Distributed Energy Storage Scheduling Optimization Based on Improved Multi-agent Deep Deterministic Policy Gradient Algorithm

Yueli Zhou*, Shaohua Zhao, Jiasheng Wu, Qihua Lin,
Xiaodong Zheng and Hanfeng Bai

CGS Power Generation(Guangdong) Energy Storage Technology Co., Ltd, Guangzhou, 510630, China
E-mail: cnkjzhouyueli@126.com
**Corresponding Author*

Received 23 March 2026; Accepted 23 July 2026

Abstract

The core of current energy storage scheduling optimization is to achieve multi-timescale collaborative decision-making through intelligent algorithms to improve system economy and reliability, and accelerate its evolution towards market-oriented and large-scale applications. This study is designed to develop a scenario-adaptive intelligent scheduling system. It skips the need for accurate long-term future time-series predictions, and directly leverages real-time observable information at the current decision point, such as renewable energy output, power load, electricity price and energy storage status, to complete dynamic optimal scheduling. For a single independent microgrid, an electric-hydrogen hybrid architecture is constructed, and an improved deep deterministic policy gradient algorithm with attenuated random noise is proposed. The scheduling strategy is optimized through online interaction with the target network and a soft update mechanism. For multiple interconnected microgrids, a centralized training and decentralized execution

Distributed Generation & Alternative Energy Journal, Vol. 41_5, 1485–1512.
doi: 10.13052/dgaej2156-3306.41510

framework is adopted to achieve multi-agent collaborative optimization and autonomous decision-making. The results show that in a single microgrid scenario, the research method achieves a renewable energy utilization efficiency of 98.77% and a average operating cost of 0.381 yuan/kWh; in a multi-microgrid scenario, the average operating cost is 0.389 yuan/kWh. The research indicates that the two types of algorithms are respectively adapted to single-microgrid internal optimization and multi-microgrid collaborative scheduling, providing scenario-based solutions for distributed energy storage optimization.

Keywords: Distributed energy storage dispatch, DDPG, electric-hydrogen hybrid microgrid, MAPPO, multi-microgrid collaboration.

1 Introduction

During the transition to high-proportion renewable energy, the intermittency of wind and solar power generation and the randomness of load demand form a strong counterbalance, leading to a severe time-series supply and demand imbalance problem in microgrid systems [1]. Traditional battery energy storage is limited by capacity and adjustment time, making it difficult to achieve large-scale energy transfer across time periods. Existing scheduling methods either rely on long-term predictions with unreliable accuracy or cannot efficiently handle the time coupling constraints of energy storage dynamics, resulting in a large waste of clean electricity and significantly increasing the system's dependence on external power purchases and operating costs [2, 3]. As the energy internet advances further, multi-microgrid interconnection has become a trend, and new demands such as cross-microgrid dynamic energy trading and carbon emission collaborative management have emerged, further exacerbating scheduling complexity. Single-agent algorithms are prone to the dimensionality curse, and traditional distributed algorithms are difficult to cope with environmental non-stationarity [4, 5]. Therefore, this research employs a Reinforcement Learning (RL) framework and introduces a Deep Deterministic Policy Gradient (DDPG) algorithm with attenuated random noise for single microgrid scenarios to enhance exploration capabilities and optimize the coordinated scheduling of electricity and hydrogen. The research aims to develop intelligent scheduling methods adaptable to different scenarios, utilizing only the measured state information at the current decision moment to achieve dynamic optimization decision-making.

The innovation of this research lies in innovatively constructing two architectures: an electric-hydrogen hybrid single microgrid and a carbon capture + peer-to-peer trading multi-microgrid. It employs a Centralized Training with Decentralized Execution (CTDE) framework, and then applies a Multi-Agent System (MA) Proximal Policy Optimization (PPO) algorithm (C-MAPPO) to guide local decision-making in each microgrid through global collaborative optimization. The aim is to form a scenario-adaptive intelligent scheduling system, breaking through the bottlenecks of traditional methods.

2 Related Works

The volatility of renewable energy has intensified, and the optimization of energy storage scheduling urgently needs to improve system stability, economy and energy utilization efficiency. Lu et al. proposed a two-layer hybrid integer model based on split-blob bar optimization. The results showed that the method improved energy storage utilization [6]. Dong et al. proposed a data-driven scheduling strategy based on deep RL. The strategy considered random charging of electric vehicles. The findings demonstrated that the gradient algorithm of deep deterministic strategy with priority experience replay reduced power fluctuations and operating costs [7]. Krishna et al. introduced a two-stage energy management approach to tackle the scheduling problem caused by the uncertainty of renewable energy. The results showed that the method could effectively extend battery life [8]. Todakar et al. proposed a real-time operation strategy that combines autoregressive integral moving average model generation with mixed integer linear programming. The findings demonstrated that the method effectively optimized the layout of charging and discharging stations [9].

Yan et al. proposed a two-layer Stackelberg game model grounded in shared energy storage. The results showed that the method reduced the cost of pollutant emission penalties [10]. Sui et al. established a stochastic scheduling model for island microgrid groups compatible with the variable speed characteristics of hydrogen carrier ships. The results showed that the method effectively improved frequency stability [11]. Sun et al. proposed an improved imitation learning method based on data driving. They used a limit gradient boosting model to learn the mapping relationship between system state and scheduling decision. The results showed that the method improved scheduling efficiency and robustness [12]. Tian et al. proposed a "source-grid-load-storage" multimodal economic optimization scheduling method. They used an improved gray wolf optimization algorithm to solve the two-layer model. The results showed that the method reduced electricity costs by 5.5%–8.7% [13].

