Optimization of Dynamic Dispatch for Multiarea Integrated Energy System Based on Hierarchical Learning Method

YIJIN LI[1], HAO TANG[1], (Member, IEEE), KAI LV[1], (Graduate Student Member, IEEE), KE WANG[2], AND GANG WANG[2]

[1]School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, China
[2]China Electric Power Research Institute (Nanjing), Nanjing 210003, China

Corresponding author: Hao Tang (htang@hfut.edu.cn)

ABSTRACT The integrated energy system (IES) with various energy demands and distributed energy resources has been a significant approach to improve the efficiency of energy utilization. Considering the uncertainties of renewable energy sources and loads, the energy dispatch optimization for multiarea IESs is studied in this paper. Different from the most current studies, not only the electrical power and heat distribution in each area is optimized, but also the coordination power dispatch between areas. A hierarchical learning method is proposed inhere to improve the operation performance of the multiarea IESs. The proposed method with data-driven way is a model-free method which has no requirement for the accurate mathematical model. With the hierarchical structure, the electrical power dispatched between areas is optimized in upper layer, together with the dispatching optimization in each area at lower layer, to decrease the operation cost for the system and power demand from the power grid. Finite horizon discrete dynamic process model is adopted to simulate the data for learning. The simulation results show the effectiveness of the optimization policy can achieve an economic and stable operation for the multiarea IES.

INDEX TERMS Integrated energy system, dispatch optimization, stochastic process, hierarchical learning, dynamic programming.

NOMENCLATURE

A. PARAMETERS

P	Power, kW
M	Total number of areas
cop	Efficiency coefficient
E	Capacity of energy storage
N	Maximum discrete level
Δ	Random numbers in (0,1) interval
$a_{gt0}/a_{gt1}/a_{gt2}$	Coefficient of the cost for gas turbine
T/K	Decision period and number of decision cycles per day
α_{rl}	Learning rate

B. VARIABLES

t	Time, h
k	Index of decision epoch

n	Index of discretization level
a/D	Action and action set
s/Φ	State and state set
c/V	Cost and performance criteria
π	Policy

C. SUBSCRIPT AND INDICES

$cool/h/ele$	Cooling load, heat load and electrical load
$pv/gt/es/ths$	Photovoltaic, gas turbine, electrical storage and thermal storage
exp/stc	Expected value and standard condition
$-/\_$	Maximum/minimum limit
$ch/disch$	Charge or discharge action of storage

I. INTRODUCTION

Increasing energy consumption is inevitable with the development of society. Integrated energy system (IES) can

produce onsite electricity and the byproduct heat to satisfy heat, cooling and electrical demands with high efficiency and increase the renewable energy penetration level, which has been rapid developed in recent years [1], [2]. The development and studies of IES have received widespread attention in different countries [3], [4]. In China, many projects of IESs have been successfully established in recent years, and many researchers focus on the construction and operation of the IES [5], [6]. Compared with traditional individual energy system, cascade utilization of energy in IES could improve the energy efficiency and decrease the operation cost to meet the demand of different types of energy [7]. Together with the development of renewable energy and distributed energy technologies, the energy efficiency and benefit for environment can be further improved [8]. References [9] and [10] investigate the energy performance of distributed energy systems and renewable energy source integrated with IES and indicate that the energy saving is efficiently improved and the fossil fuels is decreased.

In order to facilitate energy sharing and improve system flexibility, the coordinated operation of multiple IESs has received more and more attentions by combining the complementarity of energy and load stochastic characteristic in different areas [11]. Interaction mechanism of district electricity and heating systems is analyzed in [12], and [13] proposes a technique to dispatch simultaneously energy and reserve to improve the efficiency of the interconnected multi-area system. Considering traditional generation of interconnected power grid, [14] develops a real-time coordinated energy control of the multiarea. As the power system is still the main structure for supplying energy and its stable operation is of great significance for the whole society, we study a multiarea IES with electrical power cooperation evaluation in this paper. And the optimization of the dynamic energy dispatch is studied to decrease the daily operation cost. The multiarea IES is operating in the power grid connected mode that each area is connecting to the power grid. And the power coordinated between areas which can further improve the overall performance of the multiarea IES and reduce the random mismatch of renewable sources and loads. While for the joint operation for the multiarea IES, there are some difficulties to obtain the optimal energy dispatch policy. First, compared with the individual IES, the composition and energy coupling of the multiarea IES is more complicated. Second, the state dimension and control variables are several times than that of the independent IES, which leads to an exponential growth in computation scale. Third, the randomness of the renewable energy sources and loads in different areas further increase the difficulty of energy dispatch optimization. Finally, the optimization for the energy dispatch and operation of both the overall multiarea IES and each area are need to be considered simultaneously, which is rarely mentioned in many related studies.

There have been many studies related to the optimization for the configuration and operation of IES with distributed energy resources. For the multiple IESs, the structure and operation constraints are more complicated with energy interaction. However, for the multiarea IES, most studies focus on the overall economic operation of the multiarea IES with coordinated energy scheduling, while ignoring the performances of each area. Due to the multiple energy coupling and the uncertainties of resources and loads, the simplified mathematic model is used in many related optimization problems for IES. For example, a coordinated dispatching method for IESs is proposed in [15] in order to improve the consumption rate of renewable energy with considering the differences of multiple functional areas. Reference [16] establishes a gas-electricity joint operation model and finds an optimal scheduling method for integrated energy systems. In [17], a mixed integer linear programming model of multi-district IES is established to achieve the optimal operation. Reference [18] describes the modeling method of the IES with a generalized network flow model of IES incorporates the production, storage, and transportation of energy in a single mathematical framework, and [19] summarizes the modeling tools and simplifications of the IES including the operation constraints and transmission networks. As far as the authors knowledge, mixed integer linear or nonlinear programming model and corresponding optimization method is the most widely used for the optimal operation or design of the IES. For example, the pipeline flow equation is linearized in [20] and the optimization model is transformed into mixed linear programming formulation to achieve the optimal dispatch strategy for IES, and refer to the mixed integer nonlinear programming model, [21] proposes an optimization procedure for the optimal design of a trigenerative system to satisfy the energy needs. The structure and configuration of the IES is closely related to its operation efficiency and economic. Combining with the mixed integer linear and nonlinear programming method, [22] and [23] adopt a two-stage optimization algorithm to optimize the capacity design, energy production and distribution of the IES. The two-stage algorithm is widely applied in the optimization problem of the IES with different optimization methods for the complexity of the IES [24], [25]. For example, dominance based evolutionary algorithm and fuzzy c-means clustering method are used in [26] with two-stage algorithm to find the best compromise solutions for the economic emission dispatch of the IES. However, most of these studies are for the single IES, and the optimization in each stage are concerning about different aspects of structure and operation. Based on these, a hierarchical optimization structure is proposed in this paper for the operation optimization of the multiarea IES considering the uncertainties and power interactions. The uncertainties and randomness of the renewable resources and loads further improve the complexity for optimization of the IES and have many corresponding studies. For example, the optimization for the IES with wind power and stochastic demands is studied in [27] and [28] to enhance the absorption and economic dispatch. And many algorithms have been applied to the optimization for the IES. In [29], three different control approaches for integrated energy system is compared with

renewable uncertainty. Particle swarm optimization algorithm and generation algorithm are used in the optimization problems of the IES in [30] and [31]. The improved particle swarm optimization algorithms are studied in [32] and [33], and [34] developed a multi-objective group search optimizer with adaptive covariance and Lévy flights to optimize the power dispatch of the large-scale IES.

These algorithms above are all applied to the operation optimization of the IES with simplified mathematic model or scenario, while it can be hardly achieved in practical precisely. Besides, the constraints and criterions of power connection between areas, such as interactive power, climbing constraints and the peak-valley difference, are important criterion to evaluate the stability of the system under grid-connected mode, but are considered rarely in previously studies. The main limitation of static optimization is that static values are not the best option due to the changing environment. Therefore, the best option for the optimization for the real operation process is to obtain the optimum values for the decision variables in each time-step. One of the effective methods for dealing with the dynamic programming problem is reinforcement learning, which is a model-free optimization method by interaction with the environment and achieve many successful results [35]–[37]. For the IES, the model-free optimization methods are also applied in recent years. A comprehensive review of the application of reinforcement learning to developing autonomous building energy management systems is studied in [38], [39], and [39] reported that energy savings are of greater than 20% for complex building energy management problems when implementing reinforcement learning. Considering the system operator's operating cost, Reference [40] proposes a dynamic energy conversion strategy for the energy management of an IES with renewable energy. However, the multiarea IES problems with both energy dispatching in areas and collaboration optimization between areas are rarely mentioned with learning methods together with the impact for the power grid. Thus, in this paper, we propose a model-free hierarchical learning method based on the reinforcement learning for the dynamic dispatch optimization of the multiarea IESs with considering the uncertainties of both the loads and the renewable energy sources to decrease the daily operational cost. The heat and power dispatch of areas are optimized simultaneously in the lower layer, and the power dispatch areas is optimized at the upper layer. The information between the upper layer and the lower layer is interacting in the whole learning process, that the operation cost of the overall system and each area can be both decreased. In this study, we simulated a multiarea IES to generate the samples for learning. Both simulation and real-world data can be available and applicable for the learning method. The information interactions between the upper layer and the lower layer occur in the whole learning process, which lead the agents to achieve the optimal goal of the system.

