

Received August 20, 2020, accepted August 25, 2020, date of publication August 27, 2020, date of current version September 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3019929

Multi-Agent Deep Reinforcement Learning for Sectional AGC Dispatch

JIAWEN LI¹, TAO YU¹, (Member, IEEE), HANXIN ZHU¹,
FUSHENG LI¹, DAN LIN¹, AND ZHUOHUAN LI¹

College of Electric Power, South China University of Technology, Guangzhou 510640, China

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

This work was supported in part by the National Natural Science Foundation of China under Grant 51777078.

ABSTRACT Aiming at the problem of coordinating system economy, security and control performance in secondary frequency regulation of the power grid, a sectional automatic generation control (AGC) dispatch framework is proposed. The dispatch of AGC is classified as three sections with the sectional dispatch method. Besides, a hierarchical multi-agent deep deterministic policy gradient (HMA-DDPG) algorithm is proposed for the framework in this paper. This algorithm, considering economy and security of the system in AGC dispatch, can ensure the control performance of AGC. Furthermore, through simulation, the control effect of the sectional dispatch method and several AGC dispatch methods on the Guangdong province power grid system and the IEEE 39 bus system is compared. The result shows that the best effect can be achieved with the sectional dispatch method.

INDEX TERMS Automatic generation control, hierarchical multi-agent deep deterministic policy gradient, sectional AGC dispatch, reinforcement learning.

I. INTRODUCTION

Automatic generation control (AGC) is an important operation task of interconnected power grid, which can maintain the system frequency and tie-line exchange power to the expected values [1]. The current regular AGC, if still adopting methods such as engineering actual experience or simple generation unit capacity and regulation speed, so that fixed dispatch can dispatch the total AGC generation power commands of the system to each generation unit, it will not satisfy the control performance standard (CPS) appraisal in control area with a high renewable energy penetration rate and insufficient regulated sources. Moreover, the short-term random fluctuation and corresponding system power deficiency are relatively small in the interconnected power grid with traditional hydropower power and coal-fired power as the main power supply, the small power regulation variation of the AGC unit has little effect to the security and generation cost of the power grid. For these reasons, the regular AGC dispatch methods ignore the effect on security constraint and generation cost. However, the connection among various new energy and the development of Ultra High Voltage (UHV) technology have greatly made the increase in uncertainty

power disturbance inevitable. Especially when mono-polar blocking fault of one or even multiple DC transmission lines occurs, the huge load disturbance will reduce the system frequency significantly. Correspondingly, the AGC unit regulation output, to meet the requirements of the frequency's stability, will increase significantly as well. As a result, line overload may occur, thus seriously affecting the system frequency stability and security. Therefore, in an interconnected power grid with large-scale new energy and DC transmission lines, its AGC dispatch method should consider not only control performance but also the effect on power grid security and unit generation costs [2].

Some experts and scholars have already proposed some improvement measures for the deficiency of the regular AGC dispatch methods [2]–[5]. In detail, a model is proposed in [2] that takes into generation cost and the regulation cost in the AGC as well as the system security constraint, however, ignoring the control performance optimization during the AGC dispatch, thereby resulting in CPS index deterioration and poor control performance. Besides, the AGC dispatch coefficient in the AGC system is optimized with the hierarchical Q-learning method in [3]. However, hierarchical Q-learning needs to discretize its state and action and cannot regulate the AGC generation power command continuously and smoothly. Even more, the algorithm ignores the

The associate editor coordinating the review of this manuscript and approving it for publication was Huai-Zhi Wang.

effect on system security and operation economy in AGC dispatch. In [4], a dynamic optimization dispatch strategy, instead of the original regular AGC dispatch, is adopted to create a new AGC dynamic optimization dispatch model [4]. It considers power balance constraint but ignores the system security constraint and generation cost. It is difficult to ensure the regulation control performance in the case of sudden large-scale load disturbance due to the too long AGC dispatch interval. Reference [5] proposes a predictive control method of the interconnected power grid model that takes participation factors into account for the interconnected power system with multiple units in a single area. This method links the two regulation and control frameworks with different time scales, namely, economical scheduling and AGC, considering the system power balance constraint [5], however, ignoring the system security power flow constraint and control performance.

To sum it up, the current AGC dispatch method has the following deficiencies:

Foremost, AGC control performance, system operation economy and the security constraint are not considered simultaneously. In the electricity power market environment, the power grid scheduling center tries to ensure operation economy, causing the power grid operating within the boundary of the system security constraint range. At this time, adjusting the AGC unit output may cause the power grid to operate beyond the security constraint range. Besides, it is impossible to realize AGC control performance, system operation economy and the network security constraint at the same time. The preference must be determined based on needs.

In response of the above-mentioned problem, a sectional AGC dispatch model that considers AGC control performance, the regulation cost and system economy and security simultaneously is proposed in this paper. Under the guidance of the CPS appraisal rule in China [1], the AGC dispatch problem is classified as three sections according to different CPS1 indices. This ensures the fast recovery of the system frequency, allowing to quickly restore the system generation cost to the optimal state of the current system, and the system security constraint is considered during AGC dispatch. When finding the solution to this model, the HMA-DDPG is proposed for higher solution-finding efficiency and solution quality.

the main motivations and novelties of this paper are given as follows:

1) The previous studies of AGC dispatch don't consider the coordination of AGC control performance, generation cost and security constraints, especially for collaborative optimization in a power grid that has large-scale of new energy and distributed energy. To fill up this gap, a sectional AGC dispatch is proposed to simultaneously solve the coordination problem, when it dispatches the real-time total power regulation command to all the controllable regulation resources.

2) HMA-DDPG proposed in this paper uses hierarchical framework and Multi-agent optimization, which makes the

strategy more effective, have the advantages of fast convergence, not tending to have a local optimum as well as realizing continuous AGC dispatch for multiple objects.

II. MULTI-AGENT DEEP DETERMINISTIC POLICY GRADIENT (MADDPG)

In 2017, DeepMind proposed the MADDPG in [6], and its basic framework is as follows:

A. BASIC ASSUMPTIONS

Assume that there are N agents in environment E , and each agent has its own strategy. The total strategy set of the N agents is:

$\pi = \{\pi_1, \pi_2, \dots, \pi_N\}$ neural network is used to represent each strategy. Its parameter set $\theta = \{\theta_1, \theta_2, \dots, \theta_N\}$. The environment system satisfies the following assumptions [6]:

Assumption 1: each agent's strategy only depends on the state it observes and has nothing to do with the states that other agents observe, that is, $a_i = \pi_i(o_i)$.

Assumption 2: in an unknown environment, the reward value of each agent and the next state are unpredictable after an action is taken. The reward comes from the feedback from the environment, and its own action only depends on the strategy.

Assumption 3: during training, the agents do not communicate with each other, or the communication content is only a component of their respective observation.

B. MADDPG TRAINING METHOD

The training framework of the MADDPG is shown in Figure 1. All agents in the environment consist of an actor network, a critic network, a target actor network and a target critic network. To facilitate illustration demonstration, agent i is used as an example in Figure 1, and other agents are represented by squares.

The method of decentralized implementation and concentrated training is adopted during training. That is to say, each agent obtains the action to be performed for the current state based on its own strategy: $a_i^j = \pi_i^j(o_i^j)$ and interacts with the environment to obtain the experience sample $(\sigma_i^j, a_i^j, o_i^{j+1}, r_i^j)$, then store it into its own experience buffer pool. After all agents have interacted with the environment, each agent randomly sample experience from its experience pool to train its own neural network. To accelerate the learning process of an agent, the critic network input must include the observed states of other agents and action taken by them, that is, $Q = (s_j, a_1, a_2, \dots, a_n, \theta^Q)$, where $s_j = (o_1^j, o_1^j, \dots, o_N^j)$, the critic network parameter is updated by minimizing the strategy loss. The strategy loss calculation formula as [6]:

$$L = \frac{1}{K} \sum_{j=1}^K \left(y_j - Q \left(s_j, a_1, a_2, \dots, a_N, \theta^Q \right) \right)^2 \quad (1)$$

Afterwards, calculate the parameter for updating the action network through the gradient descent method. The gradient

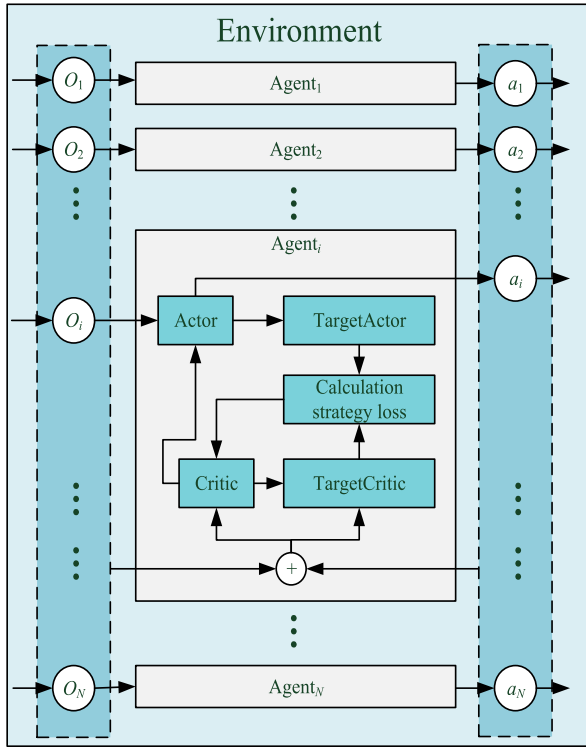


FIGURE 1. Diagram of MADDPG training framework.

calculation formula as [6]:

$$\nabla_{\theta} \pi J = \frac{1}{K} \sum_{j=1}^K \nabla_{\theta} \pi \pi(o, \theta^{\pi}) \nabla_a Q(s, a_1, a_2, \dots, a_N, \theta^Q) \quad (2)$$

C. SECTIONAL AGC DISPATCH ALGORITHM-HMA-DDPG

In the MADDPG, the agents are independent from each other. Based on the assumption 3 that the agents do not communicate with each other or have hierarchical relations, the author found in actual simulation that it is easy for the algorithm to be trapped in a local optimum if it is up to the critic of each agent to obtain extra information (such as the actions of other agents) to directly perform concentrated training without correlation between each agent’s reward function or being able to cover global information. Besides, there will be the problem of difficult convergence by using the traditional MADDPG when there are too many agents, and the strategy estimation method in the traditional MADDPG may result in overestimation [7]: that is, an agent will produce a strong strategy against a competing agent through over-fitting. However, such strong strategy is very fragile, as it is difficult for the strategy to adapt to the opponent’s new strategy with the updating of the opponent’s strategy.

