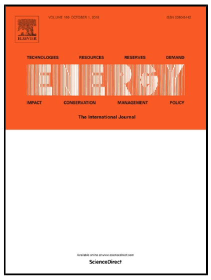
Accepted Manuscript



A cyber-physical-social system with parallel learning for distributed energy management of a microgrid

Xiaoshun Zhang, Tao Yu, Zhao Xu, Zhun Fan

PII:	S0360-5442(18)31833-4
DOI:	10.1016/j.energy.2018.09.069
Reference:	EGY 13763
To appear in:	*Energy*
Received Date:	24 April 2018
Accepted Date:	09 September 2018

Please cite this article as: Xiaoshun Zhang, Tao Yu, Zhao Xu, Zhun Fan, A cyber-physical-social system with parallel learning for distributed energy management of a microgrid, *Energy* (2018), doi: 10.1016/j.energy.2018.09.069

A cyber-physical-social system with parallel learning for distributed energy management of a microgrid

Xiaoshun Zhang[a,b], Tao Yu[c,*], Zhao Xu[b], Zhun Fan[a]

[a] *College of Engineering, Shantou University, 515063 Shantou, China*

[b] *Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong*

[c] *College of Electric Power, South China University of Technology, 510640 Guangzhou, China*

***Abstract*—A novel cyber-physical-social system (CPSS) with parallel learning is presented for distributed energy management (DEM) of a microgrid. CPSS is developed by extending the conventional cyber-physical system to the social space with human participation and interaction. Each energy supplier or each energy demander is regarded as a human in the social space, who is able to learn the knowledge, co-operate with others, and make a decision with various preference behaviors. The correlated equilibrium (CE) based general-sum game is employed for realizing the human interaction on the complex optimization subtask, while the novel adaptive consensus algorithm is used for achieving that on the simple optimization subtask with multi-energy balance constraints. A real-world system and multiple virtual artificial systems are introduced for parallel and interactive execution based on the small world network, thus a higher quality optimum of DEM can be rapidly emerged with a high probability. Case studies of a microgrid with 11 energy suppliers and 7 energy demanders demonstrate that the proposed technique can effectively achieve the human-computer collaboration and rapidly obtain a higher quality optimum of DEM compared with other centralized heuristic algorithms.**

***Keywords* – Cyber-physical-social system; Parallel learning; Correlated equilibrium; Adaptive consensus algorithm; Distributed energy management**

* Corresponding author. Tel.: +86 13002088518.
E-mail address: taoyu1@scut.edu.cn (T. Yu).

Nomenclature

Variables

P_{dg}	electricity energy output of conventional DER
H_{h}	heat energy output of heat-only unit
P_{chp}	electricity energy output of CHP unit
H_{chp}	heat energy output of CHP unit
P_{tie}	tie-line power
ΔD	responding power
a_j	action of the *j*th agent
$\boldsymbol{Q}_j^k$	the knowledge matrix of the *j*th agent
$\boldsymbol{\pi}_j$	probability distributions of state-action pairs
R_j	feedback reward of the *j*th agent
λ_p	incremental cost of the *p*th agent
x_j^k	solution of the *j*th agent at the *k*th iteation
x_p^{c}	energy consensus value of the *p*th agent
$\boldsymbol{x}^{ik}$	current solution of the *i*th VAS
$\boldsymbol{x}_{\mathrm{b}}^{rk}$	current best solution of the real-world system
p_{iw}	interaction probability between the *i*th VAS and the *w*th VAR
h^i	current best VAS in the *i*th VAR's interactive network
f_{cost}	total operating cost
v	current wind speed
S	current irradiance
T	ambient temperature

Parameters

v_{r}	rated wind speed
v_{in}, v_{out}	cut-in and cut-out wind speeds
α_{pv}	temperature coefficient
α_{dg}, β_{dg}, γ_{dg}	fuel cost coefficients of conventional DER
α_{h}, β_{h}, γ_{h}	operating cost coefficients of heat-only unit
α_{chp}, β_{chp}, γ_{chp}	operating cost coefficients of CHP unit
δ_{chp}, θ_{chp}, ξ_{chp}	operating cost coefficients of CHP unit
C_{buy}, C_{sell}	electricity buying price and selling price
a^{lin}, b^{lin}	coefficients of linear demand versus price expression
N_{dg}	number of DG
N_{h}	number of heat-only units
N_{chp}	number of CHP units
N_{dr}	number of energy demanders
N_{wt}	number of WT
N_{pv}	number of PV units
η	maximum allowable power curtailment portion
α	knowledge learning factor
γ	discount factor
ε	exploitation rate
μ	adjustment factor of energy mismatch
C_1, C_2	feedback reward coefficients
$k_{\max}$	maximal iteration number

Abbreviations

CE	correlated equilibrium
CPS	cyber-physical system
CPSS	cyber-physical-social system
DERs	distributed energy resources
EMS	energy management system
DEM	distributed energy management
RL	reinforcement learning
DG	diesel generator
VASs	virtual artificial systems
WT	wind turbine
PV	photovoltaic
CHP	combined heat and power
GA	genetic algorithm
PSO	particle swarm optimization
ABC	artificial bee colony
GSO	group search optimizer

1. Introduction

With the fast increasing renewable energy, the energy internet [1] which aims to realize the coordination among various generations, storage devices, and loads in a wide area by the internet technology, has gained extensive studies in recent years [2]. At present, energy internet is essentially a tight integration between cyber and physical resources, i.e., an application of cyber-physical system (CPS) on integrated energy systems [3]. Although CPS [4] can offer many potential benefits to the energy internet, including faster response, higher control precision, larger scale distributed coordination, and so on, but it almost ignores the human participation and interaction. In fact, the energy internet is highly coupled with the human and social characteristics [5], thus CPS may not satisfy different optimal operations of integrated energy systems in some cases, e.g., the demand response management without considering the social characteristics of different consumers. As a result, the cyber-physical-social system (CPSS) [6] was developed by logically extending CPS to the social space with human participation and interaction. As a promising system architecture of industry, CPSS is well available as a core part of future intelligent energy systems [7], rightfully including the multi-energy microgrid.

Microgrid is usually a small-scale multi-energy system with a low-voltage distribution network [8], which can effectively integrate various distributed energy resources (DERs), storage devices, and controllable loads in the grid-connected or islanded mode. In general, energy management of a microgrid seeks to minimize the total operating cost via an optimal dispatch strategy of energy balance among DERs, storage devices, tie-line power from the main grid, and controllable loads under various constraints [9]. In order to address this problem, the centralized optimization is the most commonly used type of methods, e.g., mixed integer linear programming [10] and gravitational search algorithm [11], which often produce a satisfactory result with a low total operating cost. However, it easily leads to a communication bottleneck in a microgrid with a larger number of controllable devices since the centralized energy management systems (EMS) needs to collect and process all the corresponding information from each one, which also cannot ensure the security and privacy of each owner [12]. In terms of the optimization performance, the centralized optimizer is apt to trap in a relatively high computation burden or a low-quality optimum with the great increasing controllable devices. Furthermore, it cannot satisfy the requirement of high operation reliability because the operation of an entire microgrid completely depends on the only centralized optimizer.

For the sake of handling these issues, the distributed control architecture is more suitable for the practical energy management [13], thus various distributed optimization algorithms have been proposed for distributed energy management (DEM) of a microgrid. The consensus algorithms have been deeply researched for DEM due to its remarkable self-organizing ability, significant robustness, and easy scalability [14]-[16], in which the performance influence by the time delays in communication network is strictly investigated in [17]. Besides, the sub-gradient based distributed optimization was also successfully designed for minimizing the total operating cost of a microgrid [18],[19]. In order to effectively realize the complex interaction among independent agents [20], the game theory, e.g., Stackelberg game [21] and bargaining game [22], were introduced to combine different optimization techniques for DEM. Unfortunately, all of these algorithms mainly suffer from the following four problems:

●*High dependence on the mathematical model*: The core optimizers are essentially the gradient-based algorithms, the performance of which are fully determined by the initial solution of DEM. Therefore, it easily leads to a low quality local-optimum if nonlinearities, nonconvexity, discontinuous and nondifferentiable objective functions (e.g., purchase energy cost or sell energy benefit according to direction of tie-line power), and complex constraints (e.g., the heat-power feasible operation region of combined heat and power (CHP) units [23]) exist.

●*Incapability of knowledge learning and single decision strategy:* Each game agent is constructed with only a single decision strategy and is incapable of knowledge learning, which is not consistent with the intelligent human in real-world system.

●*Invalid multi-energy interactions with consensus algorithm:* The consensus based human interaction is only suitable for only single energy interaction among the local energy supplier and demander, which cannot satisfy the multi-energy interaction with multiple energy balance constraints.

●*Inefficient optimization with a single execution system:* The traditional game theory based human interaction usually seeks an optimal equilibrium with a single execution system, which easily results in a long computation time as the iterations may involve repeated games.

In order to simultaneously address these problems, this paper proposes a CPSS with parallel learning for DEM of a microgrid, which has the following features and novelties:

- CPSS is firstly introduced to DEM of a microgrid, which fully considers the human (energy supplier or energy demander) participation and interaction, thus the obtained dispatch strategy is more practical for an optimal operation**.**
- The model-free Q-learning can effectively enable each agent to flexibly handle the nonconvex nonlinear DEM with complex constraints and nondifferentiable objective functions, while each agent can learn the knowledge from the continuous interactions with the environment. Instead of a single decision strategy, the correlated equilibrium (CE) based general-sum game [24] with multiple decision strategies is used for increasing the decision diversity of each human.
- By improving the original consensus algorithm, the adaptive consensus algorithm is proposed for effectively achieving the multi-energy interactions with multiple energy balance constraints, thus they can reach a consensus (optimum) on the incremental cost.
- Multiple virtual artificial systems (VASs) are built to guide the real-world system for DEM with a single execution system, thus the game interaction efficiency of the real-world system can be dramatically improved without any adverse trials according to the guidance by all the VASs. Besides, the small world network is adopted for constructing the interaction network among different VASs, which can properly balance the exploitation and exploration of VASs.

The remaining of this paper is organized as follows. Section 2 presents the mathematical model of DEM of a microgrid. Section 3 gives the design of CPSS with parallel learning for DEM. Case studies are carried out in Section 4. Finally, Section 5 concludes the paper.

2. Mathematical model of DEM of a microgrid

Traditionally, a microgrid adopts the centralized energy management for an optimal operation under different scenarios. It aims to determine the optimal dispatch scheme for all the energy suppliers or demanders via a centralized optimizer. In contrast, The proposed DEM allows each energy supplier or demander to calculate its own optimal energy output or demand via a coordination with interactive agents. Moreover, the primary task of DEM is to achieve the energy balance between the supply and demand, as shown in Fig. 1. In general, the supply side can consist of various energy suppliers, including wind turbine (WT), photovoltaic (PV) unit, CHP unit, diesel generator (DG), and so on. On the other hand, the demand side usually contains three types of energy demanders, i.e., residential building, factory, and commercial building.

2.1. Energy suppliers

(i) *Renewable energy resources:* for improving the generation power outputs of renewable energy resources, both WT and PV unit are operated at the maximum power points under different weather conditions, which can be expressed as follows [25],[26]:

$$P_{\mathrm{wt}} = \begin{cases} 0, & \text{for } v < v_{\mathrm{in}} \text{ and } v > v_{\mathrm{out}} \\ P_{\mathrm{wt}}^{\mathrm{r}} \dfrac{v - v_{\mathrm{in}}}{v_{\mathrm{r}} - v_{\mathrm{in}}}, & \text{for } v_{\mathrm{in}} \le v \le v_{\mathrm{r}} \\ P_{\mathrm{wt}}^{\mathrm{r}}, & \text{for } v_{\mathrm{r}} < v \le v_{\mathrm{out}} \end{cases} \tag{1}$$

$$P_{\mathrm{pv}} = P_{\mathrm{pv}}^{\mathrm{r}} \left(1 + \alpha_{\mathrm{pv}} \cdot \left(T - T_{\mathrm{ref}}\right)\right) \cdot \frac{S}{S_{\mathrm{ref}}} \tag{2}$$

where P_{wt} and P_{pv} are the current maximum power points of WT and PV unit, respectively; $P_{\mathrm{wt}}^{\mathrm{r}}$ and $P_{\mathrm{pv}}^{\mathrm{r}}$ are the rated power of WT and PV unit, respectively; v_{r} is the rated wind speed; v is the current wind speed; v_{in} and v_{out} are the cut-in and cut-out wind speeds, respectively; S is the current irradiance; S_{ref} is the reference irradiance; T is the ambient temperature; T_{ref} is the reference temperature; and α_{pv} is the temperature coefficient.

(ii) *Conventional DER:* this type of suppliers is essentially a dispatchable synchronous generator, e.g., diesel or natural gas generators, which can easily regulate its electricity energy output. In general, the fuel cost of the conventional DER can be expressed via a typical quadratic function, as follows [27]:

$$f_{\mathrm{dg}}\left(P_{\mathrm{dg}}\right) = \alpha_{\mathrm{dg}} + \beta_{\mathrm{dg}} P_{\mathrm{dg}} + \gamma_{\mathrm{dg}} P_{\mathrm{dg}}^2 \tag{3}$$

where P_{dg} is the electricity energy output of conventional DER; α_{dg}, β_{dg}, and γ_{dg} are the fuel cost coefficients of conventional DER.