The main contributions of this study are listed: First, a model-free hierarchical learning method by data-driven is

proposed in order to avoid developing an accurate mathematical model and handle the curse of dimensionality. The power dispatch between areas is optimized in upper layer, together with the dispatch optimization in each area at lower layer. Second, the dynamic energy dispatch problem is formulated as a finite dynamic programming decision process and the stochastic and uncertainties of renewable resources and loads are both considered in the optimization. The simulation model to achieve data for learning is established considering operation characteristics and constraints of different devices. Last, we systematically evaluate the performance of the method proposed with numerical analysis. With the optimization by hierarchical learning method, not only the daily operation cost of the overall system and each area are reduced, but also the policy stability in the random environment is improved. Besides, the influence of the power interaction between areas and the power grid is analyzed, and the different grid-connection modes are compared to explain the advantages of the power interconnection between areas.

The remainder of this paper is organized as follows: Section 2 describes the multiarea IES and the simulation models of the various components in detail. The hierarchical learning method is described in Section 3. The results of the simulations are presented and discussed in Section 4. Finally, Section 5 presents the conclusions of the study.
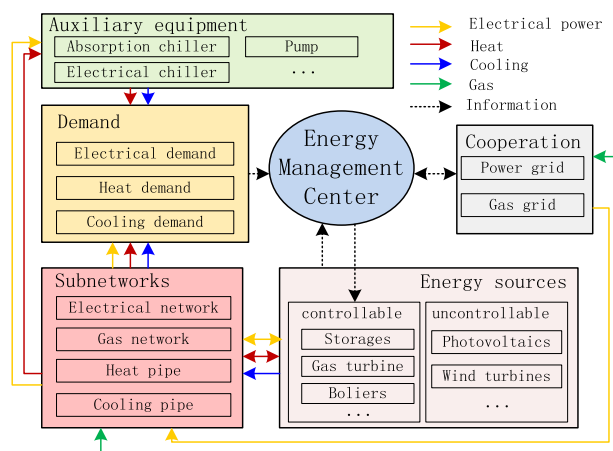

FIGURE 1. Energy management and transfer of an IES.

II. STRUCTURE OF THE MULTIAREA IES
The structure of the IES which composed of electrical, gas, and thermal subnetworks and an energy management center (EMC) is described in Fig. 1. The subnetworks in the IES transfer the energy to meet the different types of loads. In the operation mode, the distributed energy sources, storages and other auxiliary equipment are mapped with subnetworks. When the IES is working in the grid connected mode, it can trade electricity with the power grid. The EMC can monitor and control the controllable energy resources with optimization policy to achieve the economic and stable operation of the IES.
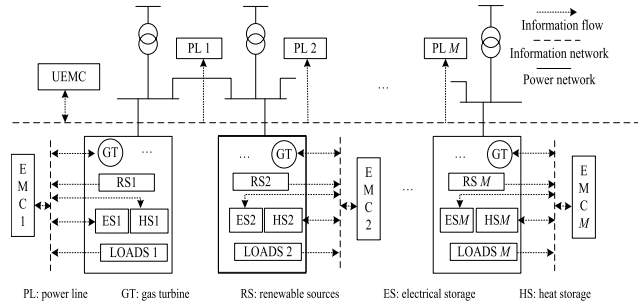
FIGURE 2. Energy management for the multiarea IES with power interaction.

In this paper, the multiarea IES studied is consisted with multiple IECs in different areas, as shown in Fig. 2. The UEMC is the energy management center for power dispatch between areas. Electricity in each area can be traded not only with the power grid, but also with other areas. The energy coupling relationship between subnetworks in each area is complicated. For example, the cooling load can be met with both electricity and heat and the gas turbines can generate available heat while produce electricity by consuming gas. And some auxiliary equipment consumes both heat and electrical power for working. Besides, the stochastic and dynamic characteristics of distribution resources and loads can directly influence the economic operation and power transactions for areas.

Consider the joint dispatching optimization problem of multiple IES in Fig. 2, three types of loads in area m at time t are taken into account, i.e., electricity load $P_{ele,m}(t)$, heat load $P_{h,m}(t)$ and cooling load $P_{cool,m}(t)$. Note that the cooling load in each area can be met with both electricity and heat, as described in (1), where $P_{cool,m}^{ele}(t)$ represents the cooling load satisfied by electricity and $P_{cool,m}^{th}(t)$ represents that satisfied by heat.

$$P_{cool,m}(t) = P_{cool,m}^{ele}(t) + P_{cool,m}^{th}(t) \quad (1)$$

The real-time energy supply and demand balance in each area are considered as shown below:

$$P_{th,m}(t) + P_{cool,m}^{th}(t)/cop_{thc,m}$$
$$= cop_{gt,m}(t)P_{gt,m}(t)\eta_{th,ele} + P_{ths,m}(t) \quad (2)$$
$$P_m(t) + P_{m,grid}(t) - \sum_{m,i\in\{1,...,M\},m\neq i} P_{mi}(t) + P_{es,m}(t)$$
$$= P_{ele,m}(t) + P_{cool,m}^{ele}(t)/cop_{elec,m} \quad (3)$$

Here $\eta_{th,ele}$ is the thermoelectric ratio of the gas turbine, and $P_{mi}(t)$ is the power translated from area m to area i, where $P_{mi}(t) = -P_{im}(t)$. As discussed in Section 1, the simulation model described in this section is used to simulate the data for leaning, which can be replaced with practical operation data. Since each area contains different types of equipment in following description, the subscript of area m is omitted for simplification.

A. PHOTOVOLTAIC POWER GENERATION (PV)

According to [41], [42], the output power of PV can be simulated with statistic power and stochastic perturbation. Define $P_{pv}^{\max}(t)$ as simulated value of the statistic maximum output of PV under clear condition (without clouds) at time t. In practical, it can be determined by its capacity, location, the temperature, the illumination of the sun and changes regularly over time in a given day. Thus, $P_{pv}^{\max}(t)$ can be calculated by (4) in theoretical, in which f_{pv}, G_{pv}, α_{pv} and T_{pv} represent the power derating factor, illuminance, temperature coefficient, and panel temperature of the PV module, respectively.

$$P_{pv}^{\max}(t) = f_{pv}P_{pv,cap}(G_{pv}^{\max}(t)/G_{pv,stc})$$
$$\qquad * (1 + \alpha_{pv}(T_{pv}^{\max}(t) - T_{pv,stc})) \quad (4)$$

Here, we consult the output power data in a typical clear day in summer in Anhui, China to achieve the maximum output curve of $P_{pv}^{\max}(t)$ in Section 4.

In this way, $P_{pv}(t)$ can be achieved by $P_{pv}^{\max}(t)$ and random reduce caused by the environment, such as clouds or temperature changing. We consider the randomness with the reduction rate $\Delta_{pv}(t) \in [0, 1]$ as a Gaussian distribution, and the $P_{pv}(t)$ is described as (5):

$$P_{pv}(t) = (1 - \Delta_{pv}(t))P_{pv\_\exp}^{\max}(t) \quad (5)$$

B. GAS TURBINE

The gas turbine can supply electricity and heat at the same time by consuming gas. The wasted heat produced during the power generation process can be reused with heat recovery equipment and absorption chiller is equipped to produce chilled water with heat for cooling load. The maximum and minimum power constraints are to ensure the operational stability and efficiency, and the power output must be in keeping with the ramping constraints, which are shown in (6).

$$\underline{P}_{gt} \leq P_{gt}(t) \leq \bar{P}_{gt} \quad (6)$$
$$\begin{cases} P_{gt}(t) - P_{gt}(t - \Delta t) \leq P_{gt}^{up}, P_{gt}(t) > P_{gt}(t - \Delta t) \\ P_{gt}(t - \Delta t) - P_{gt}(t) \leq P_{gt}^{down}, P_{gt}(t) < P_{gt}(t - \Delta t) \end{cases} \quad (7)$$

