Automatic Generation Control Based on Multiple-step Greedy Attribute and Multiple-level Allocation Strategy

Lei Xi, Le Zhang, Yanchun Xu, Shouxiang Wang, *Senior Member, IEEE*, and Chao Yang

***Abstract*—The strong stochastic disturbance caused by large-scale distributed energy access to power grids affects the security, stability and economic operations of the power grid. A novel multiple-step greedy policy based on the consensus Q-learning (MSGP-CQ) strategy is proposed in this paper, which is an automatic generation control (AGC) for distributed energy incorporating multiple-step greedy attribute and multiple-level allocation strategy. The convergence speed and learning efficiency in the MSGP algorithm are accelerated through the predictive multiple-step iteration updating in the proposed strategy, and the CQ algorithm is adopted with collaborative consensus and self-learning characteristics to enhance the adaptability of the power allocation strategy under the strong stochastic disturbances and obtain the total power commands in the power grid and the dynamic optimal allocations of the unit power. The simulations of the improved IEEE two-area load-frequency control (LFC) power system and the interconnected system model of intelligent distribution network (IDN) groups incorporating a large amount of distributed energy show that the proposed strategy can achieve the optimal coordinated control and power allocation in the power grid. The algorithm MSGP-CQ has stronger robustness and faster dynamic optimization speed and can reduce generation costs. Meanwhile it can also solve the strong stochastic disturbance caused by large-scale distributed energy access to the grid compared with some existing intelligent algorithms.**

***Index Terms*—Automatic generation control, collaborative consensus, multiple-step greedy attribute, multiple-level allocation.**

I. Introduction

THE large-scale access of renewable new energy [1], such as wind power [2], solar energy [3], and the popularization of electric vehicles [4], is needed in order to deal with the energy crisis and increasing environmental pollution. However, the distributed energy sources and loads are stochastic, intermittent and difficult to accurately predict [5], [6], which produces greater challenges to the security, stability and economic operation of power systems [7]. The traditional centralized automatic generation control (AGC) can hardly meet the developing requirements and operational conditions for the power grid [8]. Therefore, the urgent task is how to solve the strong stochastic disturbance problem caused by the large-scale interconnection of distributed energy from the perspective of distributed AGC.

In recent years, many scholars have begun studying the distributed AGC strategy. According to the control mode and implementation mechanism of AGC, the existing distributed AGC strategies [9] can be divided into two categories: the AGC "control" strategy and the AGC "power allocation" strategy. Aiming at the "control" strategy of the distributed AGC, an adaptive dynamic programming strategy for load-frequency control (LFC) was proposed to obtain the optimal frequency regulation of multiple-area power system by adapting to the real-time disturbances and uncertainties [10]. And a multiple-area AGC strategy based on artificial neural network (ANN) was presented in [11], which trained with the back propagation algorithm to achieve the optimal control of the area control error (ACE). Similarly, the authors have also completed some studies on distributed AGC control strategies based on the multiple-agent reinforcement learning (MARL) proposed DCEQ (λ) [12], DWoLF-PHC (λ) [13] and WPH [14]. The intelligence of the system is improved in these algorithms by sharing information experience and mutual cooperation among agents, thus the problem of the optimal cooperative control in AGC is effectively solved [15]–[17].

An improved algorithm hierarchical Q-learning (HQL) was proposed in [18] for the "power allocation" strategy of the distributed AGC to deal with the curse of dimensionality in the dynamic optimization of power allocation under control performance standards (CPS). And a new consensus transfer Q-learning (CTQ) was developed in [19] to achieve the effect of decentralized autonomy and centralized collaboration, which can effectively utilize historical optimization information for fast dynamic power allocation. However, the two categories of the AGC strategies only separately take into account the "control" or "power allocation" in the distributed AGC systems, without considering both of them, so they really cannot

Manuscript received June 19, 2020; revised September 12, 2020; accepted October 25, 2020. Date of online publication November 20, 2020; date of current version May 8, 2021. This work was supported in part by the National Natural Science Foundation of China (No. 51707102).

L. Xi (corresponding author, e-mail: xilei2014@163.com; ORCID: https://orcid.org/0000-0002-5564-9953) is with the College of Electrical Engineering and New Energy, and the Yichang Key Laboratory of Defense and Control of Cyber-Physical Systems, China Three Gorges University, Yichang 443002, China.

L. Zhang, Y. C. Xu and C. Yang are with the College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China.

S. X. Wang is with the Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China.

DOI: 10.17775/CSEEJPES.2020.02650

achieve the intelligence from the whole to the branch [20].

The authors observed the long-term simulation data of these control algorithms and found that their convergence rate still was slow with the continuous access of large-scale distributed energy, thus new methods of AGC "control" need to be explored. A multiple-step greedy policy (MSGP) was proposed in [21], which is a reinforcement learning (RL) algorithm with multiple-step greedy attributes. It updates the greedy strategy of selecting multiple-step action by several iterations, and quickly converges into an optimal strategy, so as to obtain a good performance of coordinated control.

Meanwhile, a collaborative consensus algorithm was utilized in the AGC "power allocation" part of the EPCC strategy [22] to dynamically optimize the power allocation of generating units. However, the collaborative consensus algorithm is only a simple first-order consensus algorithm, which is strongly dependent on the model and is easy to fall into the local optimal solution. Therefore, it needs to explore the new algorithms of dynamic optimization allocation. A consensus Q-learning (CQ) algorithm is proposed in this paper in order to improve the adaptability of the consensus algorithm in a dynamic stochastic environment, which integrates a collaborative consensus algorithm and classical Q-learning algorithm [23] to solve the shortcomings of the first-order consensus algorithm.

Therefore, a novel multiple-step greedy policy based on consensus Q-learning (MSGP-CQ) strategy is proposed in this paper, that is, the algorithm MSGP is used in the "control" part of AGC and the algorithm CQ is used in the "power allocation" part of the unit. The cooperative control strategy, with multiple-step greedy attributes, is combined with the optimal power allocation method with self-learning ability to form a MARL algorithm. The proposed strategy is adopted to obtain the total power command and the dynamic optimal allocation of the unit for the regional power grid. Then the optimal cooperative control of the distributed AGC is obtained from the whole to the branch, which solves the strong stochastic problem caused by the large-scale access of distributed energy to the grid. An improved IEEE two-area LFC power system is employed and the interconnected system model of intelligent distribution network (IDN) groups incorporating a large amount of distributed energy is constructed to simulate and verify the effectiveness of the proposed strategy. The results show that the proposed strategy MSGP-CQ can reduce the generation cost of generating units, has stronger robustness and faster dynamic optimization speed, and can solve the strong stochastic disturbance caused by large-scale access of distributed energy compared with the existing intelligent algorithms.

II. AGC Framework Based on MSGP-CQ Strategy

The distribution networks incorporated with a large number of distributed energy resources can be virtually divided into several small area grids according to the cut-set method of graph theory, and the control framework is shown in Fig. 1. Each small area grid is regarded as a territorial power grid, which maintains the power frequency stability by controlling the power exchange at the regional boundary. The territorial power grids can be actively decoupled and enter the island operation mode while serious faults happened in the power