(iii) *Heat-only unit:* such as gas furnace or heater exchanger, it can only provide the heat energy for local demanders in a microgrid. Similarly, its operating cost can be constructed as a quadratic function, as follows [27]:

$$f_{\mathrm{h}}\left(H_{\mathrm{h}}\right) = \alpha_{\mathrm{h}} + \beta_{\mathrm{h}} H_{\mathrm{h}} + \gamma_{\mathrm{h}} H_{\mathrm{h}}^2 \tag{4}$$

where H_{h} is the heat energy output of heat-only unit; α_{h}, β_{h}, and γ_{h} are the operating cost coefficients of heat-only unit.

(iv) *CHP unit:* as a co-generation unit, it can significantly increase the thermal efficiency and reduce the environment emissions [28] by reusing the heat, thus both the electricity and heat energy can be simultaneously generated, where the total operating cost can be calculated as follows [29]:

$$f_{\mathrm{chp}}\left(P_{\mathrm{chp}}, H_{\mathrm{chp}}\right)=\alpha_{\mathrm{chp}}+\beta_{\mathrm{chp}} P_{\mathrm{chp}}+\gamma_{\mathrm{chp}} P_{\mathrm{chp}}^{2}+\delta_{\mathrm{chp}} H_{\mathrm{chp}}+\theta_{\mathrm{chp}} H_{\mathrm{chp}}^{2}+\xi_{\mathrm{chp}} H_{\mathrm{chp}} P_{\mathrm{chp}} \tag{5}$$

where P_{chp} is the electricity energy output of CHP unit; H_{chp} is the heat energy output of CHP unit; α_{chp}, β_{chp}, γ_{chp}, δ_{chp}, θ_{chp}, and ξ_{chp} are the operating cost coefficients of CHP unit.

(v) *Main grid:* when the microgrid is operated in the grid-connected mode, the main grid can be regarded as an electricity energy supplier if the total electricity energy output of all the DERs is insufficient to balance the total electricity energy demand of all the loads in the microgrid, otherwise it will become an electricity energy demander. Hence, the operating cost from the main grid can be determine by the direction of tie-line power and the current electricity price, as follows:

$$f_{\mathrm{mg}}\left(P_{\mathrm{tie}}\right)=\begin{cases} C_{\mathrm{buy}} P_{\mathrm{tie}}, & \text{if } P_{\mathrm{tie}} \geq 0 \\ C_{\mathrm{sell}} P_{\mathrm{tie}}, & \text{otherwise} \end{cases} \tag{6}$$

where C_{buy} and C_{sell} are the electricity buying price and selling price, respectively; and P_{tie} is the tie-line power, while a negative P_{tie} will result in a negative operating cost, i.e., the electricity selling profit from the microgrid to the main grid.

2.2. Energy demanders

In order to reduce the peak-valley difference of total power load for an electric power system, the electricity company generally employs a time-of-use pricing strategy to allow the demanders to automatically adjust their electric power consumptions. In general, this process is well known as demand response (DR). Based on the linear demand versus price expression [30], the cost function of each energy demander can be calculated according to the responding power (power curtailment) and his or her sensitiveness of power loads, as follows:

$$f_{\mathrm{dr}}(\Delta D)=\frac{-1}{b^{\mathrm{lin}}} \Delta D^{2}+\frac{D_{0}-a^{\mathrm{lin}}}{b^{\mathrm{lin}}} \Delta D \tag{7}$$

where ΔD is the responding power; D_0 is the current initial electric power; a^{lin} and b^{lin} are the coefficients of linear demand versus price expression.

2.3. Social welfare and constraints

In this paper, DEM aims to maximize the social welfare (i.e., minimize the total operating cost) of the microgrid while satisfying all the constraints, including energy balance constraint, capacity limits of all energy sources, feasible operating region constraints of CHP units, and minimum demand constraint of each energy demander for the must-run loads. Hence, the mathematical model of DEM of a microgrid can be described as follows [19]:

$$\min f_{\mathrm{cost}}=\sum_{i=1}^{N_{\mathrm{dg}}} f_{\mathrm{dg}}^{i}\left(P_{\mathrm{dg}}^{i}\right)+\sum_{j=1}^{N_{\mathrm{h}}} f_{\mathrm{h}}^{j}\left(H_{\mathrm{h}}^{j}\right)+\sum_{k=1}^{N_{\mathrm{chp}}} f_{\mathrm{chp}}^{k}\left(P_{\mathrm{chp}}^{k}, H_{\mathrm{chp}}^{k}\right)+\sum_{m=1}^{N_{\mathrm{dr}}} f_{\mathrm{dr}}^{m}\left(\Delta D^{m}\right)+f_{\mathrm{mg}}\left(P_{\mathrm{tie}}\right) \tag{8}$$

subject to

$$\sum_{i=1}^{N_{\mathrm{dg}}} P_{\mathrm{dg}}^{i} + \sum_{k=1}^{N_{\mathrm{chp}}} P_{\mathrm{chp}}^{k} + \sum_{l=1}^{N_{\mathrm{wt}}} P_{\mathrm{wt}}^{l} + \sum_{d=1}^{N_{\mathrm{pv}}} P_{\mathrm{pv}}^{d} + P_{\mathrm{tie}} - \sum_{m=1}^{N_{\mathrm{dr}}} \Delta D^{m} = 0 \tag{9}$$

$$\sum_{j=1}^{N_{\mathrm{h}}} H_{\mathrm{h}}^{j} + \sum_{k=1}^{N_{\mathrm{chp}}} H_{\mathrm{chp}}^{k} - H_{\mathrm{demand}} = 0 \tag{10}$$

$$P_{\mathrm{dg}}^{i,\min} \le P_{\mathrm{dg}}^{i} \le P_{\mathrm{dg}}^{i,\max}, \quad i = 1,2,\ldots,N_{\mathrm{dg}} \tag{11}$$

$$H_{\mathrm{h}}^{j,\min} \le H_{\mathrm{h}}^{j} \le H_{\mathrm{h}}^{j,\max}, \quad j = 1,2,\ldots,N_{\mathrm{h}} \tag{12}$$

$$P_{\mathrm{chp}}^{k,\min}\left(H_{\mathrm{chp}}^{k}\right) \le P_{\mathrm{chp}}^{k} \le P_{\mathrm{chp}}^{k,\max}\left(H_{\mathrm{chp}}^{k}\right), \quad k = 1,2,\ldots,N_{\mathrm{chp}} \tag{13}$$

$$H_{\mathrm{chp}}^{k,\min}\left(P_{\mathrm{chp}}^{k}\right) \le H_{\mathrm{chp}}^{k} \le H_{\mathrm{chp}}^{k,\max}\left(P_{\mathrm{chp}}^{k}\right), \quad k = 1,2,\ldots,N_{\mathrm{chp}} \tag{14}$$

$$P_{\mathrm{tie}}^{\min} \le P_{\mathrm{tie}} \le P_{\mathrm{tie}}^{\max} \tag{15}$$

$$0 \le \Delta D^{m} \le \eta D_{0}^{m}, \quad m = 1,2,\ldots,N_{\mathrm{dr}} \tag{16}$$

where the superscripts i, j, k, m, l, and d represent the ith DG, the jth heat-only unit, the kth CHP unit, the mth energy demander, the lth WT, and the dth PV unit, respectively; the superscripts *min* and *max* represent the lower and upper limits, respectively; N_{dg} is the number of DG; N_{h} is the number of heat-only units; N_{chp} is the number of CHP units; N_{dr} is the number of energy demanders; N_{wt} is the number of WT; N_{pv} is the number of PV units; and η denotes the maximum allowable power curtailment portion of the current initial electric power, which can ensure the normal operation of must-run loads for the energy demanders.

Since both WT and PV units are operated at the maximum power points under different weather conditions [31], their maintenance costs are fixed for each optimization task. Hence, the operating costs of WT and PV unit are not considered in the total operating cost f_{cost} due to their inherent zero fuel consumption. Moreover, the feasible operating region constraints of CHP units (13) and (14) indicate that the electricity energy output and heat energy output are tightly coupled. In general, the shape of the feasible operating region is mainly determined by the struct of CHP units, e.g., the primary mover. It usually consists of two types, including one segment shape and two segment shape [32], as illustrated in Fig. 2. In fact, both of them belong to convex and nonconvex feasible operating regions, respectively. For example, the back-pressure CHP unit with condensing and auxiliary cooling options, gas turbines, and combined gas and steam cycles can result in the nonconvex feasible operating region [33]. It can be observed from Fig. 2 that the energy outputs of CHP units should be enclosed by the boundary curves ABCD or ABCDEF [23], where both the lower and upper limits of electricity energy output are determined by different heat energy outputs and vice versa. In order to reduce the optimization difficulty, DEM of a microgrid is decomposed into two optimization subtasks. The first one is responsible for optimizing the tie-line power and the heat energy outputs of CHP units, which is relatively complex with a nondifferentiable operating cost from the main grid (6) and the feasible operating region constraints. Based on the decision results of the first optimization subtasks, the second one with the rest of operating cost and constraints is essentially a convex optimization and relatively easy to be addressed.

3. CPSS with parallel learning for DEM

3.1. CPSS framework for DEM of a microgrid

As illustrated in Fig. 3, CPSS is a complex system with three dimensions, including physical space, cyberspace, and social space, and all of them are tightly connected via the cyberspace [6]. For DEM of a microgrid, the main task of CPSS is to maximize the social welfare and to react to the physical space. Compared with CPS, the major improvement part of CPSS is the social space with human beings, such as human behaviors and human interactions. For each optimization task, each energy supplier or energy demander firstly acquires current operating parameters of the corresponding distributed device from the physical space, then each of them will autonomously make a dispatch decision through the interaction with others in social space based on the communication and computation in cyberspace with parallel learning, finally the optimal dispatch strategy will be issued to each distributed device for optimal control in the physical space.

For effectively searching a high quality dispatch strategy, a CE based general-sum game with model-free Q-learning [24] is used for achieving the human interaction on the complex optimization subtask, while the novel adaptive consensus algorithm is implemented for human interaction on the simple optimization subtask.

3.2. Parallel learning with multiple parallel systems

According to the real-world system, multiple parallel VASs [34] are constructed for different evolutions of DEM in a microgrid. In this paper, the real-world system mainly provides the optimization model (8)-(16), the current best solution, and the energy management knowledge of each agent to multiple VASs, then each VAS can generate an optimal dispatch strategy via the human interactions and risk-free trial-and-error, while the energy management knowledge of each agent will be updated. Consequently, the parallel n-VASs will vote for n optimal dispatch strategies and provide their energy management knowledge to the real-world system, while each VAS will improve its dispatch strategy and energy management knowledge through learning from its interactive VASs based on small world network, as shown in Fig. 4.

3.2.1 CE based human interaction on complex optimization subtask

(i) CE based general-sum game

For a general-sum game, a CE is more general than a Nash equilibrium as the set of Nash equilibria is wholly included in the set of correlated equilibria [24]. Generally speaking, a CE is a probability distribution of joint actions from which no agent is motivated to deviate unilaterally, which can be combined with Q-learning, as follows:

$$\begin{cases} \sum_{\vec{a}_{-j} \in \boldsymbol{A}_{-j}(s_k)} \boldsymbol{\pi}_j(s_k, \vec{a}) \boldsymbol{Q}_j^k\left(s_k, (\vec{a}_{-j}, a_j)\right) \geq \sum_{\vec{a}_{-j} \in \boldsymbol{A}_{-j}(s_k)} \boldsymbol{\pi}_j(s_k, \vec{a}) \boldsymbol{Q}_j^k\left(s_k, (\vec{a}_{-j}, a_j^{\text{o}})\right) \\ \boldsymbol{A}_{-j} = \prod_{p \neq j} \boldsymbol{A}_p,\ \vec{a}_{-j} = \prod_{p \neq j} a_p,\ \vec{a} = (\vec{a}_{-j}, a_j),\ a_j^{\text{o}} \neq a_j \end{cases} \tag{17}$$

where $\boldsymbol{\pi}_j$ is the probability distributions of state-action pairs of the jth agent, which can be called a CE when it satisfies the inequality constraint (17); $\boldsymbol{Q}_j^k$ is the knowledge matrix of the jth agent at the kth iteration, which represent the knowledge values of station action pairs; s_k is the state of the multi-agent system at the kth iteration; $\vec{a}$=[a_1, …, a_j, …, a_J] is the joint action of all the agents; a_j is the action of the jth agent; J is the number of agents; $\vec{a}_{-j}$ is the joint action of all the agents except the jth agent; $\boldsymbol{A}(s_k)$ is the agents' set of available joint actions in state s_k; $\boldsymbol{A}_j$ is the ith agent's set of pure actions; and a_j^{o} is the jth agent's any other action except a_j.

(ii) Knowledge learning

According to the state-action-reward-state data via continuous interactions with the environment, each agent can update its own knowledge of different state-action pairs with the feedback rewards by reinforcement learning. In this paper, Q-learning is used for achieving this learning process, thus the knowledge can be stored by the Q-value matrix, as follows [35]:

$$V_j(s_{k+1}) = \sum_{\vec{a} \in \boldsymbol{A}(s_{k+1})} \boldsymbol{\pi}_j(s_{k+1}, \vec{a}) \boldsymbol{Q}_j^k(s_{k+1}, \vec{a}) \tag{18}$$

$$\boldsymbol{Q}_j^{k+1}(s_k, \vec{a}) = \boldsymbol{Q}_j^k(s_k, \vec{a}) + \alpha\left[(1-\gamma) R_j(s_k, \vec{a}) + \gamma V_j(s_{k+1}) - \boldsymbol{Q}_j^k(s_k, \vec{a})\right] \tag{19}$$

where $V_j(s_{k+1})$ denotes the state value-function of the jth agent for state s_{k+1}; α is the knowledge learning factor; γ is the discount factor; and $R_j(s_k, \vec{a})$ is the feedback reward after implementing a joint action $\vec{a}$ at the state s_k.