The operating cost of the gas turbine from time t to $t + \Delta t$ can be described with the quadratic function.

$$c_{gt}(t) = \int_t^{t+\Delta t} a_{gt2}P_{gt}^2(t) + a_{gt1}P_{gt}(t) + a_{gt0}dt \quad (8)$$

C. ENERGY STORAGES

The electrical storage and heat storage are applied in each area to deal with the randomness of the load and sources. The simulation of the electrical storage refers to the dynamic characteristics of the vanadium redox flow battery (VRB) in this paper. The charging and discharging power of the VRB is limited in the range of $[P_{es}^{\min}, P_{es}^{\max}]$. The charging /discharging power constraint is described as follow:

$$P_{es}^{min} \leq |P_{es}(t)| \leq P_{es}^{max} \quad (9)$$

in which $P_{es}(t) < 0$ represents that the electrical storage is discharging with power $|P_{es}(t)|$. The state of the electrical storage is defined as SOC_{es} and it represents the percentage of the remaining energy of the electrical storage. The dynamic charging/discharging process is described as below:

$$SOC_{es}(t + \Delta t) = SOC_{es}(t) + \int_{t}^{t+\Delta t} P_{es}(t)/E_{es}dt \quad (10)$$

In the practical operation, SOC_{es} has influence on the operation performance of the electrical storage. In order to avoid the damage caused by the charging/discharging power and improve the efficiency and service time, a two-stage charge/discharge rule is adopted to charge/discharge for the VRB [43]. And the SOC_{es} is used as the condition to switch the charging/discharging mode. It is discharged/charged within the maximum and minimum power constraint when $SOC \in (0.2, 0.8)$; otherwise, it is discharged/charged with constant current I_{const}. As the power is determined with terminal current and voltage, we approximate the discharging/charging process with a constant minimum power P_{es}^{const} in the simulation as the terminal voltage U_{es} of the VRB is almost unchanged in this range.

$$P_{es}^{const} = I_{const} * U_{es} \quad (11)$$

Thus, the discharging/charging power with constant power can be approximated as a minimum value, which is defined in (10):

$$P_{es}(t) = P_{es}^{const}, \quad SOC_{es}(t) \notin (0.2, 0.8) \quad (12)$$

The power loss caused by the resistance in practical is considered with efficiency cop_{es}, and the loss cost $c_{es}^{loss}(t)$ is adopted to evaluate the performance of the electrical storage, which is defined as below, in which β_{es} is the cost coefficient for the electrical storage.

$$c_{es}^{loss}(t) = \begin{cases} \int_{t}^{t+\Delta t} -\beta_{es}P_{es}(t)(1-cop_{es})dt, \text{ charge} \\ \int_{t}^{t+\Delta t} \beta_{es}P_{es}(t)(1-cop_{es})/cop_{es}dt, \text{ discharge} \end{cases} \quad (13)$$

The state of the heat storage is defined as SOC_{ths} and it represents the percentage of the remaining heat of the storage. The dynamic charge/discharge characteristic of the heat storage is described as:

$$SOC_{ths}(t + \Delta t) = SOC_{ths}(t) + \int_{t}^{t+\Delta t} P_{ths}(t)/E_{ths}dt \quad (14)$$

where $SOC_{ths}(t) = e_{ths}(t)/E_{ths}$, $0 < SOC_{ths}(t) \leq 1$ and $e_{ths}(t)$ represents the remaining energy in storages and E represents the capacity for storages. P_{ths} represents the charge/discharge power of the heat storage. $P_{ths}(t) > 0$ means the heat storage is charging and the constraint that $P_{ths}^{min} \leq |P_{ths}(t)| \leq P_{ths}^{max}$. And the cost of the heat storage is defined as below:

$$c_{ths}^{loss}(t) = \begin{cases} \int_{t}^{t+\Delta t} -\beta_{ths}P_{ths}(t)(1 - cop_{ths})dt, \text{ charge} \\ \int_{t}^{t+\Delta t} \beta_{ths}P_{ths}(t)(1 - cop_{ths})/cop_{ths}dt, dis \end{cases} \quad (15)$$

D. ELECTRICAL DISTRIBUTION NETWORK

Each area can exchange electricity with the other areas and the power grid in this paper. The power exchanging between areas is realized through inter-area power lines and must meet the operation constraints of lines. Denote $P_{grid,m}(t)$ and $P_{im}(t)$ as the power exchanged between area m with the power grid and area j respectively. The power exchanged between areas should meet the constraints as follow: First, the exchanging power $P_{im}(t)$ from area i to area m need to be within the allowable range of the power line, which is described as $P_{im}^{min} \leq |P_{im}(t)| \leq P_{im}^{max}$. Second, area m is not allowed to purchase power from one area and sell power to other areas at same time. This is to avoid unnecessary electricity loss between areas. Last, (16) are adopted as the constraint for power charging between areas:

$$|P_{im}(t) - P_{im}(t - \Delta t)| \leq P_{im}^{c} \quad (16)$$

Then the total cost for the electricity trading of area m can be expressed as Eq. (17) shows. If $P_{grid,m}(t)$ or $P_{im}(t)$ is negative, it means that area m sells power to the power grid or area i

$$c_m(t) = P_{grid,m}(t)c_{grid,m}(t) + P_{im}(t)c_{mi}(t) \quad (17)$$

Here, $c_{grid,m}(t)$ and $c_{im}(t)$ denote the prices that area m charges the public grid and area i, respectively. When $c_m(t) < 0$, area m earns a profit from the public grid or other areas.

E. LOADS

There are three types of loads in each IES: electrical load, heat load and cooling load. The electrical load not only includes user's demand, but also the energy conversion equipment and other devices. The heating load includes the heating ventilating in winter and hot water demands, while the cooling load is the air conditioning demand in hot days. The cooling load can be met with electricity by electrical chiller or with heat by absorption chiller. The performance of the absorption chiller and electrical chiller can be expressed using the linear functions.

$$0 \leq P_{ac,out}(t) = cop_{ac}P_{ac,in}(t) \leq P_{ac,out}^{max} \quad (18)$$
$$0 \leq P_{ec,out}(t) = cop_{ec}P_{ec,in}(t) \leq P_{ec,out}^{max} \quad (19)$$

Assume that the statistical value of electrical load at time t is $P_{ele,exp}(t)$. The value of $P_{ele}(t)$ at time t can be expressed as shown below, where $\Delta_{ele}(t) \in [-\underline{\Delta}_{ele}, \overline{\Delta}_{ele}]$ represents the random coefficient of the electrical load and the Gaussian randomness is adopt to describe the stochastic distribution of $\Delta_{ele}(t)$. $P_{th}(t)$ and $P_{cool}(t)$ can be achieved similar to $P_{ele}(t)$.

$$P_{ele}(t) = (1 + \Delta_{ele}(t))P_{ele,exp}(t) \quad (20)$$

This study set a multiarea IES with the following three areas ($m = 1, 2, 3$) for simulation: a residential area, a commercial area and an industrial area. As the optimization process considers both heat and power balance in each area, the energy imbalance will not happen with all above constraints.

III. ENERGY DISPATCH OPTIMIZATION WITH HIERARCHICAL LEARNING METHOD

The optimization of multiarea IES include the dispatch policy of power and heat in each area and the coordinated power dispatch of power between areas. The coordinated power dispatch policy between areas is not only related to the operation cost of each area, but also to the overall optimization performance. Without the optimization of coordination policy, each area wants an interarea power dispatch policy that benefit to itself, even it may increase the overall operation cost of the multiarea system. Thus, we consider the energy dispatch optimization problem of the multiarea IES as a hierarchical stochastic dynamic programming problem with an EMC at upper layer (UEMC) and EMCs at the lower layer. The EMCs at lower layer decide the energy dispatch policy for electricity and heat in each area and UEMC at upper layer decides the coordinated power dispatch policy between areas.

A. FORMULATION OF THE ENERGY DISPATCH OPTIMIZATION PROBLEM

The decision cycle between the upper and lower layers are different, and it is more meticulous at lower layer. We discretize a day into K_{up} for upper layer and K_m for lower layer, where $K_m = K_{up} * L$, $L \in N^+$. The dynamic decision process and information exchange between layers are shown in Fig. 3.

(a)   Decision epoch and state translation

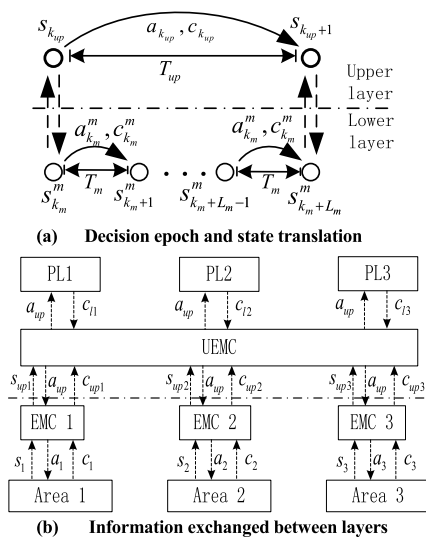**(b)   Information exchanged between layers**

FIGURE 3. Dynamic decision process of multiarea IES.