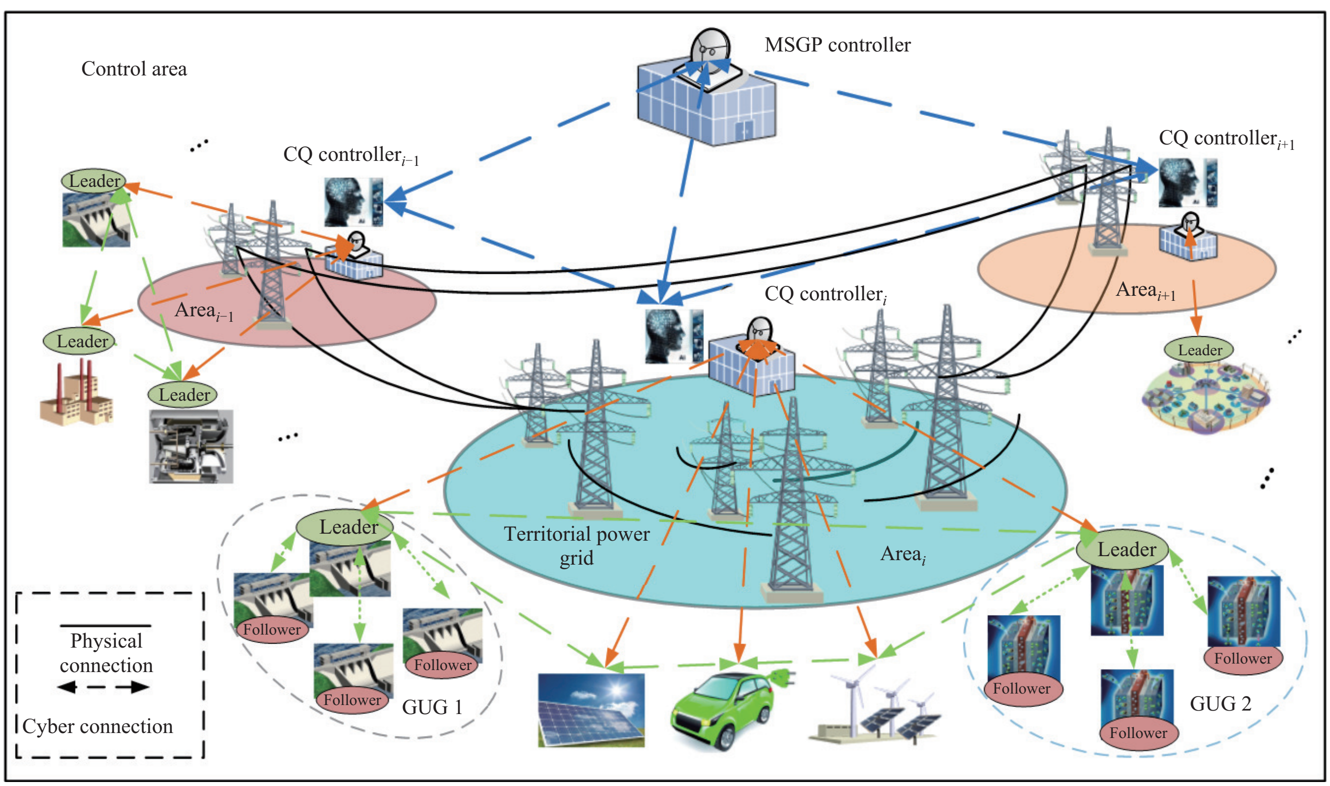

Fig. 1. The AGC framework based MSGP-CQ strategy.

system, so that the collapse of the whole power grid can be avoided. And each territorial grid consists of several generator unit groups (GUG) of the same type.

The MSGP-CQ strategy, based on the multiple-level control mode, is proposed to achieve the optimal coordinated control and power allocation of the AGC system in the framework of distribution networks composed of multiple territorial power grids. The so-called "multiple-level" is based on the multiple-level control strategy consisting of the AGC "control" part and unit "power allocation" part. The "control" part of AGC is the first level in the proposed multiple-level control mode. The MSGP control algorithm with heterogeneous attributes of the multiple-agent stochastic game is adopted to obtain the total power command in each territory power grid. And the unit "power allocation" of AGC is the second to third level in the control mode. The first level of the CQ algorithm is applying a new HQL (λ) algorithm with the multiple-step backtracking eligibility trace to dynamically allocate the total power command obtained in each territorial grid to the GUGs. The second level is that each GUG uses a consensus algorithm with homogeneous attributes of multiple-agent system collaborative consensus (MAS-CC) to dynamically allocate the power command of each GUG to each generator unit.

III. MSGP-CQ Strategy

The MSGP-CQ strategy is proposed by combining the acquisition of total power commands and the dynamic optimal power allocation of the units to obtain the optimal solution of the distributed AGC, and to solve the problem of the stochastic disturbance caused by large-scale distributed energy access to the power grid.

A. MSGP Control Algorithm

The algorithm MSGP is proposed as the "control" part of AGC to obtain the total power commands of the territorial power grids, which incorporates multiple-step greedy policy iteration with faster convergence speed. It has a higher convergent multiple-step look-ahead greedy attribute.

However, the monotonic improvement of action selection policy is not guaranteed by multiple-step greedy policy below a certain step-size value. Therefore, the one-step greedy policy with the largest immediate reward is introduced as the action selection policy while the multiple-step greedy policy is not monotonously improved.

The action of one-step greedy policy $a_{\text{1-step}}$ and the action of multiple-step greedy policy a_k are as follows, based on the above two action selection policies.

$$a_{\text{1-step}}(s) = \arg\max_{a\in A} Q(s,a) \tag{1}$$

$$a_k(s) = \arg\max_{a\in A} Q_k(s,a) \tag{2}$$

where $Q(s,a)$ is the Q value function of one-step greedy policy under the state s and the action a; $Q_k(s,a)$ is the Q_k value function of the multiple-step greedy policy under the state s and the action a, and A is the set of actions.

The agent calculates the current value function errors of the one-step greedy policy and multiple-step greedy policy through the reward value R_1 obtained in the current exploration, which is given as follows.

$$\begin{aligned}\delta_n = {} & R_1(s_n, s_{n+1}, a_n) + \\ & \gamma_1 Q_n(s_{n+1}, a_{\text{1-step}}) - Q_n(s_n, a_n)\end{aligned} \tag{3}$$

$$\begin{aligned}\delta_{k,n} = {} & R_1(s_n, s_{n+1}, a_n) + \gamma_1(1-k)V_n^{\pi}(s_{n+1}) + \\ & k\gamma_1 \max_{a_k} Q_{k,n}(s_{n+1}, a_k) - Q_{k,n}(s_n, a_n)\end{aligned} \tag{4}$$

where $R_1(s_n, s_{n+1}, a_n)$ is the reward function of the agent from the state s_n to the state s_{n+1} under the selected action a_n; γ_1 is the reward discount factor; k is the step-size factor; $V^{\pi}(s_{n+1})$ is the Q value function expectation under the state s_{n+1} and the decision-making policy π; δ_n is the Q value function error of the agent at the nth iteration, and $\delta_{k,n}$ is the Q_k value function error of the agent at the nth iteration.

The algorithm MSGP is updated iteratively as follows:

$$Q_{n+1}(s_n, a_n) = Q_n(s_n, a_n) + \alpha_1 \delta_n \tag{5}$$

$$Q_{k,n+1}(s_n, a_n) = Q_{k,n}(s_n, a_n) + \alpha_1 \delta_{k,n} \tag{6}$$

where α_1 is the learning rate of the value function.

The decision-making policy with a multiple-step greedy attribute is iteratively obtained by updating the calculated Q and Q_k value functions, and the convergent optimal strategy is finally achieved, which is iteratively updated as follows:

$$\begin{aligned}& \pi_{n+1}(s_n) = \pi_n(s_n) + \\ & \beta_1 \left(b_{s_n}\left(Q_{n+1}, Q_{k,n+1}, \pi_n\right) - \pi_n(s_n)\right)\end{aligned} \tag{7}$$

where $\pi_n(s_n)$ is the value function of the decision-making policy under the state s_n at the nth iteration; β_1 is the learning rate of policy iteration, and $b_s(Q, Q_k, \pi)$ is the determinant equation to guarantee the monotonic improvement of the decision-making policy π, which is given as follows.

$$b_s(Q, Q_k, \pi) = \begin{cases} a_k(s), & Q(s, a_k) \geqslant V^{\pi}(s) \\ a_{\text{1-step}}(s), & \text{otherwise} \end{cases} \tag{8}$$

It is shown in (8) that the value of $b_s(Q, Q_k, \pi)$ chooses the action of multiple-step greedy policy a_k if the Q value function under the state s and the action of multiple-step greedy policy a_k is not less than the Q value function expectation under the state s and the decision-making policy π. That is the multiple-step greedy policy is adopted to ensure monotonic improvement. Otherwise, the action of one-step greedy policy $a_{\text{1-step}}$ is chosen.

With the sufficient trial-and-error iterations, the value function of the decision-making policy will gradually converge to a stable optimal action to accelerate the convergence speed of the RL policy, then an optimal control policy will be finally acquired.

B. CQ Allocation Algorithm

The algorithm CQ is adopted to optimize the power allocation of each unit in the territorial power grid after the total power command of the AGC control is acquired. It is primarily composed of two power allocation levels and their corresponding allocation algorithms.

