

Research Article

Research on Flexible Resource Dynamic Interactive Regulation Technology for Microgrids with High Permeable New Energy

Songsong Chen,¹ Lutao Zhang ,¹ Ying Zhou,¹ Ke Chen,¹ Zhongdong Wang,² and Lihui Xie ³

¹China Electric Power Research Institute, Beijing 100192, China

²State Grid Jiangsu Electric Power Co, Jiangsu, Nanjing 210024, China

³College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China

Correspondence should be addressed to Lutao Zhang; zhanglutao0502@163.com and Lihui Xie; xielihui2022@163.com

Received 8 December 2022; Revised 13 January 2023; Accepted 22 March 2023; Published 15 April 2023

Academic Editor: Xueqian Fu

Copyright © 2023 Songsong Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The efficiency of on-site consumption of new energy and the economy of dispatching strategy for that in modern microgrids are increasingly concerning, which are closely related to the microgrid control model with source-load uncertainty. To this end, this paper proposes the multiagent hierarchical IQ (λ)-HDQC regulation strategy to realize the source-load-storage-charging collaborative control of the microgrid model with high-permeable new energy. The first layer adopts the IQ (λ) strategy, which avoids the overestimation and underestimation problems of traditional reinforcement learning by the coupled estimation method. The second layer adopts the HDQC allocation strategy, which solves the problem of low utilization of new energy in the proportional allocation method and improves the adaptability of the regulation strategy in the complex stochastic environment. The interaction of the two-layer strategies realizes the source-network-load-storage-charge global dynamic interactive regulation of microgrids. Indicators of energy efficiency are constructed in this paper to measure the simulation results. And the superiority of the proposed strategy is verified through the simulation results of the microgrid system.

1. Introduction

Under the pressure of energy demand and environmental protection, renewable energy generation is gradually attracting attention. As an effective carrier of renewable energy, the application of microgrids [1–3] reduces the impact of the randomness of renewable energy output on the stability of the power systems, which is an effective way to improve the utilization and penetration rate of new energy. However, due to the lack of grid support and environmental uncertainty, the energy autonomy of microgrids faces many challenges [4–6], and how to solve the energy autonomy of microgrids becomes a hot issue.

For microgrids to realize the efficient consumption of new energy, they must realize the priority consumption of new energy in the microgrid. With the development of artificial intelligence, the research on the automatic generation control [7–9] (AGC) realizes the global dynamic interactive

adjustment of source-network-load-storage-charging (SNLSC) for microgrids. Wu et al. [10] propose an extreme Q-learning algorithm to parameterize the sag control of the microgrid, thus achieving the integration of frequency regulation and economic dispatch. However, the above method has the problem of “overestimation” of the action value in the exploration process in a strong stochastic environment. To solve this problem, Xi and Zhou [11] propose the DQ forecast (σ, λ) algorithm, which improves the fast and stable power regulation of AGC units but generates a new “underestimation” problem. The allocation of the total power command is achieved by a fixed proportion of adjustable capacity, and the new energy generation model has a strong nonlinearity, which makes it easy to fall into the local optimal solution and leads to the curse of dimensionality.

For this reason, multiagent reinforcement learning is applied to AGC to achieve a dynamic allocation of

conventional units and new energy output of microgrid systems. The paper [12] proposed an ecological population cooperative control strategy with a win-lose criterion and the space-time tunneling idea, which can converge to Nash equilibrium quickly, which is based on the stochastic consensus of a multiagent system (MAS). This strategy is based on the multiagent system stochastic consensus game framework to achieve frequent information exchange among multiagent. The paper [13] establishes a three-level architecture MAS to achieve coordinated control of AGC and automatic voltage control and utilizes the characteristics of independent autonomy and collaboration of agents to achieve coordinated control in physical distribution control while maintaining logical unity. However, in the case of large-scale distributed energy access to a microgrid, the convergence speed of the above method decreases, and the main problems it faces are the low backup capacity of the microgrid system, the difficulty of local consumption of new energy, and the decreased convergence accuracy of the previous algorithm.

Therefore, to solve the above problems, this paper improves on both the two-layer strategy and model. The distributed AGC strategy can be divided into an AGC control strategy and an AGC allocation strategy. In order to mitigate the impact of adding large amounts of new energy to the grid, interleaved Q-Learning [14] (IQ) is introduced in the control algorithm, which avoids both the “overestimation” problem generated by the maximum estimator (ME) and the “underestimation” problem generated by the double estimator (DE) and incorporates eligibility traces [15] to reduce the control bias. In the allocation part, a multiagent hierarchical strategy of hierarchical double Q-learning consensus (HDQC) is formed using a consistency algorithm with isomorphic properties [16], incorporating the double Q-learning (DQ) algorithm [17]. The simulation of the

microgrid with EVs incorporating large-scale new energy sources shows that, compared with previous agent algorithms, the proposed scheme can fully utilize the new energy sources and realize the global dynamic interactive regulation of SNLSC.

2. High Penetration New Energy Microgrid Control Framework

2.1. Microgrid Control Architecture. As shown in Figure 1, microgrid units that incorporate a large amount of new energy have great differences in ramp rate and spatial distance. HDQC allocation strategy uses clustering to divide the generating units into different power generation groups (PGG) and selects the generating unit with the largest content in the PGG as the dominant unit. The IQ (λ)-HDQC regulation strategy uses IQ (λ) to obtain the total power generated in the microgrid system, and then the HDQC strategy allocates the total power command to each unit in the PGG to realize the global dynamic interactive regulation of the microgrid SNLSC.

2.2. Microgrid Distribution Model. The HDQC allocation strategy takes area control error, creep time, and energy efficiency as three objectives and constructs two multi-objective functions within the microgrid. Among them, the objective function h_1 is to minimize the sum of area control error (ACE) and the maximum creep time of all generating units in the microgrid. The objective function h_2 is to minimize the ratio of carbon emission (CE) and non-renewable energy generation. Therefore, the mathematical model of the power command allocation process of the microgrid based on the HDQC allocation strategy is as follows:

$$\left\{ \begin{array}{l} \min h_1 = \sum_{i=1}^m A^2 + \sum_{i=1}^m \sum_{w=1}^{W_i} \frac{\Delta P_{iw}}{P_{iw}^{\text{rate}}}, \\ \min h_2 = \sum_{i=1}^m C_{\text{total}} + \sum_{i=1}^m \frac{P_n}{P}, \\ \text{s.t. } A = \Delta P_{\text{tie}} - 10B\Delta f, \\ \Delta P_i = \eta_i \Delta P_{\Sigma}, \\ L_{iw} \leq P_{iw}^{\text{rate}} \leq U_{iw}, \\ \Delta P_{iw}^{\text{min}} \leq \Delta P_{iw} \leq \Delta P_{iw}^{\text{max}}, \\ \sum_{i=1}^m \eta_i = 1, 0 \leq \eta_i \leq 1, \\ i = 1, 2, \dots, m; w = 1, 2, \dots, W_i, \end{array} \right. \quad (1)$$