For the UEMC, the precise state information of each area will cause the dimension curse of the optimization. Hence, the comprehensive information of each area is summarized to the UEMC at the beginning of decision epoch. Define the state at the upper layer as $s_k^{up} = (k_{up}, s_1, s_2, s_3) \in \Phi_{up}$, where s_m is defined as the comprehensive state of loads and PV integration information $p_{up,m}(k_{up})$. $p_{up,m}(k_{up})$ is used to approximately estimate the average total electrical demand of area m at k_{up}th decision epoch, and we use a linear function

to describe $p_{up,m}(k_{up})$ as below.

$$p_{up,m}(k_{up}) = p_{pv,m}(k_{up}) - \eta_m(p_{ele,m}(k_{up})$$
$$+ \lambda_{m,cool} p_{cool,m}(k_{up})/cop_{cool,m}^{ele}) \quad (21)$$

where $\lambda_{m,cool} = 0.5$, $cop_{cool,m}^{ele}$ is the efficiency of electrical chiller in area m and η_m is the discount factor for decreasing estimation error. $a_{up} = (a_{up,12}, a_{up,13}, a_{up,23}) \in D_{up}$ represents the transferred power level between the areas, and $a_{up,ij} > 0$ means the power is transferred from area i to j. The information related to m in a_{up} will be integrated in EMC of area m and taken as a part of state information. $s_{k_m}^m = (k_m, n_{pv,m}, n_{gt,m}, n_{es,m}, n_{ths,m}, n_{ele,m}, n_{heat,m}, n_{cool,m}, n_{pl,m}) \in \Phi_m$ is the state for EMC m and Φ_m is the state set $n_{pl,m}$ is the state of power transaction of area m achieved with a_{up}.

$$n_{pl,m} = \sum_{j=m+1}^{M} a_{up,mj} + \sum_{i=1}^{m-1} a_{up,im} \quad (22)$$

In Fig. 3 (b), action $a_m = (a_{gt,m}, a_{es,m}, a_{ths,m}) \in D_m$ at the lower layer is for controlling the gas turbines and storages. Thus, the operation cost for area m during a decision epoch includes the cost of the gas turbines and storages and exchanging power with the grid and other areas.

$$c_m(k_m) = c_{m,gt}(k_m) + c_{m,es}(k_m) + c_{m,ths}(k_m)$$
$$+ c_{m,grid}(k_m) + \sum_{j>m} c_{m,j}(k_m) + \sum_{i<m} c_{i,m}(k_m) \quad (23)$$

The power dispatch policy at upper layer influence the economic operation of all areas. As the UEMC concerns about the overall operation cost, and the power interaction cost between areas are not included. Thus, the cost for the upper layer in one decision epoch as below:

$$c_{up}(k_{up}) = \sum_{m=1}^{3} \sum_{k_m=k_{up}*L_m}^{(k_{up}+1)*L_m} c_m^{gt}(k_m)$$
$$+ c_m^{es}(k_m) + c_m^{ths}(k_m) + c_m^{grid}(k_m) \quad (24)$$

In this study, the dynamic energy dispatch optimization problem is a decision-making frame-work and modeled as a discrete finite horizon dynamic programming problem. For the EMC at lower layer, the optimal goal is to decrease the daily operation cost for its own area with UEMC decision, which described as below.

$$\pi_{low,m}^* = \arg \min_{\pi_{low,m}} \sum_{k_m=1}^{K_m} c_m^{\pi_{low,m}}(k_m|\pi_{up}) * \gamma^{k_m-1} \quad (25)$$

Because the cost at the same state with same action can be different for the randomness, the discount factor $\gamma \in [0, 1]$ is adopt to decrease the influence for a better converge, and we set $\gamma = 0.95$ in this paper. Similar to the lower layer, the optimal policy for UEMC is shown below.

$$\pi_{up}^* = \arg \min_{\pi_{up}} \sum_{k_{up}=1}^{K_{up}} c_{up}^{\pi_{up}}(k_{up}) * \gamma^{k_{up}-1} \quad (26)$$

TABLE 1. Calculation scale of hierarchical and nonhierarchical learning method.

Method	State Num.	Action Num.	Policy size
Hierarchical	$K_{up}N_{up} + K_{low}\sum_{m=1}^{M}N_m$	$A_{up} + \sum_{m=1}^{M}A_m$	$N_{up}^s N_{up}^a + \sum_{m=1}^{M}N_{low,m}^a$
Nonhierarchical	$K\prod_{m=1}^{M}N_m$	$A_{up}\prod_{m=1}^{3}A_m$	$(K\prod_{m=1}^{M}N_m)(A_{up}\prod_{m=1}^{3}A_m)$

B. HIERARCHICAL LEARNING METHOD WITH SIMULATED ANNEALING-BASED ALGORITHM

The hierarchical learning algorithm proposed in this paper is based on the reinforcement learning with hierarchical framework. The convergence of the reinforcement learning is based on the theory of the Bellman equation, and for the discrete finite horizon dynamic programming problem, there exists a policy that achieves the optimal goal for all the initial states [44]. State-action value function $Q(s, a)$ is adopted in the hierarchical model-free learning algorithm proposed to evaluate the chosen actions for states. The basic idea behind hierarchical reinforcement learning is to decompose the overall task into subtasks, whose solutions can be learned more tractably to deal with the "curse of dimensionality" and the low convergence speed in reinforcement learning [45]. In [46], it proves that for the hierarchical reinforcement learning with bounded rewards, the state-action value will converge to the optimal state action value with probability 1. And the convergence of the hierarchical reinforcement learning depends less on the environmental changing if the update process of the policy is restricted in the its space [47]. Thus, based on the theorical analysis, the policy will eventually converge or approximately converge to the optimal or suboptimal solution with simulation.

There is an iterative process of updating $Q(s, a)$ by exploring the environment. The exploration process is converged by simulated annealing (SA) algorithm. The agent will select action a_{min} by a greedy policy based on the current $Q(s, a)$ and action a_{rand} with randomness. If (27) holds, a_{min} is chosen as the action a_k for the state; otherwise, we choose a_{rand}.

$$e^{(Q(s,a_{greedy}) - Q(s,a_{rand}))/K_b * temp} < random\,(0, 1) \quad (27)$$

where K_b is the Boltzmann constant, and *temp* is the simulated annealing temperature, which is updated every $N_{learning}$ learning steps during the optimization process as $temp = \gamma_t * temp$.

As shown in Fig. 3 (a), the subsequent state s_{k+1} and immediate cost c_k for the episode k^{th} of the system is achieved after a decision epoch. The data for learning can be achieved from this processing. Note the data as (s_k, a_k, c_k, s_{k+1}), and the $Q(s_k, a_k)$ updates according to (23) with the learning factor α if s_{k+1} isnottheendstate.Otherwise,updates$Q(s_k, a_k)$

with (29).

$$Q(s_k, a_k) = Q(s_k, a_k) + \alpha(c_k - Q(s_k, a_k) + \min_a Q(s_{k+1}, a)) \quad (28)$$

$$Q(s_k, a_k) = Q(s_k, a_k) + \alpha(c_k - Q(s_k, a_k)) \quad (29)$$

The flowchart in Fig. 4 shows the hierarchical learning applied to the energy dispatch optimization problem of the multiarea IES. The EMCs at the lower layer learn the optimized policy parallelly based on the policy of the upper layer policy, and send the cost and state information to the UEMC every L_m steps. With the information exchanging, which is described in Fig. 3 (b) in detail, not only the overall optimization performance of the multiarea IES, but the economic operation of each area can also be guaranteed.

Besides, the hierarchical learning method greatly reduces the memory space and calculation time as shown in Table 1. Theoretically, the learning process must visit all state-action pair multiple times to obtain enough experience for optimization, and the learning time is proportional problem scale. Define the total number of state and action for area are N_m and A_m, and the total number of state and action for UEMC are N_{up} and A_{up}. For the nonhierarchical learning method, the state in all three areas should be taken into account and the policy is used to control both the power interaction between areas and the energy dispatch in each area. First, the space complexity is proportional to the policy size. It's easy to tell that the storage space for nonhierarchical learning method is much huger than that of hierarchical learning from Table 1. Second, the time complexity for one step learning is $o(A)$, which means the nonhierarchical learning consume longer time for one learning step. However, it also needs much more learning steps to converge for the scale of policy space. Besides, the EMC in lower layer of the hierarchical learning method can learning parallelly, which would further decrease the calculation time.