1) HQL (λ) Algorithm

The introduction of the HQL (λ) algorithm can achieve the process of interactive learning and self-learning between GUGs, that is, the initial allocation of power instructions between GUGs. The linear weighted sum of ΔP_{error} and C_{total} are selected as the reward function considering the optimal allocation effect of this power allocation level. Due to the inconsistency of physical dimensions between ΔP_{error} and C_{total}, the order magnitude of C_{total} is divided by 1,000, which is classified into the same order of ΔP_{error}, so that the two control objectives of ΔP_{error} and C_{total} are balanced. The reward function is chosen as follows.

$$R_2\left(s_n, s_{n+1}, a_n\right) = -\Delta P_{\text{error}}^2 - \frac{C_{\text{total}}}{1000} \tag{9}$$

The eligibility trace based SARSA (λ) is selected for the multiple-step iteration updating process to obtain the multiple-step information updating mechanism of the RL algorithm, which is expressed as shown in (10).

$$e_{n+1}(s,a) = \begin{cases} 1, & (s,a) = (s_n, a_n) \\ \gamma_2 \lambda e_n(s,a), & \text{otherwise} \end{cases} \tag{10}$$

where $e_n(s,a)$ is the eligibility trace at the nth iteration under the state s and the action a; γ_2 is the reward discount factor, and λ is the eligibility trace attenuation factor. The eligibility trace records the frequency of each joint action strategy in detail, and updates the iteration Q value of each action strategy. While the current action state pair is the same as the next step, the action state pair will be given a higher backtracking reward [24]. Therefore, the eligibility trace value of the action state pair is usually set to 1.

The agent evaluates the calculation of the value function errors, which is given as follows:

$$\begin{aligned} \delta_n = {} & R_2(s_n, s_{n+1}, a_n) + \gamma_2 Q_n(s_{n+1}, a_{\text{1-step}}) - \\ & Q_n(s_n, a_n) \end{aligned} \tag{11}$$

$$\begin{aligned} M_n = {} & R_2(s_n, s_{n+1}, a_n) + \gamma_2 Q_n(s_{n+1}, a_{\text{1-step}}) - \\ & Q_n(s_n, a_{\text{1-step}}) \end{aligned} \tag{12}$$

where δ_n is the Q value function error of the agent at the nth iteration, and M_n is the evaluation of the function error.

The algorithm HQL (λ) is updated iteratively as follows:

$$Q_{n+1}(s,a) = Q_n(s,a) + \alpha_2 M_n e_n(s,a) \tag{13}$$

$$Q_{n+1}(s_n, a_n) = Q_{n+1}(s_n, a_n) + \alpha_2 \delta_n \tag{14}$$

where α_2 is the value function learning factor.

The total power command of each territorial grid is taken as a state variable. And it is discreted into $(-\infty, -650]$, $(-650, 20)$, $(20, 850)$, and $[850, +\infty)$.

2) Collaborative Consensus Algorithm

In the collaborative consensus algorithm, the ramp time is selected as the consensus variable, and each GUG is regarded as a hierarchical multiple-agent network [25]. The collaborative consensus algorithm for ramp time is to make each agent update its own information state timely based on the ramp time of its neighboring agents, so that the information state of all agents in the network converges to a common value. Supposing that the GUG is a multiple-agent network with the p agents, which are respectively indicated by $p (p = 1, \ldots, P)$. The relationship between the interactive agents is indicated by the graph $G = (V, E, A)$. The node set is $V = (V_p, p = 1, \ldots, P)$, and each node indicates an agent. The edge set is $E \in V \times V$, and the element indicates the relationship between the interactive agents by a directed or undirected communication connection [26].

Assume that the connection between the interactive agents v_p and v_q is resolved by the probability b_{pq}. The laplace matrix $\boldsymbol{L} = [l_{pq}]$ can reflect the topology of the multiple-agent network [27], which is described as follows:

$$\begin{cases} l_{pp} = \sum_{q=1, p \neq q}^{P} b_{pq} \\ l_{pq} = -b_{pq} \end{cases}, \quad \forall p \neq q \tag{15}$$

The ramp time of each generator unit in GUG is selected as the consensus variable, and more disturbances are assumed by the leader of the units with a larger ramp rate in this paper. The power command of the wth unit of the GUG_i in the territorial power grid is described as shown in (16).

$$\Delta P_{iw} = t_{iw} \times \Delta P_{iw}^{\text{rate}} \tag{16}$$

where t_{iw} and $\Delta P_{iw}^{\text{rate}}$ are the ramp time and ramp rate of the wth generator unit of the GUG_i. And $\Delta P_{iw}^{\text{rate}}$ is described as follows:

$$\Delta P_{iw}^{\text{rate}} = \begin{cases} \Delta P_{iw}^{\text{rate}+}, & \Delta P_i > 0 \\ \Delta P_{iw}^{\text{rate}-}, & \Delta P_i < 0 \end{cases} \tag{17}$$

The ramp time of each follower is updated by (18) in GUG.

$$t_{iw}[k+1] = \sum_{v=1}^{W_i} d_{wv}[k] t_{iv}[k] \tag{18}$$

where W_i is the generator units number of the GUG_i, and $d_{wv}[k]$ indicates the term $[w, v]$ of the random row matrix $\boldsymbol{D} = d_{wv}[k] \in R^{Wi \times Wi}$ of discrete time k, which is described as follows:

$$d_{wv}[k] = \frac{|l_{wv}|}{\sum_{v=1}^{W_i} |l_{wv}|}, w = 1, 2, \ldots, W_i \tag{19}$$

And the ramp time of the leader is selected as follows according to [28]:

$$t_{iw}[k+1] = \begin{cases} \sum\limits_{v=1}^{W_i} d_{wv}[k] t_{iv}[k] + \sigma_i \Delta P_{\text{error-}i}, & \text{if } \Delta P_i > 0 \\ \sum\limits_{v=1}^{W_i} d_{wv}[k] t_{iv}[k] - \sigma_i \Delta P_{\text{error-}i}, & \text{if } \Delta P_i < 0 \end{cases} \tag{20}$$

where σ_i is the factor of the power regulation in GUG_i, and $\Delta P_{\text{error-}i}$ is the power deviation for GUG_i, which is shown in (21).

$$\Delta P_{\text{error-}i} = \Delta P_i - \sum_{w=1}^{W_i} \Delta P_{iw} \tag{21}$$

Meanwhile, the power commands and the maximum ramp time are described in (22) and (23).

$$\Delta P_{iw} = \begin{cases} \Delta P_{iw}^{\max}, & \Delta P_{iw} > \Delta P_{iw}^{\max} \\ \Delta P_{iw}^{\min}, & \Delta P_{iw} < \Delta P_{iw}^{\min} \end{cases} \tag{22}$$

$$t_{iw} = t_{iw}^{\max} = \begin{cases} \dfrac{\Delta P_{iw}^{\max}}{\Delta P_{iw}^{\text{rate}+}}, & \Delta P_{iw} > \Delta P_{iw}^{\max} \\ \dfrac{\Delta P_{iw}^{\min}}{\Delta P_{iw}^{\text{rate}-}}, & \Delta P_{iw} < \Delta P_{iw}^{\min} \end{cases} \tag{23}$$

where $\Delta P_{iw}^{\max}$ and $\Delta P_{iw}^{\min}$ are the maximum and minimum capacity of the wth units in GUG_i.

The weighed factor can be selected as shown in (24) while the power command ΔP_{iw} over the limitation.

$$b_{wv} = 0, v = 1, 2, \cdots, W_i \tag{24}$$

C. Reward Function Selection

With selecting the weighted sum of ACE and C_{total} as the reward function of the AGC control algorithm, a larger weighted sum will lead to a smaller reward. However, considering the physical dimension inconsistency between ACE and C_{total}, the C_{total} is divided into 50,000, that is, ACE and C_{total} are classified into the same order of magnitude, which reflects the significance of the comprehensive objective function. And finally, the reward function of MSGP is defined as follows:

$$R_1(s_n, s_{n+1}, a_n) = -\mu [ACE(n)]^2 - \frac{(1-\mu) C_{\text{total}}(n)}{50000} \tag{25}$$

where *ACE* (n) is the absolute instantaneous value of the ACE at the nth iteration; $C_{\text{total}}(n)$ is the actual generation cost of all generator units at the nth iteration; μ and $1-\mu$ are the reward weighting ratios of ACE and C_{total}. The parameter μ is set to 0.5 because of the two control performance indexes are of the same importance, that is, ACE and C_{total} can be regarded as having the same weight.