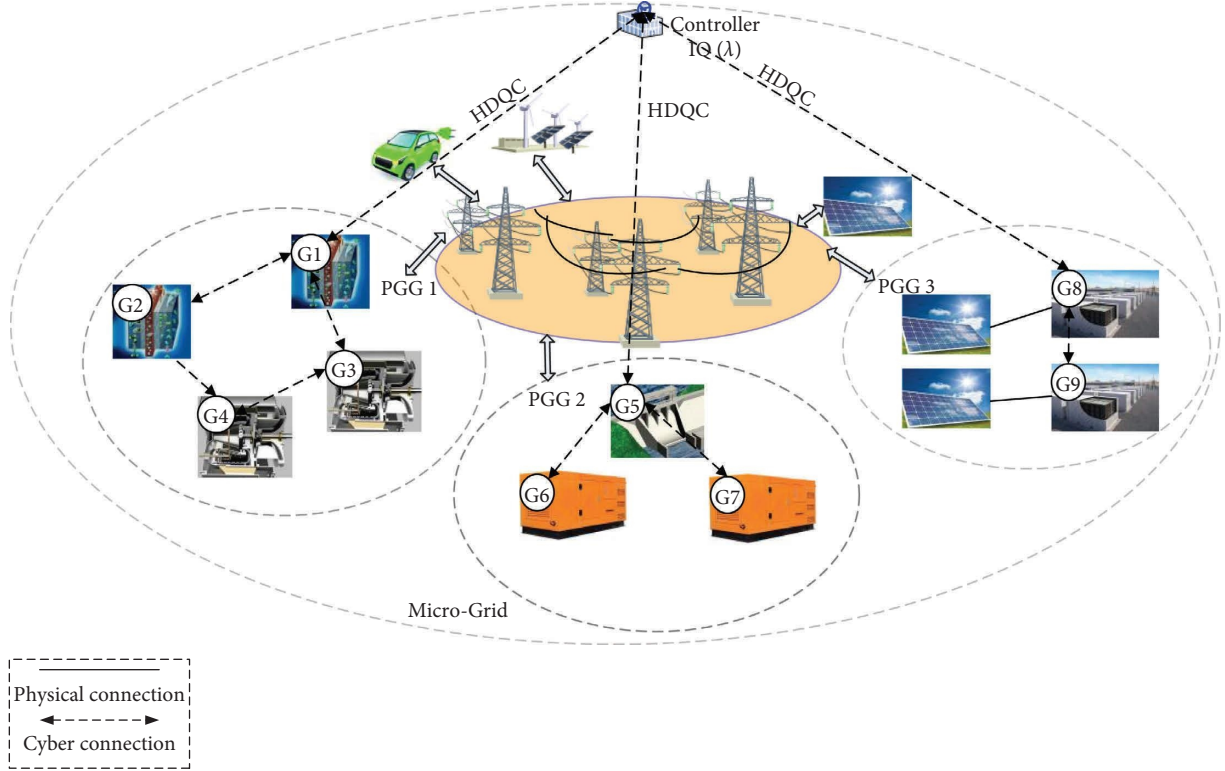


FIGURE 1: A microgrid control framework based on an IQ(λ)-HDQC strategy.

where A is the ACE of the microgrid system; C_{total} denotes the sum of CE for all units in the microgrid system; P_{iw} and P_{iw}^{rate} are the power command and regulation rate of the w th unit in PGG $_i$, respectively; P_n and P are the non-new energy generation power and total generation power of the microgrid, respectively; P_{fie} is the contact line exchange power; f is the frequency deviation and B is the frequency response coefficient; P_i is the power command of PGG $_i$, which is the product of the distribution factor η_i and the total regulation power command P_{Σ} of the system; U_{iw} and L_{iw} are the upper and lower limits of the power regulation rate of the w th unit in PGG $_i$, respectively; P_{iw}^{max} and P_{iw}^{min} are the upper and lower limits of the power regulation capacity of the w th unit in PGG $_i$; m is the number of PGG; and W_i is the number of units in PGG $_i$.

3. IQ(λ)-HDQC Regulation Strategy

The IQ(λ)-HDQMP regulation strategy requires both control and allocation of AGC. For control, the IQ(λ) algorithm is used to improve the fast convergence and control performance of Q-learning in a strongly stochastic environment; for allocation, the HDQC algorithm is used to solve the “dimensional catastrophe” problem caused by the proliferation of large-scale units using a novel hierarchical Q-learning strong consistency algorithm. The method can achieve fast convergence in the two-tier power allocation.

3.1. IQ(λ) Control Strategy. In traditional reinforcement learning, the maximum expectation estimation represented by Q-learning excessively pursues the maximum long-term discounted reward. It tends to choose actions corresponding to the maximum Q value, leading to an overestimation of action values during the strategy exploration process. The dual estimation method represented by DQ learning uses a more conservative strategy, which gave rise to the underestimation of action values. Both methods affect the optimal strategy exploration by the agents. For this reason, this paper incorporates eligibility traces based on the IQ algorithm using the coupled estimation method and then proposes a new IQ(λ) algorithm with fast convergence properties by reducing the difference of Q values.

3.1.1. Maximum Expectation Estimation. Q-learning always picks the action with the highest Q value, called a greedy strategy π^* , as in the following equation:

$$\pi^*(s) = \operatorname{argmax} Q^k(s, a), \quad (2)$$

where s is the current state; Q^k is the k th iteration of the optimal value function Q^* . $Q(s, a)$ is the Q-value function under the state s and strategy a . Based on the greedy strategy, the Q-learning algorithm uses iterative computation to find the optimal Q-value function, and the Q-value iteration is communicated as follows:

$$Q^{k+1}(s_k, a_k) = Q^k(s_k, a_k) + \alpha_1 \left[R(s_k, a_k) + \gamma \max_{a_{k+1} \in A} Q^k(s_{k+1}, a_{k+1}) - Q^k(s_k, a_k) \right], \quad (3)$$

where α is the learning rate; $R(s_k, a_k)$ is the reward function under the state s_k and policy a_k ; γ is the discount factor. Always choosing the action with the highest Q value will result in agents that always follow the same path and do not adequately search for other actions in the space, often converging to a local optimum.

