IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 25, 2018, accepted June 27, 2018, date of publication July 16, 2018, date of current version August 7, 2018.

*Digital Object Identifier 10.1109/ACCESS.2018.2853263*

# Multi-Agent Bargaining Learning for Distributed Energy Hub Economic Dispatch

**XIAOSHUN ZHANG[1,2], (Member, IEEE), TAO YU[1], (Member, IEEE), ZHIYI ZHANG[1], AND JIANLIN TANG[1]**

[1]College of Electric Power, South China University of Technology, Guangzhou 510640, China
[2]Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong

Corresponding author: Tao Yu (taoyu1@scut.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grants 51477055 and 51777078.

**ABSTRACT** This paper proposes a novel multi-agent bargaining learning (MABL) for the distributed energy hub economic dispatch (EHED) of multiple energy carrier systems (MECS). Distributed EHED is developed by extending the conventional economic dispatch (ED) into MECS in a distributed manner, in which each energy hub is regarded as a learning agent for self-scheduling. The classical Q-learning with associative memory is employed for knowledge learning of each agent, while the non-uniform mutation operator is adopted for handling the continuous control variables. To maximize the total payoff of all the energy hubs, the bargaining game is presented for achieving an effective coordination between the buyer agents and a seller agent, where the slack energy hub is designed as the seller agent and the others are the buyer agents. MABL has been thoroughly evaluated for the distributed EHED on a high-complex 39-hub MECS with 29 energy hub structures and 76 energy production units. Case studies verify the superior performance of MABL for the distributed EHED compared with six centralized heuristic optimization algorithms.

**INDEX TERMS** Multi-agent bargaining learning, distributed energy hub economic dispatch, multiple energy carrier systems, knowledge learning, bargaining game.

## I. INTRODUCTION

In recent years, multiple energy carrier systems (MECS) has intrigued many scholars as it is an inventible solution for the future energy network [1], which will lead to a high energy utilizing efficiency, a low $CO_2$ emission, a high operation economy and flexibility [2]. Accordingly, an attendant energy hub, i.e., a functional unit where multiple energy carriers [3] are converted, stored, and dissipated, is employed for achieving the interaction between different energy carriers.

Based on this intuition, a novel energy hub economic dispatch (EHED) was developed in [4], which aims to minimize the operation cost by optimizing the energy inputs of all the energy hubs to balance different types of energy demands (i.e., electricity, heat, cool, and compressed air) while satisfying various operation constraints. Consequently, EHED is essentially an extension version of the conventional economic dispatch (ED), however, it is more difficult to handle the highly nonlinear, non-convex, non-smooth, non-differential, and high-dimensional EHED with more equality and inequality constraints. To tackle this issue, the heuristic optimization algorithms are commonly considered as priority techniques since they are highly independent on the mathematical model, compared with the conventional gradient-based algorithms (e.g., interior-point method [5]), which usually obtain a local optimum for the highly nonlinear and non-convex optimization.

To obtain a high-quality optimal solution with a lower objective function for EHED, a self-adoptive learning with time varying acceleration coefficient-gravitational search algorithm (SAL-TVAC-GSA) was proposed in [4] by designing three fundamental modifications based on the original GSA, which has been proven excellent performance on solution quality and computation efficiency compared that of GSA [6], enhanced GSA (EGSA) [6], particle swarm optimization (PSO) [7], and genetic algorithm (GA) [8]. No doubt it can be effectively addressed by other heuristic optimization algorithms, e.g., artificial bee colony (ABC) [9] and differential evolution (DE) [10]. Nevertheless, most of these algorithms essentially belong to the centralized optimization algorithms, which easily leads to a high computation burden and poor performance with a low-quality optimal solution as the scale and complexity of MECS increases [11], while

it is also difficult to satisfy the demand of high privacy and security [12].

Generally speaking, the distributed optimization algorithms can effectively handle these problems. In [13], an event-triggered communication-based method was presented for distributed ED. Based on the consensus theory, a distributed primal-dual dynamic multiagent system [14] was proposed for dynamical ED. Moreover, a bargaining-based alternating direction method of multipliers (ADMM) [15] was designed for distributed EHED. However, all of them are only available for convex and smooth ED, which cannot address the non-convex, non-smooth ED. On the other hand, the distributed heuristic optimization algorithms, including cooperative co-evolutionary differential evolution [16], distributed PSO [17], distributed multi-step Q(λ) learning [18], and so on, are more suitable for the distributed EHED, but all of them can only search a local optimal solution in most cases.

Over the years, multi-agent reinforcement learning (MARL) has attracted extensive investigations and real-world applications due to its distributed nature of the multi-agent solution [19], in which each agent can maximize its payoff by competing or cooperating with other agents, e.g., a deep communication for obtaining a high-quality optimal solution of the multi-agent system (MAS). Motivated from this benefit, a novel MARL was developed for an optimal reactive power dispatch by combining the consensus theory with the distributed Q-learning [20]. Besides, a game theory based hierarchical correlated Q-learning was designed for multi-layer optimal generation command dispatch [21] and reactive power optimization [22], respectively. All of these algorithms can achieve a satisfactory optimal solution, but ineluctably suffer from three flaws, as follows:

- *Slow convergence rate*: each agent will consume a large amount of computation time to acquire the optimal Q-value matrix, especially for the equilibrium computation [23], e.g., the Nash equilibria computation is polynomial parity arguments on directed graphs (PPAD)-hard [24].
- *Curse of dimension*: the joint action space will increase exponentially with the growing controllable variables [25], thus they are incapable of solving EHED in a large-scale MECS.
- *Weak disposal ability for the continuous controllable variables*: action space needs to be discretized into a finite number of cells for each controllable variable, thus the control accuracy will be contradictory with the computation time [26], i.e., the higher the control accuracy, the longer computation time and vice versa.

In order to resolve these problems, this paper proposes a novel MARL called multi-agent bargaining learning (MABL) for rapidly searching a high-quality optimal solution of distributed EHED.

The remaining of this paper is organized as follows. The mathematical model of distributed EHED is presented in Section II. Section III provides the basic principle of MABL. The detailed design of MABL for distributed EHED is given in Section IV. Simulation results and conclusion are given in Section V and Section VI, respectively.

## II. MATHEMATICAL MODEL OF DISTRIBUTED EHED

### A. GENERAL MODEL OF ENERGY HUB

As illustrated in Fig. 1, energy hub can be treated as an interface between energy infrastructures and loads [27], which usually consists of connection devices (e.g., transformer), converters (e.g., combined heat and power (CHP) units), and storage devices. In practice, CHP units, combined heat, cool, and power (CHCP) units, big building complexes, small isolated systems, and so on, can be regarded as different energy hubs [28]. For a general energy hub, the relation model between the output energy carriers and the input energy carriers can be written as follows [29]:

$$\underbrace{\begin{bmatrix} E_\alpha^{\text{out}} \\ E_\beta^{\text{out}} \\ \vdots \\ E_\zeta^{\text{out}} \end{bmatrix}}_{\boldsymbol{E}^{\text{out}}} = \underbrace{\begin{bmatrix} C_{\alpha\alpha} & C_{\beta\alpha} & \cdots & C_{\zeta\alpha} \\ C_{\alpha\beta} & C_{\beta\beta} & \cdots & C_{\zeta\beta} \\ \vdots & \vdots & \ddots & \vdots \\ C_{\alpha\zeta} & C_{\beta\zeta} & \cdots & C_{\zeta\zeta} \end{bmatrix}}_{\boldsymbol{C}} \underbrace{\begin{bmatrix} E_\alpha^{\text{in}} \\ E_\beta^{\text{in}} \\ \vdots \\ E_\zeta^{\text{in}} \end{bmatrix}}_{\boldsymbol{E}^{\text{in}}} \tag{1}$$

where the subscripts $\{\alpha, \beta, \ldots, \zeta\}$ represents various energy carriers, including electricity, gas, heat, and so on; $\boldsymbol{E}^{\text{out}}$, $\boldsymbol{E}^{\text{in}}$ denote the vectors of energy outputs and energy inputs, respectively; and $\boldsymbol{C}$ is the coupling matrix, in which each element (e.g., $C_{\alpha\alpha}$) denotes the coupling factor.