D. Parameter Setting

The parameters of the system need to be set properly in the design of the distributed AGC. It proves that pretty good effects can be obtained by setting the parameters through enough simulations and trial-and-errors as shown in Table I.

TABLE I
AGC PARAMETER SETTINGS

Parameters	Quantity	Setting value
α_1	learning rate of the MSGP algorithm	0.1
α_2	learning rate of the CQ algorithm	0.5
β	learning rate of the policy iteration	0.3
γ_1	reward discount factor of the MSGP algorithm	0.9
γ_2	reward discount factor of the CQ algorithm	0.9
k	step-size factor of the MSGP algorithm	0.88
λ	eligibility trace attenuation factor	0.9

1) The learning rate of value function α is set between 0 and 1, which weights the stability of the MSGP-CQ strategy. A larger value α can accelerate the policy iteration learning rate of the Q_k and Q value function, while a smaller α can enhance the system stability. And the learning rate α_1 of the MSGP algorithm and the learning rate α_2 of the CQ algorithm are selected as 0.1 and 0.5, respectively according to trial and error.

2) The learning rate of policy iteration β is set between 0 and 1, which weights the impact of the action selection policy on policy iteration updating. And a larger value β can accelerate the convergence rate of policy improvement, while a smaller β can ensure that the system can fully explore the other actions in the action space. Finally, the value β is chosen as 0.3 through trial and error.

3) The reward discount factor γ is also set between 0 and 1, which weights the importance of immediate rewards and future rewards. The more far-sighted the agent is, the more attention it attaches to the long-term rewards in the future when γ is larger. Meanwhile, more attention is attached to the immediate rewards when γ is close to 0. The reward discount factor γ_1 of the MSGP algorithm and the reward discount factor γ_2 of the CQ algorithm are both selected to be 0.9 through the simulations.

4) The step-size factor of multiple-step greedy policy k is between 0 and 1, which calculates the step-size number of greedy policies for the action selection. The control performance is the best while the step-size is set to 5. Therefore the parameter k is calculated to be 0.88.

5) The eligibility trace attenuation factor λ allocates the credits among state-action pairs. Usually, the parameter is between 0 and 1, which determines the convergence rate and the non-Markov decision process effects for the large time delay system. And the factor λ is chosen as 0.9 after many simulations.

E. MSGP-CQ Procedure

The overall MSGP-CQ procedure of the execution steps is described in Fig. 2.

IV. ANALYSIS OF EXAMPLES

A. The Improved IEEE Two-Area LFC Power System

One equivalent generator in each area is replaced by three types of GUGs: the thermal power (TP), the liquefied natural gas (LNG) and large hydropower (LH) based on the IEEE standard two-area LFC power system model [29], [30]. Meanwhile, the type, and number and parameters of the generator units in area B are exactly the same as area A. The structure of the improved model of the IEEE two-area LFC power system is shown in Fig. 3. And the parameters of the model and the AGC generator units are set according to [16].

The real-time data of areas A and B are collected during the operation of the AGC system, and the current state of each area is observed to determine the action. The two areas share and exchange the information through AGC controllers to avoid the control inconsistency between the two areas. Thereby the optimal strategy can be quickly achieved.

1) Pre-learning

The algorithm MSGP-CQ needs sufficient pre-learning to optimize the Q_k and Q functions before on-line operations, that is, the algorithm is optimized and the optimal set of actions is acquired through random exploration and off-line trial-and-error training. In the pre-learning stage, a continuous sinusoidal load disturbance with the period of 5,000 s, the amplitude of ± 1,000 MW and the duration of 30,000 s is applied to areas A and B.

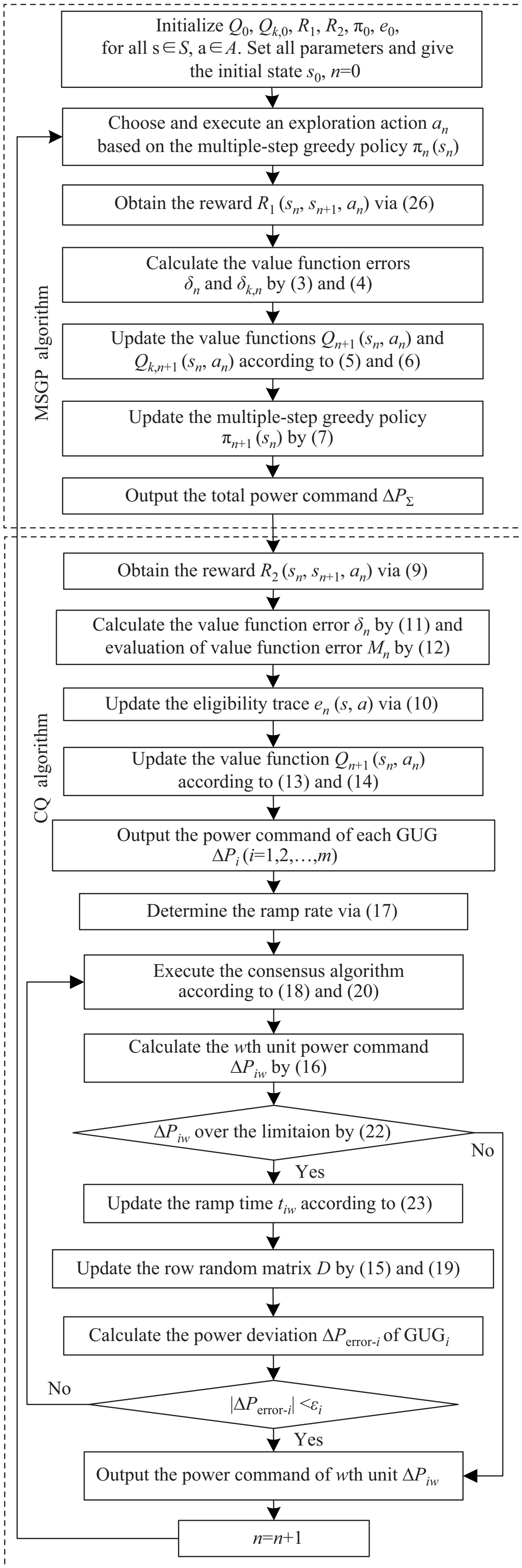

Fig. 2. Execution steps of the MSGP-CQ strategy.

The algorithm will approach a deterministic optimal strategy after enough explorations, so that it can execute optimal action quickly as the system obtains the real-time operation information. The AGC system also outputs the adjustment command of the total power generation in real-time. The pre-learning effect of MSGP-CQ is shown in Fig. 4, and the output of AGC controller based on the MSGP-CQ strategy has completely tracked the load disturbance after only experiencing the optimization of about 2,000 s through trial-and-error. It means that MSGP-CQ is a certain optimal strategy with a faster convergence rate approximately and can be run in real environments after massive training exploration.

The four algorithms MSGP-CQ, EPCC [22], HQL [18], and WPH [14] are put into the AGC system for pre-learning to compare the convergence effects. The 2-norm Q function matrix $\|Q_k(s,a) - Q_{k-1}(s,a)\|^2 \leqslant \varsigma$ ($\varsigma = 0.01$) is employed as a criterion for pre-learning to achieve an optimal strategy. Both the Q values and look-up table will be saved after pre-learning to ensure the application of MSGP-CQ into a real power system. The results of the Q-function differential convergence in area A during the pre-learning of the four algorithms are shown in Fig. 5. The convergence results show that the convergence speed of MSGP-CQ is improved by 81.25%–94.44% compared with other three algorithms. It illustrates that the convergence speed in the MSGP-CQ is much better, and the update stability of the value function is also improved very well compared with other algorithms.

2) White Noise Load Disturbance

The load disturbance of stochastic white noise is introduced into the improved two-area LFC power system to simulate the random disturbance caused by distributed energy access, and further verify the control performance in the proposed strategy after the pre-learning. It can be seen that the output power of the AGC controller can track the load disturbance accurately and quickly and maintain stable following in Fig. 6, while the frequency deviation Δf and ACE are always in the ideal range. And the total power command and the unit actual output can balance the load disturbance placidly. It indicates that the MSGP-CQ controller has a faster and smoother tracking effect on the load change, which fits the character that the smoother the unit's secondary frequency modulation process, the higher the generation efficiency. It also shows that the MSGP-CQ strategy has stronger adaptability in the distributed energy grid-connected environment.