(iii) Decision strategies

In the complex optimization subtask, the strategy decision of each agent is divided into two processes. Firstly, each agent choose a pure action strategy (i.e., interval of optimization) according to its preference behavior, then an accurate solution can be determined by the non-uniform mutation operator based on the local optimum of the corresponding interval. In this paper, four human decision strategies are introduced to each agent for selecting a pure action, as [24]

- *Utilitarian behavior*: maximize the sum of all agents' benefits, as follows:

$$\max f_{\text{b}}(\boldsymbol{\pi}_j) = \sum_{j=1,2,\ldots,J} \sum_{\vec{a} \in \boldsymbol{A}(s_k)} \boldsymbol{\pi}_j(s_k, \vec{a}) \boldsymbol{Q}_j^k(s_k, \vec{a}) \tag{20}$$

- *Egalitarian behavior*: maximize the minimum of all agents' benefits, as follows:

$$\max f_{\text{b}}(\boldsymbol{\pi}_j) = \min_{j=1,2,\ldots,J} \sum_{\vec{a} \in \boldsymbol{A}(s_k)} \boldsymbol{\pi}_j(s_k, \vec{a}) \boldsymbol{Q}_j^k(s_k, \vec{a}) \tag{21}$$

- *Plutocratic behavior*: maximize the maximum of all agents' benefits, as follows:

$$\max f_{\text{b}}(\boldsymbol{\pi}_j) = \max_{j=1,2,\ldots,J} \sum_{\vec{a} \in \boldsymbol{A}(s_k)} \boldsymbol{\pi}_j(s_k, \vec{a}) \boldsymbol{Q}_j^k(s_k, \vec{a}) \tag{22}$$

- *Dictatorial behavior*: maximize the maximum of any individual agent's benefits, as follows:

$$\max f_{\mathrm{b}}\left(\boldsymbol{\pi}_j\right)=\sum\nolimits_{\vec{a}\in \boldsymbol{A}\left(s_k\right)}\boldsymbol{\pi}_j\left(s_k,\vec{a}\right)\boldsymbol{Q}_j^k\left(s_k,\vec{a}\right) \tag{23}$$

where f_{b} is the behavior function, in which the maximum and the corresponding optimal CE can be calculated by linear programming with the inequality constraints (17) and the following constraints, as

$$\sum\nolimits_{\vec{a}\in \boldsymbol{A}\left(s_k\right)}\boldsymbol{\pi}_j\left(s_k,\vec{a}\right)=1,\ 0\le \boldsymbol{\pi}_j\left(s_k,\vec{a}\right)\le 1 \tag{24}$$

After acquiring the optimal CE $\boldsymbol{\pi}_j^*$, a pure action of each agent and an accurate dispatch strategy can be determined. Aiming at a proper trade-off between exploration and exploitation, the ε-Greedy rule [36] is used for interval selection, as

$$a_j=\begin{cases}\arg\max\limits_{a_j\in \boldsymbol{A}_j}\boldsymbol{\pi}_j\left(s_k,\left(a_{-j},a_j\right)\right), & \text{if } q_0\le \varepsilon\\ a_{\mathrm{rand}}, & \text{otherwise}\end{cases} \tag{25}$$

$$x_j^k=\begin{cases}x_j^{\mathrm{best}}\left(a_j\right)+\Delta\left[k,x_j^{\mathrm{ub}}\left(a_j\right)-x_j^{\mathrm{best}}\left(a_j\right)\right], & \text{if rand(0,1)<0.5}\\ x_j^{\mathrm{best}}\left(a_j\right)-\Delta\left[k,x_j^{\mathrm{best}}\left(a_j\right)-x_j^{\mathrm{lb}}\left(a_j\right)\right], & \text{otherwise}\end{cases} \tag{26}$$

$$\begin{cases}x_j^{\mathrm{ub}}\left(a_j\right)=x_j^{\min}+a_j\cdot\left(x_j^{\max}-x_j^{\min}\right)/\left|\boldsymbol{A}_j\right|\\ x_j^{\mathrm{lb}}\left(a_j\right)=x_j^{\min}+\left(a_j-1\right)\cdot\left(x_j^{\max}-x_j^{\min}\right)/\left|\boldsymbol{A}_j\right|\end{cases} \tag{27}$$

$$\Delta\left[k,y\right]=y\cdot\left(1-r^{(1-k/k_{\max})^b}\right) \tag{28}$$

where q_0 is a uniform random value from [0, 1]; ε is the exploitation rate which represents the probability of exploitation; a_{rand} denotes a random action (exploration) chosen from the action space $\boldsymbol{A}_j$; $x_j^{\mathrm{best}}(a_j)$ is the previous best optimal solution at the action (a_j) interval of the jth controllable variable; $x_j^{\mathrm{ub}}(a_j)$ and $x_j^{\mathrm{lb}}(a_j)$ are the upper and lower bounds of the action (a_{j}) interval, respectively ; $x_j^{\min}$ and $x_j^{\max}$ are the minimum and maximum values of the jth controllable variable, respectively; $\Delta[k,y]$ is a decay function as the iteration k increases; r is a uniform random value from [0,1]; b is the system parameter which determines the degree of non-uniformity; and $k_{\max}$ is the maximal iteration number.

3.2.2 Adaptive consensus algorithm based human interaction on simple optimization subtask

(i) Graph theory of interaction network

The interaction network among humans can be typically built with a directed graph G=(V, E, A), where $V=\{v_1, v_2, \ldots, v_N\}$ is the set of nodes (agents); $E\subseteq V\times V$ denotes the edges (interactions); and $A=[a_{pq}]\in R^{N\times N}$ is a weighted adjacency matrix [37]. Based on these most basic elements, the Laplacian matrix $L=[l_{pq}]\in R^{N\times N}$ and row stochastic matrix $D=[d_{pq}]\in R^{N\times N}$ of the graph G can be calculated as follows:

$$l_{pp}=\sum_{p=1,q\neq p}^{N}a_{pq},l_{pq}=-a_{pq},\forall p\neq q \tag{29}$$

$$d_{pq}[k]=\left|l_{pq}\right|\bigg/\sum_{q=1}^{N}\left|l_{pq}\right|,\quad p=1,2,\ldots,N \tag{30}$$

(ii) Adaptive consensus algorithm on incremental cost

The adaptive consensus algorithm inherently represents a herd behavior of human interactions, i.e., each agent will regulate its own state to reach a consensus with the adjacent agents after acquiring their current states. In this paper, the first-order adaptive consensus algorithm is adopted for this consensus process, as follows [38]:

$$s_p[k+1]=\sum_{q=1}^{N}d_{pq}[k]s_q[k] \tag{31}$$

where s_p is the state of the pth agent.

Note that the simple optimization subtask only has a unique minimum point as it is a strictly convex optimization, thus its global optimum can be obtained when all the agents can reach a consensus on the incremental cost while satisfying various constraints. Hence, the incremental cost is taken as the consensus state for human interactions, which can be written as [14]

$$\lambda_p=\frac{\partial f_p\left(x_p\right)}{\partial x_p}=\kappa_p x_p+\varphi_p \tag{32}$$

where λ_p is the incremental cost of the pth agent; x_p is the controllable variable (energy output or demand) of the pth agent; κ_p and φ_p are the incremental cost coefficients of the pth agent, respectively, which can be determined by the corresponding cost coefficients; and f_p is the operating cost of the pth agent.

In order to satisfy the energy balance constraints (9)-(10), the electricity energy mismatch ΔE and heat energy mismatch ΔH between the energy suppliers and energy demanders are introduced in adaptive consensus algorithm, as follows:

$$\Delta E=\sum_{i=1}^{N_{\mathrm{dg}}}P_{\mathrm{dg}}^{i}+\sum_{k=1}^{N_{\mathrm{chp}}}P_{\mathrm{chp}}^{k}+\sum_{l=1}^{N_{\mathrm{wt}}}P_{\mathrm{wt}}^{l}+\sum_{d=1}^{N_{\mathrm{pv}}}P_{\mathrm{pv}}^{d}+P_{\mathrm{tie}}-\sum_{m=1}^{N_{\mathrm{dr}}}\Delta D^{m} \tag{33}$$

$$\Delta H=\sum_{j=1}^{N_{\mathrm{h}}}H_{\mathrm{h}}^{j}+\sum_{k=1}^{N_{\mathrm{chp}}}H_{\mathrm{chp}}^{k}-H_{\mathrm{demand}} \tag{34}$$

It can be found from (32) that an increasing incremental cost will lead to an increasing energy output and an decreasing energy demand, thus the consensus interaction should be carefully designed to satisfy the energy balance constraints following this changing rule, as follows:

- *Unified consensus*: if the signs of ΔE and ΔH are consistent, i.e., $\Delta E\cdot\Delta H\geq 0$, then all the agents can update their incremental cost state in an unified interaction network, as

$$\lambda_p[k+1]=\begin{cases}\sum_{q=1}^{N}d_{pq}[k]\lambda_q[k]-\mu\Delta E, & p\in\Omega_{\mathrm{E}}\\ \sum_{q=1}^{N}d_{pq}[k]\lambda_q[k]-\mu\Delta H, & p\in\Omega_{\mathrm{H}}\end{cases} \tag{35}$$

- *Independent consensus*: if the signs of ΔE and ΔH are inconsistent, i.e., $\Delta E\cdot\Delta H<0$, then the electricity agents and heat agents need to be separated to update their incremental cost state in two independent interaction network, as

$$\lambda_p[k+1]=\begin{cases}\sum\limits_{q\in\Omega_{\mathrm{E}}} d_{pq}^{\mathrm{E}}[k]\lambda_q[k]-\mu\Delta E,\ \ p\in\Omega_{\mathrm{E}}\\ \sum\limits_{q\in\Omega_{\mathrm{H}}} d_{pq}^{\mathrm{H}}[k]\lambda_q[k]-\mu\Delta H,\ p\in\Omega_{\mathrm{H}}\end{cases}\tag{36}$$

where Ω_{E} and Ω_{H} represent the sets of electricity agents and heat agents, respectively; $d_{pq}{}^{\mathrm{E}}$ is the (p,q) entry of the row stochastic matrix of the interaction network among the electricity agents; $d_{pq}{}^{\mathrm{H}}$ is the (p,q) entry of the row stochastic matrix of the interaction network among the heat agents; and μ denotes the adjustment factor of energy mismatch, $\mu>0$.

Therefore, each agent will regulate its incremental cost between these two consensus mode according to the sign of $(\Delta E\cdot\Delta H)$, as illustrated in Fig. 5.

By fully considering the lower and upper limits of each controllable variable (11)-(16), the energy output or demand of each agent can be determined based on (32), as follows [14]:

$$x_p^{\mathrm{c}}=\frac{\lambda_p[k]-\varphi_p}{\kappa_p}\tag{37}$$

$$x_p^k=\begin{cases}x_p^{\min},\ \ \text{if}\, x_p^{\mathrm{c}}<x_p^{\min}\\ x_p^{\mathrm{c}},\ \ \ \ \text{if}\, x_p^{\min}\le x_p^{\mathrm{c}}\le x_p^{\max}\\ x_p^{\max},\ \text{if}\, x_p^{\mathrm{c}}>x_p^{\max}\end{cases}\tag{38}$$

where $x_p{}^{\mathrm{c}}$ denotes the energy consensus value of the pth agent; $x_p{}^{\min}$ and $x_p{}^{\max}$ are the minimum and maximum values of the pth controllable variable, respectively.

3.2.3 Interaction between different parallel systems

(i) Interaction between VASs and the real-world system

In the initial phase, the real-world system will provide the prior energy management knowledge $\boldsymbol{Q}_j{}^{\mathrm{p}*}$($j$=1,2,…, J) and the optimal incremental cost λ* of a similar optimization task to each VAS, which can be regarded as the initial knowledge matrices $\boldsymbol{Q}_j{}^{i0}$ and the initial incremental cost λ of each agent. On the other hand, the agent of real-world system will update its current best solution and the knowledge matrix according to the current solutions of VASs, as follows:

$$h=\arg\min_{i=1,2,\ldots,n} f_{\mathrm{cost}}\left(\boldsymbol{x}^{ik}\right)\tag{39}$$

$$\boldsymbol{x}_{\mathrm{b}}^{\mathrm{r}k}=\begin{cases}\boldsymbol{x}^{hk},\ \text{if}\ f_{\mathrm{cost}}\left(\boldsymbol{x}_{\mathrm{b}}^{\mathrm{r}k}\right)\ge f_{\mathrm{cost}}\left(\boldsymbol{x}^{hk}\right)\\ \boldsymbol{x}_{\mathrm{b}}^{rk},\ \text{otherwise}\end{cases}\tag{40}$$

$$\boldsymbol{Q}_j^{\mathrm{r}k}=\begin{cases}\boldsymbol{Q}_j^{\mathrm{r}k}+r_{\mathrm{Q}}\cdot\left(\boldsymbol{Q}_j^{hk}-\boldsymbol{Q}_j^{\mathrm{r}k}\right),\ \text{if}\ f_{\mathrm{cost}}\left(\boldsymbol{x}_{\mathrm{b}}^{\mathrm{r}k}\right)\ge f_{\mathrm{cost}}\left(\boldsymbol{x}^{hk}\right)\\ \boldsymbol{Q}_j^{\mathrm{r}k},\qquad\qquad\qquad\qquad\text{otherwise}\end{cases}\tag{41}$$

where $\boldsymbol{x}^{ik}$ represent the current solution of the ith VAS, which consists of all the controllable variables; h denotes the current best VAS with the smallest total operating cost; $\boldsymbol{x}_{\mathrm{b}}^{rk}$ is the current best solution of the real-world system; r_{Q} is the random matrix from [0,1] with the same scale of knowledge matrix; and $\boldsymbol{Q}_j^{\mathrm{r}k}$ is the current knowledge matrix of the jth agent in the real-world system.