To address the above-mentioned problem, the author proposed HMA-DDPG based on the thinking of hierarchical reinforcement learning, and its framework is shown in Table 1.

TABLE 1. HMA-DDPG process.

<ol style="list-style-type: none"> 1. Divide the original task into Y layers. Number the N agents from 1 to N. The highest layer agent is numbered 1 and the rest can be deduced in this way. The numbering must go from higher layers to lower layers until all agents are numbered.
<p>FOR1 $i=1:N$</p> <ol style="list-style-type: none"> 2. Randomly initialize the critic network parameter θ^Q and the actor network parameter θ^{π} in the strategy network of the ith agent; 3. Initialize the target network parameter of the ith agent; 4. Initialize the experience pool D_i of the ith agent; <p>END FOR1</p>
<p>FOR2 episode = 1:M</p> <ol style="list-style-type: none"> 5. Initialize the environment to obtain initial state of each agent; <p>FOR3 $t=1:T$</p> <p>FOR4 $i=1:N$.</p> <p>Start from the agents of the highest layer agents select action a_i</p> $a_i = \mu_{\theta}(s_i) + Noise_i$ <p>END FOR4</p> <ol style="list-style-type: none"> 7. Perform action $a_i=(a_1, a_2, \dots, a_n)$, and observe and report r_t and the state s_{t+1} of the next moment; 8. Store the sample (s_t, a_t, r_t, s_{t+1}) in one's own experience pool D_i; 9. Make $s_t=s_{t+1}$. <p>FOR5 $i=1:N$</p> <ol style="list-style-type: none"> 10. Select m samples (s_i, a_i, r_i, s_i') from one's own experience pool; 11. Update the critic network parameter through minimizing loss and the strategy loss calculation as formula (1). 12. Calculate the parameter of the update action network through the gradient descent method. The gradient as formula (2). 13. Update the target network parameter θ^Q, θ^{π}. <p>END FOR5</p> <p>END FOR4</p> <p>END FOR2</p>

This method has the following characteristics:

1) changes the original distributed optimum-seeking method of the MADDPG into the centralized-distributed optimum-seeking method through a hierarchical method and adopts centralized training as well as decentralized execution. During training, the critic of each agent can observe the actions of all other agents, and each agent’s actor only needs to observe the local state to make a decision during testing.

2) During training, each agent’s convergence is more directional. As the hierarchical method is used, the agents on the lowest layer converge first, and agents of all layers converge in order from the bottom layer to the top layer.

3) Compared with the hierarchical reinforcement learning method, this method adopts the centralized training and decentralized execution of the MADDPG and can observe global information during centralized training, which is more conducive to global optimum seeking and easier to obtain an optimal solution.

4) The hierarchical method disintegrates the original task into many sub-tasks, and the state and action space of each sub-task are greatly reduced. Compared with that of the single-agent DDPG method, the possibility of curse of dimensionality is reduced, which favors algorithm convergence.

5) Hierarchical learning is adopted to force the final result of the problem to be related to each agent’s decision, and therefore the cooperation and game relationship among agents is strengthened.

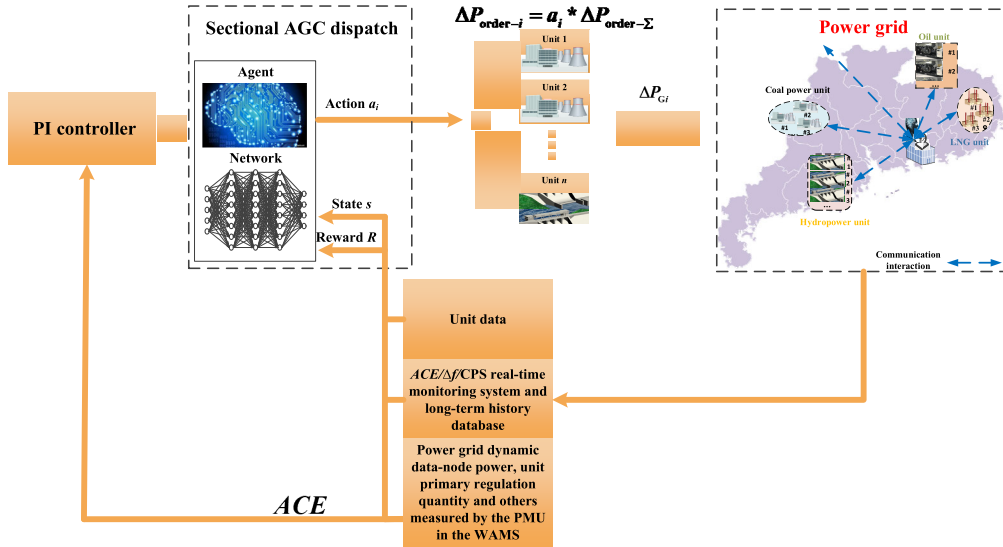


FIGURE 2. Diagram of sectional AGC dispatch system.

III. SECTIONAL AGC DISPATCH MODEL

A. SECTIONAL AGC DISPATCH MODEL BASED ON REAL-TIME MEASUREMENT DATA

1) REAL-TIME MEASUREMENT TECHNOLOGY-SYNCHRONOUS PHASOR MEASUREMENT UNIT

The phasor measurement unit (PMU) in the wide area measurement system (WAMS) is a phasor measurement unit using the global positioning system pulse as its synchronous clock.

The PMU can directly measure generator power angle, generator outlet frequency, active power, voltage, power flow of important buses of the converting station to synchronously collect data of each bus in the power grid.

It can provide dynamic power grid data with millisecond-level precision for the scheduling main station. Therefore, to more comprehensive measurement information about the power grid, the author uses the synchronous PMU to measure the voltage, the current, power, power grid frequency, real-time power flow state, primary frequency regulation variation of units and other information of power grid buses for the AGC to use. As a result, during the AGC dispatch, frequency stability, the dynamic power flow of the current power grid and the satisfaction condition of the real-time security constraint can be all considered at the same time.

2) SECTIONAL AGC DISPATCH SYSTEM

Figure 2 shows a sectional AGC dispatch system. In each AGC control interval, the power grid scheduling center obtains the real-time CPS index value of the current moment, the power plant generation plan and other historical values from the SCADA database of the energy management system [8]. At the same time, the synchronous PMU in the

WAMS database measures the real-time bus voltage and current, frequency, the primary regulation output of each unit, the real-time power flow state and sends them to the sectional AGC dispatch system. The controller of the sectional AGC dispatch system is a regular PI controller which calculates the total AGC generation power command of the unit $\Delta P_{order-\Sigma}$. The power grid scheduling center then dispatches the total AGC generation power commands to each AGC unit through the related generation power command optimization algorithm and calculates the regulation output of each AGC unit $\Delta P_{order-n}$.

The AGC generation power command of each unit is sent to the generation control system of each power plant through an information transmission system [9], [10]. The WAMS collects the regulation output of each unit and the operation information of other units then sends them to the Sectional AGC dispatch of the scheduling center. The control cycle is 8s.

The AGC dispatch algorithm provides the optimal dispatch strategy and outputs sets of continuous data, which are the participation factors allocated to n units. The AGC generation power commands of n units are the product of the total AGC generation power command output by the PI controller multiplied by each participation factor. In Figure 2, $\Delta P_{order-\Sigma}$ is the total AGC power regulation command of the scheduling center, $\Delta P_{order-i}$ is the AGC generation power command of the i th unit, and a_i is the participation factor of the i th unit. It Satisfy the formula as $\Delta P_{order-i} = a_i * \Delta P_{order-\Sigma}$. To satisfy the power balance constraint, the participation factor satisfies the constraint of formula (3).

$$0 \leq a_i \leq 1, \quad \sum_{i=1}^n a_i = 1 \quad (3)$$

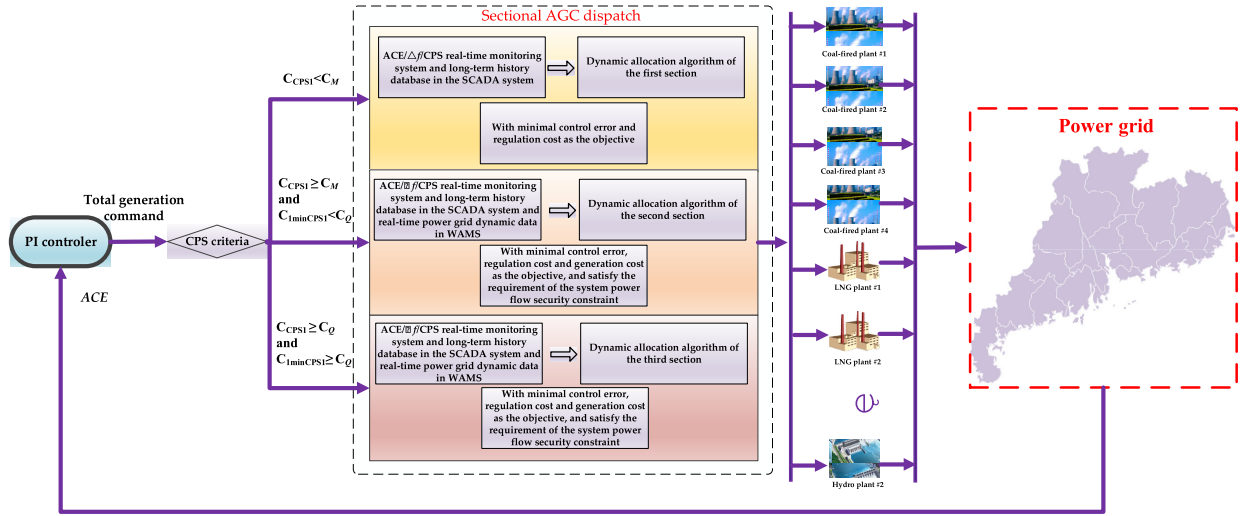


FIGURE 3. Diagram of sectional AGC dispatch model.