IV. SIMULATION RESULTS

The multiarea IES studied includes three areas with different types load characteristics: a residential area, a commercial area and an industrial area. And the simulation of the loads and PV in a sunny working day in summer are shown in Fig. 5 [48], [49]. The PV is equipped on the roof in residential and industrial areas. The profiles of the PV in these
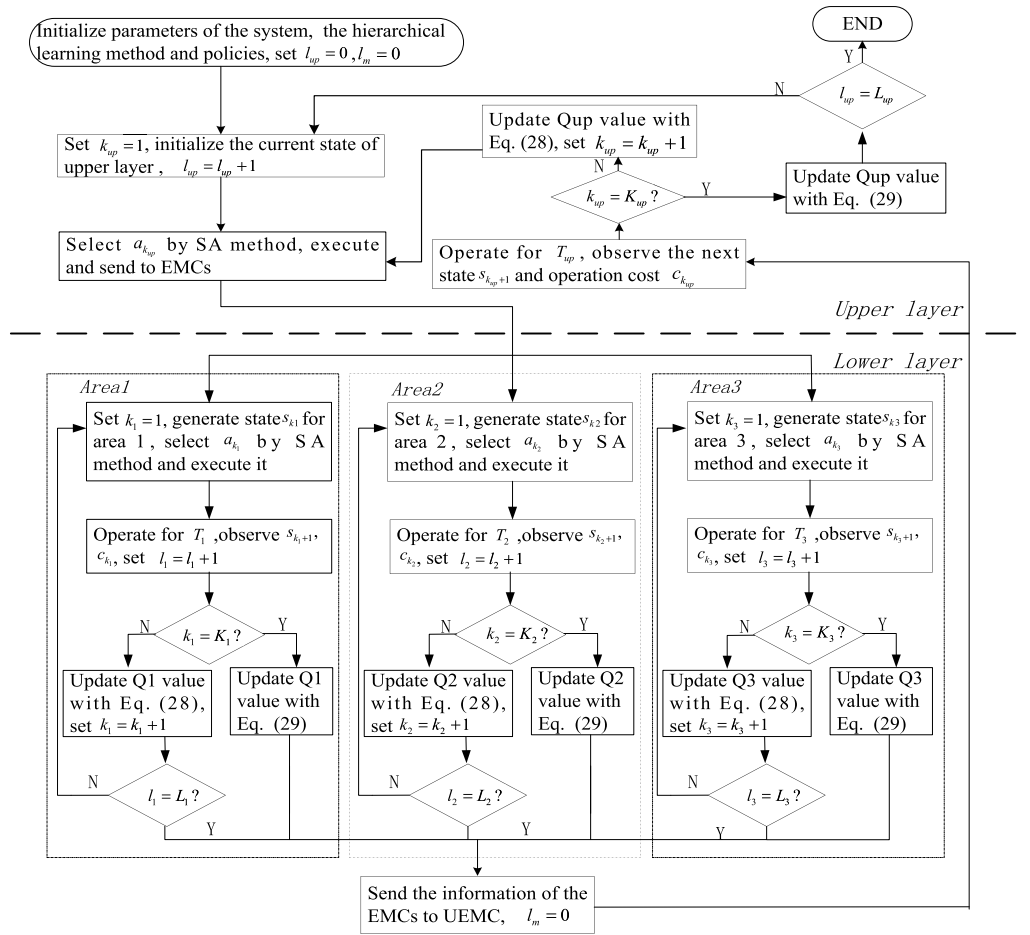
FIGURE 4. Flowchart for hierarchical learning process for optimization of multiarea IES.
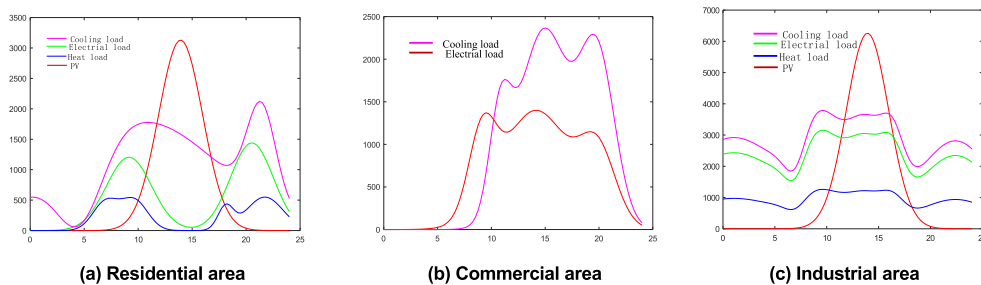

FIGURE 5. Statistic curves for PV and multi-type loads in various areas.

two areas have a basically consistent trend over time except for capacity difference as they are located in the same city. As Figure 5 shows, cooling load accounts for a large proportion in each area in summer, while the heat load varies greatly in different areas. The heat load in the resident area is only for hot water in daily life and large heat demand is needed in the production process in industrial area. In this paper, the randomness of the load is described with the Gaussian distribution within the range [−0.3, 0.5].

The time-of-use electricity prices for residential, commercial, and industrial areas in Anhui, China are listed in

Table 2 and parameters setting are described in Table 3. In this section, we described three groups of analysis to present the effect of the hierarchical learning method. First, the optimization process of hierarchical learning is shown with convergence and effectiveness to decrease the operation costs of the overall system and each area, and the stability of policy is also improved with optimization. Secondly, in order to explain the effectiveness of the optimal policy on the system, we analyze the performance of the optimal strategy obtained in the random environment of the multiarea IES. Finally, the comparison with the optimization of non-interconnected

TABLE 2. Time-of-use electricity prices in different areas.

Residential area	Time interval	8:00-22:00	23:00-8:00(next day)	
	Electrical price/RMB	0.593	0.3153	
	Time interval	9:00-12:00 17:00-22:00	8:00-9:00,12:00-17:00 22:00-23:00	0:00-8:00 23:00-24:00
Commercial area	Electrical price/RMB	1.2892	0.8084	0.5051
Industrial area	Electrical price/RMB	1.0261	0.6474	0.4085

TABLE 3. Parameter settings.

Parameter	Area	Setting
Gas turbine capacity*number, thermoelectric ratio	1	600kW*3, 1.1
	2	600kW*3, 1.1
	3	1000kW*3, 1.3
Electricity storage capacity and efficiency	1	1000kWh, 0.9
	2	1000kWh, 0.9
	3	2000kWh, 0.85
Heat storage capacity and efficiency	1	1000kWh, 0.9
	2	1000kWh, 0.9
	3	2000kWh, 0.85
K_{up}, K_m	\	24, 96
Initial value of *Temp*	\	600
γ, K_b, α	\	0.95, 0.3, 0.5

areas validates the role of interconnection in improving system operation economics.

In the learning process, we evaluate the optimal policy during the learning processing every 50000 learning steps. The optimal policy is obtained by the state-action value, which consist of the action corresponding to the minimum value for each state. Due to the randomness of PV and loads, the policy is evaluated with the average operation cost in 1000 days, as shown in Fig. 6.
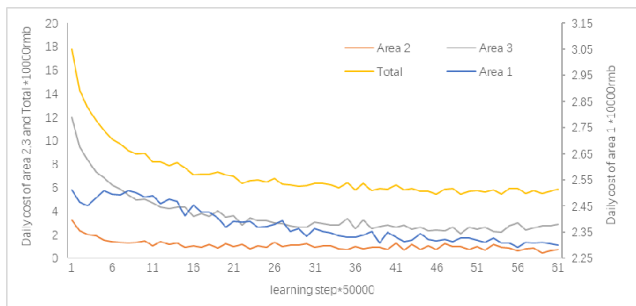

FIGURE 6. Optimization process of multiarea IES.

The proposed algorithm is trained for 60*50000 epochs for the optimal energy dispatching policy with state-action value iterations. It can be observed that due to the initial perception of the stochastic environment, the UEMC and EMCs have no experience of choosing a reliable action to achieve a low cost. However, with the continuous interaction with the environment, the operation cost of the overall and each area

decrease and converge to the minimum values. It illustrates that the optimum policy to minimize the operating cost of the multiarea IES and each area has obtained with the hierarchical learning method proposed.
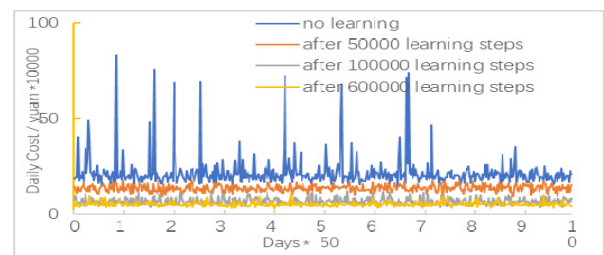

FIGURE 7. Operation cost under different policies.