(ii) Interaction among VASs

Generally speaking, the larger otherness between different VASs will lead to more diverse dispatch strategies, which can effectively avoid the low-quality local optimum, but it will consume more computation time to search the potential global optimum. To properly balance them, the small world network is used for constructing the interaction network among VASs, in which each VAS can stochastically interact with any other VASs with a decreasing probability, as follows [39]:

$$\rho_{iw} = \left(1 - \frac{k}{k_{\max}}\right) \cdot C_{\mathrm{p}}, \quad w = 1,2,\ldots,n \tag{42}$$

where p_{iw} is the interaction probability between the ith VAS and the wth VAR; $k_{\max}$ is the maximal iteration number; and C_{p} is the probability coefficient, with $0<C_{\mathrm{p}}<1$.

Similarly, each VAS will update its current best solution and the knowledge matrix according to the current solutions of its interactive VASs, as follows:

$$h^i = \arg\min_{w \in \Omega_i} f_{\mathrm{cost}}\left(\boldsymbol{x}^{wk}\right) \tag{43}$$

$$\boldsymbol{x}_{\mathrm{b}}^{ik} = \begin{cases} \boldsymbol{x}^{h^i k}, & \text{if } f_{\mathrm{cost}}\left(\boldsymbol{x}_{\mathrm{b}}^{ik}\right) \geq f_{\mathrm{cost}}\left(\boldsymbol{x}^{h^i k}\right) \\ \boldsymbol{x}_{\mathrm{b}}^{ik}, & \text{otherwise} \end{cases} \tag{44}$$

$$\boldsymbol{Q}_j^{ik} = \begin{cases} \boldsymbol{Q}_j^{ik} + r_{\mathrm{Q}} \cdot \left(\boldsymbol{Q}_j^{h^i k} - \boldsymbol{Q}_j^{ik}\right), & \text{if } f_{\mathrm{cost}}\left(\boldsymbol{x}_{\mathrm{b}}^{ik}\right) \geq f_{\mathrm{cost}}\left(\boldsymbol{x}^{h^i k}\right) \\ \boldsymbol{Q}_j^{ik}, & \text{otherwise} \end{cases} \tag{45}$$

where h^i denotes the current best VAS in the ith VAR's interactive network; Ω_i is the VAR set in the ith VAS's interactive network, which can be determined by (42); $\boldsymbol{x}_{\mathrm{b}}^{ik}$ is the current best solution of the ith VAS; and $\boldsymbol{Q}_j^{ik}$ is the current knowledge matrix of the jth agent in the ith VAS.

3.3 Application design for DEM

3.3.1 Communication information in each learning system

As shown in Fig. 4, all the agents will communicate with the microgrid EMS, in which each agent will transmit its current optimal energy output or demand to the microgrid EMS. For the complex optimization subtask, the microgrid EMS is regarded as an external environment for each learning agent, thus each agent can acquire the state and feedback reward after implementing an optimal CE action. Besides, each learning agent can access the current actions and knowledge matrices of other agents at any time. For the simple optimization subtask, the microgrid EMS will continuously issue the energy mismatches to each agent, thus the consensus collaboration among the agents can be achieved.

3.3.2 Design of feedback reward

To maximize the social welfare, the feedback reward should be designed to match the total operating cost f_{cost} in (8), i.e., a smaller f_{cost} encourages a larger feedback reward, which can be calculated as follows:

$$R_j\left(s_k,\vec{a}\right)=C_1-\left(f_j\left(x_j^k\right)+\frac{1}{J}\sum_{p=1,p\neq j}^{N}f_p\left(x_p^k\right)\right)\Big/C_2\quad j=1,2,\ldots,J \tag{46}$$

where f_j is the operating cost of the jth learning agent; C_1 and C_2 are the feedback reward coefficients.

3.3.3 Execution procedure

In summary, the CPSS with parallel learning for DEM of a microgrid is given in Fig. 6. Note that the convergence criteria of adaptive consensus algorithm is that both the electricity and heat energy mismatches (ΔE and ΔH) can simultaneously satisfy the energy mismatch tolerance τ, i.e., i.e., $|\Delta E|<\tau$ & $|\Delta H|<\tau$, where τ is set to be 0.0001 in this paper. Beside, each agent will prefer to produce a decision behavior from (20)-(23) based on a probability distribution, which is set to be [0.7, 0.1, 0.1, 0.1] for utilitarian, egalitarian, plutocratic, and dictatorial, respectively.

4. Case studies

4.1 Simulation model

For the microgrids, the best available energy system should be selected, capable of satisfying the demand requirements for a particular area. During this process, the design engineers need to determine the optimal generation units selection, sizing, and siting for the microgrid. It is usually built with a minimization of planning cost under various constraints [40], such as technical, environmental, geographical, social and regulatory constraints. In order to obtain the optimal planning scheme, various optimization technique can be used for handling this problem. Since this paper mainly focuses on the microgrid operation rather than the microgrid planning, the testing system simply consults from the built microgrid in [19] and [41].

The testing microgrid is operated in grid-connected mode, which contains 11 energy suppliers, 7 energy demanders, where the suppliers consists of 3 PV units, 2 WTs, 2 DGs, 2 CHPs, 1 heat-only unit, and the main grid, as shown in Fig. 7. Besides, the complex optimization subtask is composed of tie-line power and the heat energy outputs of CHP units, while the simple optimization subtask consists of the rest controllable variables. The main parameters of testing microgrid are given in Tables 1 to 6 and Figs. 7-8. In particular, CHP_1 is enclosed by the two segment shape in Fig. 2, where the boundary nodes ABCDEF are (0, 1), (0.15, 1), (0.6, 0.85), (0.3, 0.05), (0.08, 0.2), and (0, 0.2), respectively. Besides, CHP_2 is enclosed by the one segment shape with the boundary nodes ABCD, which are (0, 0.6), (0.6, 0.5), (0.35, 0.05), and (0, 0.1), respectively. Furthermore, the switching cyber connection is designed around the heat unit (Heat_1), i.e., if $\Delta E\cdot\Delta H\geq 0$, then Heat_1 will simultaneously connect with L_1, L_2, and DG_1, otherwise, Heat_1 will disconnect with them, while L_1 and L_2 will connect with each other. To construct the human participations in real-world system, thirteen experimenters are employed as the controllers for each energy supplier or demander, respectively. Through trial-and-error, the main parameters of parallel learning are given in Table 7.

In order to verify the performance of the proposed technique, four commonly used heuristic algorithms are introduced for comparisons, including genetic algorithm (GA) [42], particle swarm optimization (PSO) [43], artificial bee colony (ABC) [44] , group search optimizer (GSO) [45], where population size and maximum iteration number are set to be 50 and 250, respectively. Moreover, all the simulations are undertaken in Matlab R2016a by a small server with Intel(R) Xeon (R) E5-2670 v3 CPU at 2.3 GHz with 64 GB of RAM.

4.2 Study of convergence

Figs. 9-11 provide the convergence of parallel learning under scenario 3 (at 14:00) with peak price, where the number of VASs is set to be 5. It can be found from Fig. 10 that each system can converge to a high-quality optimal solution with a lower total operating cost via an effective interaction with other systems, especially for the real-world system. At the same time, three energy suppliers (tie-line power, CHP_1, and CHP_2) can achieve an optimal CE among them after around 200 CE based game iterations, while the feedback reward of each one will increase as its energy output decrease and vice versa, as illustrated in Fig. 10. After each game iteration, twelve energy agents (suppliers and demanders) will interact their incremental costs for reaching a consensus by adaptive consensus algorithm. As shown in Fig. 11(a), the interactive incremental cost will update between unified consensus mode and independent consensus mode according to the dynamic energy mismatches, in which the heat-only unit $Heat_1$ cannot reach a consensus with other electricity energy agents due to the heat energy balance constraint (10). Besides, the actual incremental costs of some energy agents have reached their limits after a fewer interactions, thus their energy outputs or responding power can strictly satisfy their capacity lower and upper limits, as shown in Fig. 11(b)-(e). Finally, both the electricity and heat energy mismatches (ΔE and ΔH) simultaneously satisfy the energy mismatch tolerance after a series of corrections (See Fig. 11(f)), i.e., $|\Delta E|<\tau$ and $|\Delta H|<\tau$.

4.3 Comparative results and discussions

Fig. 12 shows the convergence of total operating costs obtained by different algorithms under scenario 3. It is clear that both parallel learning and PSO outperform other three algorithms because each of them can obtain a lower total operating cost in a fewer iterations. Moreover, the detailed optimal dispatch strategies of all the energy suppliers and demanders obtained by different algorithms are listed in Table 8. It verifies that the proposed parallel learning can not only realize the distributed optimization of DEM with human participation and interaction, but also can guarantee the quality of the obtained optimal solution compared with the commonly used centralized heuristic algorithms, which results from the deep exploitations and explorations with various decision strategies (20)-(23) in multiple VASs.

In order to further test the performance of parallel learning, all the algorithms are implemented for three different scenarios (See Table 6) with 50 independent runs. Fig. 13 shows the statistical results obtained by them, where parallel learning I, parallel learning II, and parallel learning III represent the parallel learning with different numbers of VASs, i.e., 5, 10, and 15 respectively. It also proves that parallel learning can search the highest quality optimal solution with a lowest total operating cost under each scenario, especially parallel learning with more VASs. This obviously demonstrates that the increasing VASs can generate a potential higher quality optimal solution via a deeper exploitation and exploration with various decision strategies. However, it requires more computation capability and consumes more execution time, as shown in Fig. 13(d). Furthermore, the execution time of parallel learning is slightly larger than that of GA, PSO, and GSO since each game agent requires a linear programming computation to search an optimal CE policy at each iteration.

4.4 Study of influence by renewables uncertainty

For the DEM of a microgrid, the uncertainty of the input parameters is ubiquitous due to the unavoidable forecast error, especially for the intermittent energy out of renewables. In order to figure out the influence by the renewables uncertainty, Fig. 14 provides the statistical results obtained by proposed system under different uncertainty degrees in 50 runs, where the uncertainty degree represent the forecast error of the total energy output of renewables; both the total operating cost and the execution time are the average of 50 runs. It is clearly that the obtained total operating cost decreases as the forecast error increases under each scenario, which results from that a larger energy output of renewables is beneficial to reduce the fuel cost of the microgrid. On the other hand, the execution time of parallel learning is almost not affected by the renewable uncertainty (See Fig. 14(d)), which also verifies the high convergence stability of the proposed method. Moreover, there is only little difference on the total operating costs obtained among all the parallel learning with different numbers of VASs. This reveals that the testing microgrid only requires a small number of VASs for parallel learning based DEM.

6. CONCLUSION

In this paper, a novel CPSS with parallel learning has been proposed for DEM of a microgrid. The main contributions can be summarized as follows

(i) The proposed CPSS is constructed by considering the human participation and interaction in social space, which can widely yield potential optimal distributed strategies for DEM in a real-world microgrid.

(ii) The CE based human interaction with various decision strategies can efficiently search an optimal dispatch strategy of a complex optimization subtask of DEM, including the tie-line power with a nondifferentiable objective function, and the CHP units with complex feasible operating regions constraints.

(iii) The adaptive consensus algorithm based human interaction can effectively achieve the self-organization of each human with local communications, while the switch between unified consensus and independent consensus can simultaneously satisfy the consensus requirement and the multi-energy balance constraints.

(iv) The real-world system can continuously learn the knowledge from multiple virtual VASs, in which a deep exploitation and exploration can be implemented in each VAS. Hence, the quality of the obtained optimal solution of DEM can be guaranteed for the real-world system without any adverse trials, i.e., a lower total operating cost (higher social welfare) of a microgrid can be obtained.

(v) The proposed CPSS with parallel learning is not only highly independent on the mathematical model for a specific optimization, but also can achieve a high-quality optimum in a distributed manner. Hence, it can also be applied to other optimizations of the complex energy system.

Acknowledgment

The authors gratefully acknowledge the support of National Natural Science Foundation of China (51477055, 51777078), Hong Kong RGC Theme based Research Scheme Grants No.T23-407/13N and T23-701/14N, Guangdong Key Laboratory of Digital Signal and Image Processing, Project of Educational Commission of Guangdong Province of China (2017KZDXM032) and Project of International, as well as Hong Kong, Macao & Taiwan Science and Technology Cooperation Innovation Platform in Universities in Guangdong Province (2015KGJH2014).

References

[1] Huang AQ, Crow ML, Heydt GT, Zheng JP, Dale SJ. The future renewable electric energy delivery and management system: The energy internet. Proc IEEE 2011;99(1):133–48.

[2] Zhou K, Yang S, Shao Z. Energy internet: The business perspective. Appl Energy 2016;178:212-22.

[3] Palensky P, Widl E, Elsheikh A. Simulating cyber-physical energy systems: Challenges, tools and methods. IEEE Trans Syst Man Cybern Syst 2014;44(3):318–26.

[4] US National Science Foundation. Cyber-physical systems (CPS). NSF08-611, 2008, http://www.nsf.gov/pubs/2008/nsf08611/nsf08611.htm.

[5] Deng JL, Wang FY, Chen YB, Zhao XY. From industry 4.0 to energy 5.0: Concept and framework of intelligent energy systems. ACTA Autom Sinica 2015;41(12):2003-16.