B. SECTION

To solve the problem of integrated control of power flow and frequency in the interconnected power grid with large-scale new energy, AGC optimization model for sectional dynamic allocation of generation power commands is proposed in this paper. This model not only considers the power deficiency of the power grid, the ramp rate constraint for units [11], the frequency quality constraint [12] that need to be considered in regular AGC systems, but also unit generation cost [13] and the security constraint of regular optimal power flow models that need to be considered in economical scheduling [14]. The model adopts the power flow formula that considers static characteristics to represent the relationship between power/frequency change [15] and the power flow as well as security constraint [16], [17].

The sectional AGC dispatch process is divided into three parts according to the CPS1 instantaneous value C_{CPS1} and the average value of CPS1 within one-minute $C_{1minCPS1}$. That is, a dispatch method of three sections is adopted, as shown in Figure 3.

1) FIRST SECTION-FREQUENCY STABILITY CONTROL SECTION

The judgment standard is the CPS1 instantaneous [18] value $C_{CPS1} < C_M$, with C_M being a constant smaller than 200%. When the area meets this standard the AGC dispatch will use the method for the first section.

If the CPS1 instantaneous value Satisfy the following formulas: $C_{CPS1} \geq 180\%$, the CPS appraisal in this period is excellent [19]. As the load prediction technology is mature, the author suggests that $C_M = 180\%$ be set. The main consideration of control for the first section is the CPS index and the regulation cost, and the linear weight obtained from the product of the two multiplied by the

weight coefficient as the objective function is used to find the cumulative minimum. The purpose of this section is to quickly recover frequency and the CPS index to a normal level.

$$\begin{cases} \min E = \sum_{t=1}^T \left(\mu_1 \sum_{n=1}^N \Delta P_{error-n}^2(t) + \mu_2 \sum_{n=1}^N C_n(t) \right) \\ s.t. \\ \Delta P_{error-n}(t) = \Delta P_{order-n}(t-1) - \Delta P_{Gn}(t) \\ \Delta P_{order-\Sigma} = \sum_{n=1}^N \Delta P_{order-n} \\ DR_n \leq \Delta P_{order-n}(t) - \Delta P_{order-n}(t-1) \leq UR_n \\ \Delta P_{Gn}^{\min} \leq \Delta P_{Gn}(t) \leq \Delta P_{Gn}^{\max} \end{cases} \quad (4)$$

where t is the discrete time; $\Delta P_{error-n}$ is the difference between the unit generation power command received by the n th unit and this unit's actual output (hereinafter referred to as control error), (MW); C_n is the AGC regulation cost of the n th unit (\$); E is the cumulative value of the square of the control error ($\Delta P_{error-n}$) and the regulation cost C_n of each unit within the time period of T , N is the total number of units; μ_1 and μ_2 are constants that are set to solve the different dimension problem between different optimization targets in the multi-target optimization problem and that are the weights of the square of the control error and the regulation cost in the control target respectively. $\Delta P_{order-\Sigma}$ is the total AGC system generation power command value (MW); $\Delta P_{order-n}$ is the AGC generation power command dispatch to the n th unit (MW); UR_n and DR_n are the ramp rate upper limit and lower limit of the n th unit (MW); ΔP_{Gn} is the actual regulation output of the n th unit (MW); ΔP_{Gn}^{\max} and ΔP_{Gn}^{\min} are the regulation capacity upper and lower limits of the n th unit respectively (MW).

2) SECOND SECTION-TRANSITION CONTROL SECTION

The judgment standard is

$$\begin{cases} C_{CPS1} \geq C_M \\ C_{1minCPS1} < C_Q \end{cases} \quad (5)$$

where C_{CPS1} is the instantaneous value of CPS1. $C_M = 180\%$, $C_Q = 195\%$. $C_{1minCPS1}$ is the average CPS1 value within one minute. Those meet this standard will use the generation power command dispatch method for the second section.

This section is set mainly to prevent the direct change of AGC dispatch method from the first section AGC generation power command dispatch method to the third section, causing generation power command change and therefore sudden change of the CPS index and generator output, which will affect frequency stability and the system CPS appraisal index. When the transitional second section dispatch method is set between the first section dispatch method and the third section dispatch method, the objective function needs to simultaneously consider the generation cost, the CPS index and the regulation cost, with the generation cost as the main consideration index. Besides, the system power flow constraint needs to be considered at the same time.

$$\begin{cases} \min E = \sum_{t=1}^T \left(\mu_1 \sum_{n=1}^N \Delta P_{error-n}^2(t) \right. \\ \quad \left. + \mu_2 \sum_{n=1}^N C_n(t) + \mu_3 \sum_{n=1}^N C_{nG}(t) \right) \\ s.t. \Delta P_{error-n}(t) = \Delta P_{order-n}(t-1) - \Delta P_{Gn}(t) \\ \Delta P_{order-\Sigma} = \sum_{n=1}^N \Delta P_{order-n} \\ DR_n \leq \Delta P_{order-n}(t) - \Delta P_{order-n}(t-1) \leq UR_n \\ \Delta P_{Gn}^{\min} \leq \Delta P_{Gn}(t) \leq \Delta P_{Gn}^{\max} \\ P_{Gi}(t) - P_{Di}(t) - P_i(t) = 0 \\ Q_{Gi}(t) - Q_{Di}(t) - Q_i(t) = 0 \\ \left. \begin{cases} P_i(t) = \sum_{j=1}^n \left[e_i(t) (G_{ij}e_j(t) - B_{ij}f_j(t)) \right. \\ \quad \left. + f_i(t) (G_{ij}f_j(t) + B_{ij}e_j(t)) \right] \\ Q_i(t) = \sum_{j=1}^n \left[f_i(t) (G_{ij}e_j(t) - B_{ij}f_j(t)) \right. \\ \quad \left. - e_i(t) (G_{ij}f_j(t) + B_{ij}e_j(t)) \right] \\ C_{nG}(t) = (a_i P_{Gi}^2(t) + b_i P_{Gi}(t) + c_i) \\ U_{i\min}^2 \leq e_i^2(t) + f_i^2(t) \leq U_{i\max}^2 \\ -S_{ij\max} \leq S_{ij}(t) \leq S_{ij\max} \end{cases} \right. \end{cases} \quad (6)$$

where $C_{nG}(t)$ is the generation cost of the n th unit at moment t , $P_{Gi}(t)$ and $Q_{Gi}(t)$ are the active and passive power output by the i th generator, which is the sum of base point power of the i th generator and AGC regulation power; $P_{Di}(t)$ and $Q_{Di}(t)$ are the active and passive loads of the i th generator at

moment t ; $P_i(t)$ and $Q_i(t)$ are the injected active and passive power of the i th generator at moment t ; G_{ij} and B_{ij} are the real part and virtual part of the i th row and the j th column elements in the system bus admittance matrix respectively; $e_i(t)$ and $f_i(t)$ are the actual part and virtual part of the voltage component of bus i at moment t respectively, a_i , b_i and c_i are the cost coefficient of the i th unit. $S_{ij}(t)$ is the apparent power transmitted by i and j buses of the line at moment t . The subscripts “max” and “min” represent the upper and lower limits of the corresponding variables respectively in this paper. As AGC control is a dynamic process, it is difficult to satisfy the last two constraints of formula (7) within the first couple of control cycles after the section entry. However, the two constraints will be gradually satisfied after AGC control. Therefore, the two constraints can be achieved as a control target.

In the model of this paper, the generation units are divided into two categories, the first category is regular generation units that only participate in the primary frequency regulation, and the second category is AGC units that participate in both primary and secondary frequency regulation. The relationship between the generation power and the frequency of the regular units is as follows:

$$P_{Gi}(t) = P_{Gi0}(t) - K_{Gi} (f(t) - f_N) \quad (8)$$

where $f(t)$ is the current frequency of the system; f_N is the system's rated frequency; $P_{Gi0}(t)$ is the base point of the i th unit at moment t ; K_{Gi} is the active-frequency static characteristic coefficient of the i th generator. The relationship between the generation power and the frequency of the AGC units is as follows:

$$P_{Gi}(t) = P_{Gi0}(t) - K_{Gi} (f(t) - f_N) + \Delta P_{Gi}(t) \quad (9)$$

If the static frequency characteristic is considered, the expression [14] of the active and passive loads of each mode is:

$$\begin{cases} P_{Di}(t) = [1 + K_{Pfi} (f(t) - f_N) / f_N] P_{DNi}(t) \\ Q_{Di}(t) = [1 + K_{Qfi} (f(t) - f_N) / f_N] Q_{DNi}(t) \end{cases} \quad (10)$$

where f_N is the rated frequency: 50HZ; $P_{DNi}(t)$ and $Q_{DNi}(t)$ are the active load and passive load of bus i under the rated voltage and frequency at moment t respectively; K_{Pfi} and K_{Qfi} are all static frequency characteristic parameters of the load model.