In order to evaluate the effect of the hierarchical learning method on the stable operation of the system, we test and compare the performance of the policies in the random environment, as shown in Fig. 7. We test the optimum policies after learned for 0, 250000, 500000 and 3000000 (converged) steps respectively. The polices are tested in the random environment for 500 days, and the curves in Fig. 7 show that with the learning process, not only the system operation cost is decreased, but also the stability and robustness is improved. Eventually, the optimal policy can make the system run stably at a lower operating cost in the random environment.

The specific data analysis for the comparison of policies in Fig, 7 is shown in Table 4. It can be seen that compared with operation by the random policy, the daily running cost of the system is reduced by about 68.3% after optimization with hierarchical learning method. At the same time, the stability of the overall operation of the system is greatly improved. In terms of variance, the fluctuation of the operating cost in a random environment has been reduced by 80.26%.

TABLE 4. Average and variance of daily cost of different policies.

Optimized steps	Average cost/RMB	Variance
0	177518.28	39165.20
250000	136976.72	15076.42
500000	69351.25	12971.65
3000000	56279.32	7727.60

Table 5 shows the percentage of cooling load met by the absorption chiller with heat. The cooling load in summer is

the mainly energy consumption, which can be satisfied by both electricity and power. The huge demand for electricity also brings pressure for the power grid and may cause the problem of stability for power supply. The optimization for distribution of cooling load can effectively decrease the demand of electricity for the multiarea IES and its operation cost a lot. As Table 5 shows that the percentages of cooling load satisfied with heat are all increased in three areas, especially in area 2. It can indicate that the optimization makes each area benefit from energy dispatch and improve the efficiency of usage of the wasted heat from gas turbines, which is the main heat source in areas.

TABLE 5. Percentage of cooling load satisfied with heat.

Areas	Area 1	Area 2	Area 3
Before optimization	23%	3%	12%
After optimization	26%	29%	14%

(a) power between area 1 and grid

(b) power between area 2 and grid

(c) power between area 3 and grid

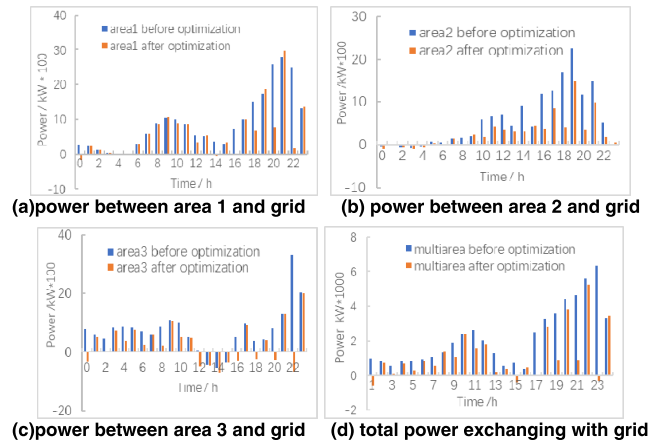(d) total power exchanging with grid

FIGURE 8. Power exchanging between areas and the public grid.

Since each area operates in the grid-connected mode, the optimization can also have the influence on the power grid. Excessive peak-valley difference of load will affect the stability and economy of the power grid. At the same time, the IES should achieve self-sufficiency in energy as far as possible while ensuring that load demands are met. We observe the peak-valley difference and power purchasing from the power grid before and after optimization, which are the important indicators for the grid. In general, decreasing of the power purchased from the power grid and the peak-valley difference is benefit for the economic and stable operation. As can be seen in Fig. 8, the peak-valley difference in each area is reduced effectively.

We can see form Table 6 that the overall peak-valley difference is reduced by 1.8% from 5966.2kW to 5856.6 kW with optimization. The amount of electricity purchased from the grid in each area has obviously decreased, and the total amount of electricity purchased has decreased from 52609.0 kWh to 27844.5 kWh by 47.07%. The data

TABLE 6. Peak-valley and power purchasing from grid.

Power from grid/kW	Max	Min	Peak-valley	Total/kWh
Area1 before opt	2795.3	-7.1	2802.4	21157.0
Area1 after opt	2956.9	-179.9	3136.8	14749.9
Area2 before opt	2249.1	-79.3	2328.4	13759.2
Area2 after opt	1497.1	-107.5	1604.6	6757.9
Area3 before opt	3315.4	-561.3	3876.8	17692.6
Area3 after opt	2008.5	-722.0	2730.5	6336.6
Total before opt	6330.3	364.1	5966.2	52609.0
Total after opt	5243.3	-613.3	5856.6	27844.5

in Table. 6 shows that the self-sufficiency rate of each area and the entire system has been improved which can be benefit for the operation of the IES and the power grid both.

TABLE 7. Partial state-action pairs in optimal upper layer policy.

Decision time	1	6	10	14	22
State	(0,0,2)	(2,1,0)	(0,2,2)	(0,2,0)	(0,0,2)
Action	0,-4,-2	0,-3,-2	2,2,0	5,0,-3	0,0,3

The description of the power interaction between areas by optimal policy under some random states are shown in Table 7. The state is the comprehensive discrete level of load and PV of each area, and the action is the interaction power level between area 1 and 2, area 1 and 3, and area 2 and 3 respectively. Take state (0, 0, 2) as example, it means the comprehensive load level in area 1 to 3 are 0, 0 and 2 respectively. And the action (0, −4, −2) means no power transmitted between area 1 and 2, 4*50 kW power is transmitted from area 3 to area 2 and 2*50kW power is transmitted from area 3 to area 1. This is because the load in area 3 is at a peak level at 1:00, while in area 1 and 2 the loads are lower than that in day time. The extra power from area 3 can be support for the load in area 1 and 2 with a higher efficient operation of gas turbine. Power is purchased from area 1 and 3 to area 2 at the state (0,2,0), and the decision time is 14:00 with sufficient PV output.

The power interaction between the areas and the gas turbine output are observed under two different weather conditions as shown in Fig. 9. The average output power of gas turbine and the power interaction between areas are carried out in different cases. The output of PV in case 1 is at high level in all day while case 2 is the opposite. The loads are changing randomly in both cases.

It's obviously to see that the outputs of gas turbines of all areas in case 1 are lower than that in case 2 for the higher output of PV. The difference for area 3 in the two cases is not significant because of the heat demand for area 3 is much larger than others. The ratio of electrical load to heat load keeps the output level of the gas turbine at the high level in area 3. At the same time, the total amount of interaction power between the areas in case 1 is significantly higher than that in case 2, and the electrical power sold from area 1 and area 3 is
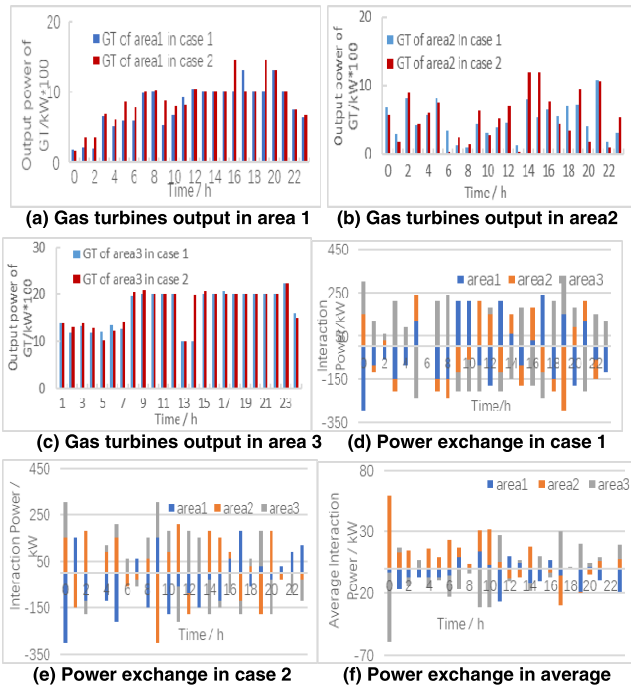
(a) Gas turbines output in area 1

(b) Gas turbines output in area2

(c) Gas turbines output in area 3

(d) Power exchange in case 1

(e) Power exchange in case 2

(f) Power exchange in average

FIGURE 9. Operation of multiarea IES with the optimal policy.