3.1.2. Double Estimation. DQ learning uses two disjoint value functions Q_A and Q_B instead of a single value function Q . The

behavioral strategies for Q_A and Q_B are chosen as π_B^* and π_A^* , respectively, as follows:

$$\pi_A^*(s) = \operatorname{argmax} Q_B^*(s, a), \quad (4)$$

$$\pi_B^*(s) = \operatorname{argmax} Q_A^*(s, a). \quad (5)$$

DQ learning splits the strategy selection and estimation process to avoid overestimation of Q . The DQ learning iteration is updated as follows:

$$\begin{aligned} Q_A(s_k, a_k) &= Q_A(s_k, a_k) + \alpha_1 [R(s_k, a_k) + \gamma_1 Q_B(s_{k+1}, a_{k+1}) - Q_A(s_k, a_k)], \\ Q_B(s_k, a_k) &= Q_B(s_k, a_k) + \alpha_1 [R(s_k, a_k) + \gamma_1 Q_A(s_{k+1}, a_{k+1}) - Q_B(s_k, a_k)]. \end{aligned} \quad (6)$$

DQ learning completely decouples the selection and estimation processes not only avoiding overestimation of the true value but also introducing an underestimation problem that slows down the convergence of the algorithm.

3.1.3. IQ(λ) Learning. IQ learning avoids the above problems of overestimation and underestimation of reinforcement learning by coupling the sample set. The Q-value function error estimate for IQ learning is given by the following equation:

$$M_k^A = R(s_k, a_k) + \sigma * \gamma_2 (Q_A(s_{k+1}, a_{k+1}^A) + Q_B(s_{k+1}, a_{k+1}^B)) - Q_B(s_k, a_k), \quad (7)$$

$$M_k^B = R(s_k, a_k) + \sigma * \gamma_2 (Q_A(s_{k+1}, a_{k+1}^A) + Q_B(s_{k+1}, a_{k+1}^B)) - Q_A(s_k, a_k), \quad (8)$$

where M_k^A and M_k^B are the evaluation errors of Q_A and Q_B functions, respectively; a_{k+1}^A and a_{k+1}^B are based on the actions selected by equations (4) and (5), respectively; σ ($0 < \sigma < 0.5$) is the coupling ratio, which reflects the proportion of shared states of Q_A and Q_B . The closer the value is to 0, the more IQ learning is underestimated state when $\sigma = 0.5$, IQ learning is totally overestimated state. The simulation comparison study shows that $\sigma = 0.25$ has a better effect. Incorporating eligibility traces based on IQ learning to retrace past information, the SARSA eligibility trace algorithm is selected in this paper as follows:

$$e_{k+1}(s, a) = \begin{cases} \gamma \lambda e_k(s, a) + 1, & (s, a) = (s_k, a_k), \\ \gamma \lambda e_k(s, a), & \text{otherwise,} \end{cases} \quad (9)$$

where $e_k(s, a)$ is the k th step iteration under state s and action a . The IQ (λ) algorithm is updated as follows:

$$\begin{aligned} Q_A(s_k, a_k) &= Q_A(s_k, a_k) + \alpha_2 M_k^A e_k^B(s, a), \\ Q_B(s_k, a_k) &= Q_B(s_k, a_k) + \alpha_2 M_k^B e_k^A(s, a). \end{aligned} \quad (10)$$

3.2. HDQC Algorithm. The introduced HQL algorithm [18] enables interactive learning and self-learning process among PGGs. Since the algorithm is based on the Q (λ) algorithm, the time tunneling method is iteratively updated as in

equation (7) and incorporates DQ learning to form the HDQC algorithm.

Suppose there are N agents in PGG $_i$, denoted by $p = \{1, \dots, N\}$. $G = (V, E)$ denotes the multiagent communication undirected graph, and $V = \{v_1, \dots, v_p\}$ is the set of nodes and $E \subseteq V \times V$ is the set of edges. The Laplacian matrix $L = [l_{ij}]$ reflects the topology of the multiagents body network [19], which is represented as follows:

$$\begin{cases} l_{ii} = \sum_{j=1, j \neq i}^N b_{ij}, \\ \forall i \neq j, l_{ij} = -b_{ij}, \end{cases} \quad (11)$$

where b_{ij} is the probability of communication between agents v_i and v_j ($i \neq j$; $i, j = 1, \dots, N$).

The ramp time is chosen as a consistent variable for PGG. The ramp time of the w th unit of the i th PGG in the regional grid is expressed as follows:

$$t_{iw} = \frac{\Delta P_{iw}}{\Delta P_{iw}^{\text{rate}}}, \quad (12)$$

where ΔP_{iw} and $\Delta P_{iw}^{\text{rate}}$ denote the power generation command of the w th unit within the i th PGG and the ramp rate of the unit, respectively, where $\Delta P_{iw}^{\text{rate}}$ is expressed as follows:

$$\Delta P_{iw}^{\text{rate}} = \begin{cases} \Delta P_{iw}^{\text{rate}+}, \Delta P_i > 0, \\ \Delta P_{iw}^{\text{rate}-}, \Delta P_i < 0. \end{cases} \quad (13)$$

The ramp time of the w th agent within the PGA is updated as follows:

$$t_{iw}[k+1] = \sum_{j=1}^{N_i} d_{jw}[k] t_{iw}[k], \quad (14)$$

$$d_{jw}[k] = \frac{|l_{jw}|}{\sum_{w=1}^{N_i} |l_{jw}|}, \quad j = 1, \dots, N_i.$$

Meanwhile, power correction is required according to whether the power is balanced within the microgrid by judging $\Delta P_{\text{error}-i}$, which is the power correction command for the i th PGG and is expressed as follows:

$$\Delta P_{\text{error}-i} = \Delta P_i - \sum_{w=1}^{W_i} \Delta P_{iw}. \quad (15)$$

Under the condition of frequent information interactions between agents and constant gain b_{ij} , collaborative consistency of agents can be achieved when and only when the directed graph is strongly connected [20].

When the boundary condition is reached, the generated power command ΔP_{iw} with the maximum ramp time is as follows:

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

$$t_{iw} = t_{iw}^{\text{max}} = \begin{cases} \frac{\Delta P_{iw}^{\text{max}}}{\Delta P_{iw}^{\text{rate}+}}, \Delta P_{iw} > \Delta P_{iw}^{\text{max}}, \\ \frac{\Delta P_{iw}^{\text{min}}}{\Delta P_{iw}^{\text{rate}-}}, \Delta P_{iw} < \Delta P_{iw}^{\text{min}}, \end{cases} \quad (16)$$

where $\Delta P_{iw}^{\text{max}}$ and $\Delta P_{iw}^{\text{min}}$ are the maximum and minimum power reserve capacity of the w th unit of the GUG _{i} , respectively.