The control performance of the four intelligent algorithms MSGP-CQ, EPCC, HQL, and WPH, is tested with the stochastic white noise disturbance for 24 hours. In this case, $|\Delta f|$ is the average value of the absolute values for the frequency deviation, |ACE| is the average value of the absolute value for ACE, and all the indicators are the average values in the simulation time. CPS1 evaluates the effect of ACE changes on system frequency, and CPS2 evaluates the ACE amplitude. The CPS index considers the distribution of CPS1 and CPS2 indicators, which are mainly used to evaluate the control performance of the entire AGC system. The statistical results are shown in Table II. The algorithm MSGP-CQ can decline $|\Delta f|$ by 15.51%–67.95% and |ACE| by 15.21%–50.83% compared with the other intelligent algorithms in area

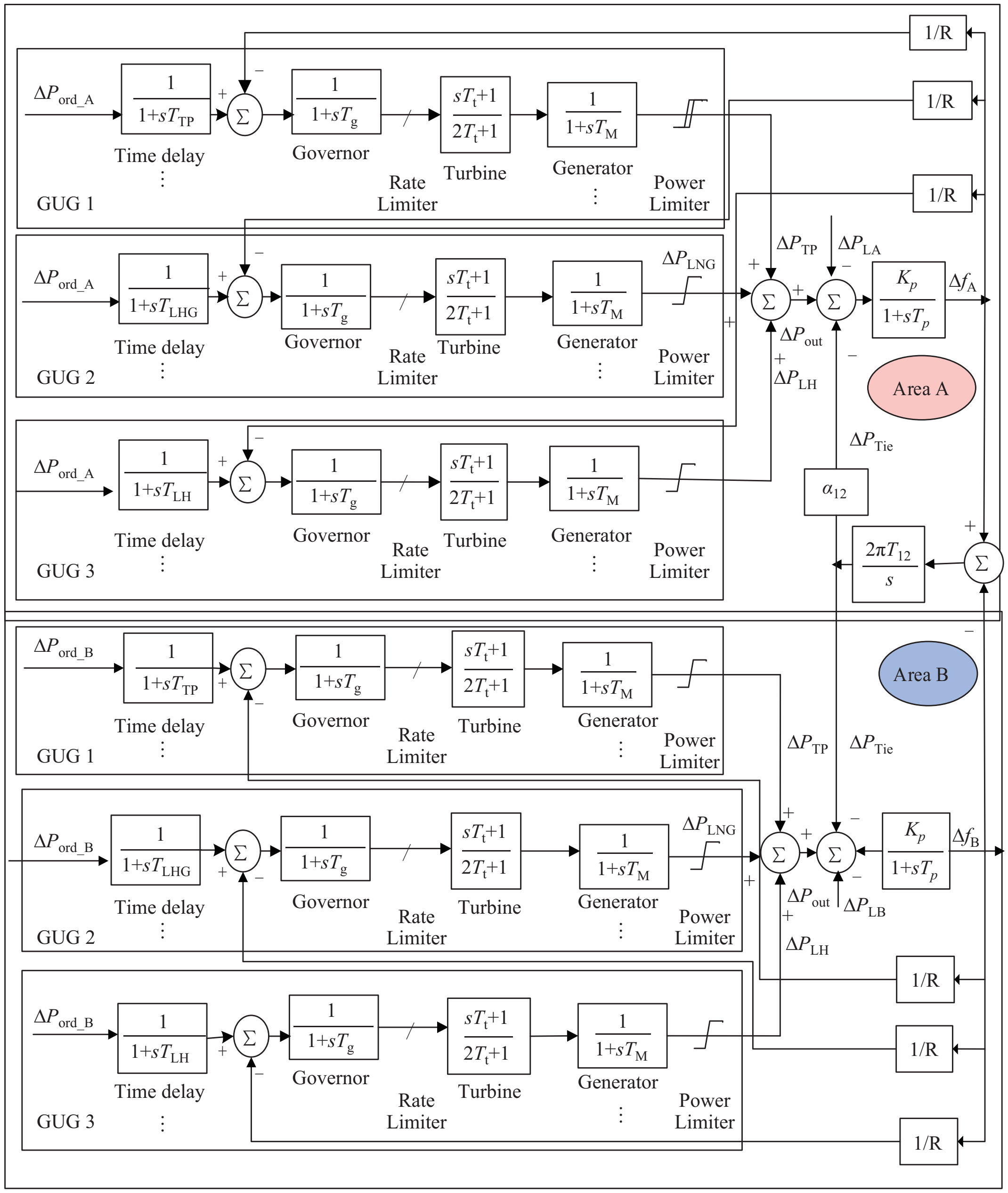

Fig. 3. The improved model of the IEEE Two-area LFC power system.

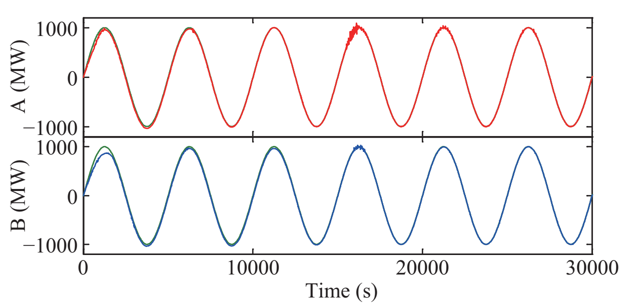

Fig. 4. The pre-learning of the MSGP-CQ strategy in area A.

TABLE II
THE CONTROL PERFORMANCE STATISTICS OF THE FOUR ALGORITHMS UNDER WHITE NOISE DISTURBANCE

Area	Algorithm	$\lvert\Delta f\rvert$ (Hz)	|ACE| (kW)	CPS1 (%)	CPS2 (%)
Area A	MSGP-CQ	0.0158	8.64	199.04	98.28
	EPCC	0.0187	10.19	198.27	97.04
	HQL	0.0404	15.67	198.01	97.26
	WPH	0.0493	17.57	196.87	96.18
Area B	MSGP-CQ	0.0175	9.06	198.84	99.10
	EPCC	0.0189	10.57	198.05	97.25
	HQL	0.0398	15.94	197.59	96.96
	WPH	0.0452	18.29	196.03	96.07

A. And the value of CPS1 and the value of CPS2 are also improved by 0.77%–2.17% and by 1.24%–2.10% respectively. It can be seen that the MSGP-CQ controller has a significant control effect on frequency with less output fluctuation, which can provide better control performance for AGC units in the condition of reducing control costs and the unit abrasion.

B. The Interconnected Power System of IDN Groups

In this paper, a model of the three-area interconnected power

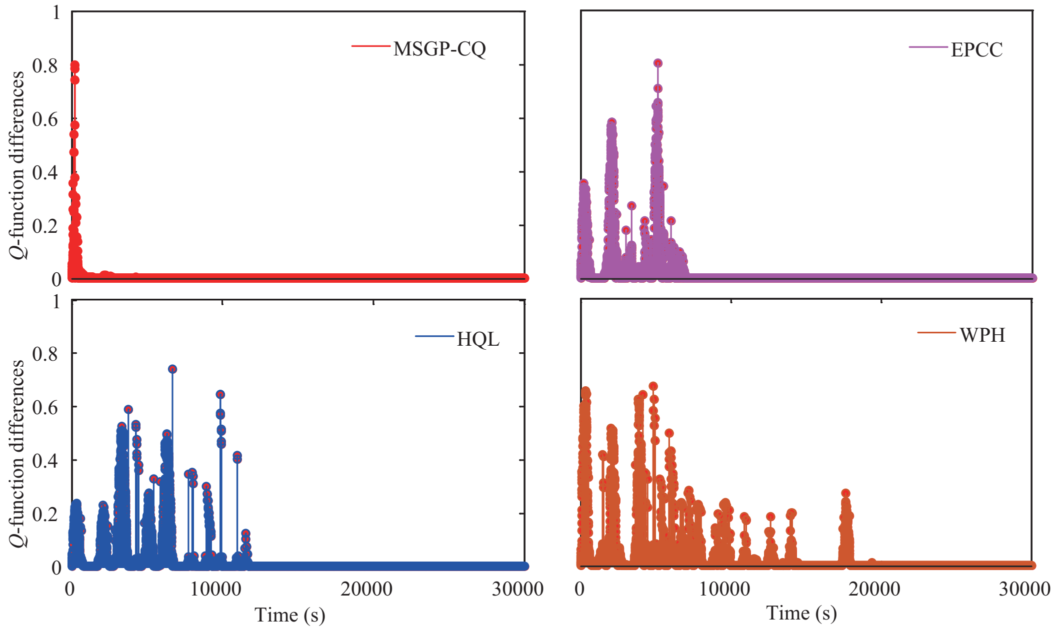

Fig. 5. Convergence results of four algorithms.