3) THIRD SECTION-OPTIMAL POWER FLOW SECTION

The judgment standard

$$\begin{cases} C_{CPS1} \geq C_Q \\ C_{1minCPS1} \geq C_Q \end{cases} \quad (11)$$

$C_Q = 195\%$, only considers the generation cost as the objective function in this section. The constraint is the security constraint, including the equality constraint and the inequality

constraint in formula (7).

$$\left\{ \begin{array}{l} \min E = \sum_{t=1}^T \left(\sum_{n=1}^N C_{nG}(t) \right) \\ s.t. \\ \Delta P_{order-\Sigma} = \sum_{n=1}^N \Delta P_{order-n} \\ DR_n \leq \Delta P_{order-n}(t) - \Delta P_{order-n}(t-1) \leq UR_n \\ \Delta P_{Gn}^{\min} \leq \Delta P_{Gn}(t) \leq \Delta P_{Gn}^{\max} \\ C_{nG}(t) = (a_i P_{Gi}^2(t) + b_i P_{Gi}(t) + c_i) \end{array} \right. \quad (12)$$

IV. SETTINGS FOR SECTIONAL AGC DISPATCH ALGORITHM

As the HMA-DDPG has the advantage of fast response and being not easy to get into a local optimum in the online testing operation after pre-training, this algorithm needs to be used in the dispatch processes of the first section and the second section that require fast response. The security constraint and the inequality constraint of the power grid need to be strictly satisfied after frequency stability during the dispatch process of the third section. Besides, considering the continuous non-linear characteristic of the optimization model [10], the primal-dual interior point method with good convergence is selected for solution finding in [20].

A. FIRST SECTION DISPATCH AND SECOND SECTION DISPATCH ALGORITHM-HMA-DDPG

The HMA-DDPG used in first section and second section dispatch.

First, the units with little difference in the secondary frequency regulation delay time are categorized. Afterwards, the categorized unit groups are further divided into several layers of unit sets based on other regulation characteristics of the units. Each set corresponds to one agent i . The hierarchical dispatch method is shown in Figure 4.

1) ACTION SPACE

To satisfy the constraint requirement of formula (3), during each AGC dispatch process, for a certain agent i , assume that there are n allocated units, and the AGC dispatch algorithm only needs to output the participation factors for $n-1$ units [18]. The participation factor for the n th unit is:

$$a_{in} = 1 - \sum_{j=1}^{n-1} a_{ij} \quad (13)$$

The n th unit defined in this paper is the balance unit, and the unit with the greatest adjustable capacity is selected as the balance unit.

For any agent i at moment t , the participation factors of the first $n-1$ units are agent actions, $n-1$ in total, as shown in

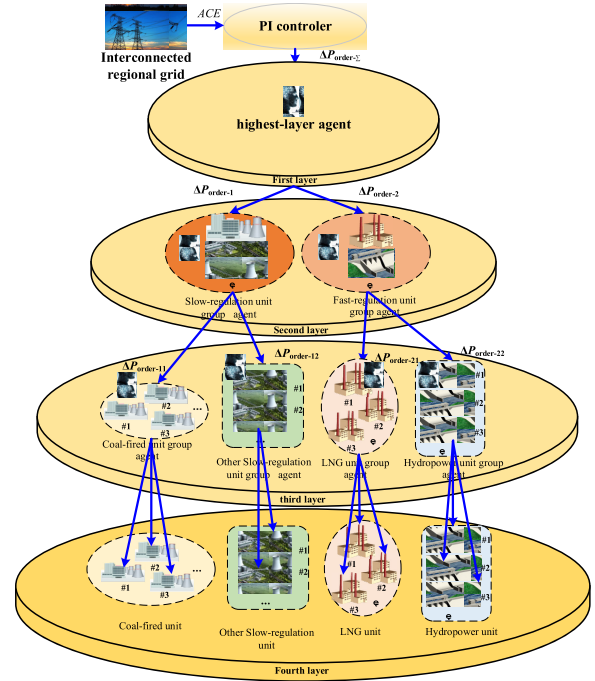


FIGURE 4. Diagram of sectional AGC hierarchical dispatch.

formula (14).

$$\left\{ \begin{array}{l} [a_{i1} a_{i2} \dots a_{in-1}], \sum_{j=1}^{n-1} a_{ij} < 1 \\ a_{in} = 1 - \sum_{i=1}^{n-1} a_{ij} \end{array} \right. \quad (14)$$

2) STATE SPACE

The state of the highest-layer agents is similar to that of sub-agents: as the HMA-DDPG is used for first section dispatch and second section dispatch, the CPS index C_{CPS1} in the SCADA must be observed in the state space. The total AGC generation power command dispatch by the scheduling center must also include power grid dynamic data in the WAMS. N buses, M units (including X new energy units), and $(7+M)$ state space dimensions are set for the grid topology, so the actor state input by the agent corresponding to the n th category unit group includes: the generation power command value allocated to the n th unit group from the previous layer algorithm $\Delta P_{order-n}$, the system CPS index C_{CPS1} , the actual active output value of the n th category of unit groups — ΔP_{G-n} , the power difference of the n th category of unit groups — $\Delta P_{error-n}(t)$, the power grid load value — L_A , the real-time active output of each unit in the power grid — $(P_{G1}, P_{G2}, \dots, P_{Gn})$, and the overload index of the power grid line — $Line_{out}$ which equals M when there is any overload line in the power grid, or otherwise, equals 0.

$$\left\{ \begin{array}{l} Line_{out} = M \text{ There is line overload} \\ Line_{out} = 0 \text{ There is no line overload} \end{array} \right. \quad (15)$$

The critic state input during centralized training also includes the actions of all agents.

3) REWARD FUNCTION

Based on the requirement of first section dispatch and second section dispatch, it is necessary to ensure continuous and smooth change of generation output and the action of each agent. The reward function is divided into two parts based on the CPS1 instantaneous value. When $C_{CPS1} < 180\%$, the first and the sub-layers' learning objective is to control the total power deviation and regulation cost of this section. When $180\% \leq C_{CPS1} \leq 195\%$ and $C_{1minCPS1} < 195\%$, the first layer and the sub-layers' learning task is to control the total power deviation and regulation cost, the generation cost and the system security constraint.

The highest-layer agent reward function is designed as follows:

$$\begin{cases} \text{if } C_{CPS1} < 180\% \\ R_{-h}(t) = -[\mu_1 \Delta P_{\text{error}-h}^2(t) + \mu_2 C_{h-\Sigma}(t)] \\ \text{if } 180\% \leq C_{CPS1} < 195\% \text{ and } C_{1minCPS1} < 195\% \\ R_{-h}(t) = -[\mu_1 \Delta P_{\text{error}-h}^2(t) + \mu_2 C_{h-\Sigma}(t) + \mu_3 C_{hG}(t)] \\ + W + Line_{out} \\ \Delta P_{\text{error}-h}(t) = \Delta P_{\text{order}-h}(t-1) - \Delta P_{G-h}(t) \end{cases} \quad (16)$$

$$W = \begin{cases} 0 & C_{CPS1} \leq 180\% \\ w & C_{CPS1} > 180\% \end{cases} \quad (17)$$

where $\Delta P_{\text{error}-h}(t)$ is the difference between the CPS command and the total unit output (MW); W is a positive constant term. To ensure that the algorithm finds the gradient, the positive constant term and $\mu_3 C_{hG}(t)$ are added for the reward function when the CPS enters the second-section and third-section dispatch state; $C_{h-\Sigma}(t)$ is the total regulation cost (\$); μ_1, μ_2, μ_3 are weight coefficients. μ_3 shall be greater than μ_1 and μ_2 to achieve the objective of giving more emphasis on generation cost reduction during the second-section dispatch process.

The sub-agent reward function is designed as:

$$\begin{cases} \text{if } C_{CPS1} < 180\% \\ R_{-h}(t) = -[\mu_1 \Delta P_{\text{error}-n}^2(t) + \mu_2 C_{n-\Sigma}(t)] \\ \text{if } 180\% \leq C_{CPS1} < 195\% \text{ and } C_{1minCPS1} < 195\% \\ R_{-h}(t) = -[\mu_1 \Delta P_{\text{error}-n}^2(t) + \mu_2 C_{n-\Sigma}(t) + \mu_3 C_{nG}(t)] \\ + W + Line_{out} \\ \Delta P_{\text{error}-n}(t) = \Delta P_{\text{order}-n}(t-1) - \Delta P_{Gn}(t) \end{cases} \quad (18)$$

where $\Delta P_{\text{error}-n}(t)$ is the control error of the n th category unit group (MW); $C_{n-\Sigma}(t)$ is the sum of regulation cost of all units of the n th category unit group (\$); μ_1, μ_2 and μ_3 are all constants, which are the weight coefficients of the control error, the regulation cost and the generation cost respectively; $\Delta P_{\text{order}-n}(t)$ is the generation power command value dispatch from the n th category unit group of this

layer (MW); $\Delta P_{Gn}(t)$ is the actual output of the n th category unit group (MW).

B. THIRD-SECTION DISPATCH ALGORITHM-PRIMAL-DUAL INTERIOR POINT METHOD

The algorithm adopts the general primal-dual interior point method. The basic principle for finding the optimal solution is as follows: set the slack variable to equalize the inequality constraint. Set the disturbance factor and the punishment term to change the original optimization problem into a new optimization problem. Use the Kuhn-Tucker conditions to obtain a series of non-linear equations. Finally, use the Newton-Raphson method to solve the non-linear equations and judge convergence through the duality gap.