4. Simulation Design

Different regional grids play a multi-intelligent body dynamic game through an IQ (λ) control strategy to obtain the total power of the region. Within each regional grid, according to the spatial distance and generator type, the microgrid system is virtually partitioned into multiple PGGs using the graph theory cut-set method, and each PGG is regarded as a multi-intelligent body system, which dynamically allocates the total regulation power command to each unit through the HDQC strategy and implements the regional boundary power exchange control to jointly maintain the global dynamic interactive regulation of the microgrid SNLSC.

4.1. Reward Function Design. In order to judge the control performance of the regional grid system, the three main performance evaluation criteria of AGC (area control error (ACE), interconnection grid frequency deviation (Δf) and control performance standard (CPS) [21]) and energy efficiency are used as the input of reward function, which can evaluate whether the current decision can obtain long-term benefits and avoid large power fluctuations. The agent calculates and updates the state quantity and reward function of the system in real time and outputs the optimal control signal $\Delta P_{\text{ord}-i}$ (the power regulation command of the i th unit).

4.1.1. IQ (λ) Reward Function Design. After dimensionless processing, $A(i)$ (the instantaneous values of ACE) and $\Delta f(i)$ (the instantaneous values of Δf) are normalized linearly weighted to obtain the target reward function as follows:

$$R_1 = -100\mu|\Delta f(i)|^2 - 0.1 * (1 - \mu)|A(i)|^2, \quad (17)$$

where μ is the weighting factor taken as 0.5.

4.1.2. HDQMP Reward Function Design. The dimensionless processed $A(i)$ and the linear weighting and energy efficiency are selected as the reward function, which is shown as follows:

$$R_2 = -\omega_1|A(i)|^2 - 0.001 * (1 - \omega_1) \left[C_{\text{total}} + \frac{P_n}{P} \right], \quad (18)$$

where ω_1 is the weighting factor and ω_1 is taken as 0.7.

The IQ(λ) control strategy outputs the total regulation power command, which HQDC uses as a state quantity, discrete as $(-\infty, -850]$, $(-850, -400]$, $(-400, -20]$, $(-20, 20]$, $[20, 400]$, $[400, 850]$, and $[850, +\infty)$; the set of action strategies is $A_i = [\eta_1, \eta_2, \dots, \eta_j] = [(\eta_{11}, \eta_{12}, \dots, \eta_{1j}), (\eta_{21}, \eta_{22}, \dots, \eta_{2j}), \dots, (\eta_{n1}, \eta_{n2}, \dots, \eta_{nj})]$, η_{nj} is the allocation factor of PGG _{j} within the regional grid n .

4.2. Reward Function Design. In the IQ(λ)-HDQC regulation strategy, five system parameters are set as follow.

- (1) The learning factor α_1, α_2 ($0 < \alpha_1, \alpha_2 < 1$) weigh the stability of the algorithm. The larger α can accelerate the convergent speed, smaller α can improve the system stability. α_1 is taken as 0.9 for faster learning convergence. Considering the strong randomness of load perturbation after the high proportion of high-capacity new energy access, α_2 is taken as 0.1.
- (2) The discount factor γ_1 and γ_2 ($0 < \gamma_1, \gamma_2 < 1$) weigh the importance of current and future reward. The closer the value is to 1, the more emphasis is placed on long-term rewards. γ_1 is taken as 0.8, γ_2 is taken as 0.9.
- (3) The attenuation factor of the eligibility trace λ ($0 < \lambda < 1$), reflect the degree of influence on convergence rate and non-Markov effect. The larger λ is, the slower the eligibility trace of the previous

historical state-action pair will decay, and the more reputation will be allocated. λ is taken as 0.9.

4.3. Strategy Process. The IQ algorithm is introduced into the qualification trace as the control strategy, and the HQL algorithm is introduced into the double-Q learning strategy to form the IQ(λ)-HDQC regulation strategy. The IQ(λ)-HDQC process is shown in Figure 2, combined with the parameter settings described.

5. Simulation Studies

5.1. Microgrid System Model Simulation. In order to realize the global dynamic interactive regulation of microgrid SNLSC, a microgrid model is built in this paper, including microgas turbines, small hydropower, electric vehicles [22], solar energy storage power plant [23], wind farm and cooling, heating and power storage model [24], as shown in Figure 3, and the model parameters [25] are shown in Table 1. Among them, wind farms and electric vehicles are involved in only one FM, PV is simulated with 24-hour light intensity [26] adding a small perturbation, and unit parameters are shown in Table 2. Meanwhile, the work period of AGC is 4 sec. In the figure, the controller in each region shares data through the interconnection between regions, obtains the dynamic information of the AGC performance index, realizes the coordinated control of the system in the continuous trial and error optimization, effectively obtains the AGC optimal total power command in each region, and optimizes the active power output of the frequency modulation units.

Electric vehicles have replaced a part of fuel vehicles. Plug-in electric vehicles (PEVs) are equipped with energy storage batteries that can be charged and discharged, and when a large number of PEVs are connected to the grid as a cluster, they can participate in the frequency regulation of the grid to replace the frequency regulation of traditional thermal power units. The block diagram of the transfer function of a single PEV is shown in Figure.4. Where I_{chj} is the constant charging current; SOC and SOC₀ are the EV battery charge state and initial charge state, respectively; K_C is the sag control gain; T_C is the sag control time constant; E_r is the rated capacity of the energy storage battery; R_s and R_t are the series resistance and parallel resistance values of the battery, respectively; and C_t is the shunt capacitance capacity.

The combined cooling, heating, and power (CCHP) system [24] incorporates solar power including solar collectors, solar PV power generation equipment, gas boilers, heat exchange equipment, and centrifugal chillers. This system can realize the complementary and collaborative optimal operation of multiple energy sources. It uses the waste heat of solar power and gas boilers to produce electric energy and meet heating and cooling requirements. The purpose is to improve energy utilization efficiency and reduce the emission of carbon and harmful gases, which can greatly improve energy utilization efficiency.