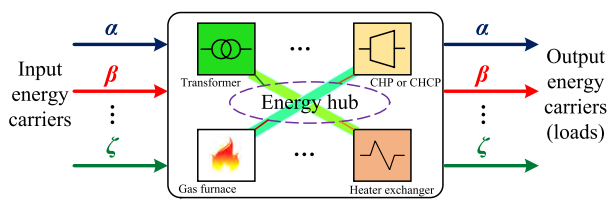


**FIGURE 1.** A general model of energy hub.

Note that the coupling matrix $\boldsymbol{C}$ is completely determined by the specific configuration of the corresponding energy hub, while each coupling factor $C$ is equal to the product of the dispatch factor $v$ and the converter efficiency $\eta$, i.e., $C = v\eta$ [30]. For an energy hub with multiple devices, the dispatch factors represent the distribution ratios of an input energy carrier to these devices, which can be considered as the controllable variables, while the converter efficiency is usually perceived as a constant.

### B. DISTRIBUTED EHED

#### 1) OBJECTIVE FUNCTION

To maximize the economic benefits of MECS, both the total energy costs $f_{\text{C}}$ and the total energy losses $f_{\text{L}}$ are taken as two basic objective functions for distributed EHED, respectively. In order to minimize the total energy costs and the total energy losses simultaneously, they can be transformed into a single

objective function $f_{\mathrm{I}}$, as follows [4]:

$$\min f_{\mathrm{C}}(\boldsymbol{x}) = \sum_{m=1}^{M} f_{\mathrm{C}}^{m}(\boldsymbol{x}_m), f_{\mathrm{L}}(\boldsymbol{x}) = \sum_{m=1}^{M} f_{\mathrm{L}}^{m}(\boldsymbol{x}_m),$$
$$f_{\mathrm{I}}(\boldsymbol{x}) = f_{\mathrm{C}}(\boldsymbol{x})\left(1 + \frac{f_{\mathrm{L}}(\boldsymbol{x})}{\sum_{p\in\boldsymbol{P}} E_{\mathrm{demand}}^{p}}\right) \tag{2}$$

where $\boldsymbol{x}$ is the vector of the controllable variables of an entire system, including the energy production of each input energy carrier and each dispatch factor; $\boldsymbol{x}_m$ is the vector of the controllable variables for the $m$th energy hub; the subscript $m$ and the superscript $p$ represent the $m$th energy hub and the $p$th energy carrier, respectively; $M$ is the number of energy hubs; $\boldsymbol{P}$ denotes the set of energy carriers, i.e., $\boldsymbol{P}=\{\alpha, \beta, \ldots, \zeta\}$; $E_{\mathrm{demand}}^{p}$ denotes the energy demand of the $p$th output energy carrier of MECS; $f_{\mathrm{C}}^{m}$ and $f_{\mathrm{L}}^{m}$ represent the energy cost and the energy loss of the $m$th energy hub, respectively, which can be calculated as [4], [31]

$$f_{\mathrm{C}}^{m}(\boldsymbol{x}_m) = \sum_{p\in\boldsymbol{P}}\sum_{j=1}^{n_m^p}\left(a_{mj}^{p}\left(E_{mj}^{p,\mathrm{in}}\right)^2 + b_{mj}^{p}E_{mj}^{p,\mathrm{in}} + c_{mj}^{p}\right)$$
$$+ \sum_{j=1}^{n_m^{\mathrm{e}}}\left|d_{mj}^{\mathrm{e}}\sin\left(e_{mj}^{\mathrm{e}}\left(E_{mj,\min}^{\mathrm{e,in}} - E_{mj}^{\mathrm{e,in}}\right)\right)\right| \tag{3}$$
$$f_{\mathrm{L}}^{m}(\boldsymbol{x}_m) = \sum_{p\in\boldsymbol{P}}\left(E_{m}^{p,\mathrm{in}} - E_{m}^{p,\mathrm{out}}\right) \tag{4}$$

where $n_m^p$ is the number of energy sources associated with the $p$th input energy carrier for the $m$th energy hub; $n_m^{\mathrm{e}}$ is the number of electrical generators with the valve-point effects of the $m$th energy hub; $a_{mj}^{p}$, $b_{mj}^{p}$, and $c_{mj}^{p}$ are the cost coefficients of the $j$th energy source; $d_{mj}^{\mathrm{e}}$ and $e_{mj}^{\mathrm{e}}$ are the cost coefficients of the additional rectified sinusoidal component by considering the value-point effects of the electrical generators; $E_{mj}^{p,\mathrm{in}}$ is the input energy production of the $j$th energy source; $E_{mj,\min}^{\mathrm{e,in}}$ is the minimum energy production of the $j$th electrical generator; $E_{m}^{p,\mathrm{in}}$ and $E_{m}^{p,\mathrm{out}}$ are the energy input and energy output of the $p$th input energy carrier for the $m$th energy hub, respectively.

#### 2) CONSTRAINTS

In general, the renewable energy resources, e.g., solar or wind energy, are usually operated at the maximum power points (MPPs) at different weather conditions in a thermal power dominated system. Hence, no renewable energy resource is not considered for set points optimizations in this paper.

In the pursuit of different economic objective functions in (9), each energy hub needs to satisfy various constraints for an optimal operation, including energy balance constraints, capacity limits of all energy sources, prohibited operation zones (POZ) constraints of conventional thermal generators for faults prevention [32], and dispatch factor limits, as follows:

$$E_{\mathrm{demand}}^{p} = \sum_{m=1}^{M} E_{m}^{p,\mathrm{out}}, p \in \boldsymbol{P} \tag{5}$$
$$E_{mj,\min}^{p,\mathrm{in}} \le E_{mj}^{p,\mathrm{in}} \le E_{mj,\max}^{p,\mathrm{in}}, p \in \boldsymbol{P};$$
$$m = 1, 2, \ldots, M; j = 1, 2, \ldots, n_m^p \tag{6}$$
$$\begin{cases} E_{mj,\min}^{\mathrm{e,in}} \le E_{mj}^{\mathrm{e,in}} \le E_{mj,\min\_1}^{\mathrm{e,in}} \\ \text{or } E_{mj,\max\_(z-1)}^{\mathrm{e,in}} \le E_{mj}^{\mathrm{e,in}} \le E_{mj,\min\_z}^{\mathrm{e,in}}, \\ \qquad z = 2, \ldots, Z_{mj}^{\mathrm{e}} \\ \text{or } E_{mj,\max\_z}^{\mathrm{e,in}} \le E_{mj}^{\mathrm{e,in}} \le E_{mj,\max}^{\mathrm{e,in}} \\ m = 1, 2, \ldots, M; j = 1, 2, \ldots, n_m^{\mathrm{e}} \end{cases} \tag{7}$$

where $E_{mj,\min\_z}^{\mathrm{e,in}}$ and $E_{mj,\max\_z}^{\mathrm{e,in}}$ are the low bound and upper bound of the $z$th POZ of $j$th electrical generator for the $m$th energy hub, respectively; and $Z_{mj}^{\mathrm{e}}$ is the POZ number of $j$th electrical generator for the $m$th energy hub.

## III. MULTI-AGENT BARGAINING LEARNING

### A. BARGAINING GAME

In a basic two-player bargaining game, the seller agent will firstly make an offer to the buyer agent, if the offer is accepted by the buyer agent, then an bargaining equilibrium (i.e., the strategy of the offer) can be determined, otherwise, the bargaining role will be shifted to be on the buyer agent in the next period until they reach an agreement on the offer [33]. Enlighted by this game, a novel cooperative one-seller and $n$-buyer bargaining game is proposed for achieving an efficient coordination between different players.