(a) Optimization of area 1

(b) Optimization of area 2

(c) Optimization of area 3

(d) Optimization of multiarea

FIGURE 10. Optimization of multiarea IES without power interaction.

significantly higher in case 1. Fig. 9 (f) shows the statistic power interacted between areas. The average value of power interaction between areas is small, which indicates that the power interaction is mainly for improving the operating efficiency in random environment, rather than to obtain benefits with selling power. Although area 2 sells more power than others, the average interactive power is still much smaller compared to load.

Fig.10 is the optimal profiles of multiarea IES without electrical interaction between areas. Fig. 10 (d) shows the optimization process of all three areas without power interactions between areas by learning algorithm. In order to describe the optimization process in detail, the daily operation cost of each area is shown in Fig. 10(a), (b), and (c) separately. Similar to the hierarchical learning processes, the number of

states and actions are consistent with that at the lower layer in the power interaction mode. Evaluate the optimized policy every 50000 learning steps using the average operation cost for 1000 days.

TABLE 8. Comparison of daily operation costs.

/rmb	Case1	Case2	Case3	Case4
Area 1	25120.63	23093.38	7349.37	6940.72
Area 2	32526.65	9766.75	56324.32	50158.17
Area 3	119870.99	23419.18	49801.27	48945.43
Total	177518.28	56279.32	113474.96	106044.32

It can be seen from Fig. 6 and Fig.10 that the optimization results with power interaction are much better. Table 8 shows the comparison of average daily operation costs for multiarea IES with and without power interaction respectively. Case 1 and 2 represent the daily costs of interaction system before and after optimization respectively, and case 3 and 4 represent that without power interaction. The daily operation cost of the system with power interaction between areas is higher before the optimization for the uneconomic power transmitted between areas. However, it decreases a lot and becomes less than the operation cost for the multiarea IES without power interaction after the optimization. The daily operation cost for multiarea IES with power interaction has reduced 68.3% with optimization, while it only decreases 6.55% without power interaction. In addition, the final daily cost of the multiarea IES with power interaction is 46.9% lower than that without power interaction with the optimization at the upper layer.
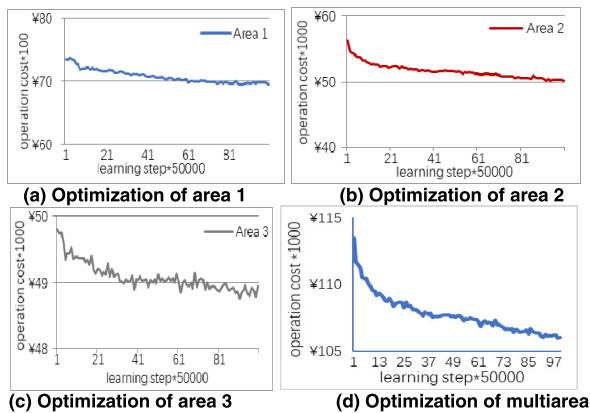
V. CONCLUSION

This study investigated a model-free hierarchical learning method for the dynamic energy dispatch optimization of a multiarea IES considering the randomness of the renewable sources and loads. The hierarchical learning method can effectively decrease the memory space and calculation time. The stochastic and dynamic operation characteristics of the multiarea IES is analyzed and the corresponding simulation model is established to obtain learning data. Compared with the multiple IES without power coordination, it can operate with better economic, flexibility and stability. With the information exchanging and hierarchical learning, both the economic operation of the overall areas and each area are optimized with better performance. The numerical simulation results show that the hierarchical method proposed can effectively improve the economic and stability for the operation of the multiarea IES. Not only the operation cost of the multiarea IES is decreased, but also the peak-valley difference is reduced with the optimization.

In the future, based on the research of the structure of the hierarchical learning method, the competitive relationship between multiple areas and different markets, e.g. gas market and electrical market, will be further considered. Besides, the analysis and optimization of the power flow is also important

in the dispatch of the IES, which is to be further studied. At the same time, advanced artificial intelligence methods should be combined with learning method to further improve the algorithm efficiency and reduce sample requirements.

REFERENCES

[1] J. Wu, J. Yan, and H. Jia, "Integrated Energy Systems," *Appl. Energy*, vol. 167, pp. 155–157, Feb. 2016.

[2] Y. Zhang, P. E. Campana, A. Lundblad, W. Zheng, and J. Yan, "Planning and operation of an integrated energy system in a Swedish building," *Energy Convers. Manage.*, vol. 199, Nov. 2019, Art. no. 111920.

[3] M. Qadrdan, M. Abeysekera, M. Chaudry, J. Wu, and N. Jenkins, "Role of power-to-gas in an integrated gas and electricity system in great Britain," *Int. J. Hydrogen Energy*, vol. 40, no. 17, pp. 5763–5775, May 2015.

[4] A. Zaltash, A. Y. Petrov, D. T. Rizy, S. D. Labinov, E. A. Vineyard, and R. L. Linkous, "Laboratory R&D on integrated energy systems (IES)," *Appl. Thermal Eng.*, vol. 26, pp. 28–35, Jan. 2006.

[5] Y. Liu, H. Li, K. Peng, C. Zhang, H. Hua, and L. Wang, "Demonstration projects of integrated energy system in China," *Energy Procedia*, vol. 145, pp. 88–96, Jul. 2018.

[6] X. Zhu, J. Yang, Y. Liu, C. Liu, B. Miao, and L. Chen, "Optimal scheduling method for a regional integrated energy system considering joint virtual energy storage," *IEEE Access*, vol. 7, pp. 138260–138272, 2019.

[7] C. Mu, T. Ding, Z. Zeng, P. Liu, Y. He, and T. Chen, "Optimal operation model of integrated energy system for industrial plants considering cascade utilisation of heat energy," *IET Renew. Power Gener.*, vol. 14, no. 3, pp. 352–363, Feb. 2020.

[8] P. C. Loh, L. Zhang, and F. Gao, "Compact integrated energy systems for distributed generation," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1492–1502, Apr. 2013.

[9] J. Gao, J. Kang, C. Zhang, and W. Gang, "Energy performance and operation characteristics of distributed energy systems with district cooling systems in subtropical areas under different control strategies," *Energy*, vol. 153, pp. 849–860, Jun. 2018.

[10] H. Wang, W. Yin, E. Abdollahi, R. Lahdelma, and W. Jiao, "Modelling and optimization of CHP based district heating system with renewable energy production and energy storage," *Appl. Energy*, vol. 159, pp. 401–421, Dec. 2015.

[11] N. Liu, J. Wang, and L. Wang, "Hybrid energy sharing for multiple microgrids in an integrated heat–electricity energy system," *IEEE Trans. Sustain. Energy*, vol. 10, no. 3, pp. 1139–1151, Jul. 2019.

[12] Z. Pan, Q. Guo, and H. Sun, "Interactions of district electricity and heating systems considering time-scale characteristics based on quasi-steady multi-energy flow," *Appl. Energy*, vol. 167, pp. 230–243, Apr. 2016.

[13] A. Ahmadi-Khatir, A. J. Conejo, and R. Cherkaoui, "Multi-area energy and reserve dispatch under wind uncertainty and equipment failures," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4373–4383, Nov. 2013.

[14] X. Zhang and T. Yu, "Fast stackelberg equilibrium learning for real-time coordinated energy control of a multi-area integrated energy system," *Appl. Thermal Eng.*, vol. 153, no. 5, pp. 225–241, May 2019.

[15] L. Li, W. Jin, M. Shen, L. Yang, F. Chen, L. Wang, C. Zhu, H. Xie, Y. Li, and T. Zhang, "Coordinated dispatch of integrated energy systems considering the differences of multiple functional areas," *Appl. Sci.*, vol. 9, no. 10, pp. 2103–2118, May 2019.

[16] S. Ma, S. Sun, H. Wu, D. Zhou, H. Zhang, and S. Weng, "Decoupling optimization of integrated energy system based on energy quality character," *Frontiers Energy*, vol. 12, no. 4, pp. 540–549, Dec. 2018.

[17] W. Gu, S. Lu, and J. Wang, "Modeling of the heating network for multidistrict integrated energy system and its operation optimization," *Proc. CSEE*, vol. 37, no. 5, pp. 1305–1315, Mar. 2017.

[18] A. Quelhas, E. Gil, J. D. McCalley, and S. M. Ryan, "A multiperiod generalized network flow model of the US integrated energy system: Part I—Model description," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 829–836, May 2007.

[19] S. Collins, J. P. Deane, K. Poncelet, E. Panos, R. C. Pietzcker, E. Delarue, and B. P. Ó Gallachóir, "Integrating short term variations of the power system into integrated energy system models: A methodological review," *Renew. Sustain. Energy Rev.*, vol. 76, pp. 839–856, Sep. 2017.