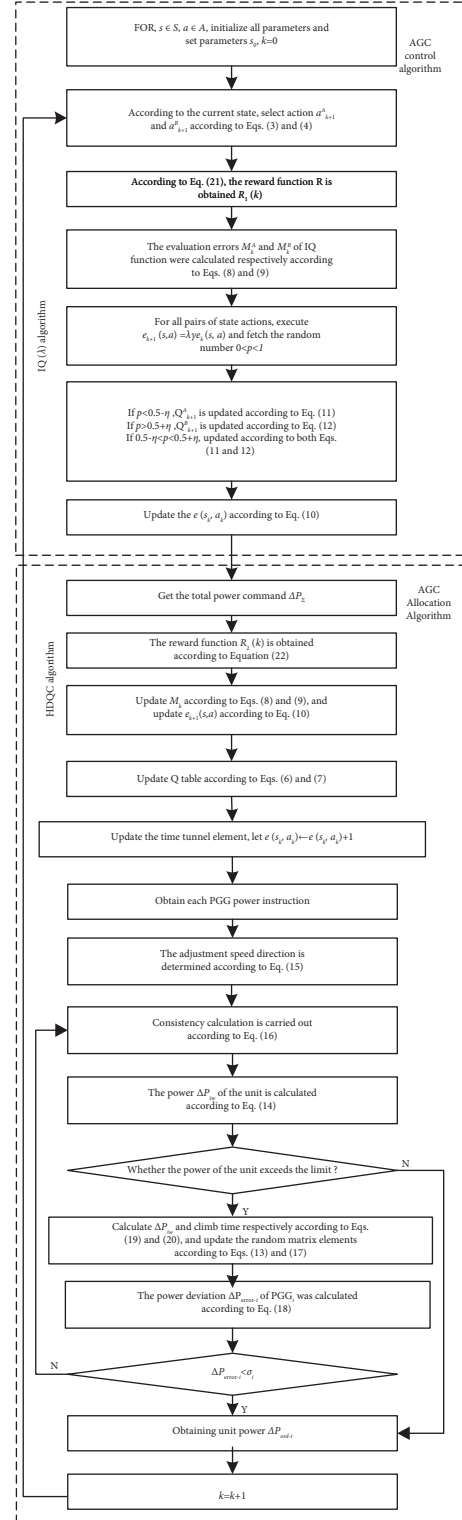


FIGURE 2: Flow chart of IQ(λ)-HDQC-based regulation strategy.

5.2. Prelearning Simulation. Before online operation, a large amount of prelearning training is required for the IQ(λ)-HDQC regulation strategy to optimize the set of strategic state actions and the action selection set. A continuous sinusoidal load disturbance with a period of 1000 sec, amplitude of 1000 MW, and duration of 10000 sec is applied to

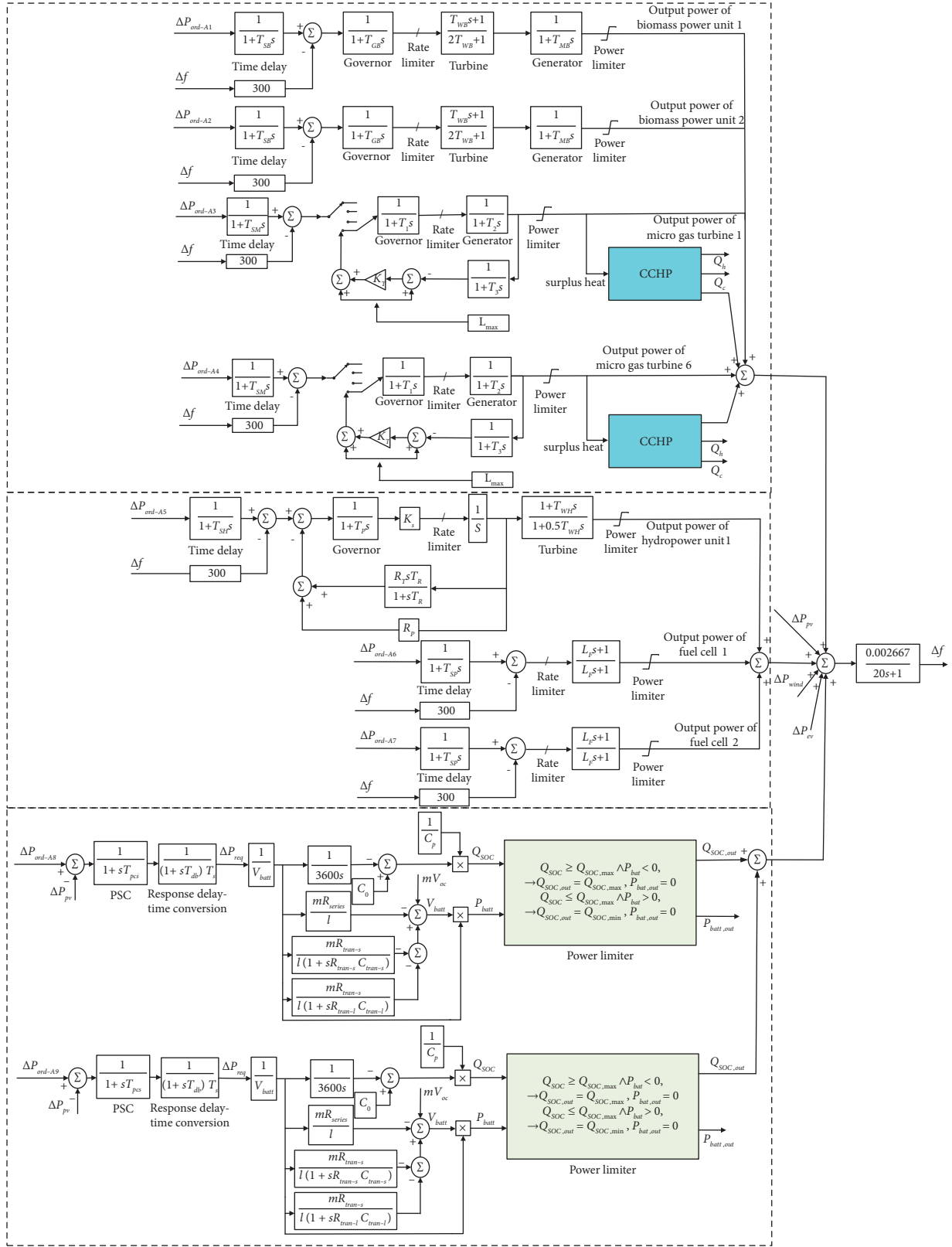


FIGURE 3: Microgrid model.

TABLE 1: The parameters of the three-area polymorphic energy system model.

Units' type	Parameters	Values
Hydropower unit	Secondary lag time, T_{SH}	3 sec
	Pilot value servomotor time constant, T_P	0.04 sec
	Servo gain, K_S	5
	Permanent droop, R_P	1
	Reset time, T_R	0.3 sec
	Temporary droop, R_T	1
	Maximum gate opening rate, $R_{maxopen}$	0.16 pu/sec
Microgas turbine	Maximum gate closing rate, $R_{maxclose}$	0.16 pu/sec
	Water starting time, T_{WH}	1 sec
	Secondary lag time, T_{SM}	5 sec
	Fuel system lag time constant 1, T_1	0.08 sec
	Fuel system lag time constant 2, T_2	0.3 sec
	Load limit time constant, T_3	3 sec
	Temperature control loop gain, K_T	1
Fuel cell	Load limit, L_{max}	1.2
	Secondary lag time, T_{SF}	2 sec
	Inverter time constant, T_F	10.056 sec
Biomass power unit	Inverter gain, K_F	9.205
	Secondary lag time, T_{SB}	10 sec
	Time constant of the governor, T_{GB}	0.08 sec
	Steam starting time, T_{WB}	5 sec
	Mechanical starting time, T_{MB}	0.3 sec

TABLE 2: Relevant parameters of three-area polymorphic energy system model.