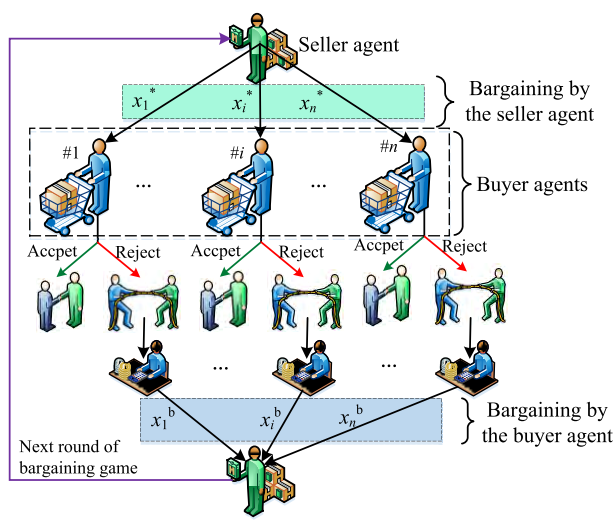


**FIGURE 2.** Principle of one-seller and $n$-buyer bargaining game.

As shown in Fig. 2, the one-seller and $n$-buyer bargaining game contains two bargaining process at each round, as

1) The seller agent firstly issues an current optimal joint action strategy $\boldsymbol{x}^* = (x_1^*, \ldots, \boldsymbol{x}_i^*, \ldots, \boldsymbol{x}_n^*)$ of the offers to all the buyer agents, which can be determined by comparing different bargaining strategies from the buyer agents,

as follows:

$$\boldsymbol{x}_k^* = \arg\max[U_1(\boldsymbol{x}_{1,k-1}^{\mathrm{b}}, \boldsymbol{x}_{1,k-1}^{\mathrm{other}}), \ldots, U_i(\boldsymbol{x}_{i,k-1}^{\mathrm{b}}, \boldsymbol{x}_{i,k-1}^{\mathrm{other}}), \ldots, U_n(\boldsymbol{x}_{n,k-1}^{\mathrm{b}}, \boldsymbol{x}_{n,k-1}^{\mathrm{other}}), U_{\mathrm{s}}(\boldsymbol{x}_{k-1}^{b}), U_{\mathrm{s}}(\boldsymbol{x}_{k-1}^{*})] \quad (8)$$

where the subscript $k$ denotes the $k$th iteration; $\boldsymbol{x}_k^*$ is the optimal joint action strategy at the $k$th iteration; $\boldsymbol{x}_{i,k-1}^{\mathrm{b}}$ is the bargaining action strategy of the $i$th buyer agent; $\boldsymbol{x}_{i,k-1}^{\mathrm{other}}$ is the the optimal joint action strategy of all the buyer agents except the $i$th buyer agent at the $(k-1)$ iteration; $\boldsymbol{x}_{k-1}^{b}$ is the joint bargaining action strategy; $U_i$ is the utility function of the $i$th buyer agent, $i = 1, 2, \ldots, n$; $n$ is the number of buyer agents; and $U_{\mathrm{s}}$ is the utility function of the seller agent.

2) If all the buyer agents accept that, the bargaining game is over, otherwise, the buyer agent will search a more optimal action strategy $x_i^{\mathrm{b}}$ to the seller agent if the following condition can be satisfied, as

$$\begin{cases} U_i(\boldsymbol{x}_{i,k}^{\mathrm{b}}, \boldsymbol{x}_{i,k}^{\mathrm{other}}) \geq U_{\mathrm{s}}(\boldsymbol{x}_k^*), & i = 1, 2, \ldots, n \\ \text{or } U_{\mathrm{s}}(\boldsymbol{x}_k^{\mathrm{b}}) \geq U_{\mathrm{s}}(\boldsymbol{x}_k^*) \end{cases} \quad (9)$$

Note that all the utility functions in (16) and (9) represent the total payoff of all the agents, in which each buyer agent can only optimize its own action strategy to increase the total payoff. After a series of bargaining game between the seller agent and the buyer agents according to (16) and (9), a high-quality bargaining equilibrium with a high total payoff of all the agents can be acquired. Hence, the key of bargaining game is to search a more optimal action strategy $x_i^{\mathrm{b}}$ for each buyer agent, which will be handled by the proposed learning method.

### B. KNOWLEDGE LEARNING

As one of the most commonly used reinforcement learning, Q-learning is a model-free learning through continuous interactions with the environment [34], which can develop the knowledge for each action strategy at each state via a reward mechanism. Hence, it is adopted for knowledge learning of each buyer agent. To effectively avoid the curse of dimension and accelerate the learning rate of conventional learning, the associative memory is used for knowledge storage [35], while a cooperative swarm with multiple intelligent individuals is employed for implementing the exploration and exploitation in the environment, as illustrated in Fig. 3. Therefore, the knowledge matrices of each buyer agent can be updated as [35]

$$\begin{cases} \boldsymbol{Q}_{ih}^{k+1}\left(s_{ih}^{kj}, a_{ih}^{kj}\right) = \boldsymbol{Q}_{ih}^{k}\left(s_{ih}^{kj}, a_{ih}^{kj}\right) + \alpha \Delta \boldsymbol{Q}_{ih}^{k} \\ \Delta \boldsymbol{Q}_{ih}^{k} = R_{ih}^{j}\left(s_{ih}^{kj}, s_{ih}^{k+1,j}, a_{ih}^{kj}\right) \\ \qquad + \gamma \max\limits_{a_{ih} \in \boldsymbol{A}_{ih}} \boldsymbol{Q}_{ih}^{k}\left(s_{ih}^{k+1,j}, a_{ih}\right) - \boldsymbol{Q}_{ih}^{k}\left(s_{ih}^{kj}, a_{ih}^{kj}\right) \\ h = 1, 2, \ldots, n_i; \quad j = 1, 2, \ldots, J \end{cases} \quad (10)$$

where $\boldsymbol{Q}_{ih}$ is the knowledge matrix of the $h$th controllable variable for the $i$th buyer agent; $\Delta \boldsymbol{Q}$ is the knowledge increment; $\alpha$ is the knowledge learning factor; $\gamma$ is the discount factor; $(s_{ih}^{kj}, a_{ih}^{kj})$ represents the state-action pair executed by the $j$th individual for the controllable variable $x_{ih}$; $R(s^k, s^{k+1}, a^k)$ is the reward function of a transition from state $s^k$ to $s^{k+1}$ used under a selected action $a^k$; $a_{ih}$ represents any alternative action strategies; $\boldsymbol{A}_{ih}$ is the action space of $x_{ih}$; $n_i$ is the number of controllable variables for the $i$th buyer agent; and $J$ is the population size of the cooperative swarm.

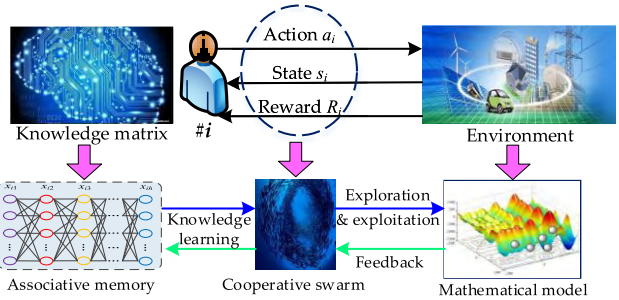


**FIGURE 3.** Principle of knowledge learning with a cooperative swarm.

### C. EXPLORATION AND EXPLOITATION

To achieve a high optimization accuracy, the exploration and exploitation of each individual is divided into two processes. For a controllable variable, each individual firstly choose an action strategy (i.e., interval of optimization) based on the corresponding knowledge matrix, then an accurate solution can be calculated by the non-uniform mutation operator [36] based on the local optimum of the corresponding interval. Aiming at a proper trade-off between exploration and exploitation, the $\varepsilon$-Greedy rule [37] is used for interval selection, as

$$a_{ih}^{kj} = \begin{cases} \arg\max\limits_{a_{ih} \in \boldsymbol{A}_{ih}} \boldsymbol{Q}_{ih}^{k}(s_{ih}^{kj}, a_{ih}), & \text{if } q_0 \leq \varepsilon \\ a_{\mathrm{rand}}, & \text{otherwise} \end{cases} \quad (11)$$

