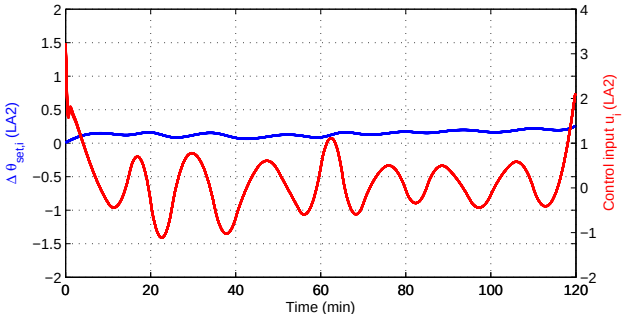

Fig. 8. The relative incremental temperature $\Delta\theta_{set,i}$ and the control input u_i in $LA2$.

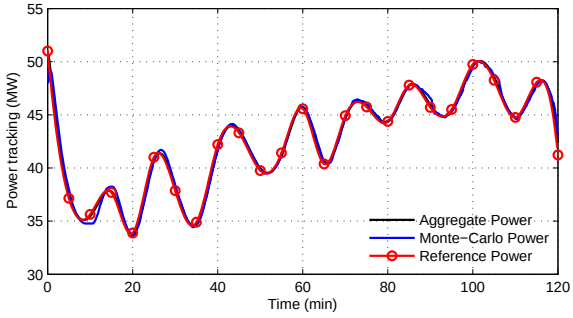

Fig. 9. The power tracking curves for $LA2$.

Remark 3. *The convergence speed for the distributed pinning control algorithm is slower than the centralized optimization algorithm, which is closely related to the communication structure among all TCL aggregators. While, it can reduce the communication burden between the load agent and each ter-*

minal aggregator. On the other hand, communication among neighbors may have some time delays, which are also present in the centralized optimization algorithm.

VI. Conclusion and Future Work

This paper presented a DR load following algorithm to coordinate the operation of generating units and TCL aggregators in smart grids. Compared with the existing load following strategies, the proposed algorithm considers the aggregate evaluation and distributed pinning operation of load aggregators. Specifically, flexible load agents gain their regulation capacities from the dispatch center through bidding with generating units, and then the regulation capacity of load agent is shared by multiple TCL aggregators in a distributed way. While, the regulation capacity of the generating units are accomplished by its local controller. We also provide an implementation of aggregate control for the TCL aggregator by tuning the temperature setpoint of the terminal loads. As one demonstration, we apply the proposed strategy to a modified IEEE-9 bus system with DR load agents. The simulation results demonstrate that the proposed load following strategy is effective in driving the population of TCLs to the reference power trajectory without considerably changing the comfort level of customers.

Future work includes utilizing the distributed pinning algorithm to realize the charge or discharge control of multiple aggregate electric vehicles with inerter [43], and management of multiple virtual power plants [44], and coordinate the operation of multiple distributed generators in Microgrid [45], [46].

Appendix

The detailed coefficient matrices for the bilinear aggregate model:

$$A(\bar{\alpha}_{on},\bar{\alpha}_{off}) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},\ B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

$$A_{11} = \begin{bmatrix} -\bar{\alpha}_{off} & & & & & & \\ \bar{\alpha}_{off} & -\bar{\alpha}_{off} & & & & & \\ & \ddots & \ddots & & & & \\ & & & \bar{\alpha}_{off} & -\bar{\alpha}_{off} & & \\ & & & & \ddots & \ddots & \\ & & & & & \bar{\alpha}_{off} & -\bar{\alpha}_{off} \end{bmatrix},$$

$$A_{22} = \begin{bmatrix} \bar{\alpha}_{on} & -\bar{\alpha}_{on} & & & & & \\ & \ddots & \ddots & & & & \\ & & \bar{\alpha}_{on} & -\bar{\alpha}_{on} & & & \\ & & & \ddots & \ddots & & \\ & & & & \bar{\alpha}_{on} & -\bar{\alpha}_{on} & \\ & & & & & \bar{\alpha}_{on} & \end{bmatrix},$$