Units type	Units no.	P^{max} (kW)	P^{min} (kW)	P^{rate+} (kW/min)	ΔP^{rate-} (kW/min)	CE parameters (kW)
Biomass power unit	G1	200	-200	3	-3	0
	G2	200	-200	3	-3	0
Microgas turbine	G3	150	-150	1.8	-1.8	0.9
	G4	150	-150	1.8	-1.8	0.9
Small hydropower unit	G5	250	-250	15	-15	0
Fuel cell	G6	200	-200	7	-7	0
	G7	200	-200	7	-7	0

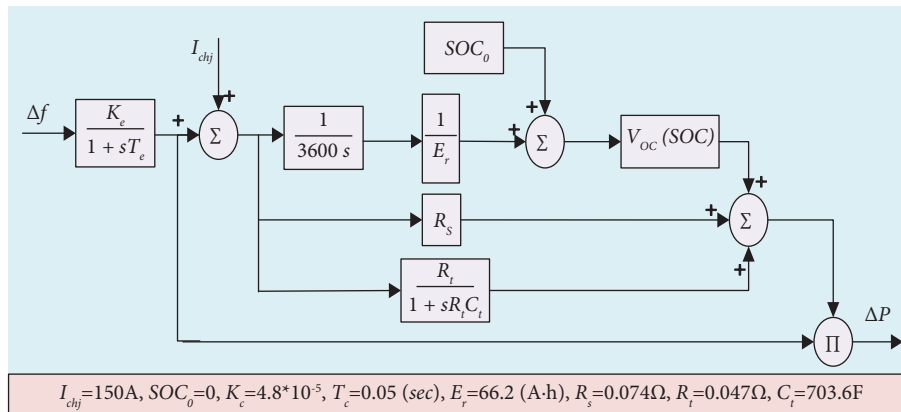


FIGURE 4: The electric vehicle model.

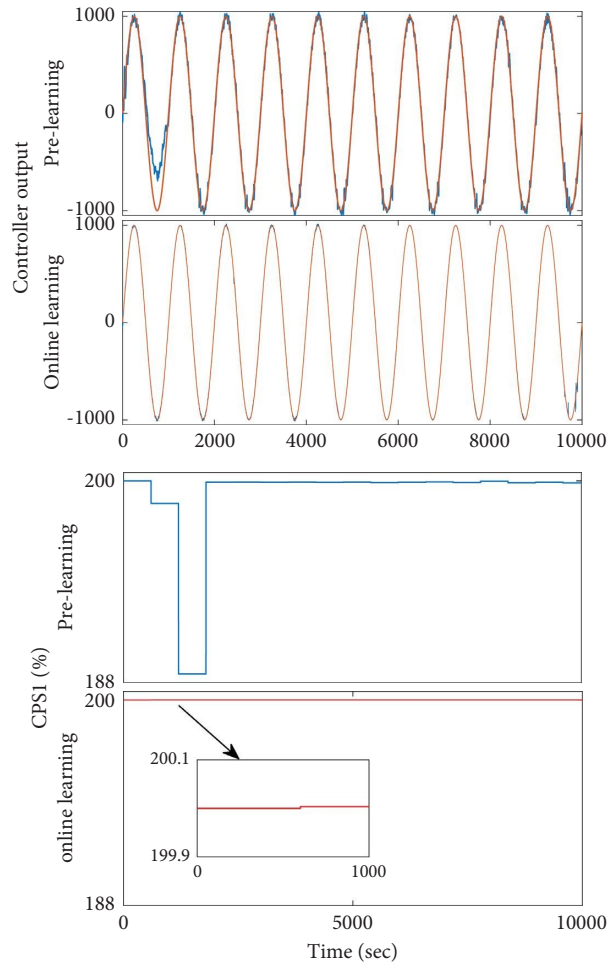


FIGURE 5: Microgrid pre-learning and online learning.

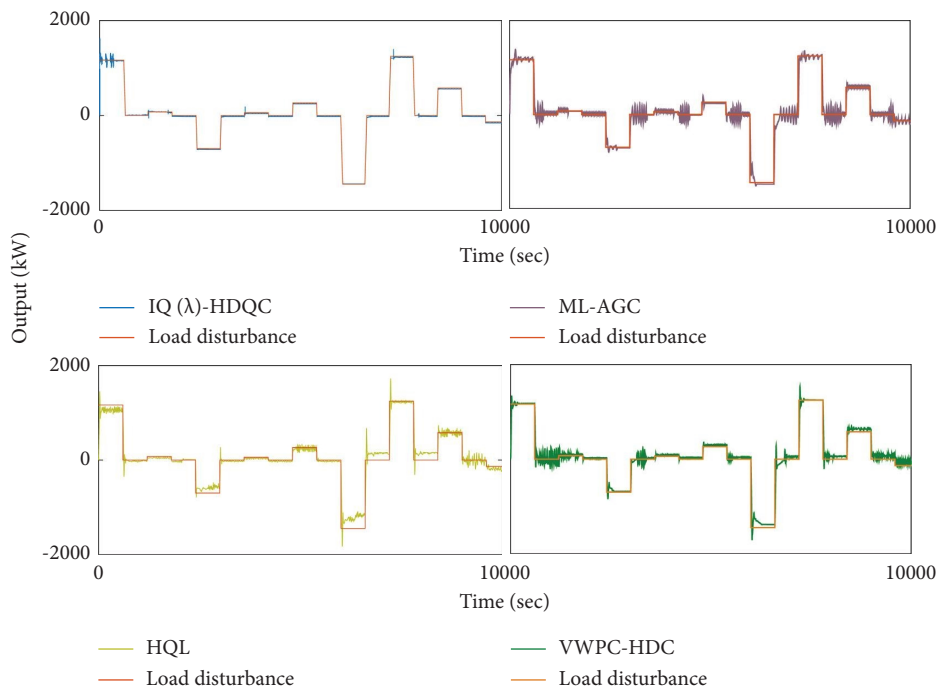


FIGURE 6: The output effect of four strategies under square wave disturbance.

TABLE 3: The control performance of four strategies under a white noise load.