$$x_{ih}^{kj} = \begin{cases} x_{ih}^{d,\mathrm{best}} + \Delta(k, x_{ih}^{d,\mathrm{ub}} - x_{ih}^{d,\mathrm{best}}), & \text{if rand}(0, 1) < 0.5 \\ x_{ih}^{d,\mathrm{best}} - \Delta(k, x_{ih}^{d,\mathrm{best}} - x_{ih}^{d,\mathrm{lb}}), & \text{otherwise} \end{cases} \quad (12)$$

$$\begin{cases} x_{ih}^{d,\mathrm{ub}} = x_{ih}^{\mathrm{lb}} + a_{ih}^{kj} \cdot (x_{ih}^{\mathrm{ub}} - x_{ih}^{\mathrm{lb}})/|\boldsymbol{A}_{ih}| \\ x_{ih}^{d,\mathrm{lb}} = x_{ih}^{\mathrm{lb}} + (a_{ih}^{kj} - 1) \cdot (x_{ih}^{\mathrm{ub}} - x_{ih}^{\mathrm{lb}})/|\boldsymbol{A}_{ih}| \end{cases} \quad (13)$$

$$\Delta(k, y) = y \cdot \left(1 - r^{(1-k/k_{\max})^b}\right) \quad (14)$$

where $q_0$ is a uniform random value from [0, 1]; $\varepsilon$ is the exploitation rate which represents the probability of exploitation; $a_{\mathrm{rand}}$ denotes a random action (exploration); $x_{ih}^{d,\mathrm{best}}$ is the previous best optimal solution at the $d$th interval of the $h$th controllable variable for the $i$th buyer agent, and $d = a_{ih}^{kj}$; $x_{ih}^{d,\mathrm{ub}}$ and $x_{ih}^{d,\mathrm{lb}}$ are the upper and lower bounds of the $d$th interval, respectively; $x_{ih}^{\mathrm{ub}}$ and $x_{ih}^{\mathrm{lb}}$ are the upper and lower bounds for the $h$th 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.

MABL can escape from the local optimum via a proper trade-off between exploration and exploitation in (11). In fact, the exploration and exploitation represent the global search and local search, respectively. Therefore, the exploration can effectively escape from the local optimum, while the exploitation can accelerate the convergence and improve the optimum quality.

## IV. DESIGN OF MABL FOR DISTRIBUTED EHED

### A. DESIGN OF BARGARINING GAME

For the distributed EHED, each energy hub can be regarded as a player of bargaining game. Moreover, any one energy with the most diverse types of output energy carriers can be selected as a slack energy hub (i.e., the seller agent), while the others are regarded as different buyer agents. Note that the seller agent is only responsible for determining the joint action strategy by (8), in contrast, each buyer agent not only needs to bargain according to (9) but also needs to search a potential more optimal solution.

### B. DESIGN OF STATE AND ACTION

For the distributed EHED, each energy hub consists of two types of controllable variables, including the input energy production of each energy source and the dispatch factor, where the value range of each controllable variable can be discretized into multiple the optimization intervals, i.e., the action space $\boldsymbol{A}_{ih}$ ($i$ =1, 2,..., $n$; $h$ = 1, 2, ..., $n_i$), while $\boldsymbol{A}_{ih}$ equals to the state space of the next controllable variables, i.e., $\boldsymbol{S}_{i,h+1} = A_{ih}$. Besides, the state of the first controllable variable for each energy hub can be determined according to the current energy gap $\Delta E$ between the energy outputs and the energy demand, as follows:

$$\begin{cases} s_{i1}^{k}, \quad \text{if } E_{i,\min\_v}^{\text{in}} \le \Delta E < E_{i,\max\_v}^{\text{in}} \\ \Delta E = \sum_{p\in\boldsymbol{p}} E_{\text{demand}}^{p} - \sum_{m=1\&m\neq i}^{M} \sum_{p\in\boldsymbol{p}} \bar{E}_{m}^{p,\text{out}} \\ E_{i,\min\_v}^{\text{in}} = \sum_{p\in\boldsymbol{P}} \sum_{j=1}^{n_m^p} E_{mj,\min}^{p,\text{in}} + (v-1) \\ \quad \cdot \left( \sum_{p\in\boldsymbol{P}} \sum_{j=1}^{n_m^p} E_{mj,\max}^{p,\text{in}} - \sum_{p\in\boldsymbol{P}} \sum_{j=1}^{n_m^p} E_{mj,\max}^{p,\text{in}} \right) \Big/ |\boldsymbol{S}_{i1}| \\ E_{i,\max\_v}^{\text{in}} = \sum_{p\in\boldsymbol{P}} \sum_{j=1}^{n_m^p} + v \\ \quad \cdot \left( \sum_{p\in\boldsymbol{P}} \sum_{j=1}^{n_m^p} E_{mj,\max}^{p,\text{in}} - \sum_{p\in\boldsymbol{P}} \sum_{j=1}^{n_m^p} E_{mj,\max}^{p,\text{in}} \right) \Big/ |\boldsymbol{S}_{i1}| \\ v = 1, 2, \ldots, |\boldsymbol{S}_{i1}| \end{cases} \tag{15}$$

where $E_{i,\min\_v}^{\text{in}}$ and $E_{i,\max\_v}^{\text{in}}$ are the total low bound and upper bound at the $v$th state of the $i$th energy hub, respectively; and $\bar{E}_{m}^{p,\text{out}}$ is the current energy output of the $p$th input energy carrier for the $m$th energy hub, which can be provided from the seller agent based on the current optimal joint action strategy.

### C. DESIGN OF REWARD FUNCTION

The reward function $R$ in (10) represents the feedback from the environment after an exploration or exploitation, which can directly influence the optimization performance of MABL. Hence, it should be designed by fully integrating the mathematical model of distributed EHED (2)-(7), as [38]

$$R_{ih}^{j}\left(s_{ih}^{kj}, s_{ih}^{k+1,j}, a_{ih}^{kj}\right) = \begin{cases} \dfrac{p_{\text{m}}}{\min\limits_{j=1,2,\ldots,J} F_i^{kj}}, & \text{if } \left(s_{ih}^{kj}, a_{ih}^{kj}\right) \in \boldsymbol{SA}_i^{\text{Best}} \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

$$F_i^{kj} = f(x_i^{kj}, x_{i,k-1}^{\text{other}}) + \sum_{u=1}^{NC_i} PF_i^u \tag{17}$$

$$PF_i^u = \begin{cases} \chi\left(Z_i^u - Z_i^{u,\text{lim}}\right)^2, & \text{if violated} \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

where $F_i^{kj}$ denotes the fitness function of the $j$th individual at the $k$th iteration, which is closely related to the utility function, i.e., $F_i$ =-$U_i$; $p_{\text{m}}$ is a positive multiplicator; $\boldsymbol{SA}_i^{\text{Best}}$ represents the state-action pairs set of the best individual at the $k$th iteration for the $i$th buyer agent; $f$ is one of three objective functions in (2); $NC_i$ is the number of constraints for the $i$th buyer agent based on (5)-(7); $PF_i^u$ is the penalty function corresponding to the $u$th constraint for the $i$th buyer agent; $\chi$ is the penalty factor; $Z_i^u$ denotes the $u$th constraint of the $i$th buyer agent; and $Z_i^{u,\text{lim}}$ is the constraint limit corresponding to $Z_i^u$.