[6] Zeng J, Yang LT, Lin M, Ning H, Ma J. A survery: Cyber-physical-social systems and their system-level design methodology. Future Gener Comp Syst (2016) https://doi.org/10.1016/j.future.2016.06.034.

[7] Wang FY. The emergence of intelligent enterprises: From CPS to CPSS. IEEE Intell Syst 2010;25(4):85-8.

[8] Guzzi F, Neves D, Silva CA. Integration of smart grid mechanisms on microgrids energy modelling. Energy 2017;129:321-30.

[9] Fang X, Ma S, Yang Q, Zhang J. Cooperative energy dispatch for multiple autonomous microgrids with distributed renewable sources and storages. Energy 2016;99:48-57.

[10] Wouters C, Fraga ES, James AM. An energy integrated, multi-microgrid, MILP (mixed-integer linear programming) approach for residential distributed energy system planning–A South Australian case-study. Energy 2015;85:30-44.

[11] Beigvand SD, Abdi H, Scala ML. Hybrid Gravitational Search Algorithm-Particle Swarm Optimization with Time Varying Acceleration Coefficients for large scale CHPED problem. Energy 2017;126:841-853.

[12] Amin SM. Smart grid security, privacy, and resilient architectures: Opportunities and challenges. In: Proc IEEE power and energy society general meeting, San Diego, CA, USA. p. 1-2.

[13] Khan MRB, Jidin R, Pasupuleti J. Multi-agent based distributed control architecture for microgrid energy management and optimization. Energy Convers Manage 2016;112:288-307.

[14] Zhang Z, Chow M Y. Convergence analysis of the incremental cost consensus algorithm under different communication network topologies in a smart grid. IEEE Trans Power Syst 2012;27(4):1761-8.

[15] Yang S, Tan S, Xu JX. Consensus based approach for economic dispatch problem in a smart grid. IEEE Trans Power Syst 2013;28(4):4416–26.

[16] Rahbari-Asr N, Ojha U, Zhang Z, Chow MY. Incremental welfare consensus algorithm for cooperative distributed generation/demand response in smart grid. IEEE Trans Smart Grid 2014;5(6):2836-45.

[17] Chen G, Zhao Z. Delay Effects on consensus-based distributed economic dispatch algorithm in microgrid. IEEE Trans Power Syst (2017) doi: 10.1109/TPWRS.2017.2702179.

[18] Dall'Anese E, Zhu H, Giannakis GB. Distributed optimal power flow for smart microgrids. IEEE Trans Smart Grid 2013;4(3):1464-75.

[19] Liu N, Wang J, Wang L. Distributed energy management for interconnected operation of combined heat and power-based microgrids with demand response. J Mod Power Syst Clean Energy 2017;5(3):478-88.

[20] Saad W, Han Z, Poor HV, Basar T. Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications. IEEE Signal Processing Mag 2012;29(5):86-105.

[21] Ma L, Liu N, Zhang J, Tushar W, Yuen C. Energy management for joint operation of CHP and PV prosumers inside a grid-connected microgrid: A game theoretic approach. IEEE Trans Ind Inf 2016;2(5):1930-42.

[22] Dehghanpour K, Nehrir H. Real-time multiobjective microgrid power management using distributed optimization in an agent-based bargaining framework. IEEE Trans Smart Grid (2017) doi:10.1109/TSG.2017.2708686.

[23] Subbaraj P, Rengaraj R, Salivahanan S. Enhancement of combined heat and power economic dispatch using self adaptive real-coded genetic algorithm. Appl Energy 2009;86(6):915-21.

[24] Greenwald A, Hall K. Correlated-Q learning. in Proc 20th Int Conf Mach Learn (ICML-03), Washington, DC, USA, Aug. 21–24, 2003:242-9.

[25] Yang B, Zhang X, Yu T, Shu H, Fang Z. Grouped grey wolf optimizer for maximum power point tracking of doubly-fed induction generator based wind turbine. Energy Convers Manage 2017;133:427-443.

[26] Brini S, Abdallah HH, Ouali A. Economic dispatch for power system included wind and solar thermal energy. Leonardo J Sci 2009;14:204-20.

[27] Aghaei J, Alizadeh MI. Multi-objective self-scheduling of CHP (combined heat and power)-based microgrids considering demand response programs and ESSs (energy storage systems). Energy 2013;55:1044-54.

[28] Karki S, Kulkarni M, Mann MD, Salehfar H. Efficiency improvements throughcombined heat and power for on-site distributed generation technologies. Cogener Distrib Gener J 2007;22:19-34.

[29] Piperagkas GS, Anastasiadis AG, Hatziargyriou ND. Stochastic PSO-based heat and power dispatch under environmental constraints incorporating CHP and wind power units. Electr Power Syst Res 2011;81(1):209-18.

[30] Yousefi S, Moghaddam MP, Majd VJ. Optimal real time pricing in an agent-based retail market using a comprehensive demand response model. Energy 2011;36(9):5716-27.

[31] Yang B, Yu T, Shu H, Dong J, Jiang L. Robust sliding-mode control of wind energy conversion systems for optimal power extraction via nonlinear perturbation observers. Appl Energy 2018;210:711-23.

[32] Salgado F, Pedrero P. Short-term operation planning on cogeneration systems: A survey. Electr Power Syst Res 2008;78(5): 835-48.

[33] Abdolmohammadi HR, Kazemi A. A Benders decomposition approach for a combined heat and power economic dispatch. Energy Convers Manage 2013;71:21-31.

[34] Wang FY, Zhang J, Wei Q, Li L. PDP: Parallel dynamic programming. IEEE/CAA J Autom Sinica 2017;4(1):1-5.

[35] Yu T, Zhang XS, Zhou B, Chan KW. Hierarchical correlated Q-learning for multi-layer optimal generation command dispatch. Int J Electr Power Energy Syst 2016;78:1-12.

[36] Bianchi RAC, Celiberto LA, Santos PE, Matsuura JP, Mantaras RL. Transferring knowledge as heuristics in reinforcement learning: A case-based approach. Artif Intell 2015;226:102-21.

[37] Godsil C, Royle G. Algebraic graph theory. New York: Springer-Verlag; 2001.

[38] Moreau L. Stability of multiagent systems with time-dependent communication links. IEEE Trans Autom Control 2005;50(2):169-82.

[39] Yan X, Yang W, Shi H. A group search optimization based on improved small world and its application on neural network training in ammonia synthesis. Neurocomputing 2012;97:94-107.

[40] Gamarra C, Guerrero JM. Computational optimization techniques applied to microgrids planning: A review. Renew Sustain Energy Rev 2015;48:413-24.

[41] Aghaei J, Alizadeh M-I. Multi-objective self-scheduling of CHP (combined heat and power)-based microgrids considering demand response programs and ESSs (energy storage systems). Energy 2013;55:1044-54.

[42] Iba K. Reactive power optimization by genetic algorithm. IEEE Trans Power Syst 1994;9(2):685-92.

[43] Clerc M, Kennedy J. The particle swarm: Explosion, stability, and convergence in a multidimensional complex space. IEEE Trans Evol Comput 2002;6(1):58-73.

[44] Secui DC. The chaotic global best artificial bee colony algorithm for the multi-area economic/emission dispatch. Energy 2015;93:2518-45.

[45] He S, Wu QH, Saunders JR. Group search optimizer: An optimization algorithm inspired by animal searching behavior. IEEE Trans Evol Comput 2009;13(5):973-90.

Table 1

Parameters of renewable energy resources.

WT		PV	
v_{in}	4 m/s	S_{ref}	1000 W/m^2
v_{out}	25 m/s	T_{ref}	25℃
v_r	15 m/s	α_{pv}	-0.47%/℃
Rated power	0.5 MW	Rated power	0.5 MW

Table 2

Parameters of DGs and heat-only units.

Units	Operating cost coefficients			Capacity (MW)	
	α_i	β_i	γ_i	Minimum	Maximum
DG_1	10.193	210.36	250.2	0	0.5
DG_2	2.305	301.4	1100	0.04	0.2
$Heat_1$	33	12.3	6.9	0	2

Table 3

Operating cost coefficients of CHP units.

Units	α_i	β_i	γ_i	δ_i	θ_i	ζ_i
CHP_1	339.5	185.7	44.2	53.8	38.4	40
CHP_2	100	288	34.5	21.6	21.6	8.8

Table 4

Electricity buying and selling prices from the main grid.

Tariff type	Time	Buying price ($/MWh)	Selling price ($/MWh)
Tariff 1 (low price)	00:00—06:59	192	180
Tariff 2 (shoulder price)	07:00—10:59 16:00—18:59 22:00—23:59	238	200
Tariff 3 (peak price)	11:00—15:59 19:00—21:59	317	260

Table 5

Operating cost coefficients of linear demand versus price expression at different tariffs.

Demanders	Tariff 1		Tariff 2		Tariff 3	
	a^{lin}	b^{lin}	a^{lin}	b^{lin}	a^{lin}	b^{lin}
L_1	0.9	-0.0028	0.95	-0.0025	1	-0.002
L_2	0.9	-0.0028	0.95	-0.0025	1	-0.002
L_3	0.9	-0.0025	0.95	-0.002	1	-0.001
L_4	0.9	-0.0025	0.95	-0.002	1	-0.001
L_5	0.9	-0.0025	0.95	-0.002	1	-0.001
L_6	0.9	-0.0042	0.95	-0.004	1	-0.0035
L_7	0.9	-0.0042	0.95	-0.004	1	-0.0035

Table 6

Forecasting results of energy demand and renewable energy outputs under three scenarios.

Demanders or renewables	Energy type	Forecasting results (MW)		
		Scenario 1 (00:00)	Scenario 2 (09:00)	Scenario 3 (14:00)
L_1	Electricity	0.13	0.38	0.5
	Heat	0.04	0.04	0.03
L_2	Electricity	0.12	0.33	0.4
	Heat	0.03	0.03	0.03
L_3	Electricity	0.14	0.42	0.6
	Heat	0.05	0.04	0.04

L_4	Electricity	0.19	0.38	0.45
	Heat	0.05	0.04	0.03
L_5	Electricity	0.2	0.44	0.55
	Heat	0.06	0.04	0.04
L_6	Electricity	0.26	0.4	0.5
	Heat	0.07	0.04	0.04
L_7	Electricity	0.07	0.18	0.35
	Heat	0.02	0.02	0.02
PV_1	Electricity	0	0.02	0.1
PV_2	Electricity	0	0.01	0.1
PV_3	Electricity	0	0.02	0.1
WT_1	Electricity	0.28	0.25	0.2
WT_2	Electricity	0.36	0.32	0.3

Table 7

The main parameters of parallel learning.

Parameter	Range	Value
α	$0<\alpha<1$	0.9
γ	$0<\gamma<1$	0.1
ε	$0<\varepsilon<1$	0.9
μ	$\mu>0$	3
C_p	$0<C_p<1$	0.9
C_1	$C_1>0$	1
C_2	$C_2>0$	1000
k_{max}	$k_{max}>0$	250

Table 8

Comparative results of optimal solutions obtained by different algorithms under scenario 3.

Demanders or Suppliers	Energy type	Optimal energy generations and consumptions (MW)				
		GA	PSO	ABC	GSO	Parallel learning
L_1	Electricity	0.419	0.410	0.422	0.432	0.412
	Heat	0.030	0.030	0.030	0.030	0.030
L_2	Electricity	0.371	0.357	0.344	0.385	0.362
	Heat	0.030	0.030	0.030	0.030	0.030
L_3	Electricity	0.597	0.600	0.581	0.600	0.600
	Heat	0.040	0.040	0.040	0.040	0.040
L_4	Electricity	0.448	0.450	0.409	0.450	0.450
	Heat	0.030	0.030	0.030	0.030	0.030
L_5	Electricity	0.540	0.550	0.542	0.550	0.550
	Heat	0.040	0.040	0.040	0.040	0.040
L_6	Electricity	0.451	0.450	0.461	0.454	0.450
	Heat	0.040	0.040	0.040	0.040	0.040
L_7	Electricity	0.279	0.270	0.282	0.270	0.270
	Heat	0.020	0.020	0.020	0.020	0.020
PV_1	Electricity	0.100	0.100	0.100	0.100	0.100
PV_2	Electricity	0.100	0.100	0.100	0.100	0.100
PV_3	Electricity	0.100	0.100	0.100	0.100	0.100
WT_1	Electricity	0.200	0.200	0.200	0.200	0.200
WT_2	Electricity	0.300	0.300	0.300	0.300	0.300
Tie-line power	Electricity	0.355	0.399	0.387	0.399	0.399
DG_1	Electricity	0.325	0.248	0.261	0.308	0.255
DG_2	Electricity	0.040	0.040	0.096	0.041	0.040
CHP_1	Electricity	0.993	1.000	0.923	0.999	1.000
	Heat	0.009	0.000	0.047	0.001	0.000
CHP_2	Electricity	0.591	0.600	0.573	0.594	0.599
	Heat	0.026	0.000	0.152	0.015	0.000
$Heat_1$	Heat	0.195	0.230	0.031	0.215	0.226
Total operating cost f_{cost} ($/h)		1182.886	1176.073	1210.318	1178.385	**1176.021**

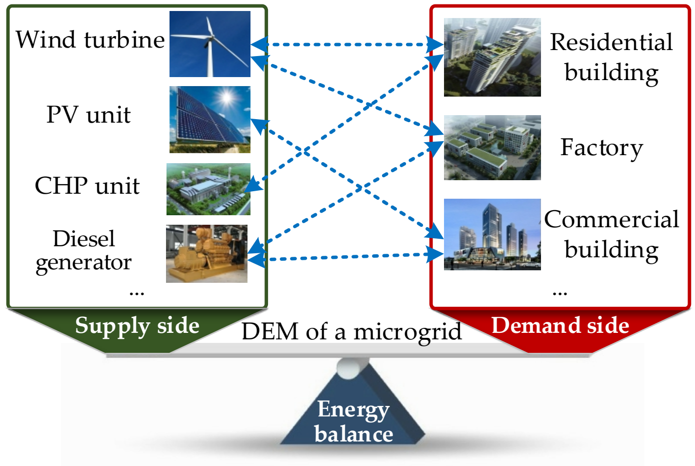


Fig. 1 DEM of a microgrid with different units.