and $A_{11} \in \mathbb{R}^{N\times N}$, $a^{(12)}_{M+1,N+1} = \bar{\alpha}_{on}$, $a^{(21)}_{P,N} = \bar{\alpha}_{off}$, and other elements of matrices A_{12}, A_{21} are zeros. Then matrix B has the same structure with A, which can be obtained by setting $\bar{\alpha}_{on} = \bar{\alpha}_{off} = -1$, i.e., $B = A(-1,-1)$. $\tilde{C} = [\tilde{C}_1, \tilde{C}_2] \triangleq [\underbrace{0,\ldots,0}_{N}, \underbrace{P/\eta,\ldots,P/\eta}_{Q-N}]$. The parameters $\bar{\alpha}_{on}$ and $\bar{\alpha}_{off}$ are the average local load transport rates, which can be calculated by the following approximate values:

$$\bar{\alpha}_{on} = \frac{1}{CR\Delta T}(\theta_a - \theta^{des}_{set} - RP_r),$$
$$\bar{\alpha}_{off} = \frac{1}{CR\Delta T}(\theta_a - \theta^{des}_{set}),$$

where ΔT is the discretization step of the temperature deadband $[\theta^-, \theta^+]$, and N, Q are given in the variables x_N and x_Q, in Fig. 10.

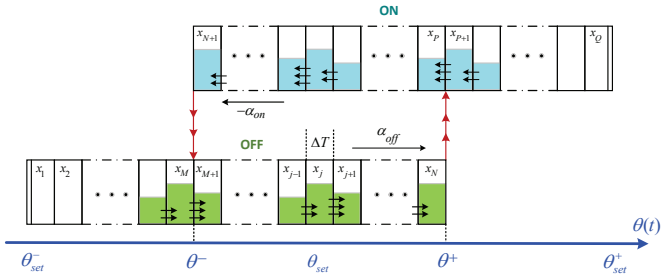

Fig. 10. Finite-difference discretization of the temperature interval in order to derive the state-space model [19].

References

[1] X. Yu and Y. Xue. Smart Grids: A Cyber-physical systems perspective. *Proceedings of the IEEE*, 104(5): 1058–1070, 2016.

[2] A.-H. Mohsenian-Rad, V.W.S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Trans. Smart Grid*, 1(3): 320–331, 2010.

[3] D. Angeli and P.-A. Kountouriotis. A stochastic approach to dynamic-demand refrigerator control. *IEEE Trans. Cont. Syst. Tech.*, 20(3): 581–592, 2012.

[4] C. Li, X. Yu, W. Yu, G. Chen, and J. Wang. Efficient computation for sparse load shifting in demand side management. *IEEE Trans. Smart Grid*, DOI: 10.1109/TSG.2016.2521377, 2016.

[5] K.-H. Ng and G.B Sheble. Direct load control-A profit-based load management using linear programming. *IEEE Trans. Power Syst.*, 13(2): 688–694, 1998.

[6] J.A. Short, D.G. Infield, and L.L. Freris. Stabilization of grid frequency through dynamic demand control. *IEEE Trans. Power Syst.*, 22(3): 1284–1293, 2007.

[7] P. Samadi, A.-H. Mohsenian-Rad, R. Schober, V.W.S. Wong, and J. Jatskevich. Optimal real-time pricing algorithm based on utility maximization for smart grid. In *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Gaithersburg, MD, Oct., pp. 415–420, 2010.

[8] J. Zhu, M. Z. Q. Chen, and B. Du. A new pricing scheme for controlling energy storage devices in future smart grid. In *J. Applied Mathematics*, Vol. 2014, Art. ID 340842, 11 pages, 2014.

[9] J. Hu, J. Cao, and T. Yong. Multi-level dispatch control architecture for power systems with demand-side resources. *IET Gener. Transm. Distrib.*, 9(16): 2799-2810, 2015.