V. SIMULATION

A. SIMULATION LOAD FREQUENCY CONTROL MODEL OF GUANGDONG POWER GRID

To elaborate the superiority of the HMA-DDPG, the Guangdong power grid under the entire large interconnected power grid of China Southern Power Grid is selected as the simulation of sectional AGC dispatch. All AGC units are approximately divided into eight groups based on the different regulation characteristics of units, and all groups participated in the primary and secondary frequency regulation. The specific parameters are shown in Table 5. T_s is the secondary frequency regulation delay time, and payment refers to the AGC regulation cost.

1) REWARD FUNCTION

The dispatch method for this simulation only used the first-section dispatch method without considering the sectional dispatch method. Therefore, the state space and the reward function are set again. The principle of first-level dispatch is to mainly ensure the CPS index, with fast regulation of the units and economical regulation also considered. That is, the difference between the total unit output and the generation power command of the scheduling end is considered first before the AGC regulation cost is considered. The reward function of each agent is shown in formula (19), where $\mu_1 = 10^{-6}$, $\mu_2 = 10^{-7}$.

$$R_{-H}(t) = -[\mu_1 \Delta P_{\text{error}-n}^2(t) + \mu_2 C_{n-\Sigma}(t)] \quad (19)$$

2) ACTION SPACE

As the HMA-DDPG is adopted, the action is set to be the participation factor of the current agent to the next layer. Based on formula (14), the sum of all participation factors needs to be 1. As it is necessary to add a punishment function in the reward function for satisfying constraint of (14) when the action space with two or more dimensions is used; when the sum of participation factors is not 1, minus a big positive number. The author found that this method would seriously affect algorithm optimum seeking and therefore cause a local optimum or difficulty in convergence, training cost increase

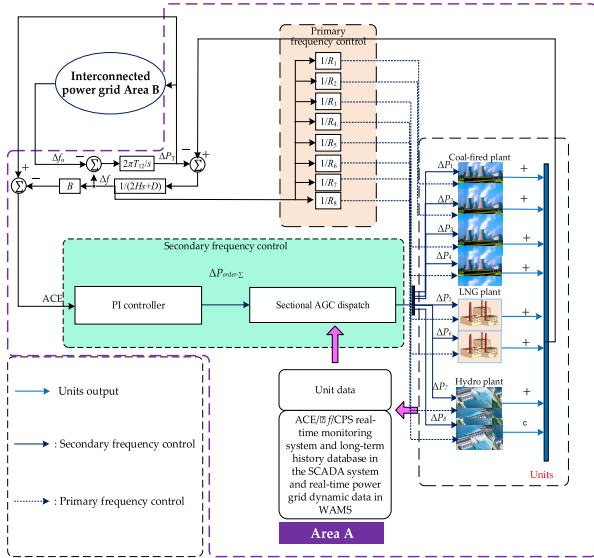


FIGURE 5. Diagram of AGC control of Guangdong power grid.

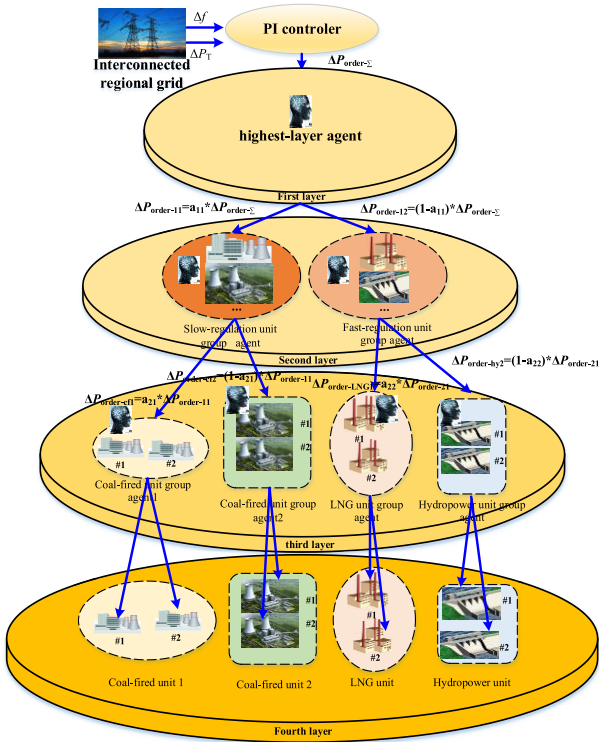


FIGURE 6. Diagram of agents.

and others. Therefore, every two units or unit groups are set as an agent, and each agent only has only one action: the participation factor of a certain unit or the participation factor of a certain category of units is a_{ij} , and the participation factor of another category of units or unit group is $1-a_{ij}$. This way, using a punishment term can be avoided. The specific hierarchical method is shown in Figure 6. The action space is a_{ij} .

TABLE 2. Hyperparameter table.

Parameter	Value
Discount rate γ	0.9
Soft updating coefficient	0.01
Actor learning rate	0.001
Critic learning rate	0.001
Minibatch	128

3) STATE SPACE

The state space is as follows: 1. $\Delta P_{order-n}$, the generation power command of the n th category unit group of this layer; 2. ΔP_{Gn1} , the actual output value of the first unit of the n th category unit group; 3. ΔP_{Gn2} , the actual output value of the second unit of the n th category unit group; 4. The state space of $\Delta P_{error-n}$, the state of the n th category unit group is as follows:

$$[\Delta P_{order-n} \ \Delta P_{Gn1} \ \Delta P_{Gn2} \ \Delta P_{error-n}] \quad (20)$$

During training, the input state of the critic also includes the participation factors of all agents and the difference values between them and 1.

$$\begin{pmatrix} a_{11}, 1 - a_{11}, a_{21}, 1 - a_{21}, a_{32}, 1 - a_{32}, a_{42}, 1 - a_{42}, \\ a_{52}, 1 - a_{52}, a_{62}, 1 - a_{62}, a_{72}, 1 - a_{72} \end{pmatrix} \quad (21)$$

The hyperparameter settings of each agent is shown in Table 2.

4) PRE-LEARNING AND ONLINE TEST

In the pre-learning stage, apply continuous step load disturbance to area A (Guangdong power grid) with a cycle of 1800s and an amplitude of ± 760 MW. After the algorithm completed pre-learning with enough iteration times, the HMA-DDPG is used in a real-life environment.

During online operation, the load disturbance is random step disturbance with an amplitude of not larger than 760 MW and a cycle of 1800s. To verify the algorithm's superiority, six AGC generation power command dispatch algorithms with different principles are used for a comparison. The simulation time is 86400s.

Figure 7 is a diagram showing the CPS1 change of the first 3200s. Based on the change curve within the range of 0s-400s, the CPS1 instantaneous value of the HMA-DDPG algorithm have already reached 199.82% at 99s, while that of other algorithms are: 196.05%, 197.96%, 197.93%, 197.42% and 198.02% respectively. At the same moment, the CPS1 of HMA-DDPG is higher, and during the stable restoration process, the CPS1 value of HMA-DDPG is all better than those of the other algorithms. Based on the curve between 2700s-3200s (the small diagram on the right of

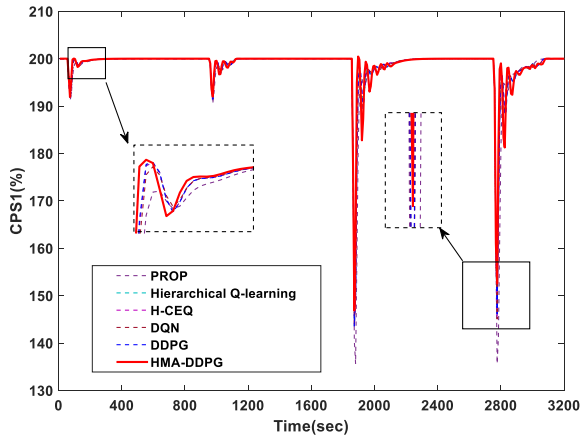


FIGURE 7. Diagram of CPS1 change.

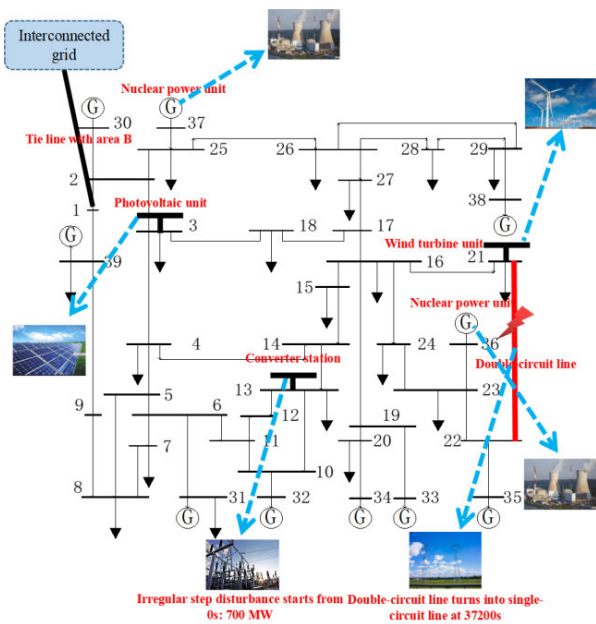


FIGURE 8. Diagram of 39-bus topology.