[20] G. Li, R. Zhang, T. Jiang, H. Chen, L. Bai, H. Cui, and X. Li, "Optimal dispatch strategy for integrated energy systems with CCHP and wind power," *Appl. Energy*, vol. 192, no. 15, pp. 408–419, Apr. 2017.

[21] P. Arcuri, P. Beraldi, G. Florio, and P. Fragiacomo, "Optimal design of a small size trigeneration plant in civil users: A MINLP (Mixed integer non linear programming Model)," *Energy*, vol. 80, pp. 628–641, Feb. 2015.

[22] Z. Huang, H. Yu, Z. Peng, and Z. Liu, "Two-stage optimization model used for community energy planning," *Energy Procedia*, vol. 75, pp. 2916–2921, Aug. 2015.

[23] C. Elsido, A. Bischi, P. Silva, and E. Martelli, "Two-stage MINLP algorithm for the optimal synthesis and design of networks of CHP units," *Energy*, vol. 121, pp. 403–426, Feb. 2017.

[24] Y. Lin and Z. Bie, "Study on the resilience of the integrated energy system," *Energy Procedia*, vol. 103, pp. 171–176, Dec. 2016.

[25] W. Lin, X. Jin, Y. Mu, H. Jia, X. Xu, X. Yu, and B. Zhao, "A two-stage multi-objective scheduling method for integrated community energy system," *Appl. Energy*, vol. 216, pp. 428–441, Apr. 2018.

[26] Y. Li, J. Wang, D. Zhao, G. Li, and C. Chen, "A two-stage approach for combined heat and power economic emission dispatch: Combining multi-objective optimization with integrated decision making," *Energy*, vol. 162, pp. 237–254, Nov. 2018.

[27] W. Gu, J. Wang, S. Lu, Z. Luo, and C. Wu, "Optimal operation for integrated energy system considering thermal inertia of district heating network and buildings," *Appl. Energy*, vol. 199, pp. 234–246, Aug. 2017.

[28] J. Yu, X. Shen, and H. Sun, "Economic dispatch for regional integrated energy system with district heating network under stochastic demand," *IEEE Access*, vol. 7, pp. 46659–46667, 2019.

[29] S. Zhou, D. He, W. Gu, Z. Wu, G. Abbas, Q. Hong, and C. Booth, "Design and evaluation of operational scheduling approaches for HCNG penetrated integrated energy system," *IEEE Access*, vol. 7, pp. 87792–87807, 2019.

[30] S. Seijo, I. del Campo, J. Echanobe, and J. García-Sedano, "Modeling and multi-objective optimization of a complex CHP process," *Appl. Energy*, vol. 161, pp. 309–319, Jan. 2016.

[31] C. Qin, Q. Yan, and G. He, "Integrated energy systems planning with electricity, heat and gas using particle swarm optimization," *Energy*, vol. 188, Dec. 2019, Art. no. 116044.

[32] Y. Wang, Y. Wang, Y. Huang, H. Yu, R. Du, F. Zhang, F. Zhang, and J. Zhu, "Optimal scheduling of the regional integrated energy system considering economy and environment," *IEEE Trans. Sustain. Energy*, vol. 10, no. 4, pp. 1939–1949, Oct. 2019.

[33] S. Soheyli, M. H. S. Mayam, and M. Mehrjoo, "Modeling a novel CCHP system including solar and wind renewable energy resources and sizing by a CC-MOPSO algorithm," *Appl. Energy*, vol. 184, pp. 375–395, Dec. 2016.

[34] J. Zhang, H. Cho, P. J. Mago, H. Zhang, and F. Yang, "Multi-objective particle swarm optimization (MOPSO) for a distributed energy system integrated with energy storage," *J. Thermal Sci.*, vol. 28, no. 6, pp. 1221–1235, Nov. 2019.

[35] V. Mnih, A. P. Badia, and M. Mirza, "Asynchronous methods for deep reinforcement learning," in *Proc. 33rd Int. Conf. Int. Conf. Mach. Learn. (ICML)*, vol. 48, Jun. 2016, pp. 1928–1937.

[36] S. Mathe, A. Pirinen, and C. Sminchisescu, "Reinforcement learning for visual object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2894–2902.

[37] M. Herman, T. Gindele, and J. J. Wagner, "Inverse reinforcement learning with simultaneous estimation of rewards and dynamics," in *Proc. 19th Int. Conf. Artif. Intell. Statist.*, vol. 51, Apr. 2016, pp. 102–110.

[38] Q.-S. Jia, J.-X. Shen, Z.-B. Xu, and X.-H. Guan, "Simulation-based policy improvement for energy management in commercial office buildings," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 2211–2223, Dec. 2012.

[39] K. Mason and S. Grijalva, "A review of reinforcement learning for autonomous building energy management," *Comput. Electr. Eng.*, vol. 78, pp. 300–312, Sep. 2019.

[40] B. Zhang, W. Hu, D. Cao, Q. Huang, Z. Chen, and F. Blaabjerg, "Deep reinforcement learning–based approach for optimizing energy conversion in integrated electrical and heating system with renewable energy," *Energy Convers. Manage.*, vol. 202, Dec. 2019, Art. no. 112199.

[41] Y. Li and J. Niu, "Forecast of power generation for grid-connected photovoltaic system based on Markov chain," in *Proc. IEEE Power Energy Eng. Conf.*, Mar. 2009, pp. 1729–1733.

[42] H. Liang, A. Kumar Tamang, W. Zhuang, and X. S. Shen, "Stochastic information management in smart grid," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1746–1770, 3rd Quart., 2014.

[43] H. Tang, C. Liu, and M. Yang, "Learning-based optimization of active distribution system dispatch in industrial park considering the peak operation demand of power grid," *Acta Automatica Sinica*, vol. 45, pp. 1–15, Dec. 2019, doi: 10.16383/j.aas.c190079.

[44] C. Ribeiro and C. Szepesvári, "Q-learning combined with spreading: Convergence and results," in *Proc. ISRF-IEE Int. Conf., Intell. Cogn. Syst. (Neural Netw. Symp.)*, 1996, pp. 32–36.

[45] A. Levy, G. Konidaris, R. Platt, and K. Saenko, "Learning multi-level hierarchies with hindsight," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2019, pp. 1–16.

[46] E. G. Schultink, R. Cavallo, and D. C. Parkes, "Economic hierarchical Q-learning," in *Proc. 23rd AAAI Conf. Artif. Intell.*, Jul. 2008, pp. 689–695.
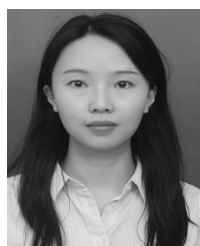
[47] C. Chen, D. Dong, H.-X. Li, and T.-J. Tarn, "Hybrid MDP based integrated hierarchical Q-learning," *Sci. China Inf. Sci.*, vol. 54, no. 11, pp. 2279–2294, Nov. 2011.

[48] S. Hou, "Analysis and application of urban customer load characteristics," South China Univ. Technol., Guangzhou, China, 2015.

[49] Y. Li, H. Tang, K. Lv, X. Guo, and D. Xu, "Modeling and learning-based optimization of the energy dispatch for a combined cooling, heat and power microgrid system with uncertain sources and loads," *Control Theory Appl.*, vol. 35, no. 1, pp. 56–64, Jan. 2018.

YIJIN LI received the B.S. degree from the Hefei University of Technology, Hefei, Anhui, China, in 2012, where she is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Automation. Her research interests include reinforcement learning methods, stochastic dynamic decision and integrated energy systems, power systems, and energy systems.

HAO TANG (Member, IEEE) received the M.S. degree from the Institute of Plasma Physics, Chinese Academy of Sciences, Hefei, China, in 1998, and the Ph.D. degree from the University of Science and Technology of China, Hefei, in 2002. He completed the Postdoctoral work with The University of Tokyo, Tokyo, Japan, from 2005 to 2007. He is currently a Professor with the School of Electrical Engineering and Automation, Hefei University of Technology, Hefei. His research interests include discrete event dynamic systems, stochastic decision, optimization theory, smart grid dispatching control, and deep reinforcement learning and its application in power grid scheduling optimization. He is a Senior Member of the China Computer Federation (CCF).

KAI LV (Graduate Student Member, IEEE) received the B.S. degree from Liaoning University, Liaoning, China, in 2012. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Automation, Hefei University of Technology, Hefei, China. His research interests include artificial intelligence and its application in power systems and energy systems.

KE WANG is currently a Senior Electrical Engineer. Her research interests include demand side response and power grid dispatch optimization.

GANG WANG is currently an Electrical Engineer. His research interests include demand side response and interaction between sources and load.

• • •