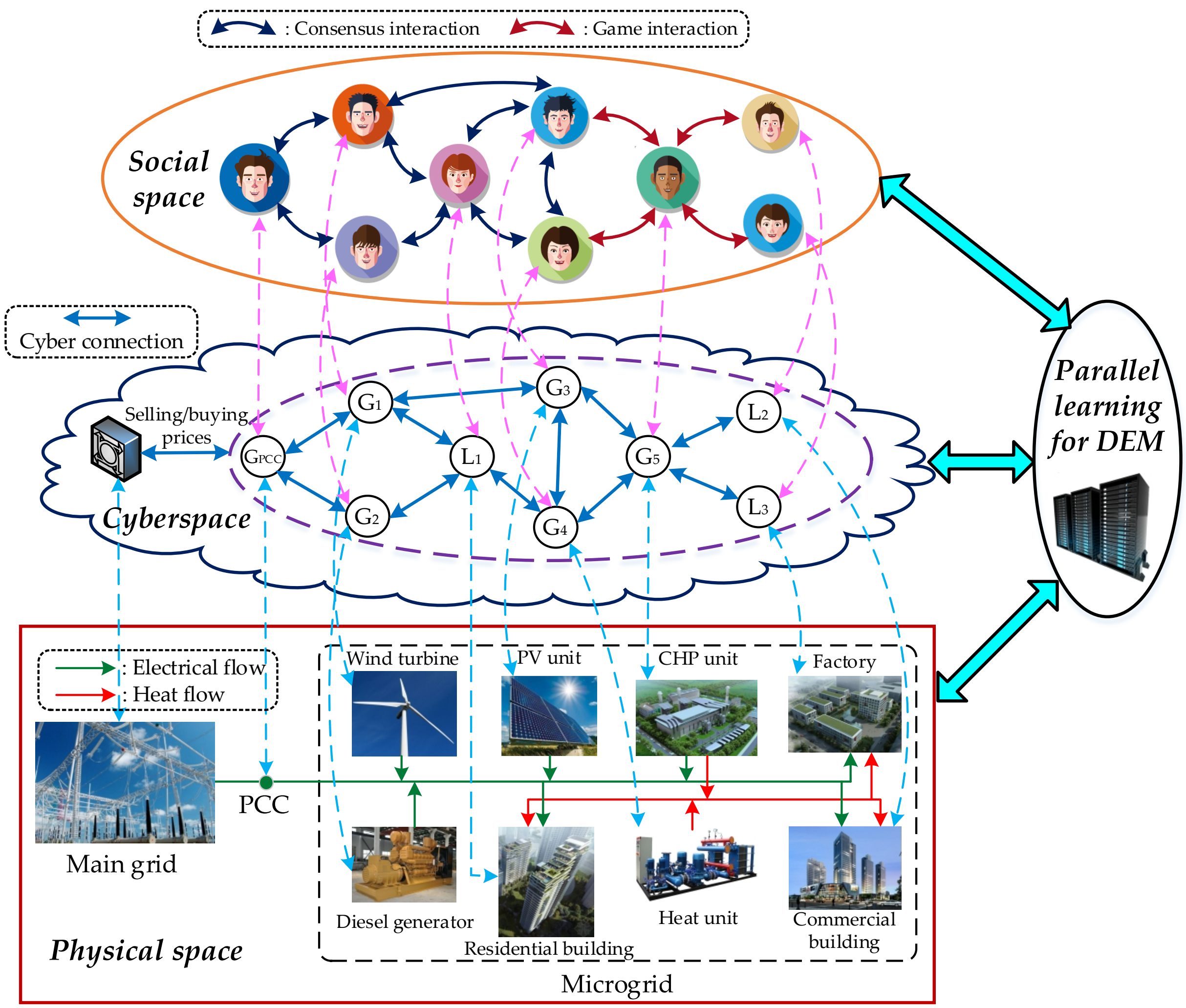
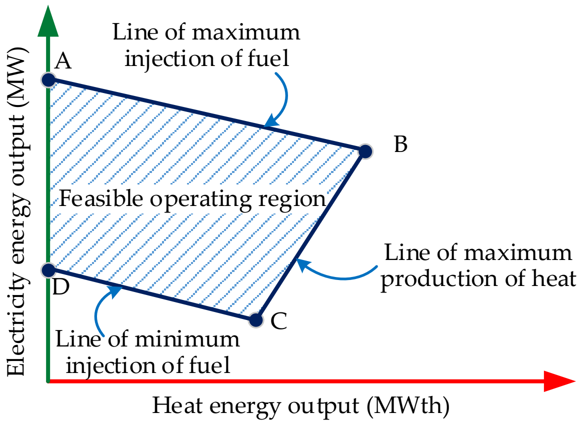
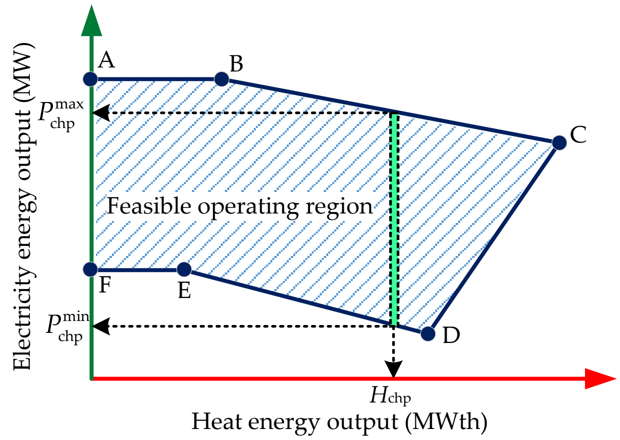


Fig. 3 CPSS framework for DEM of a microgrid.



(a) One segment shape



(b) Two segment shape

Fig. 2 Feasible operating region of CHP units.

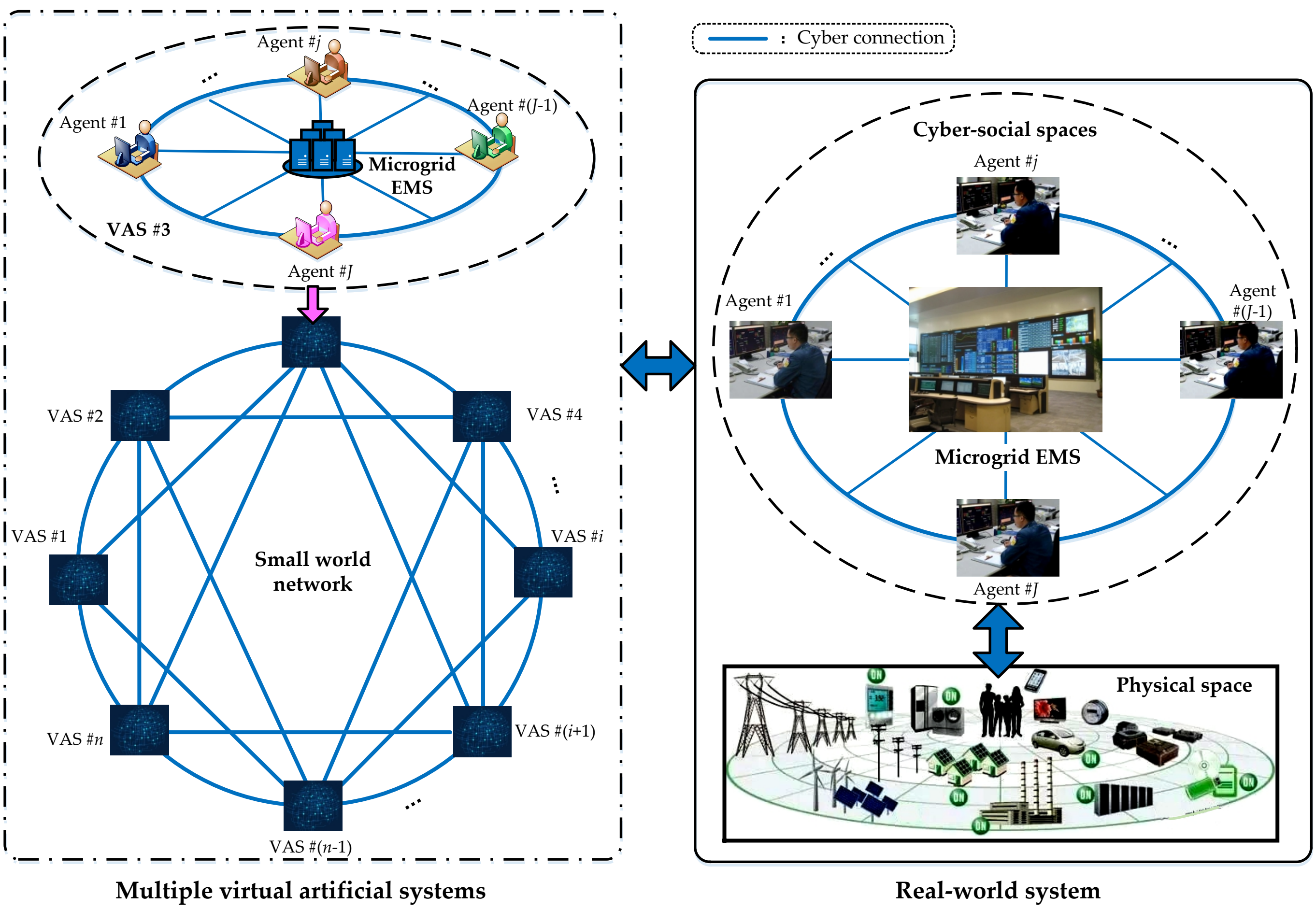


Fig. 4 Parallel learning with multiple virtual artificial systems and a real-world system.

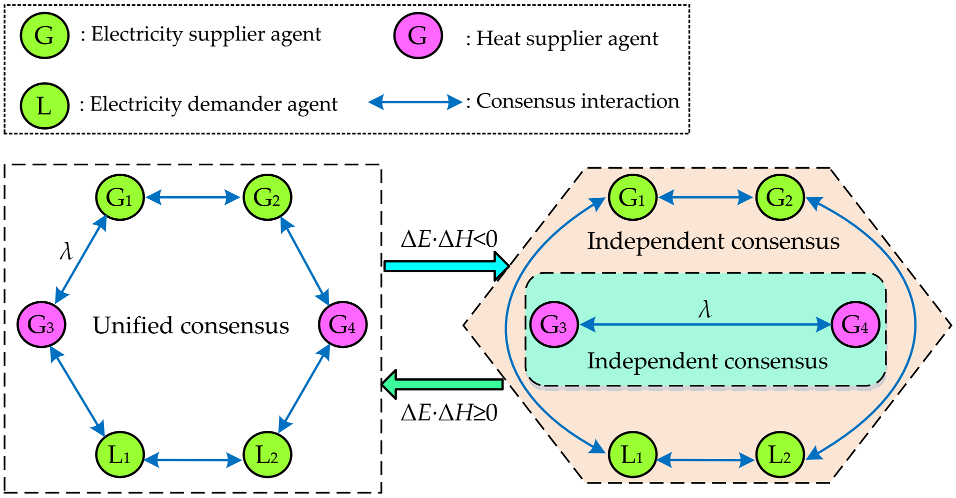


Fig. 5 Principle of human interaction by adaptive consensus algorithm.

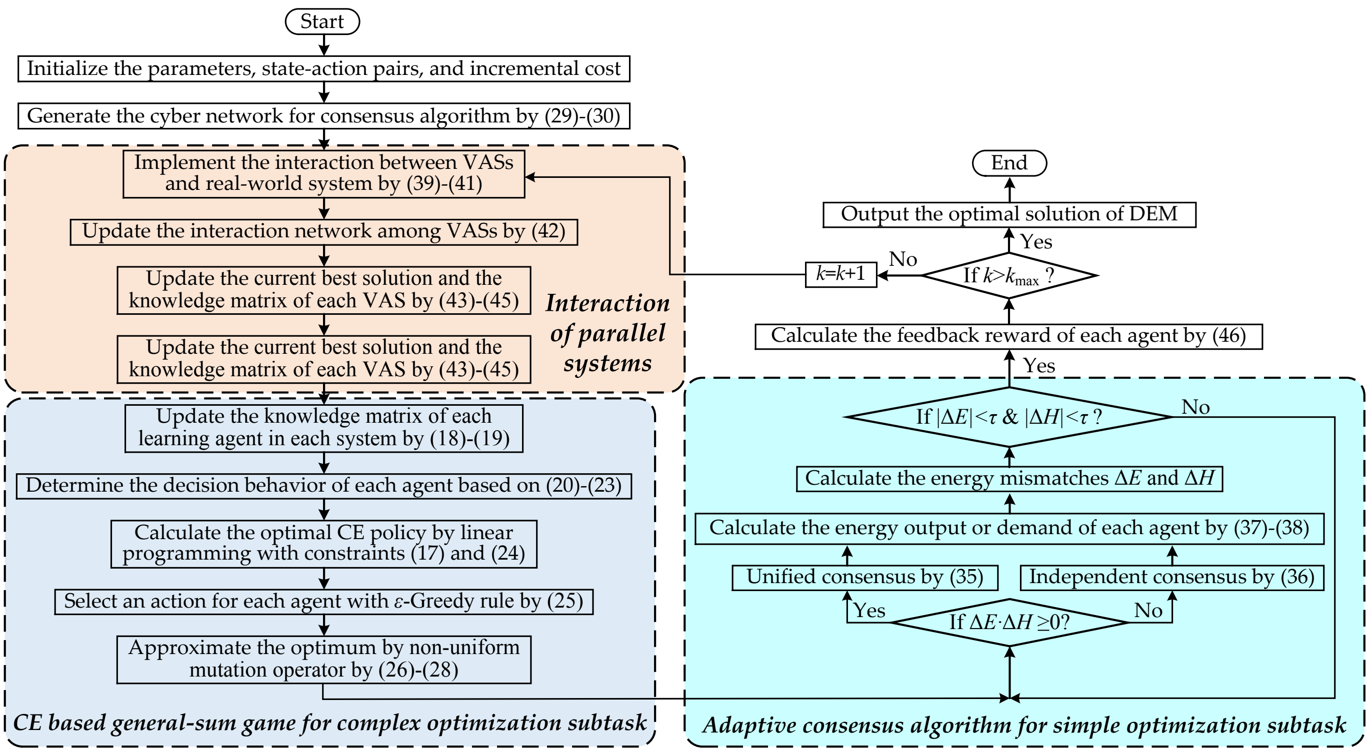


Fig. 6 Execution procedure of CPSS with parallel learning for DEM of a microgrid.