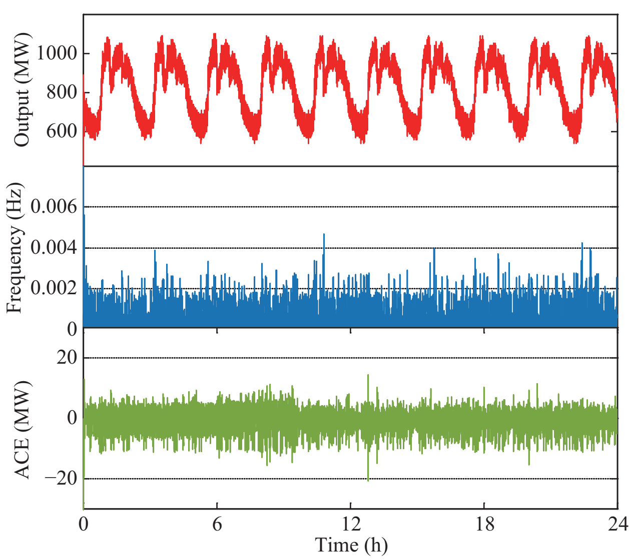

Fig. 6. The effect of MSGP-CQ under white noise disturbance.

system in the IDN groups with a large-scale distributed energy and load is built, which includes photovoltaic (PV), wind farm (WF), electric vehicle (EV), small hydropower (SH), micro-gas turbine (MT), diesel generator (DG), biomass energy (BE) and fuel cell (FC). It is shown in Fig. 7 that the model structure of the interconnected power system in IDN groups includes three areas: IDN 1, IDN 2 and IDN 3. The regulating power of the interconnected power system model in IDN groups are 2,350 kW, 2,590 kW and 2,350 kW, respectively. Each generator unit is regarded as a different agent, and the weighed factor b_{wv} of the agents is set to 1. The parameters of the model and the AGC generator units are set according to [31]–[33].

1) Impulsive Load Disturbance

The impulsive load disturbance is introduced to simulate the regular sudden increase and decrease of the load in the power system. The output power of each generator unit is regulated by the AGC controller and its own governor. The control performance of four intelligent algorithms MSGP-CQ, EPCC, HQL and WPH is tested with the impulsive load disturbance of 10,000 s. It shows the controller output in the different intelligent algorithms under the impulsive load disturbance in Fig. 8. It can be obviously seen that MSGP-CQ has smoother and faster regulation commands. The algorithm can reach better control effects, with faster dynamic optimal rate and stronger convergence.

Meanwhile, it is also shown in Fig. 9 that MSGP-CQ can decline $|\Delta f|$ by 19.62%–53.80% and |ACE| by 11.69%–56.06%. The value of CPS1 and the value of CPS2 are also improved by 0.72%–6.34% and by 2.15%–9.02% respectively. It fully proves that the control performance statistics of MSGP-CQ algorithm are better than that of the other algorithms with the impulsive load disturbance. And it also manifests that the MSGP-CQ controller has superior relaxation characteristics and better control performance.

2) The AGC Control Performance Under a Strong Stochastic Disturbance

The stochastic impulsive load disturbance is applied to the model of the interconnected power system in the IDN groups in order to verify the application effect of MSGP-CQ in a strong stochastic environment. Similarly, the long-term control performance in the four intelligent algorithms MSGP-CQ, EPCC, HQL and WPH under the strong stochastic environment is tested with the stochastic impulsive load disturbance for 24 hours.

The statistical results are shown in Table III, in which the generation cost is the sum of the total regulation cost of all generator units within 24 hours. It can also be calculated that MSGP-CQ can decline the cost of power generation by US$3,335–US$22,152 and $|\Delta f|$ by 28.24%–58.78%. It also

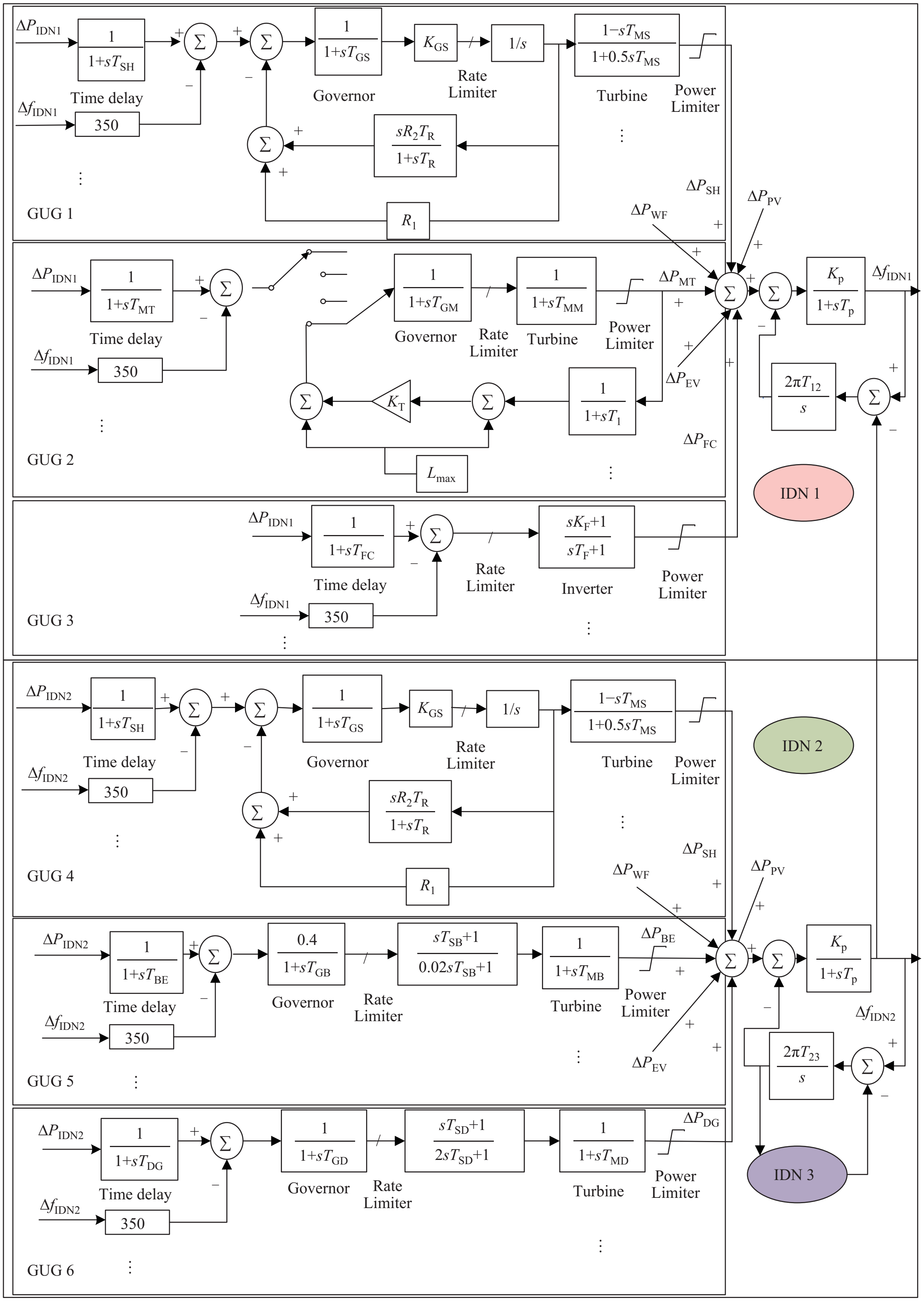

Fig. 7. The interconnected power system model of IDN groups.