**TABLE 1. Overall Execution Procedure of MABL for Distributed EHED**

```
Determine the objective function from (2);
Initialize the parameters of MABL;
While k≤k_max
    The seller agent determines the current joint action by (8);
    The seller agent issues x_k^* to all the buyer agents;
    For m:=1 to M-1
        Determine the state of the first controllable variable by (15);
        Implement exploration and exploitation using (11)-(14);
        Calculate objective and constraints obtained by each solution
          using (3)-(7);
        Calculation the reward function for each executed state-
          action by (16)-(18);
        Update the best optimal solution;
        Update the knowledge matrix by (10);
        Implement the bargaining process of each buyer agent by (9);
    End For
Set k:=k+1;
End While
Output the optimal solution of distributed EHED.
```

**TABLE 2. The Parameters Used in MABL for Distributed EHED.**

| $\alpha$ | $\gamma$ | $\varepsilon$ | $b$ | $p_{\text{m}}$ | $J$ | $k_{\max}$ |
|---|---|---|---|---|---|---|
| 0.1 | 0.05 | 0.7 | 0.5 | 0.001 | 50 | 200 |

**TABLE 3.** Performance Results with Different Objective Functions Obtained by Different Algorithms in 30 Runs

| Algorithm | | Minimization of total energy costs | | | Minimization of total energy losses | | | Minimization of total energy costs and losses simultaneously | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $f_C$ (mu) | $f_L$ (pu) | $T_C$ (s) | $f_C$ (mu) | $f_L$ (pu) | $T_C$ (s) | $f_C$ (mu) | $f_L$ (pu) | $f_I$ (mu) | $T_C$ (s) |
| GA | | 16384.4316 | 4.0788 | 27.3846 | **16622.8797** | 3.3902 | 17.2353 | 16388.6511 | 3.6722 | 18960.5487 | 29.6453 |
| PSO | | 16106.0947 | 3.7058 | 18.5038 | 16858.0050 | 2.9950 | 11.7354 | 16402.3627 | 3.1190 | 18588.6434 | 19.6453 |
| GSA | | 15870.0549 | 4.5232 | 17.3926 | 17308.0887 | 2.9385 | 9.9919 | 15964.8911 | 3.3843 | 18273.8646 | 18.8789 |
| EGSA | | 15751.6029 | 3.6376 | 17.2627 | 16823.5307 | 2.8162 | 9.6391 | 15754.6403 | 3.6154 | 18188.8198 | 18.8837 |
| TVAC-GSA | | 15747.3098 | 3.6459 | 17.3030 | 16811.3114 | 2.8317 | 9.5825 | 15751.8705 | 3.6360 | 18199.4468 | 18.6678 |
| SAL-TVAC-GSA | | 15728.3913 | 3.6706 | 17.2354 | 16824.7134 | 2.8100 | 9.2078 | **15730.4677** | 3.6284 | 18169.6515 | 18.5632 |
| MABL | Worst | 16275.4452 | **3.4500** | 2.0869 | 16877.4285 | 2.6500 | 1.7436 | 16347.2654 | 2.7939 | 18299.1442 | 2.1697 |
| | Mean | 15901.7736 | 3.4538 | 2.0887 | 17552.4483 | 2.2348 | 1.7105 | 16057.4331 | 2.4959 | 17770.5560 | 2.1550 |
| | Best | **15667.9776** | 3.5503 | **2.0810** | 17940.7631 | **1.8879** | **1.6925** | 15734.9647 | **2.3374** | **17306.6931** | **2.1479** |

### D. OVERALL EXECUTION

In summary, the overall execution of MABL for distributed EHED is provided in Table I. Note that the implementation process of each buyer agent can be can be computed simultaneously, thus the computation time can be dramatically shorten.

## V. CASE STUDIES

For evaluating the performance of MABL for distributed EHED, six centralized heuristic optimization algorithms including GA [8], PSO [7], GSA [6], EGSA [6], TVAC-GSA [6], and SAL-TVAC-GSA [4], are used for comparisons with different objective functions in (2), where the parameters of these algorithms can be referred from [4]. Through trial-and-error, the main parameters of MRBL for distributed EHED are given in Table II. Furthermore, all the simulations of the proposed algorithm are undertaken in Matlab R2016a by a personal computer with Intel(R) Xeon (R) E5-2670 v3 CPU at 2.3 GHz with 64 GB of RAM.

### A. SIMULATION MODEL

The testing MECS is composed of 39 energy hubs with 29 different structures, in which the 34th energy hub is selected as the seller agent, while others are the buyer agents. This system consists of 76 energy sources, including 27 electrical sources with the value-point effects (8 of that consider the POZ constraints), 34 gas stations, and 15 heat sources, where the types of output energy carriers contain electricity, heat, cool and compressed air. All the parameters of this system are given in [4], which has 103 controllable variables (76 of that for energy sources and 27 for dispatch factors). Besides, the optimization interval of each controllable variable is equally divided into 5 smaller intervals, thus each action space contains 5 alternative action strategies, as well as for the state space.

### B. STUDY OF CONVERGENCE

Fig. 4 illustrates the convergence process of MABL for minimizing the total energy costs $f_C$ of distributed EHED, where the total energy demand of electricity, heat, cool, and compress air are 12.0, 9.5, 0.7, and 1.2 pu, respectively. It can be found from Fig. 4(a) that the seller agent and the buyer agents can constantly search a better optimal solution with a lower total energy costs via the continuous bargaining game from each other, more concretely, each of them can approximate an

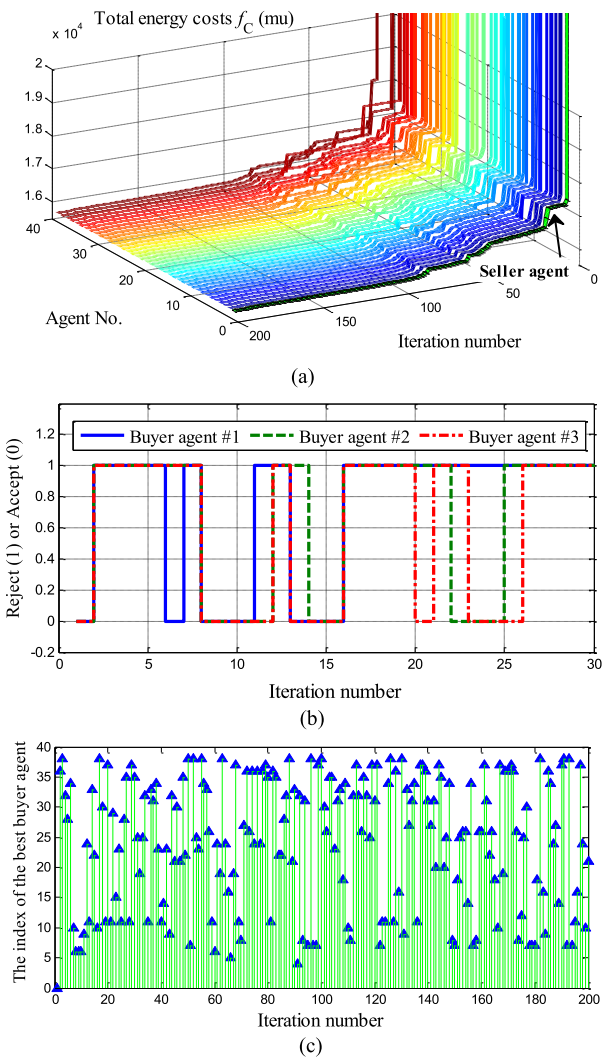


**FIGURE 4.** Convergence of MABL for minimizing the total energy costs. (a) Objective functions of different agents. (b) Bargaining processes of some agents. (c) Index of the best buyer agent with the largest payoff.