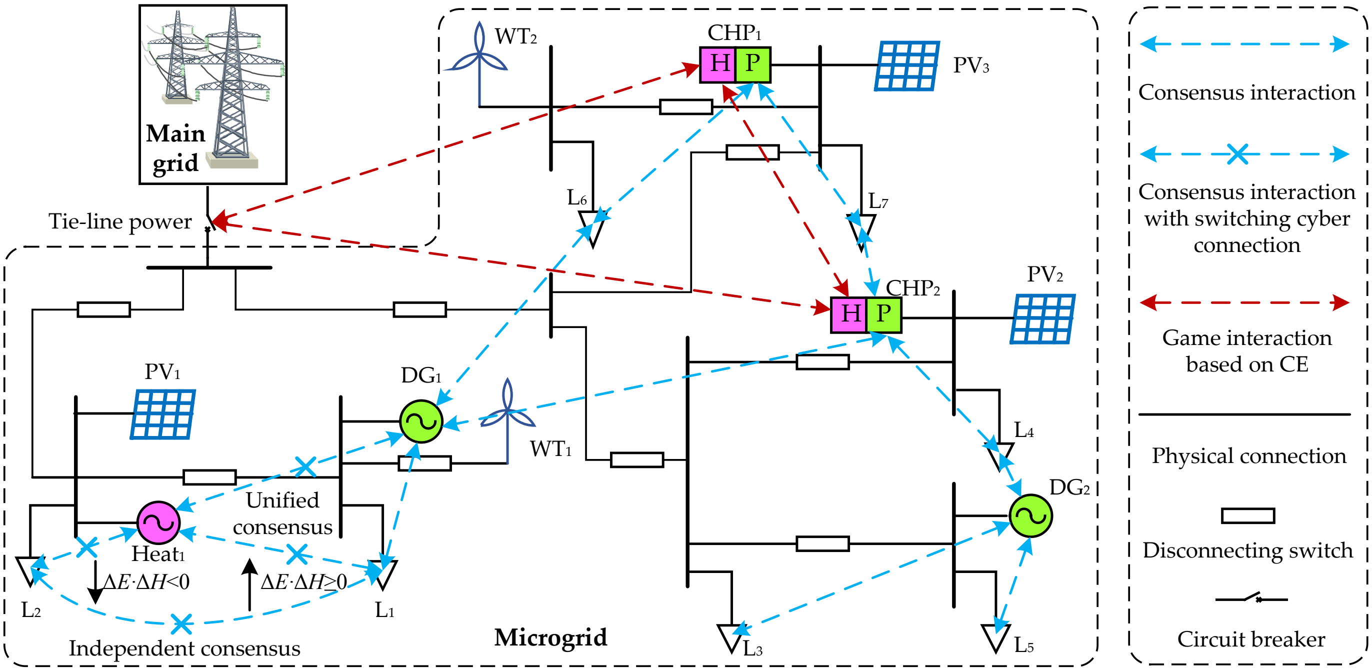
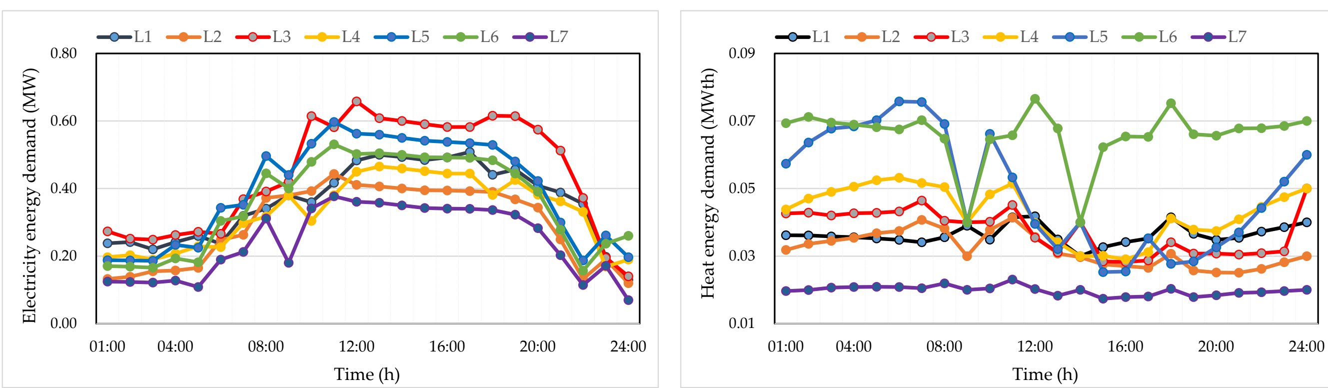


Fig. 7 CPSS of the testing microgrid.



(a) Electricity energy demand (b) Heat energy demand

Fig. 8 Energy demand profiles of energy demanders in a day.

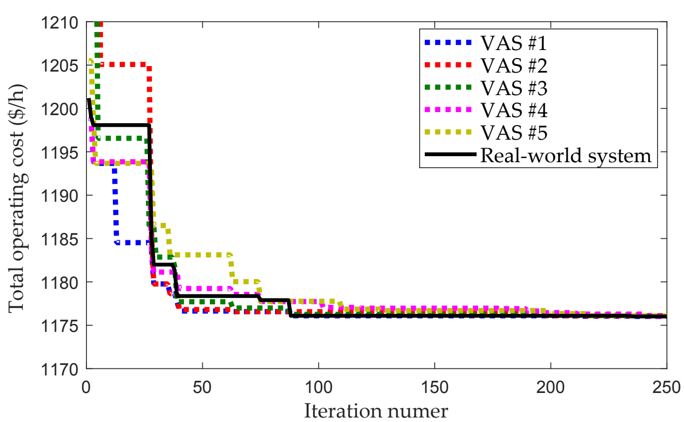
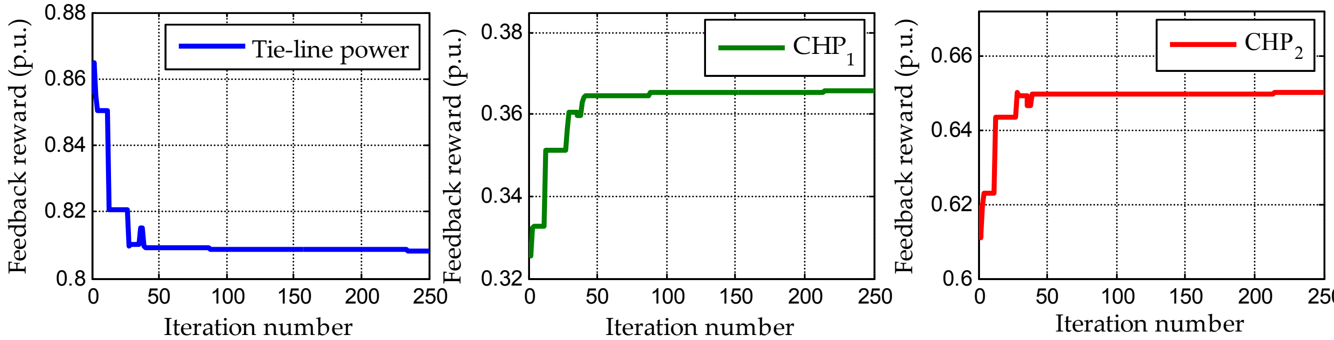
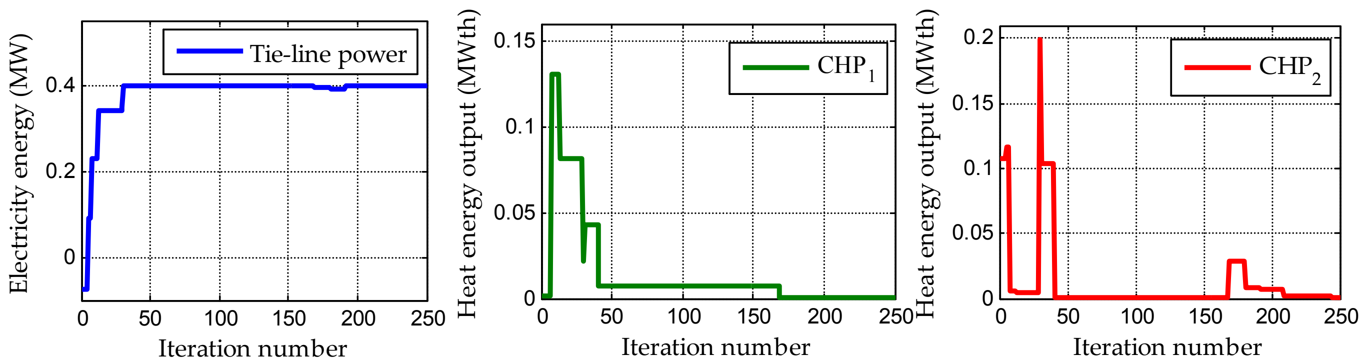


Fig. 9 Convergence of parallel learning in different systems.

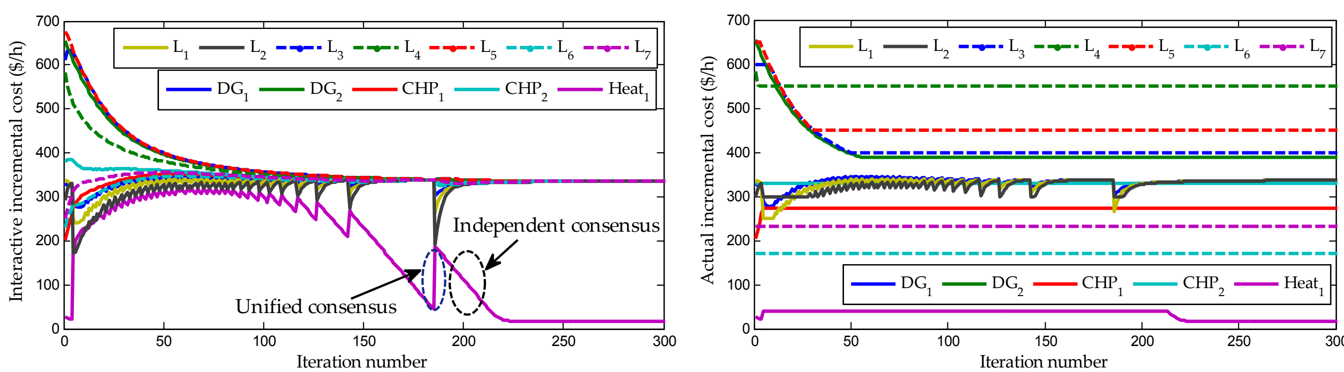


(a) Convergence of feedback reward



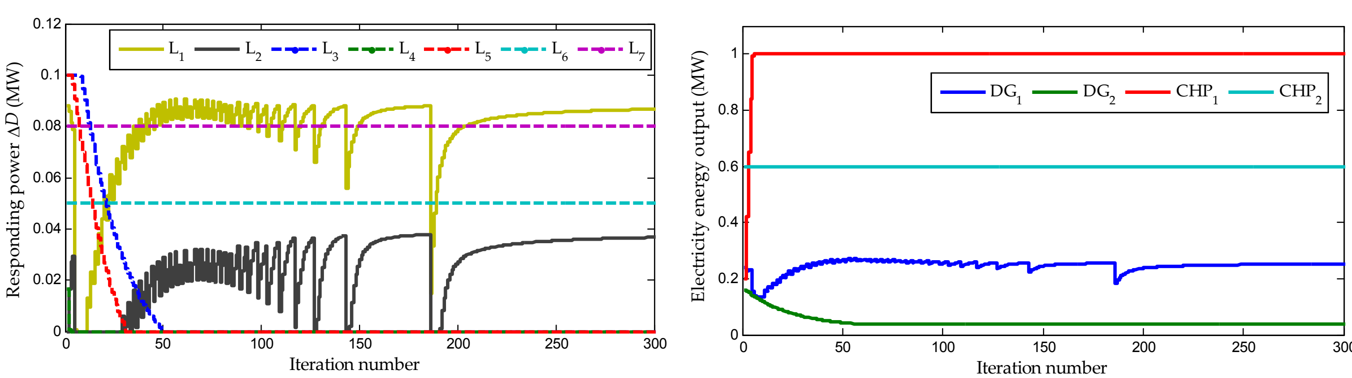
(b) Convergence of energy output

Fig. 10 Convergence of CE based human interaction in the real-world system.