Figure 7), the minimal value of CPS1 of HMA-DDPG is 152.1%, while those of the other algorithms are: PROP: 135.65%, hierarchical Q-learning: 145.75%, H-CEQ[21]: 145.66%, H-DQN[22]: 145.03%, DDPG: 146.08%. HMA-DDPG could respond to frequency change more quickly, resulting in smaller minimal CPS1 value than those of the other algorithms and ensuring the stability of the system frequency.

Based on Table 3, $|\Delta f|$, C_{CPS1} and $|E_{ACE}|$ are the average value of the absolute values of the frequency deviation, the average value of the CPS1 index and the average value of ACE absolute values. The payment is the total regulation cost. It can be concluded that the control performance of HMA-DDPG is better than that of the other five algorithms, and its regulation cost is lower than those of the other five algorithms. Therefore, HMA-DDPG is superior.

TABLE 3. Statistics of six algorithms.

Algorithm Name	$ \Delta f /\text{Hz}$	$C_{CPS1}/\%$	$ E_{ACE} /\text{MW}$	Payment(unit: $10^4\text{\$}$)
Hierarchical Q-learning	0.01077	194.45	46.18	136.5
HCEQ	0.01077	194.59	46.25	134.9
H-DQN	0.01072	194.69	46.14	135.7
DDPG	0.01071	194.75	46.13	135.2
PROP	0.01099	192.93	47.04	129.7
HMA-DDPG	0.01065	195.47	45.90	115.2

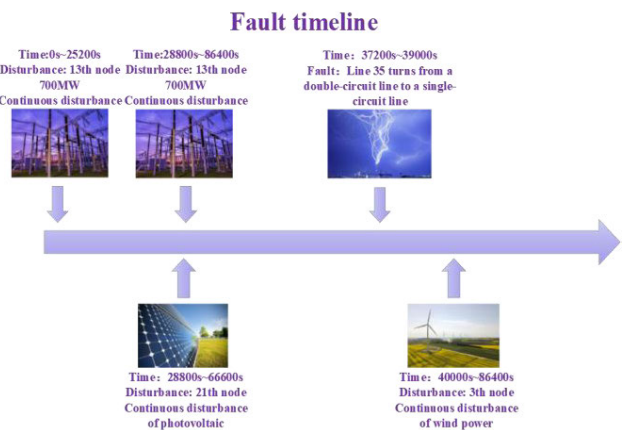


FIGURE 9. Diagram of fault timeline.

B. LOAD FREQUENCY CONTROL MODEL FOR IEEE 39 BUS SYSTEM

To elaborate the superiority of the sectional AGC dispatch method, an IEEE 39 bus system is selected as the power grid topology structure of area A. The third bus is a photovoltaic generation bus, and the 21st bus is a wind turbine generation bus. The bus topology is shown in Figure 8. The load disturbance of the 13th bus converter station is random load disturbance with an amplitude of 700 MW from 0s, and the specific information is shown in Figure 16 of Appendix. The photovoltaic generation output curve and wind turbine generation output curve settings are shown in Figure 17 and Figure 18 of Appendix. The total load increment curve is shown in Figure 19 of Appendix. Bus 1 is the connection bus of the area tie-line. Unit #7 and #8 are nuclear power units with fixed generation power and don't participate in AGC. Area B don't consider power grid topology. Unit parameters are shown in Table 6 of Appendix.

At 37,200s, the 35th line have permanent fault and changed from a double-circuit line to a single-circuit line. At 39,000s, the 35th line changed back to a double-circuit line. The result of the simulation fault timeline is shown in Figure 9.

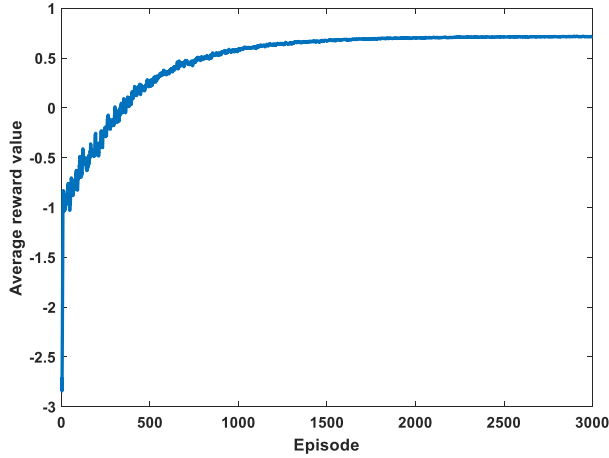


FIGURE 10. Training diagram.

To verify that the calculation result is correct, three methods commonly used in engineering are used for comparison: the PROP dispatch method, the priority AGC dispatch method and the OPF dispatch method.

1) STATE SPACE

As shown in Figure 6, the state space of the highest-layer agents and the sub-agents are similar.

The state space: the state of the agents of the n th category unit group includes: the generation power command value dispatched from the previous-layer algorithm to the n th category unit group of this layer — $\Delta P_{order-n}(t)$, the system’s instantaneous CPS1 index — C_{CPS1} , the actual output value of the n th category unit group — $\Delta P_{G-n}(t)$, the power difference of the n th category unit group — $\Delta P_{error-n}(t)$, the total power grid load — L_A , the active output of each AGC unit — ($P_{G1}, P_{G2}, P_{G3}, P_{G4}, P_{G5}, P_{G6}, P_{G9}, P_{G10}$), the new energy unit output — P_{wt}, P_{pv} , the power grid overload index — $Line_{out}$. The state of the input actor:

$$\begin{bmatrix} \Delta P_{order-n}(t) & \Delta P_{error-n}(t) & C_{CPS1} & \Delta P_{G-n} & P_{G1} \\ P_{G2} & P_{G3} & P_{G4} & P_{G5} & P_{G6} & P_{G9} & P_{G10} & P_{wt} & P_{pv} & Line_{out} \end{bmatrix} \quad (22)$$

During training, besides formula (22), the input critic state also includes: participation factors of other agents, as shown in formula (22).

2) ACTION SPACE

As shown in Figure 6, two units form one set, and the participation factor of one unit is a_{ij} and that of the other unit is $1 - a_{ij}$. The action of each agent i is the participation factor a_{ij} , and the action dimension is 1. The action space is shown in formula (22).

$$[a_{ij}] \quad (23)$$

3) REWARD FUNCTION

The reward function is shown in formula (18). The weight coefficients are $\mu_1 = 10^{-8}$, $\mu_2 = 10^{-8}$, $\mu_3 = 5 \times 10^{-8}$, and

TABLE 4. Statistical results for the four algorithms.

Algorit hm	$ \Delta f $ /Hz	$ E_{ACE} $ /MW	C_{CPS1} /%	Payme nt/ 10^4 \$	C_{NG} / 10^4 \$	$(t_o/t_n)/\%$
PROP	0.00271	7.46	199.08	198.52	7182.8	66.5
Priority	0.00311	8.57	199.17	147.87	7494.1	89.0
OPF	0.00269	7.42	199.15	173.57	7154.7	0.28
HMA-DDPG	0.00268	7.41	199.19	173.49	7154.2	0.25

in the power grid line overload index $Line_{out}$, $R = -0.08$ and $W = 0.125$.

4) PRE-LEARNING

During the pre-learning stage, continuous sinusoidal load disturbance with a cycle of 3600s, an amplitude of 900 MW and duration of 3600s is applied to the 13th bus of area A. The phase is 0.5π . At the same time, the same disturbance is applied for the total load. Sinusoidal waves with a cycle of 3600s, an amplitude of 200 MW and duration of 3600s are applied to the 3rd and 21st new energy generation unit buses. A random value selection method is adopted in each episode for the phase of the latter three types of sinusoidal waves to ensure sample diversity.

5) CONVERGENCE EFFECT

The convergence effect is shown in Figure 10. As the highest-layer agents converged at last, it can be seen from Figure 10 that the average reward value of the agents gradually approached the maximum value of the average reward value in a smooth manner.

6) RESULT OF ONLINE OPERATION TEST

After training, online simulation is used to simulate the system for 86,400s, that is, 24 hours of a day. The specific data are shown in the Table 4.

As can be seen from Figure 11, Figure 12 and Figure 13, during instant step disturbance, the CPS1 variation of the HMA-DDPG can be maintained to be smaller than those of OPF as well as PROP and close to that of the priority method. At the same time, the peak of the frequency deviation and area control error (ACE) are relatively small. As the priority algorithm adopted hydropower units which have the fastest responses, its CPS1 peak is close to that of the HMA-DDPG, but it is not as stable as the HMA-DDPG during the recovery. Besides, the CPS1 of the HMA-DDPG is always superior to that of the priority algorithm, and the $|E_{ACE}|$ of the HMA-DDPG is close to that of the priority algorithm and superior to those of the other three algorithms in the process of CPS1 approaching 200%. The HMA-DDPG is

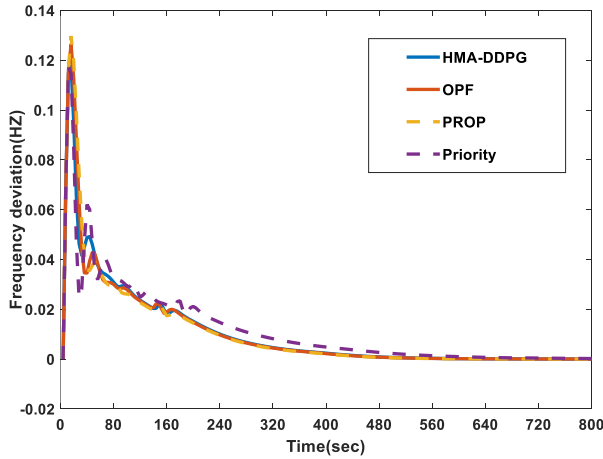


FIGURE 11. Frequency deviation curve from 0S-800S.