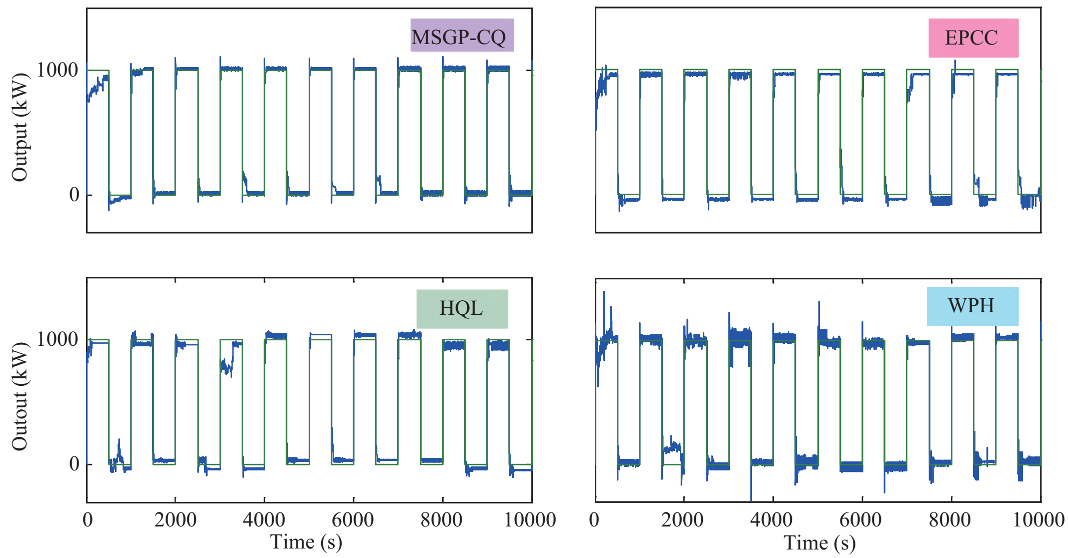

Fig. 8. The controller output in the different intelligent algorithms under the impulsive load disturbance.

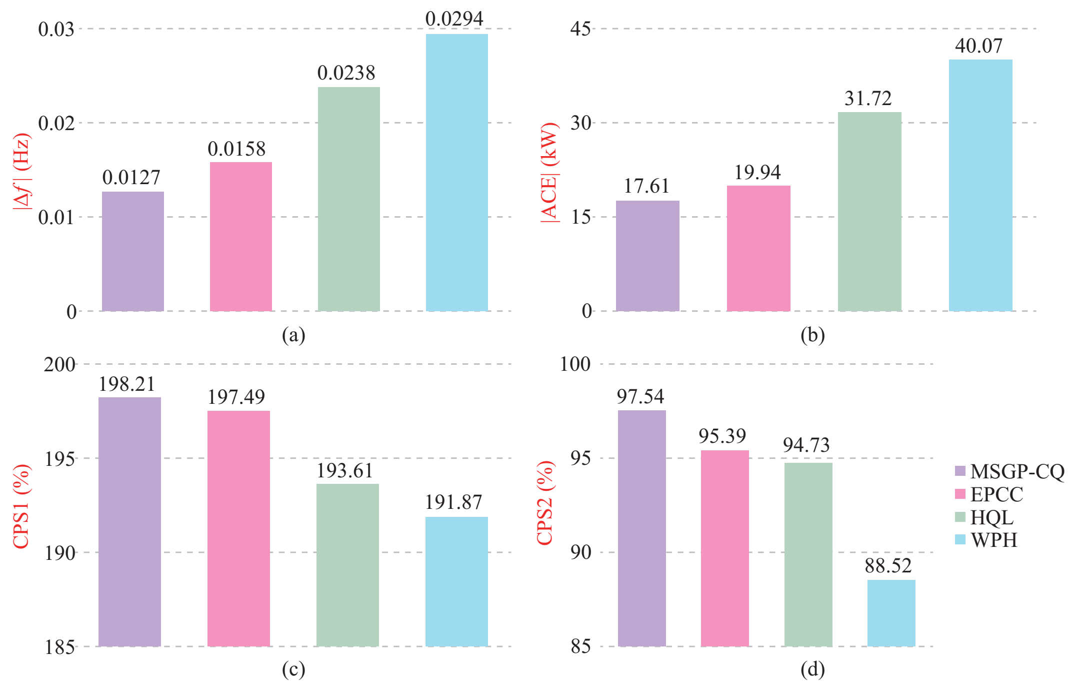

Fig. 9. The control performance in the different intelligent algorithms.

TABLE III
THE DATA STATISTICS UNDER STRONG STOCHASTIC DISTURBANCES

Area	Algorithm	$\|\Delta f\|$ (Hz)	|ACE| (kW)	CPS1%	CPS2%	Generation cost ($\times$ 10^4 US$)
IDN 1	MSGP-CQ	0.0216	16.73	199.21	99.67	15.0431
	EPCC	0.0301	23.47	196.08	95.03	15.3766
	HQL	0.0393	29.65	192.99	91.93	16.5142
	WPH	0.0524	35.82	187.81	87.22	17.2583
IDN 2	MSGP-CQ	0.0199	15.83	199.42	99.14	16.3723
	EPCC	0.0327	21.55	197.15	94.31	16.5960
	HQL	0.0399	29.98	191.57	90.42	16.9793
	WPH	0.0547	34.77	189.37	88.37	18.3902
IDN 3	MSGP-CQ	0.0231	17.29	199.01	99.01	15.0286
	EPCC	0.0382	23.02	196.36	93.22	15.4853
	HQL	0.0460	30.59	192.95	90.63	16.5540
	WPH	0.0537	36.37	185.21	86.57	17.1517

declines |ACE| by 28.72%–53.29% and improves the values of CPS1 and CPS2 by 3.31%–11.40% and by 4.64%–12.45% respectively, compared with the other intelligent algorithms in IDN 1. It can reduce the mechanical wear and economic cost caused by frequent operation in the unit. It can also be seen in the data analysis that MSGP-CQ has higher economy, stronger adaptability and better coordination and optimization control performance than that of the other three intelligent algorithms. According to the two examples, whether it is under the circumstance of impulse load disturbance or strong stochastic disturbance, MSGP-CQ can satisfy the AGC control performance standards and have more excellent control performance. The effectiveness and expandability of the algorithms referred in this paper have been verified in the experiment results.

V. Conclusion

A novel MSGP-CQ strategy is proposed to achieve the dynamic optimal control of the total power command for the distributed AGC and the optimal power allocation for generator units, so as to solve the problem of the strong stochastic disturbance caused by the large-scale distributed energy access to the power grid. Thereby, the overall coordinated optimal control of the power system is obtained.

The algorithm MSGP with the multiple-step greedy attribute is combined with the algorithm CQ with collaborative consensus in the proposed strategy. As a control algorithm in the strategy, MSGP updates the greedy policy of the action selection by several multiple-step look-ahead iterations, and quickly converges to the optimal policy. With the interactive collaboration and self-learning characteristics, CQ is used as the power allocation algorithm by constructing a hierarchical power allocation mode. It improves the adaptability of the consensus algorithm in the dynamic stochastic environment.

The improved model of the IEEE two-area LFC power system and the model of the interconnected power system in the IDN groups incorporating large amounts of distributed energy and load are constructed to simulate and verify the proposed strategy. The results illustrate that the MSGP-CQ strategy can reduce the generation cost of generator units, improve the utilization of distributed energy, and have a stronger robustness and faster dynamic optimization speed compared with that of the traditional intelligent algorithms.

References

[1] D. Xu, B. Zhou, Q. W. Wu, C. Y. Chung, C. B. Li, S. Huang, and S. Chen, "Integrated modelling and enhanced utilization of power-to-ammonia for high renewable penetrated multi-energy systems," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4769–4780, Nov. 2020.

[2] H. Lund, "Large-scale integration of wind power into different energy systems," *Energy*, vol. 30, no. 13, pp. 2402–2412, Oct. 2005.

[3] S. C. Wang, "Current status of PV in China and its future forecast," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 72–82, Mar. 2020.