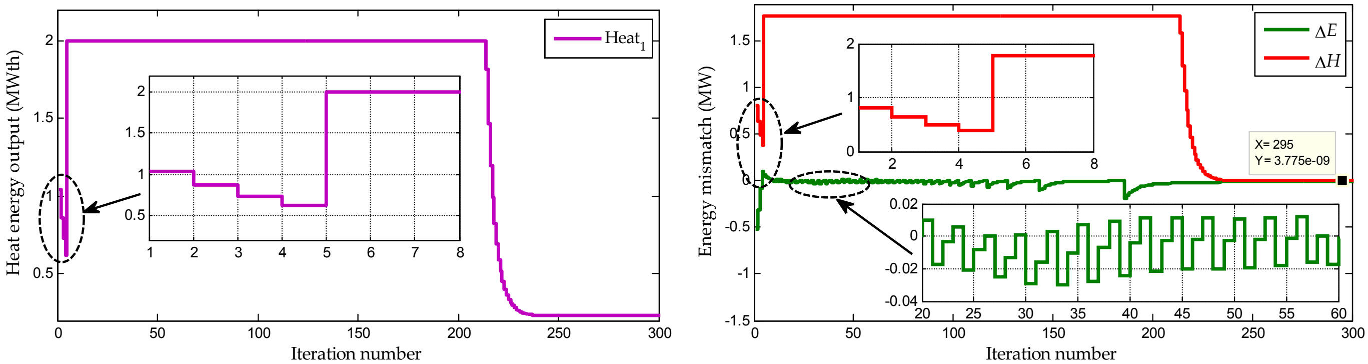
(a) Convergence of interactive incremental cost

(b) Convergence of actual incremental cost



(c) Convergence of responding power of different energy demanders

(d) Convergence of electricity energy output



(e) Convergence of heat energy output

(f) Convergence of energy mismatches

Fig. 11 Convergence of adaptive consensus algorithm based human interaction in the real-world system at the final game iteration.

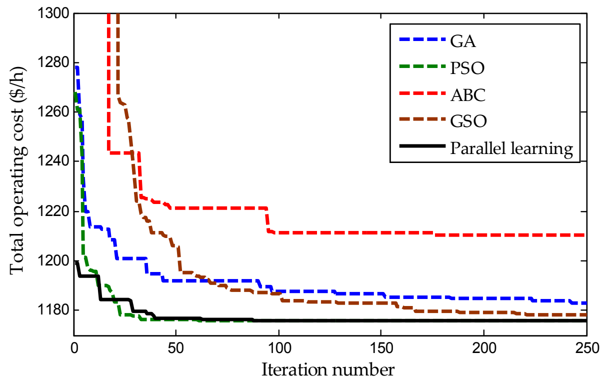
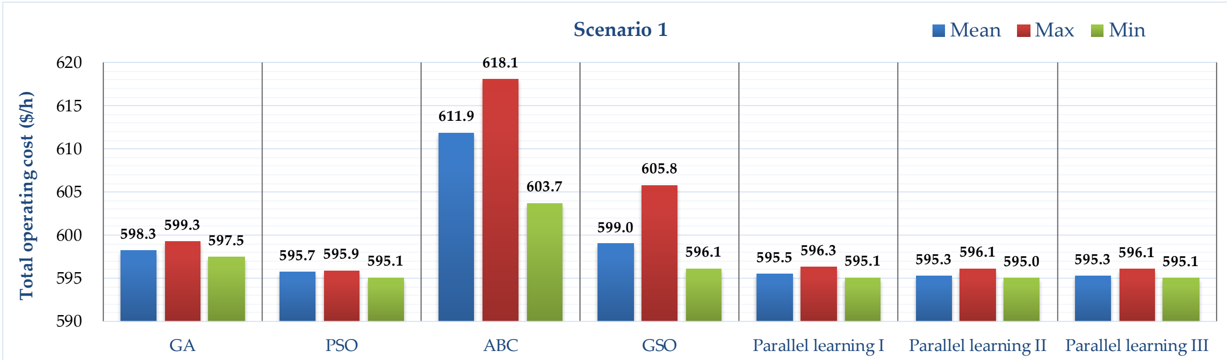
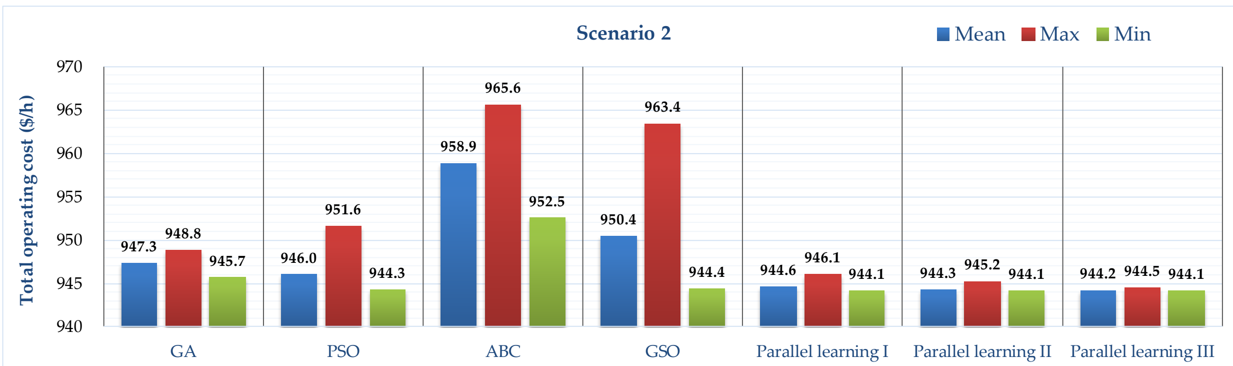


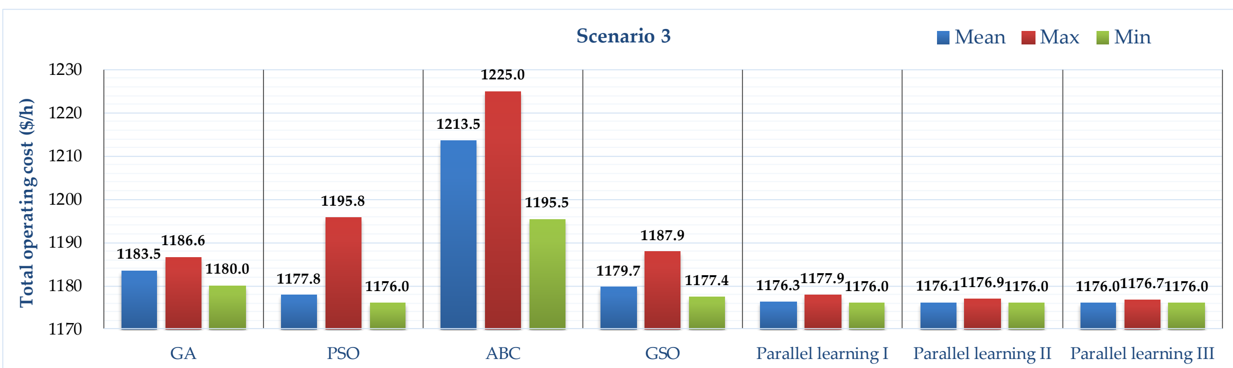
Fig. 12 Convergence of total operating costs obtained by different algorithms under scenario 3.



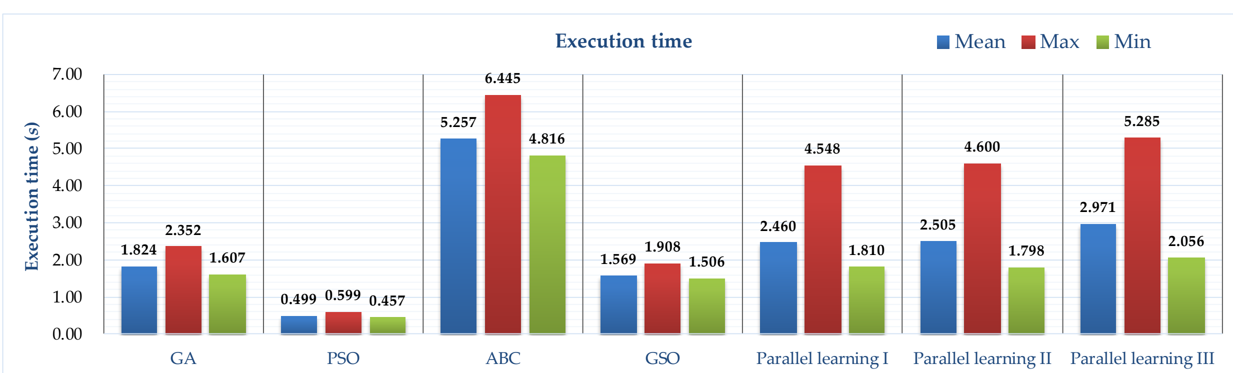
(a) Obtained total operating costs under scenario 1



(b) Obtained total operating costs under scenario 2

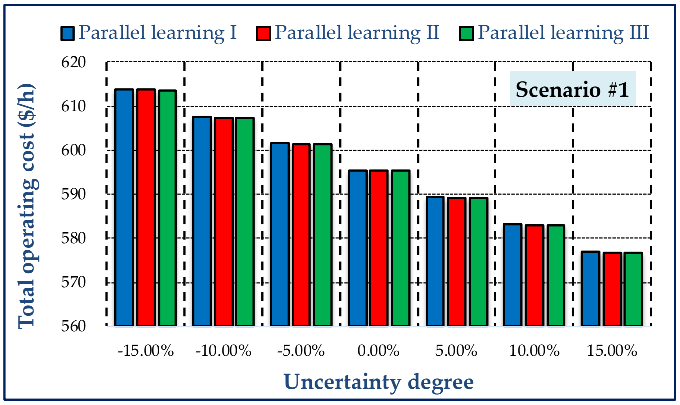


(c) Obtained total operating costs under scenario 3

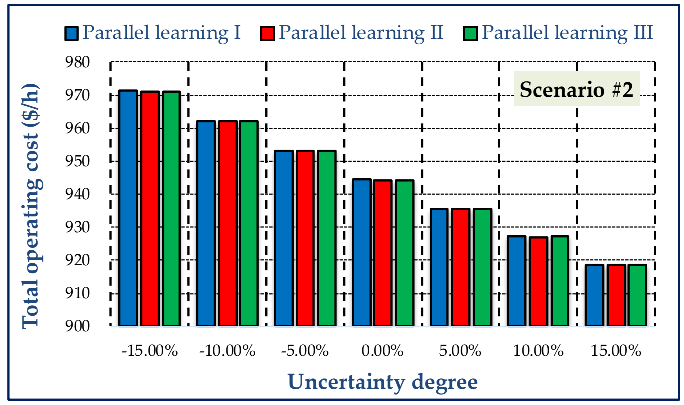


(d) Execution time of each scenario

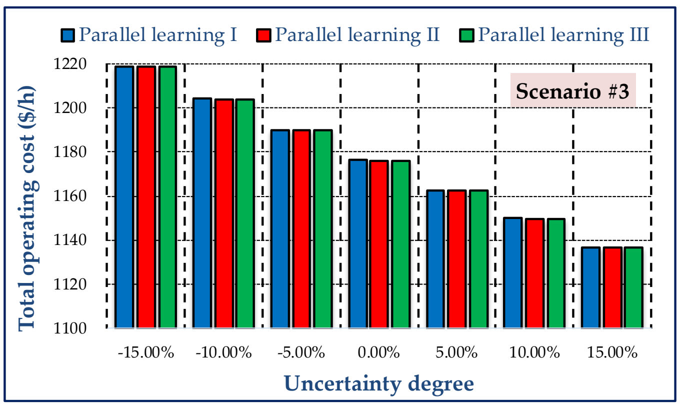
Fig. 13 Statistical results obtained by different algorithms under three scenarios in 50 runs.



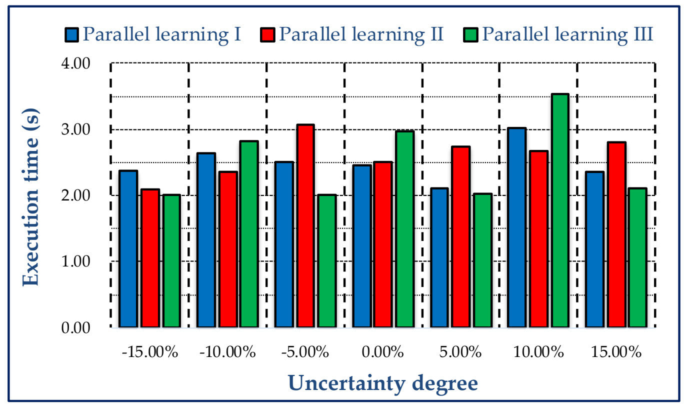
(a) Obtained total operating costs under scenario 1



(b) Obtained total operating costs under scenario 2



(c) Obtained total operating costs under scenario 3



(d) Execution time of each scenario

Fig. 14 Statistical results obtained by CPSS with parallel learning under different uncertainty degrees in 50 runs.

- *A cyber-physical-social system is constructed for distributed energy management.*
- *A game theory with various decision behaviors is proposed for human interaction.*
- *Energy suppliers or demanders can reach a consensus on the incremental cost.*
- *The parallel interactive systems can lead to a lower total operating cost.*
- *The proposed method outperforms other centralized heuristic algorithms for DEM.*