Strategy	$ ACE $ (kW)	$ \Delta f $ (Hz)	CPS1 (%)	Energy efficiency (%)
IQ (λ)-HDQMP	7.951	0.0021	200.0	81.6
ML-AGC	12.734	0.0064	197.7	73.3
VWPC-HDC	24.007	0.0117	191.8	62.6
HQL	28.997	0.0145	188.1	60.5

the microgrid system for full learning, and the effect of microgrid prelearning and online operation is shown in Figure.5. From the analysis of the figure, it can be seen that after about 1000 sec of trial and error seeking during the prelearning process, the Q value of the optimal state action is explored, and the CPS1 in the prelearning phase is stable above 188% in region A and 200% in online operation, both of which meet the qualified CPS1 range. The prelearning simulation verifies that the IQ (λ)-HDQC strategy can converge quickly in complex stochastic environments and can control the generator to meet more complex load operations.

5.3. Random Square Wave Load Disturbance. After adequate prelearning, a random square wave load disturbance is introduced into the microgrid model to simulate the random load disturbance (i.e., irregular sudden increase and decrease of load and new energy output) in the random environment of the power system, so as to analyze the performance of the proposed strategy. The load disturbance with duration of 10,000 sec was taken as the assessment and compared with three control strategies, HQL [18], ML-AGC [27], and VWPC-HDC [28], were analyzed and compared. Figure.6 for the online control effect under random square wave load disturbance, the figure shows that IQ (λ)-HQDC has a more precise instructions, faster convergence speed. Using AGC three performance evaluation standard to evaluate the effects of different control strategies, Table 3 for the intelligent strategy of 4 kinds of performance evaluation of the smart grid contrast figure, compared with other strategies, IQ(λ)-HDQC can reduce $|ACE|$ 37.6%~70.2% and reduce the $|\Delta f|$ 67.2%~85.5%. CPS1 was increased by 1.16%~6.33% and energy efficiency was increased by 11.3%~34.9%.

5.4. White Noise Load Disturbance. The 24-hour white noise load disturbance is applied to the integrated energy system model to simulate the complex condition in which the power system load changes randomly at every moment in the large-scale grid-connected environment of unknown new energy. Figure 7 shows that the IQ(λ)-HQDMP control strategy can accurately track the strong stochastic perturbation, and the system remains stable at 12:00 noon when a large number of new energy sources, mainly PV, are connected to the grid. The statistical results of the simulation experiment are shown in Figure.8, IQ(λ)-HQDMP strategy can reduce $|ACE|$ 42.8%~89.2%, reduce $|\Delta f|$ 43.7%~81.1%, improve CPS1 0.03%~1.22%, and improve energy use efficiency 30.5%~51.4%. In addition, after 40 times of simulations, the standard deviations of these three indexes of IQ(λ)-HQDMP strategy are 0.000144, 1.4277, 0.32705, respectively. The

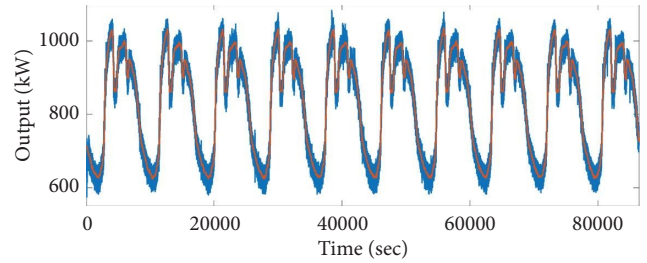


FIGURE 7: Control performance of four strategies under white noise.

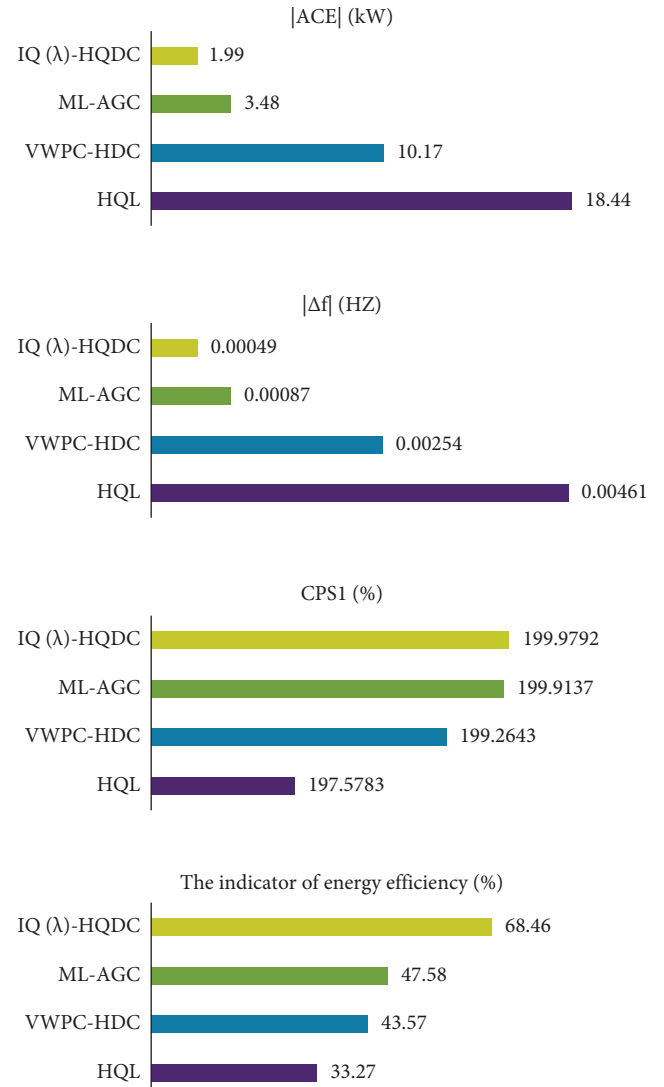


FIGURE 8: The control performance of four strategies under white noise disturbance.

IQ(λ)-HQDMP strategy has stronger antidisturbance ability and significantly improves energy efficiency compared with other strategies.

6. Conclusion

In this paper, we propose the IQ(λ)-HDQMP regulation strategy, an applicable control strategy for microgrids, to obtain the source-load-storage-charging collaborative control and optimal energy benefit of the microgrid model, thus solving the problems of strong stochastic disturbance and low utilization of new energy caused by new energy with a high proportion and large capacity connected to the grid.