[10] C.Y. Chong and A.S. Debs. Statistical synthesis of power system functional load models. In *Proc. 18th Conf. Decision Control*, pp. 264–269, 1979.

[11] R.E. Mortensen and K.P. Haggerty. A stochastic computer model for heating and cooling loads. *IEEE Trans. Power Syst.*, 3(3): 1213–1219, 1988.

[12] S. Ihara and F.C. Schweppe. Physically based modeling of cold load pickup. *IEEE Trans. power Appar. Syst.*, PAS-100(9): 4142–4150, 1981.

[13] R. Malhame and C.-Y. Chong. Electric load model synthesis by diffusion approximation of a high-order hybrid-state stochastic system. *IEEE Trans. Auto. Cont.*, 30(9): 854–860, 1985.

[14] N. Lu and D.P. Chassin. A state-queueing model of thermostatically controlled appliances. *IEEE Trans. Power Syst.*, 19(3): 1666–1673, 2004.

[15] N. Lu, D.P. Chassin, and S.E. Widergren. Modeling uncertainties in aggregated thermostatically controlled loads using a state queueing model. *IEEE Trans. Power Syst.*, 20(2): 725–733, 2005.

[16] S. Kundu, N. Sinitsyn, S. Backhaus, and I. Hiskens. Modeling and control of thermostatically controlled loads. In *Proc. 17th Power Syst. Computation Conf.*, Stokholm Sweden, Aug., pp. 1–7, 2011.

[17] 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.*, 50(5): 1389–1400, 2009.

[18] S. Bashash and H.K. Fathy. Modeling and control insights into demand-side energy management through setpoint control of thermostatic loads. *Amer. Con. Conf.*, San Francisco, CA, June, pp. 4546–4553, 2011.

[19] S. Bashash and H.K. Fathy. Modeling and control of aggregate air conditioning loads for robust renewable power management. *IEEE Trans. Cont. Syst. Tech.*, 21(4): 1318–1327, 2013.

[20] H. Hao, B.M. Sanandaji, K. Poolla, and T.L. Vincent. Aggregate flexibility of thermostatically controlled loads. *IEEE Trans. Power Syst.*, 30(1): 189–198, 2015.

[21] S. Koch, J.L. Mathieu, and D.S. Callaway. Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services. In *Proc. 17th Power Syst. Computation Conf.*, Stokholm Sweden, Aug., pp. 1–7, 2011.

[22] 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.*, 28(1): 430–440, 2013.

[23] W. Zhang, K. Kalsi, J. Fuller, M. Elizondo, and D. Chassin. Aggregate model for heterogeneous thermostatically controlled loads with demand response. In *Proc. IEEE Power and Energy Soc. Gen. Meet.*, San Diego, CA, Jul., pp. 1–8, 2012.

[24] 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.*, 28(4): 4655–4664, 2013.

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

[26] M. Liu and Y. Shi. Model predictive control of aggregated heterogeneous second-order thermostatically controlled loads for ancillary services. *IEEE Trans. Power Syst.*, 31(3): 1963–1971, 2016.

[27] X. Shi, J. Hu, J. Yu, T. Yong, and J. Cao. A novel load frequency control strategy based on model predictive control. In *2015 IEEE Power and Energy Soc. Gen. Meet.*, Denver, CO, pp. 1-5, 2012.

[28] G. Wen, G. Hu, J. Hu, X. Shi, and G. Chen. Frequency regulation of source-grid-load systems: A compound control strategy. *IEEE Trans. Ind. Informat.*, 12(1): 69–78, 2016.

[29] D. Sturzenegger, D. Gyalistras, M. Morari, and R. Smith. Model predictive climate control of a swiss office building: Implementation, sesults, and cost–benefit analysis. *IEEE Trans. Cont. Syst. Tech.*, 24(1): 1–12, 2016.