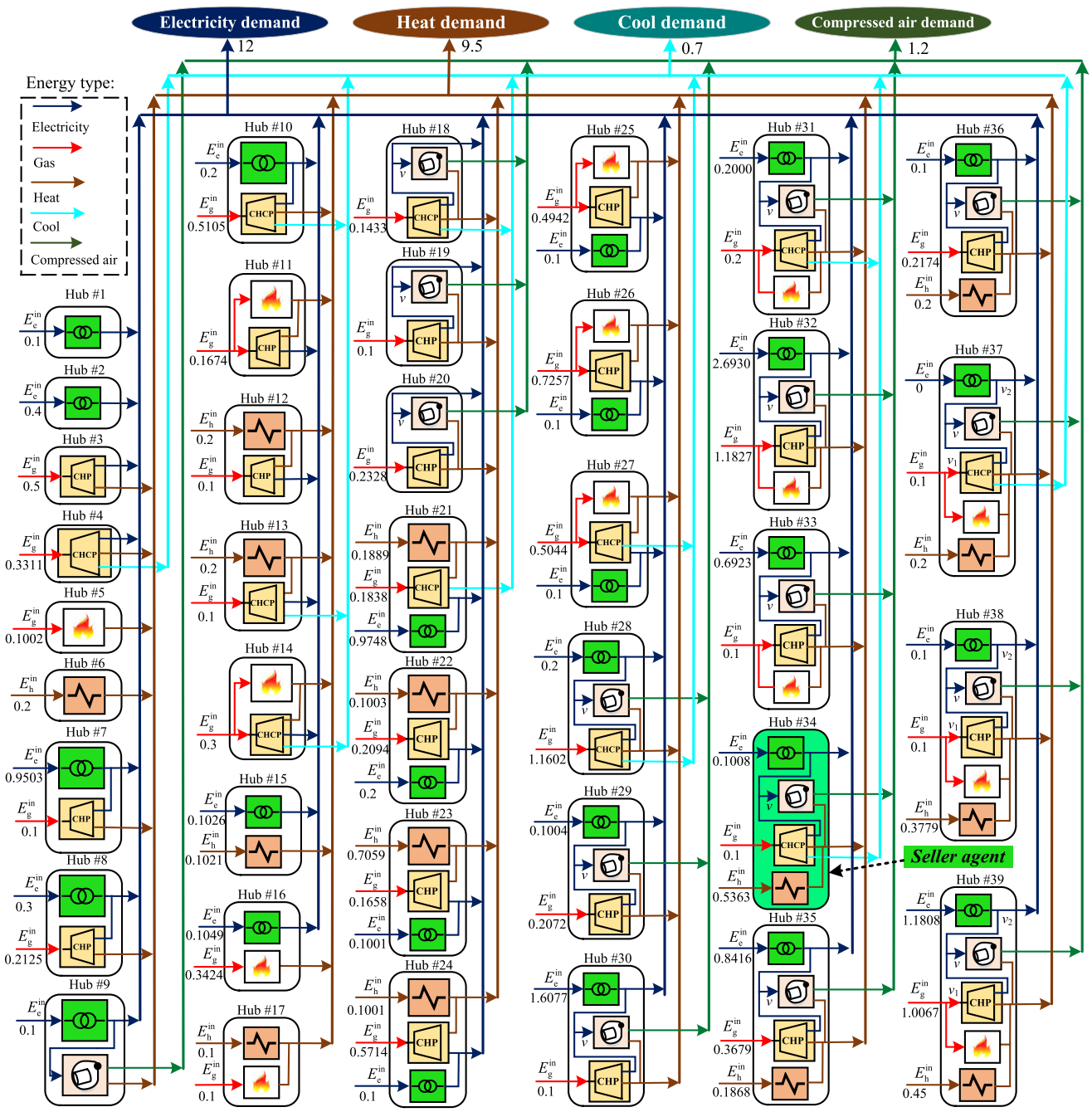


**FIGURE 5.** The 39-hub MECS with 29 energy hub structures and 76 energy sources.

optimal solution after 80 iterations, while the optimal input energy production of each energy source is given in Fig. 5. This obviously indicates that the bargaining game can effectively achieve an effective coordination between the buyer agents and a seller agent, thus the total payoff of all the energy hubs can be maximized.

Besides, to search a potential better optimal solution, each buyer agent can select an optimal action strategy for each controllable variable of the corresponding energy hub by self-scheduling based on knowledge learning, thus they can reject the seller agent (See Fig. 4(b)) and contribute to improving the total payoff of MECS (See Fig. 4(c)).

### C. COMPARATIVE RESULTS AND DISCUSSIONS

Table III shows performance results with different objective functions in (2) obtained by different algorithms in 30 independent runs, where the performance results of the former six algorithms are the best optimal solutions with the smallest objective functions in 30 runs, respectively; all the performance results are in a per-unit (pu) system and the energy

costs are in monetary-unit (mu); $T_C$ is the computation time in a run. Note that the obtained objective functions of MABL are much smaller than that of other algorithms, especially for the minimization of total energy losses. Particularly, the minimal total energy losses $f_L$ obtained by MABL in the best case is only 55.69% of that of GA, in which the optimal operation points of all the energy hubs are provided in Table IV while the obtained optimal dispatch factors are given in Fig. 6, and the detailed operation illustration of energy hub #27 from the input energy carriers to the output energy carriers is given in Fig. 7. It verifies that the obtained optimum by MABL can not only satisfy all the constraints (5)-(7), but also significantly reduce the total energy losses. Even in the worst case, it is still only 94.31% of that of SAL-TVAC-GSA. These high-quality optimal solutions clearly demonstrate the beneficial effect of knowledge learning, and the effective coordination among the seller agent and the buyer agents via the continuous bargaining games.

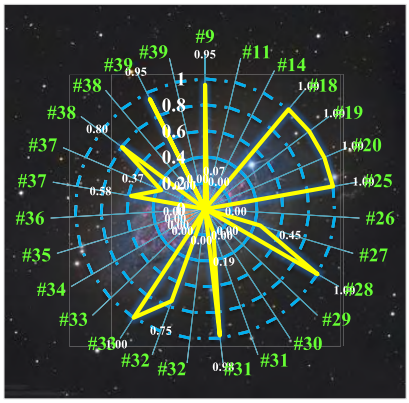


**FIGURE 6.** Dispatch factors distribution of all the energy hubs with minimization of total energy losses obtained by MABL in the best case.

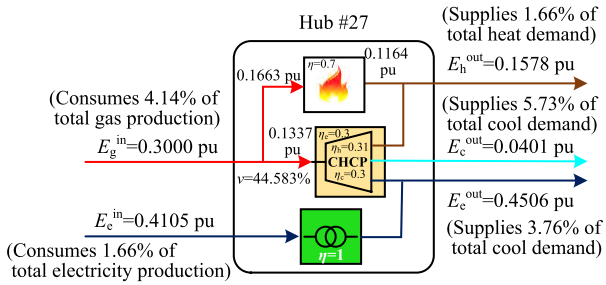


**FIGURE 7.** Best optimal results of energy hub #27 with minimization of total energy losses obtained by MABL.

Moreover, it is obviously that the computation time of MABL for distributed EHED is the shortest among all the algorithms, which is only 7.25% of that of GA for minimization of total energy costs and losses simultaneously in the best case, i.e., the computation rate can be up to 13.8 times against to GA, which is essentially benefited from the distributed optimization and associative memory based knowledge learning with a cooperative swarm.

**TABLE 4.** Optimization Results of the Best Optimal Solution with Minimization of Total Energy Losses Obtained by MABL.