[4] J. C. Mukherjee and A. Gupta, "Distributed charge scheduling of Plug-In electric vehicles using Inter-Aggregator collaboration," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 331–341, Jan. 2017.

[5] H. Z. Wang, Z. X. Lei, X. Zhang, B. Zhou, and J. C. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, pp. 111799, Oct. 2019.

[6] D. Xu, Q. W. Wu, B. Zhou, C. B. Li, L. Bai, and S. Huang, "Distributed multi-energy operation of coupled electricity, heating, and natural gas networks," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2457–2469, Oct. 2020.

[7] Q. Y. Sun, N. Zhang, S. You, and J. W. Wang, "The dual control with consideration of security operation and economic efficiency for energy hub," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 5930–5941, Nov. 2019.

[8] M. Yazdanian and A. Mehrizi-Sani, "Distributed control techniques in microgrids," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2901–2909, Nov. 2014.

[9] W. Yan, R. F. Zhao, X. Zhao, C. Wang, and J. Yu, "Review on control strategies in automatic generation control," *Power System Protection and Control*, vol. 41, no. 8, pp. 149–155, Apr. 2013.

[10] C. X. Mu, Y. F. Tang, and H. B. He, "Improved sliding mode design for load frequency control of power system integrated an adaptive learning strategy," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 8, pp. 6724–6751, Aug. 2017.

[11] B. P. Padhy and B. Tyagi, "Artificial neural network based multi area Automatic Generation Control scheme for a competitive electricity market environment," in *2009 International Conference on Power Systems*, Kharagpur, India, 2009.

[12] T. Yu, H. Z. Wang, B. Zhou, K. W. Chen, and J. Tang, "Multi-agent correlated equilibrium Q(λ) learning for coordinated smart generation control of interconnected power grids," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1669–1679, Jul. 2015.

[13] L. Xi, T. Yu, B. Yang, and X. S. Zhang, "A novel multi-agent decentralized win or learn fast policy hill-climbing with eligibility trace algorithm for smart generation control of interconnected complex power grids," *Energy Conversion and Management*, vol. 103, pp. 82–93, Oct. 2015.

[14] L. Xi, T. Yu, B. Yang, X. S. Zhang, and X. Y. Qiu, "A wolf pack hunting strategy based virtual tribes control for automatic generation control of smart grid," *Applied Energy*, vol. 178, pp. 198–211, Sep. 2016.

[15] Z. D. Zhang, D. X. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 213–225, Mar. 2020.

[16] L. Xi, L. Yu, Y. C. Xu, S. X. Wang, and X. Chen, "A novel multi-agent DDQN-AD method-based distributed strategy for automatic generation control of integrated energy systems," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2417–2426, Oct. 2020.

[17] L. Xi, J. F. Chen, Y. H. Huang, Y. C. Xu, L. Liu, Y. M. Zhou, and Y. D. Li, "Smart generation control based on multi-agent reinforcement learning with the idea of the time tunnel," *Energy*, vol. 153, pp. 977–987, Jun. 2018.

[18] T. Yu, Y. M. Wang, W. J. Ye, B. Zhou, and K. W. Chen, "Stochastic optimal generation command dispatch based on improved hierarchical reinforcement learning approach," *IET Generation Transmission & Distribution*, vol. 5, no. 8, pp. 789–797, Aug. 2011.

[19] X. S. Zhang, Q. Li, T. Yu, and B. Yang, "Consensus transfer *Q*-learning for decentralized generation command dispatch based on virtual generation tribe," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 2152–2165, May 2016.

[20] D. X. Zhang, X. Q. Han, and C. Y. Deng, "Review on the research and practice of deep learning and reinforcement learning in smart grids," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 3, pp. 362–370, Sep. 2018.

[21] Y. Efroni, G. Dalal, B. Scherrer, and S Mannor, "Multiple-step greedy policies in online and approximate reinforcement learning," in *32nd Conference on Neural Information Processing Systems (NIPS 2018)*, Montréal, Canada, 2018.

[22] L. Xi, Y. D. Li, Y. H. Huang, L. Lu, and J. F. Chen, "A novel automatic generation control method based on the ecological population cooperative control for the islanded smart grid," *Complexity*, vol. 2018, pp. 2456963, Aug. 2018.

[23] C. J. C. H. Watkins and P. Dayan, "*Q*-learning," *Machine Learning*, vol. 8, no. 3–4, pp. 279–292, May 1992.

[24] R. S. Sutton and A. G. Barto, "Reinforcement learning: an introduction," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 1054, Sep. 1998.

[25] Q. Y. Sun, R. Y. Fan, Y. S. Li, B. N. Huang, and D. Z. Ma, "A distributed double-consensus algorithm for residential WE-energy," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4830–4842, Aug. 2019.

[26] G. Merlet, T. Nowak, H. Schneider, and S. Sergeev, "Generalizations of bounds on the index of convergence to weighted digraphs," *Discrete Applied Mathematics*, vol. 178, pp. 121–134, Dec. 2014.

[27] L. H. Ji, Y. Tang, Q. Liu, and X. F. Liao, "On adaptive pinning consensus for dynamic multi-agent networks with general connected topology," in *2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, Canada, 2016.
[28] L. Sun, F. X. Yao, S. C. Chai, and Y. G, Xu, "Leader-following consensus for high-order multi-agent systems with measurement noises," in *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, China, 2016.
[29] L. Xi, J. N. Wu, Y. C. Xu, and H. B. Sun, "Automatic generation control based on multiple neural networks with Actor-Critic strategy," *IEEE Transactions on Neural Networks and Learning Systems*, 2020, doi: 10.1109/TNNLS.2020.3006080.
[30] G. Ray, A. N. Prasad, and G. D. Prasad, "A new approach to the design of robust load-frequency controller for large scale power systems," *Electric Power Systems Research*, vol. 51, no. 1, pp. 13–22, Jul. 1999.
[31] L. Xi, L. Zhang, J. C. Liu, Y. D. Li, X. Chen, L. Q. Yang, and S. X. Wang, "A virtual generation ecosystem control strategy for automatic generation control of interconnected microgrids," *IEEE Access*, vol. 8, pp. 94165–94175, May 2020.
[32] A. K. Saha, S. Chowdhury, S. P. Chowdhury, P. A. Crossley, "Modelling and simulation of microturbine in islanded and grid-connected mode as distributed energy resource," in *2008 IEEE Power and Energy Society General Meeting-conversion and Delivery of Electrical Energy in the 21st Century*, Pittsburgh, PA, USA, 2008.
[33] L. Xi, L. Zhou, L. Liu, et al., "A deep reinforcement learning algorithm for the power order optimization allocation of AGC in interconnected power grids," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 3, pp. 712–723, 2020.

Lei Xi received the M.S. degree in Control Theory and Control Engineering from Harbin University of Science and Technology, and the Ph.D. degree in Electrical Engineering from South China University of Technology, China, in 2016. Currently, he is an associate professor in the College of Electrical Engineering and New Energy, China Three Gorges University. His research interests include load frequency control, artificial intelligence techniques and automatic generation control.

Le Zhang received the bachelor degree in Electrical Engineering from Southwest University of Science and Technology, China, in 2018. He is currently pursuing his M.S. degree at the College of Electrical Engineering and New Energy, China Three Gorges University. His research interests include smart generation control and artificial intelligence techniques.

Yanchun Xu received the Ph.D degree from the Department of Electrical Engineering, Harbin Institute of Technology(HIT), Harbin, China, in 2010. She works at the China Three Gorges University, Yichang, China. Her research interests include power quality detection with distributed generation, harmonics detection of power systems, as well as matric converters applied in power systems.

Shouxiang Wang (SM'12) received the B.S. and M.S. degrees from Shandong University, Jinan, China, in 1995 and 1998, respectively, and the Ph.D. degree from Tianjin University, Tianjn, China, in 2001, all in electrical engineering. He is currently a Professor in the School of Electrical and Information Engineering at Tianjin University. His main research interests are distributed generation, microgrid and smart distribution systems.

Chao Yang received the M.S degree in Control Engineering from China Three Gorges University, China in 2017. Currently, he is a lecturer in the College of Electrical Engineering and New Energy at China Three Gorges University, Yichang, China. His research interests include Machine vision and industrial control technology.