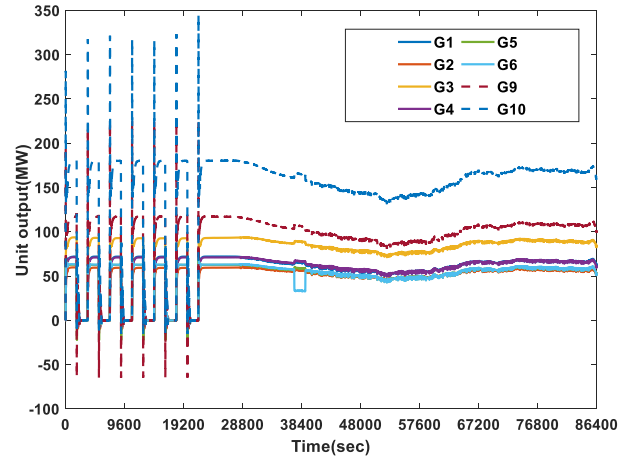


FIGURE 14. Diagram of unit output of the HMA-DDPG algorithm.

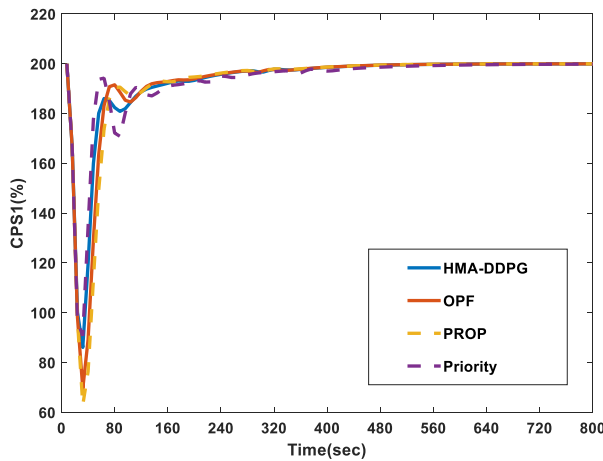


FIGURE 12. CPS1 curve from 0S-800S.

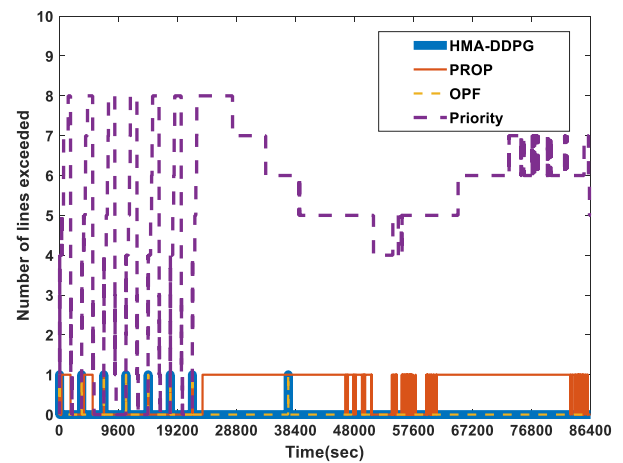


FIGURE 15. Diagram of line overload of the HMA-DDPG algorithm.

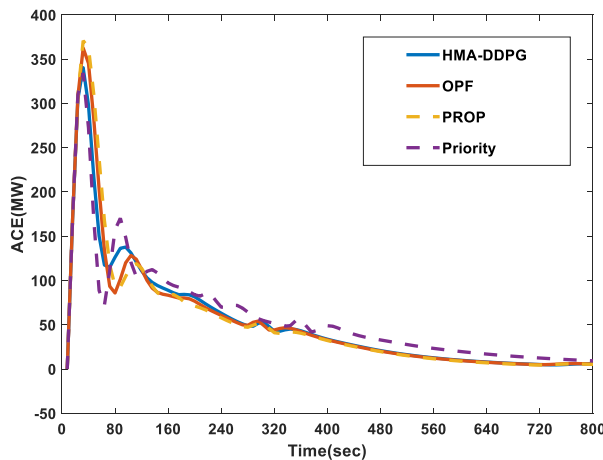


FIGURE 13. ACE curve from 0S-800S.

also superior to the priority algorithm in the recovery process. The same is true of frequency deviation. The regulation cost of the HMA-DDPG is smaller than that of the OPF algorithm,

and its overload duration is the shortest. The stability of the HMA-DDPG is better than that of the other three algorithms, as can be seen from the curves. Based on the number of the system's overload lines of Figure 15, the line overload state of the HMA-DDPG and OPF both occurred when the disturbance suddenly appeared, and after a few control intervals, they can quickly be out of the line overload state. As the HMA-DDPG have a faster early-stage response rate than the OPF algorithm, it could be out of the line overload state faster. Based on Table 4, the proportion of the line overload duration of the HMA-DDPG is smaller than that of the OPF, while the PROP and the priority algorithm could cause several lines of the system to be in a serious long-term overload state. Without artificial intervention, the secure and stable operation of the power grid will be seriously affected. Therefore, as far as control of the line overload state is concerned, the HMA-DDPG is superior to the rest three algorithms.

Based on Figure 14 and Figure 15, line 35 changed from a double-circuit line to a single-circuit line, and the maximum power limit of the line reduced by half at 37,200s. The units

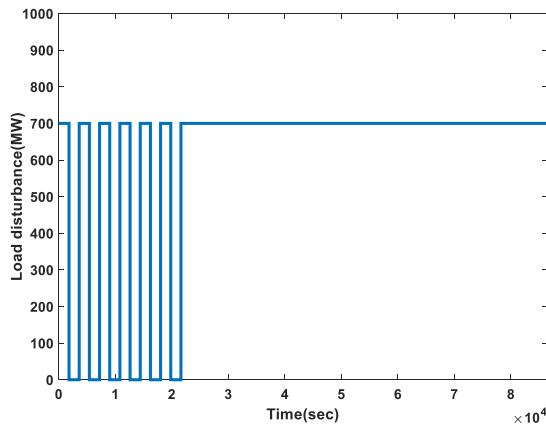


FIGURE 16. Load change of the 13th bus.

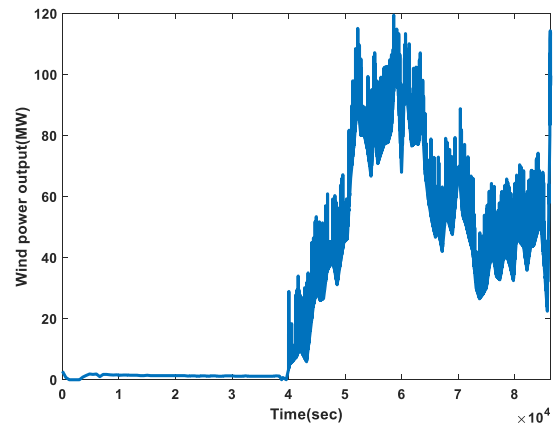


FIGURE 18. Wind power output.

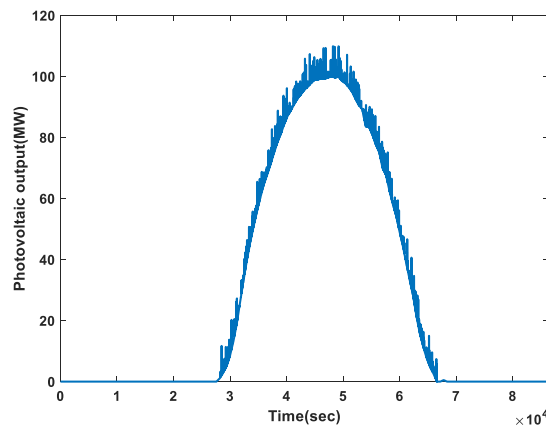


FIGURE 17. Photovoltaic output.

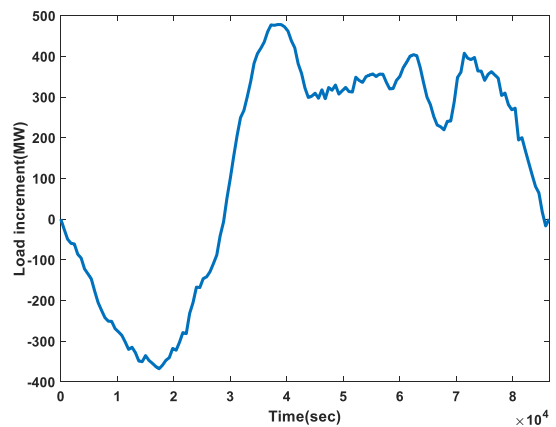


FIGURE 19. Load increment.

responded quickly, and there is a sudden change in the unit output at 37,200s. After only a short period of being overload, the line current returned to within the constraint, while that of the PROP and priority algorithms is still in a long-term overload state. At 40,000s, line 35 returned from a single-circuit line to a double-circuit line after successful forced transmission, and all the units are back to normal. The power grid is maintained in a secure and economical operation state as well as ensuring a comprehensive optimum of the control performance and the generation cost even under wind power and photovoltaic disturbance.

Based on Table 4, $|\Delta f|$, C_{CPS1} and $|E_{ACE}|$ are the average value of the frequency deviation absolute value, the average value of the CPS1 index and the average value of the ACE absolute value. The payment C_{NG} is the total regulation cost, that is, the total generation cost of the eight units, and t_o/t_n is the percentage of the line overload time compared with the total process time. Generally, it can be seen that the HMA-DDPG proposed in this paper has the minimal $|\Delta f|$ and $|E_{ACE}|$, the maximal C_{CPS1} , the lowest generation cost, the lowest overload time proportion, compared with the other three algorithms. Adopting the hydropower units that have the lowest regulation cost, the priority algorithm has the

lowest regulation cost. However, this algorithm used only one hydropower unit for frequency regulation, which caused a serious overload state of the unit's outlet as well as the nearby lines and therefore increased the grid loss, unit output then increased the generation cost. As a result, the comprehensive economic benefit is reduced. Therefore, it can be concluded that the sectional AGC dispatch of the HMA-DDPG proposed in this paper can ensure fast recovery of the frequency as well as the system's fast recovery to the optimal power flow state, and consider the system security constraint simultaneously, then achieve a comprehensive optimum of control performance, economic benefit and system security.