| Hub No. | Input Energy (pu) | | Output Energy (pu) | | | | Energy loss (pu) |
|---|---|---|---|---|---|---|---|
| | Type | Optimum | Electricity | Heat | Cool | Compressed air | |
| #1 | Electricity | 0.1000 | 0.0990 | 0.0000 | 0.0000 | 0.0000 | 0.0010 |
| #2 | Electricity | 0.4000 | 0.4000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| #3 | Gas | 0.5000 | 0.1500 | 0.2000 | 0.0000 | 0.0000 | 0.1500 |
| #4 | Gas | 0.3000 | 0.0750 | 0.1050 | 0.0600 | 0.0000 | 0.0600 |
| #5 | Gas | 0.1091 | 0.0000 | 0.0818 | 0.0000 | 0.0000 | 0.0273 |
| #6 | Heat | 0.5000 | 0.0000 | 0.4750 | 0.0000 | 0.0000 | 0.0250 |
| #7 | Electricity | 0.2000 | 0.2270 | 0.0410 | 0.0000 | 0.0000 | 0.0320 |
| | Gas | 0.1000 | | | | | |
| #8 | Electricity | 0.3000 | 0.3250 | 0.0420 | 0.0000 | 0.0000 | 0.0330 |
| | Gas | 0.1000 | | | | | |
| #9 | Electricity | 0.4186 | 0.0192 | 0.0799 | 0.0000 | 0.2795 | 0.0399 |
| #10 | Electricity | 0.2000 | 0.2500 | 0.0740 | 0.0400 | 0.0000 | 0.0360 |
| | Gas | 0.2000 | | | | | |
| #11 | Gas | 0.1500 | 0.0030 | 0.1159 | 0.0000 | 0.0000 | 0.0311 |
| #12 | Gas | 0.1000 | 0.0300 | 0.3961 | 0.0000 | 0.0000 | 0.0352 |
| | Heat | 0.3613 | | | | | |
| #13 | Electricity | 0.1000 | 0.0250 | 0.2843 | 0.0300 | 0.0000 | 0.0284 |
| | Gas | 0.2676 | | | | | |
| #14 | Gas | 0.3000 | 0.0000 | 0.2100 | 0.0000 | 0.0000 | 0.0900 |
| #15 | Electricity | 0.1001 | 0.0981 | 0.1024 | 0.0000 | 0.0000 | 0.0134 |
| | Heat | 0.1138 | | | | | |
| #16 | Electricity | 0.1000 | 0.0950 | 0.1571 | 0.0000 | 0.0000 | 0.0631 |
| | Gas | 0.2152 | | | | | |
| #17 | Gas | 0.1000 | 0.0000 | 0.1673 | 0.0000 | 0.0000 | 0.0353 |
| | Heat | 0.1025 | | | | | |
| #18 | Gas | 0.1000 | 0.0000 | 0.0370 | 0.0290 | 0.0210 | 0.0130 |
| #19 | Gas | 0.1000 | 0.0000 | 0.0460 | 0.0000 | 0.0210 | 0.0330 |
| #20 | Gas | 0.2000 | 0.0000 | 0.0861 | 0.0000 | 0.0455 | 0.0684 |
| #21 | Electricity | 1.4677 | 1.5038 | 0.4620 | 0.0240 | 0.0000 | 0.0040 |
| | Gas | 0.1001 | | | | | |
| | Heat | 0.4259 | | | | | |
| #22 | Electricity | 0.2006 | 0.2586 | 0.6323 | 0.0000 | 0.0000 | 0.0540 |
| | Gas | 0.2000 | | | | | |
| | Heat | 0.5443 | | | | | |
| #23 | Electricity | 0.1022 | 0.1272 | 0.1471 | 0.0000 | 0.0000 | 0.0407 |
| | Gas | 0.1000 | | | | | |
| | Heat | 0.1128 | | | | | |
| #24 | Electricity | 1.3101 | 1.3401 | 1.5400 | 0.0000 | 0.0000 | 0.0300 |
| | Gas | 0.1000 | | | | | |
| | Heat | 1.5000 | | | | | |
| #25 | Electricity | 0.1983 | 0.3109 | 0.1448 | 0.0000 | 0.0000 | 0.0644 |
| | Gas | 0.3218 | | | | | |
| #26 | Electricity | 0.1000 | 0.0970 | 0.2250 | 0.0000 | 0.0000 | 0.0780 |
| | Gas | 0.3000 | | | | | |
| #27 | Electricity | 0.4105 | 0.4506 | 0.1578 | 0.0401 | 0.0000 | 0.0619 |
| | Gas | 0.3000 | | | | | |
| #28 | Electricity | 0.6000 | 0.0000 | 0.8491 | 0.4118 | 0.7013 | 0.0577 |
| | Gas | 1.4200 | | | | | |
| #29 | Electricity | 0.1000 | 0.1590 | 0.0640 | 0.0000 | 0.0000 | 0.0770 |
| | Gas | 0.2000 | | | | | |
| #30 | Electricity | 0.3013 | 0.3343 | 0.0450 | 0.0000 | 0.0000 | 0.0220 |
| | Gas | 0.1000 | | | | | |
| #31 | Electricity | 0.2011 | 0.0046 | 0.2067 | 0.0129 | 0.1237 | 0.0743 |
| | Gas | 0.2211 | | | | | |
| #32 | Electricity | 0.0000 | 0.0000 | 0.0740 | 0.0000 | 0.0000 | 0.0260 |
| | Gas | 0.1000 | | | | | |
| #33 | Electricity | 0.0000 | 0.0350 | 0.0470 | 0.0000 | 0.0000 | 0.0180 |
| | Gas | 0.1000 | | | | | |
| #34 | Electricity | 1.9964 | 2.0385 | 0.1294 | 0.0365 | 0.0000 | 0.0324 |
| | Gas | 0.1405 | | | | | |
| | Heat | 0.1000 | | | | | |
| #35 | Electricity | 0.2013 | 0.2652 | 0.1568 | 0.0000 | 0.0000 | 0.0817 |
| | Gas | 0.2000 | | | | | |
| | Heat | 0.1025 | | | | | |
| #36 | Electricity | 0.1000 | 0.1478 | 0.4735 | 0.0000 | 0.0001 | 0.1999 |
| | Gas | 0.2000 | | | | | |
| | Heat | 0.5213 | | | | | |
| #37 | Electricity | 0.0002 | 0.0096 | 0.4888 | 0.0156 | 0.0031 | 0.0711 |
| | Gas | 0.1000 | | | | | |
| | Heat | 0.4880 | | | | | |
| #38 | Electricity | 2.7763 | 2.7988 | 0.4047 | 0.0000 | 0.0041 | 0.0368 |
| | Gas | 0.1000 | | | | | |
| | Heat | 0.3681 | | | | | |
| #39 | Electricity | 0.2877 | 0.3225 | 0.5510 | 0.0000 | 0.0006 | 0.0430 |
| | Gas | 0.1000 | | | | | |
| | Heat | 0.5294 | | | | | |
| **Sum** | **Electricity** | **24.6586** | **12.0000** | **9.5000** | **0.7000** | **1.2000** | **1.8879** |

## VI. CONCLUSION

This paper presents a novel MABL for distributed EHED of a large-scale MECS. The main observations are summarized as follows:

1) The one-seller and $n$-buyer bargaining game can effective achieve an effective coordination among the energy hub, thus the self-organized and distributed computation of each energy hub can be implemented for distributed EHED.

2) The associative memory and swarm intelligence can dramatically accelerate the convergence rate of knowledge learning, while the non-uniform mutation operator is conductive to a high optimization accuracy.

3) Simulation results on the complex testing system verify the superior performance of MABL compared with six centralized heuristic optimization algorithms, which can rapidly obtain a high-quality optimal solution for distributed EHED with different objective functions.

4) Compared with the gradient-based distributed methods (e.g., consensus algorithm or ADMM), MABL is more suitable for the non-convex and non-smooth EHED as it is fully independent on the mathematical model. Moreover, the performance of MABL is competitive to other distributed heuristic optimization algorithms due to its fast convergence and strong global search ability. Hence, it is adequate to be generalized to handle other distributed optimization of large-scale and complex smart energy systems.

In order to improve the applicability for the real-world system, our future works will consider both the environmental effects and the energy networks constraints.

## REFERENCES

[1] M. Geidl, P. Favre-Perrod, B. Klockl, G. Anderrson, and K. Frohlich, "Energy hubs for the future," *IEEE Power Energy Mag.*, vol. 5, no. 1, pp. 24–30, Jan./Feb. 2007.

[2] A. Parisio, C. Del Vecchio, and A. Vaccaro, "A robust optimization approach to energy hub management," *Int. J. Elect. Power Energy Syst.*, vol. 42, no. 1, pp. 98–104, 2012.

[3] A. Shabanpour-Haghighi and A. R. Seifi, "Energy flow optimization in multicarrier systems," *IEEE Trans. Ind. Informat.*, vol. 11, no. 5, pp. 1067–1077, Oct. 2015.

[4] S. D. Beigvand, H. Abdi, and M. L. Scala, "A general model for energy hub economic dispatch," *Appl. Energy*, vol. 190, pp. 1090–1111, Mar. 2017.

[5] D. Xie, Y. Lu, J. Sun, C. Gu, and G. Li, "Optimal operation of a combined heat and power system considering real-time energy prices," *IEEE Access*, vol. 4, pp. 3005–3015, Jun. 2016.