In the first layer, the IQ(λ)-HDQC regulation strategy adopts the IQ(λ) control strategy, which can simultaneously avoid overestimation and underestimation to obtain long-term dynamic stability. Compared with the ML-AGC and the VWPC-HDC, the proposed algorithm can solve the multiresolution problem effectively when the number of multiagent increases sharply. In the second layer, a consistent unit power dynamic optimization allocation strategy, namely HDQC, is used to achieve the optimal allocation of new energy sources. Sine wave, square wave, and random white noise loads are respectively introduced for simulation in microgrid model. Compared with the other four different control strategies, the results show that IQ(λ)-HQDC has better learning ability and can reach stability quickly in the prelearning stage. Even under strong random disturbances, it also has better performance and can improve the system's energy use efficiency with various new energies.

Data Availability

No data were used to support the findings of the study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the support by the head office of State Grid Co., Ltd. which manages science and technology projects (5400-202118485A-0-5-ZN).

References

- [1] X. Fu, Q. Guo, and H. Sun, "Statistical machine learning model for stochastic optimal planning of distribution networks considering a dynamic correlation and dimension reduction," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 2904–2917, 2020.
- [2] Y. Liu, C. Lin, T. Chen, J. Li, and Z. Chen, "Reactive power-voltage control strategy of AC microgrid based on adaptive virtual impedance," *Automation of Electric Power Systems*, vol. 41, no. 5, pp. 16–21, 2017.
- [3] X. Fu and Y. Zhou, "Collaborative optimization of PV greenhouses and clean energy systems in rural areas," *IEEE Transactions on Sustainable Energy*, vol. 14, no. 1, pp. 642–656, 2023.
- [4] G. Dehnavi and H. L. Ginn, "Distributed load sharing among converters in an autonomous microgrid including PV and wind power units," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4289–4298, 2019.
- [5] M. Xie, X. Ji, X. Hu, P. Cheng, Y. Du, and M. Liu, "Autonomous optimized economic dispatch of active distribution system with multi-microgrids," *Energy*, vol. 153, pp. 479–489, 2018.
- [6] X. Fu, Q. Guo, and F. Sun, "Statistical machine learning model for capacitor planning considering uncertainties in photovoltaic power," *Protection and Control of Modern Power Systems*, vol. 7, no. 1, pp. 5–63, 2022.
- [7] B. Yang, T. Yu, H. Shu, J. Dong, and L. Jiang, "Robust sliding-mode control of wind energy conversion systems for optimal power extraction via nonlinear perturbation observers," *Applied Energy*, vol. 210, pp. 711–723, 2018.
- [8] A. N. Venkat, I. A. Hiskens, J. B. Rawlings, and S. J. Wright, "Distributed MPC strategies with application to power system automatic generation control," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 6, pp. 1192–1206, 2008.
- [9] X. Fu and H. Niu, "Key Technologies and Applications of Agricultural Energy Internet for Agricultural Planting and Fisheries Industry," *Information Processing in Agriculture*, 2022.
- [10] X. Wu, J. Shi, W. Ma, and J. Chen, "Automatic generation control of micro grid based on extreme Q-learning algorithm," *The Journal of New Industrialization*, vol. 9, no. 4, pp. 22–26, 2019.
- [11] L. Xi and L. Zhou, "Coordinated AGC algorithm for distributed multi-region multi-energy micro-network group," *Acta Automatica Sinica*, vol. 46, no. 9, pp. 1818–1830, 2020.
- [12] L. Xi, Y. Li, Y. Huang, L. Lu, and J. Chen, "A Novel Automatic Generation Control Method Based on the Ecological Population Cooperative Control for the Islanded Smart Grid," *Complexity*, vol. 2018, Article ID 2456963, 17 pages, 2018.
- [13] C. Song, D. Zhao, J. Yin, and Q. Zhang, "Coordination Control of AGC and AVC Based on Multi-Agent System," in *Proceedings of the 2018 2nd IEEE Conference on Energy Internet and Energy System Integration*, Beijing, China, October 2018.
- [14] M. He and H. Guo, "Interleaved Q-learning with partially coupled training process," *Autonomous Agents and Multi-Agent Systems*, vol. 5, pp. 449–457, 2019.
- [15] Y. Li, L. Xi, Y. Guo, Y. Wang, M. Sun, and C. Jin, "Automatic generation control based on the weighted double Q-delayed update learning algorithm," *Proceedings of the CSEE*, vol. 42, no. 15, pp. 5459–5471, 2022.
- [16] A. Nedić and J. Liu, "On convergence rate of weighted-averaging dynamics for consensus problems," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 766–781, 2017.
- [17] H. Hasselt, "Double Q-learning," *Advances in Neural Information Processing Systems*, vol. 23, pp. 2613–2621, 2010.
- [18] 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 Generation, Transmission & Distribution*, vol. 5, no. 8, pp. 789–797, 2011.
- [19] L. Ji, Y. Tang, and Q. Liu, "On hybrid adaptive and pinning consensus for multi-agent networks," *Mathematical Problems in Engineering*, vol. 2016, Article ID 9127356, 11 pages, 2016.
- [20] 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, 2014.
- [21] N. Jaleeli and L. S. Vanslyck, “NERC’s new control performance standards,” *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1092–1099, 1999.
- [22] J. Lu, J. Chang, R. Liu, and J. Zhang, “Decentralized predictive control of load frequency for interconnected power grid considering electric vehicle clusters,” *Electrical Measurement and Instrumentation*, vol. 58, no. 9, pp. 96–102, 2021.
- [23] J. Huang, X. Li, Y. Cao, Q. Zhang, and W. Liu, “Battery energy storage power supply simulation model for power grid frequency regulation,” *Automation of Electric Power Systems*, vol. 39, no. 18, pp. 20–24, 2015.
- [24] Q. Chen and T. Zhao, “Heat recovery and storage installation in large-scale battery systems for effective integration of renewable energy sources into power systems,” *Applied Thermal Engineering*, vol. 122, pp. 194–203, 2017.
- [25] L. Xi, H. Li, J. Zhu, Y. Li, and S. Wang, “A novel automatic generation control method based on the large-scale electric vehicles and wind power integration into the grid,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2022.
- [26] S. Kar and G. Hug, “Distributed robust economic dispatch in power systems: a consensus + innovations approach,” in *Proceedings of the 2012 IEEE Power and Energy Society General Meeting*, pp. 1–8, San Diego, CA, USA, July 2012.
- [27] L. Xi, L. Zhang, Y. Huang, X. Chen, and Y. Xu, “Multiple level automatic generation control based on the greedy strategy,” *Proceedings of the CSEE*, vol. 40, no. 16, pp. 5204–5217, 2020.
- [28] L. Xi, Y. Li, Y. Huang, P. Yang, and Z. Xu, “Smart generation control based on the virtual wolf pack control strategy,” *Proceedings of the CSEE*, vol. 38, no. 10, pp. 2966–2979, 2018.