[30] A. Molina-Garcia, F. Bouffard, and D.S. Kirschen. Decentralized demand-side contribution to primary frequency control. *IEEE Trans. Power Syst.*, 26(1): 411–419, 2011.

[31] S.H. Tindemans, V. Trovato, and G. Strbac. Decentralized control of thermostatic loads for flexible demand response. *IEEE Trans. Cont. Syst. Tech.*, 23(5): 1685–1700, 2015.

[32] Z. Fan. A distributed demand response algorithm and its application to PHEV charging in smart grids. *IEEE Trans. Smart Grid*, 3(3): 1280–1290, 2012.

[33] R. Deng, Z. Yang, F. Hou, M.-Y. Chow, and J. Chen. Distributed real-time demand response in multiseller–multibuyer smart distribution grid. *IEEE Trans. Power Syst.*, 30(5): 2364–2374, 2015.

[34] J. Hu, Y. Li, T. Yong, J. Cao, J. Yu, and W. Mao. Distributed cooperative regulation for multiagent systems and its applications to power systems: A survey. *Sci. World J.*, 12 pages, Art. Id: 139028, 2014.

[35] S. Esmaeil Zadeh Soudjani and A. Abate. Aggregation and control of populations of thermostatically controlled loads by formal abstractions. *IEEE Trans. Cont. Syst. Tech.*, 23(3): 975–990, 2015.

[36] K.T. Tan, X.Y. Peng, P.L. So, Y.C. Chu, and M.Z.Q. Chen. Centralized control for parallel operation of distributed generation inverters in microgrids. *IEEE Trans. Smart Grid*, 3(4): 1977–1987, 2012.

[37] J. Hu, J. Cao, J. Yu, and T. Hayat. Consensus of nonlinear multi-agent systems with observer-based protocols. *Systems Control Lett.*, 72: 71–79, 2014.

[38] J. Hu, J. Cao, J.M. Guerrero, T. Yong, and J. Yu. Improving frequency stability based on distributed control of multiple load aggregators. *IEEE Trans. Smart Grid*, DOI: 10.1109/TSG.2015.2491340, 2015.

[39] Y.V. Makarov, C. Loutan, J. Ma, and P. De Mello. Operational impacts of wind generation on california power systems. *IEEE Trans. Power Syst.*, 24(2): 1039–1050, 2009.

[40] C. Li, X. Yu, W. Yu, T. Huang, and Z.-W. Liu. Distributed event-triggered scheme for economic dispatch in smart grids. *IEEE Trans. Ind. Informat.*, DOI:10.1109/TII.2015.2479558, 2015.

[41] N. Duvvuru, and K.S., Swarup. A hybrid interior point assisted differential evolution algorithm for economic dispatch. *IEEE Trans. Power Syst.*, 26(2): 541–549, 2011.

[42] K.T. Tan, P.L. So, Y.C. Chu, and M.Z.Q. Chen. Coordinated control and energy management of distributed generation inverters in a microgrid. *IEEE Trans. Power Del.*, 28(2): 704–713, 2013.

[43] M.Z.Q. Chen, Y. Hu, C. Li, and G. Chen. Performance benefits of using inerter in semiactive suspensions. *IEEE Trans. Cont. Syst. Tech.*, 23(4): 1571–1577, 2015.

[44] A.G. Zamani, A. Zakariazadeh, and S. Jadid. Day-ahead resource scheduling of a renewable energy based virtual power plant. *Appl. Energy*, 169: 324–340, 2016.

[45] H. Cai, J. Xiang, and W. Wei. Modelling, analysis and control design of a two-stage photovoltaic generation system. *IET Renew. Power Gener.*, DOI:10.1049/iet-rpg.2015.0514, 2015.

[46] K.T. Tan, P.L. So, Y.C. Chu, and M.Z.Q. Chen. A flexible AC distribution system device for a microgrid. *IEEE Trans. Energy Convers.*, 28(3): 601–610, 2013.