VI. CONCLUSION

In summary, the main contributions of this work are as follows:

- 1) The proposed sectional AGC dispatch can satisfy not only the technical, economic benefits requirements of power grid but also the system's security constraints. It thus addresses the problem of coordinating system economy, security and control performance during secondary frequency regulation in the power grid.

TABLE 5. Guangdong power grid AGC unit parameters.

Unit Type	Unit No.	T_s/s	Capacity Upper Limit/MW	Capacity Lower Limit/MW	RampRate/(MW·min ⁻¹)	Payment/(\$·(MW·h) ⁻¹)
1	Thermal power 1	45	2800	-2800	140.0	196.87
2	Thermal power 2	40	2680	-2680	134.0	126.70
3	Thermal power 3	43	1912	-1912	134.0	298.00
4	Thermal power 4	38	1788	-1788	70.5	127.04
5	Thermal power 5	12	1028	-1028	128.5	253.40
6	Thermal power 6	8	688	-688	68.8	254.08
7	Hydropower 1	5	600	600	600	93.65
8	Hydropower 2	5	400	400	400	84.29

TABLE 6. Simulation unit parameters of the IEEE 39 bus system.

Unit Group No.	Participating in AGC or Not	Unit No.	Bus	T_s/s	$\Delta P_{Gn}^{\max}/MW$	$\Delta P_{Gn}^{\min}/MW$	$UR_n/DR_n(MW^*min^{-1})$	Payment/(\$·(MW·h) ⁻¹)
1	Yes	Thermal power #1	30	45	9600	-440.00	140.0	196.87
2	Yes	Thermal power #2	31	40	4824	-279.00	134.0	126.70
3	Yes	Thermal power #3	32	43	2360	-547.00	134.0	298.00
4	Yes	Thermal power #4	33	38	2000	-466.04	70.5	127.04
5	Yes	Thermal power #5	34	12	2056	-407.00	128.5	253.40
6	Yes	Thermal power #6	35	8	1376	-382.00	68.8	254.08
7	No	Nuclear power #7	36	-	-	-	-	-
8	No	Nuclear power #8	37	-	-	-	-	-
9	Yes	Hydropower #9	38	5	5900	-1489	720.0	93.65
10	Yes	Hydropower #10	39	5	3800	-1053	480.0	84.29

2) HMA-DDPG, the hierarchical multi-agent algorithm, is employed for AGC dispatch. It reduces the difficulty and dimensions of the algorithm optimum seeking, makes the strategy more effective thus providing reference for the optimal AGC dispatch involving a large-scale power grid and multiple units.

3) The simulation analysis indicates that the sectional AGC dispatch based on HMA-DDPG can change AGC unit output with the change of system state, thus ensuring the comprehensive optimum of control performance, the economic benefit, security and stability for the power grid.

APPENDIX

See Figures 16–19 and Tables 5 and 6.

REFERENCES

- [1] X. Zhang, T. Tan, B. Zhou, T. Yu, B. Yang, and X. Huang, "Adaptive distributed auction-based algorithm for optimal mileage based AGC dispatch with high participation of renewable energy," *Int. J. Electr. Power Energy Syst.*, vol. 124, Jan. 2021, Art. no. 106371.
- [2] X. Zhang, Z. Xu, T. Yu, B. Yang, and H. Wang, "Optimal mileage based AGC dispatch of a GenCo," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 2516–2526, Jul. 2020.
- [3] T. Yu, Y. M. Wang, W. J. Ye, B. Zhou, and K. W. Chan, "Stochastic optimal generation command dispatch based on improved hierarchical reinforcement learning approach," *IET Gener., Transmiss. Distrib.*, vol. 5, no. 8, pp. 789–797, Aug. 2011.
- [4] R. Zhao, "Application research on power system distributed computing model based on inter-process communication," *IET Gener. Transm. Distrib.*, vol. 9, no. 2, pp. 575–583, Jan. 2013.
- [5] Y. Qiu, J. Lin, F. Liu, and Y. Song, "Explicit MPC based on the Galerkin method for AGC considering volatile generations," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 462–473, Jan. 2020.

- [6] R. Lowe, Y. Wu, A. Tamar, and J. Harb, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Jun. 2017, pp. 6379–6390.
- [7] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Convers. Manage.*, vol. 198, Oct. 2019, Art. no. 111799.
- [8] H. Wang, Y. Liu, B. Zhou, C. Li, G. Cao, N. Voropai, and E. Barakhtenko, "Taxonomy research of artificial intelligence for deterministic solar power forecasting," *Energy Convers. Manage.*, vol. 214, Jun. 2020, Art. no. 112909.
- [9] D. Xu, Q. Wu, B. Zhou, C. Li, L. Bai, and S. Huang, "Distributed multi-energy operation of coupled electricity, heating and natural gas networks," *IEEE Trans. Sustain. Energy*, early access, Dec. 23, 2020, doi: 10.1109/TSTE.2019.2961432.
- [10] H. Wang, G. Wang, G. Li, J. Peng, and Y. Liu, "Deep belief network based deterministic and probabilistic wind speed forecasting approach," *Appl. Energy*, vol. 182, pp. 80–93, Nov. 2016.
- [11] L. Xi, T. Yu, B. Yang, X. Zhang, and X. Qiu, "A wolf pack hunting strategy based virtual tribes control for automatic generation control of smart grid," *Appl. Energy*, vol. 178, pp. 198–211, Sep. 2016.
- [12] X. S. Zhang, T. Yu, Z. N. Pan, B. Yang, and T. Bao, "Lifelong learning for complementary generation control of interconnected power grids with high-penetration renewables and EVs," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4097–4110, Jul. 2018.
- [13] X. Zhang, Q. Li, T. Yu, and B. Yang, "Consensus transfer Q-learning for decentralized generation command dispatch based on virtual generation tribe," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2152–2165, May 2018.
- [14] X. Zhang, T. Yu, L. Guo, B. Yang, and Y. Chen, "Culture evolution learning for optimal carbon-energy combined-flow," *IEEE Access*, vol. 6, pp. 15521–15531, 2018.
- [15] 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, pp. 225–241, May 2019.
- [16] B. Yang, J. Wang, X. Zhang, T. Yu, W. Yao, H. Shu, F. Zeng, and L. Sun, "Comprehensive overview of meta-heuristic algorithm applications on PV cell parameter identification," *Energy Convers. Manage.*, vol. 208, Mar. 2020, Art. no. 112595.
- [17] L. Yin, T. Yu, B. Yang, and X. Zhang, "Adaptive deep dynamic programming for integrated frequency control of multi-area multi-microgrid systems," *Neurocomputing*, vol. 344, pp. 49–60, Jun. 2019.
- [18] X. Zhang, "Collaborative consensus transfer q-learning based dynamic generation dispatch of automatic generation control with virtual generation tribe," *Proc. CSEE*, vol. 37, no. 5, pp. 1455–1466, Mar. 2017.
- [19] L. Yin, T. Yu, L. Zhou, L. Huang, X. Zhang, and B. Zheng, "Artificial emotional reinforcement learning for automatic generation control of large-scale interconnected power grids," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 9, pp. 2305–2313, Jun. 2017.
- [20] L. M. Ramos Carvalho and A. R. Leite Oliveira, "Primal-dual interior point method applied to the short term hydroelectric scheduling including a perturbing parameter," *IEEE Latin Amer. Trans.*, vol. 7, no. 5, pp. 533–538, Sep. 2009.
- [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, Jun. 2016.
- [22] T. D. Kulkarni, K. R. Narasimhan, A. Saeedi, and J. B. Tenenbaum, "Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation," 2016, *arXiv:1604.06057*. [Online]. Available: <http://arxiv.org/abs/1604.06057>



TAO YU (Member, IEEE) 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 a Professor of power system with the School of Electric Power, South China University of Technology (SCUT), Guangzhou, China. His special fields of interest include nonlinear and coordinated control theory, artificial intelligence techniques in planning, and the operation of power systems.



HANXIN ZHU received the B.S. degree in electrical engineering from the South China University of Technology, Guangzhou, China, in 2018, where he is currently pursuing the M.S. degree in electrical engineering with the College of Electric Power. His research interest includes the optimal operation of integrated energy systems.



FUSHENG LI received the B.Eng. degree in electrical engineering from the South China University of Technology, Guangzhou, China, in 2018, where he is currently pursuing the M.S. degree with the School of Electric Power Engineering. His research interests include artificial intelligence techniques and its applications in smart grids.



DAN LIN received the B.Eng. degree in electrical engineering from the South China University of Technology, Guangzhou, China, in 2018, where she is currently pursuing the M.S. degree with the School of Electric Power Engineering. Her research interests include distribution planning and reliability assessment, and artificial intelligence techniques.



JIAWEN LI received the M.S. degree in electrical engineering from Northeast Electric Power University, Jilin, China, in 2016. He is currently pursuing the D.Eng. degree in electrical engineering with the South China University of Technology. His research interest includes automatic generation control.



ZHUOHUAN LI received the B.S. degree in electrical engineering from the South China University of Technology, Guangzhou, China, in 2018, where he is currently pursuing the M.S. degree in electrical engineering with the College of Electric Power. His research interest includes the optimal operation of power systems.

...