[6] S. D. Beigvand, H. Abdi, and M. L. Scala, "Optimal operation of multi-carrier energy systems using time varying acceleration coefficient gravitational search algorithm," *Energy*, vol. 114, pp. 253–265, Nov. 2016.

[7] G. Abbas, J. Gu, U. Farooq, A. Raza, M. U. Asad, and M. E. El-Hawary, "Solution of an economic dispatch problem through particle swarm optimization: A detailed survey—Part II," *IEEE Access*, vol. 5, pp. 24426–24445, Nov. 2017.

[8] B. B. Munyazikwiye, H. R. Karimi, and K. G. Robbersmyr, "Optimization of vehicle-to-vehicle frontal crash model based on measured data using genetic algorithm," *IEEE Access*, vol. 5, pp. 3131–3138, Feb. 2017.

[9] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Appl. Soft Comput.*, vol. 8, pp. 687–697, Jan. 2008.

[10] L. M. Zheng, S. X. Zhang, S. Y. Zheng, and Y. M. Pan, "Differential evolution algorithm with two-step subpopulation strategy and its application in microwave circuit designs," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 911–923, Jul. 2016.

[11] J. Giri, M. Parashar, J. Trehern, and V. Madani, "The situation room: Control center analytics for enhanced situational awareness," *IEEE Power Energy Mag.*, vol. 10, no. 5, pp. 24–39, Sep. 2012.

[12] S. M. Amin, "Smart grid security, privacy, and resilient architectures: Opportunities and challenges," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, San Diego, CA, USA, Jul. 2012, pp. 1–2.

[13] 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.*, vol. 12, no. 5, pp. 1775–1785, Oct. 2016.

[14] C. Li, X. Yu, T. Huang, and X. He, "Distributed optimal consensus over resource allocation network and its application to dynamical economic dispatch," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2407–2418, Jun. 2018.

[15] S. Fan, Z. Li, J. Wang, L. Piao, and Q. Ai, "Cooperative economic scheduling for multiple energy hubs: A bargaining game theoretic perspective," *IEEE Access*, vol. 6, pp. 27777–27789, May 2018.

[16] C. H. Liang, C. Y. Chung, K. P. Wong, and X. Z. Duan, "Parallel optimal reactive power flow based on cooperative co-evolutionary differential evolution and power system decomposition," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 249–257, Feb. 2007.

[17] A. Erdeljan, D. Capko, S. Vukmirovic, D. Bojanic, and V. Congradac, "Distributed PSO algorithm for data model partitioning in power distribution systems," *J. Appl. Res. Technol.*, vol. 12, no. 5, pp. 947–957, 2014.

[18] T. Yu, J. Liu, K. W. Chan, and J. J. Wang, "Distributed multi-step Q(λ) learning for optimal power flow of large-scale power grids," *Int. J. Elect. Power Energy Syst.*, vol. 42, no. 1, pp. 614–620, 2012.

[19] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 2, pp. 156–172, Mar. 2008.

[20] Y. Xu, W. Zhang, W. Liu, and F. Ferrese, "Multiagent-based reinforcement learning for optimal reactive power dispatch," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 1742–1751, Nov. 2012.

[21] T. Yu, X. S. Zhang, B. Zhou, and K. W. Chan, "Hierarchical correlated Q-learning for multi-layer optimal generation command dispatch," *Int. J. Electr. Power Energy Syst.*, vol. 78, pp. 1–12, Jan. 2016.

[22] M. Tan, C. Han, X. Zhang, L. Guo, and T. Yu, "Hierarchically correlated equilibrium Q-learning for multi-area decentralized collaborative reactive power optimization," *CSEE J. Power Energy Syst.*, vol. 2, no. 3, pp. 65–72, Sep. 2016.

[23] Y. Hu, Y. Gao, and B. An, "Accelerating multiagent reinforcement learning by equilibrium transfer," *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1289–1302, Jul. 2015.

[24] J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games," *J. Mach. Learn. Res.*, vol. 4, pp. 1039–1069, Nov. 2003.

[25] Y. Hu, Y. Gao, and B. An, "Multiagent reinforcement learning with unshared value functions," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 647–662, Apr. 2015.

[26] J. D. R. Millán, D. Posenato, and E. Dedieu, "Continuous-action Q-learning," *Mach. Learn.*, vol. 49, nos. 2–3, pp. 247–265, 2002.

[27] F. Kienzle, P. Favre-Perrod, M. Arnold, and G. Andersson, "Multi-energy delivery infrastructures for the future," in *Proc. 1st Int. Conf. Infrastruct. Syst. Services, Building Netw. Brighter Future (INFRA)*, Rotterdam, The Netherlands, Nov. 2008, pp. 1–5.

[28] P. Favre-Perrod, F. Kienzle, and G. Andersson, "Modeling and design of future multi-energy generation and transmission systems," *Eur. Trans. Elect. Power Banner*, vol. 20, no. 8, pp. 994–1008, 2010.

[29] X. Zhang, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "Optimal expansion planning of energy hub with multiple energy infrastructures," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2302–2311, Sep. 2015.

[30] M. Geidl and G. Andersson, "Optimal power flow of multiple energy carriers," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 145–155, Feb. 2007.

[31] J.-B. Park, Y.-W. Jeong, J.-R. Shin, and K. Lee, "An improved particle swarm optimization for nonconvex economic dispatch problems," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 156–166, Feb. 2010.

[32] J. Sun, V. Palade, X.-J. Wu, W. Fang, and Z. Wang, "Solving the power economic dispatch problem with generator constraints by random drift particle swarm optimization," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 222–232, Feb. 2014.

[33] J. F. Nash, Jr., "The bargaining problem," *Econometrica*, vol. 18, no. 2, pp. 155–162, 1950.

[34] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.

[35] Z. Xiaoshun, Y. Tao, Y. Bo, and C. Lefeng, "Accelerating bio-inspired optimizer with transfer reinforcement learning for reactive power optimization," *Knowl.-Based Syst.*, vol. 116, pp. 26–38, Jan. 2017.

[36] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd ed. New York, NY, USA: Springer, 1996.

[37] R. A. C. Bianchi, L. A. Celiberto, P. E. Santos, J. P. Matsuura, and R. L. de Mantaras, "Transferring knowledge as heuristics in reinforcement learning: A case-based approach," *Artif. Intell.*, vol. 226, pp. 102–121, Sep. 2015.
[38] X. Zhang, T. Bao, T. Yu, B. Yang, and C. Han, "Deep transfer Q-learning with virtual leader-follower for supply-demand Stackelberg game of smart grid," *Energy*, vol. 133, pp. 348–365, Aug. 2017.

**XIAOSHUN ZHANG** (M'17) received the B.Eng. and Ph.D. degrees in electrical engineering from the South China University of Technology, Guangzhou, China, in 2012 and 2017, respectively. He is currently a Research Associate with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong. His major research interests include game theory, reinforcement learning, and transfer learning in operation of power systems.

**TAO YU** (M'11) received the B.Eng. degree in electrical power system from Zhejiang University, Hangzhou, China, in 1996, the M.Eng. degree in hydroelectric engineering from Yunnan Polytechnic University, Kunming, China, in 1999, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2003. He is currently a Professor with the College of Electric Power, South China University of Technology, Guangzhou, China. His special fields of interest include nonlinear and coordinated control theory, artificial intelligence techniques in planning, and operation of power systems.

**ZHIYI ZHANG** received the B.Eng. degree in electrical engineering from the South China University of Technology, Guangzhou, China, in 2017, where he is currently pursuing the M.Eng. degree in electrical engineering with the College of Electric Power. His major research interests include optimization operation and control of power systems.

**JIANLIN TANG** received the B.Eng. degree in electrical engineering from the South China University of Technology, Guangzhou, China, in 2017, where he is currently pursuing the M.Eng. degree in electrical engineering with the College of Electric Power. His major research interests include application of artificial intelligence algorithm in power systems.