In summary, existing technologies suffer from problems such as reliance on prediction, difficulty in handling high-dimensional uncertainty, and lack of low-carbon coordination mechanisms for multi-microgrid scheduling. To adapt to different scenarios, this study proposes an electric-hydrogen hybrid single-microgrid architecture and a carbon capture + peer-to-peer trading multi-microgrid architecture, along with improved DDPG and C-MAPPO algorithms. The aim is to achieve dynamic optimization without requiring precise prediction, thereby improving energy efficiency and reducing operating costs.

3 Methods and Materials

3.1 Hybrid Energy Dispatch Optimization Method Based on DDPG

In the transition to high-proportion renewable energy, the intermittency of wind and solar power generation and the randomness of load demand lead to a serious time-series supply-demand imbalance in microgrids [14]. This not only wastes clean electricity but also increases the dependence on and cost of purchasing electricity from outside sources. Traditional battery energy storage is limited by capacity and regulation time, making it difficult to solve such problems [15, 16]. Meanwhile, existing scheduling methods either rely on long-term predictions that are difficult to make accurately or cannot efficiently handle the time coupling constraints of energy storage dynamics, resulting in limited optimization results [17]. Therefore, this study constructs a collaborative framework that fits the reality of energy interconnection. By simulating the heterogeneous coupling, spatiotemporal uncertainty, and inter-subject game of distributed energy, its single microgrid system architecture is shown in Figure 1.

As shown in Figure 1, the electric-hydrogen hybrid microgrid architecture takes renewable energy generation and external power purchase as inputs, supplies power to the base and flexible loads through the public AC bus, and exchanges energy bidirectionally with battery energy storage. The battery energy storage status update is shown in Equation (1).

$$SOC(t+1) = SOC(t) + \frac{k_{ch}W_{ch}(t) - \frac{W_{dis}(t)}{k_{dis}}}{E_{rated}} \cdot \Delta t \tag{1}$$

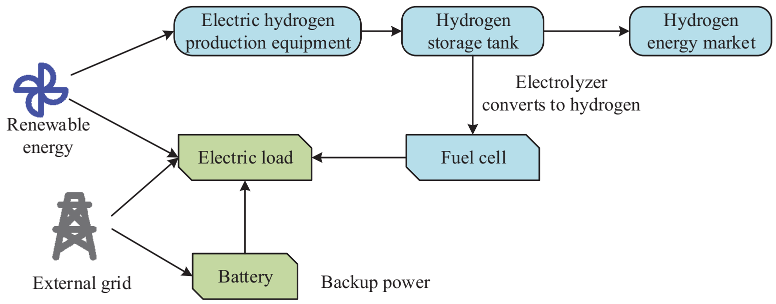


Figure 1 Single microgrid system architecture diagram.

In Equation (1), $SOC(t)$ represents the state of charge of the battery at time t; k_{ch} represents the charging efficiency; $W_{dis}(t)$ represents the charging power; k_{dis} represents the discharging efficiency; $W_{dis}(t)$ represents the discharging power; E_{rated} represents the rated capacity of the battery;and Δt represents the time step. The external power grid serves as a backup power source to ensure power supply security and also guides the optimized operation of the microgrid through real-time electricity price signals [18]. Surplus electrical energy is converted into hydrogen through an electrolyzer and stored in a hydrogen storage tank. The dynamic equation for the hydrogen storage capacity of the hydrogen storage tank is shown in Equation (2).

$$Q_{t+1} = Q_t + X_t - E_t^S - E_t^P, \quad \forall t. \tag{2}$$

In Equation (2), Q_t represents the amount of hydrogen at the beginning of the time slot t; Q_{t+1} represents the amount of hydrogen at the end of the time slot t; X_t represents the amount of hydrogen produced and injected into the hydrogen storage tank during the time slot t. E_t^S represents the amount of hydrogen sold during the time slot t; E_t^P represents the amount of hydrogen consumed for power generation during the time slot t. Energy storage and conversion equipment is the core of the system's flexible adjustment. Batteries cope with short-term power fluctuations, while the hydrogen energy chain realizes energy transfer across time periods [19]. All equipment is connected to the intelligent dispatch center through a communication network, forming an organic whole that can dynamically respond to electricity prices, balance renewable energy fluctuations, and maximize economic efficiency. The objective function J for minimizing operating costs is shown in Equation (3).

$$J = \min \sum_{t=1}^{T} (C_{grid}(t) + C_{batt}(t) + C_{H_2}(t) + C_{penalty}(t)) \tag{3}$$

In Equation (3), T is the total scheduling time; $C_{grid}(t)$ is the cost of purchasing electricity from the external power grid; $C_{batt}(t)$ is the operating loss cost of battery energy storage; $C_{H_2}(t)$ is the operating cost of the hydrogen energy system; and $C_{penalty}(t)$ is the penalty cost. The study defines the state variable $s_t = [SOC(t), Q_t, P_w(t), P_{pv}(t), L(t), price(t)]$, where $P_w(t)$ and $P_{pv}(t)$ represent the actual power output of wind power and photovoltaic power respectively, $L(t)$ represents the load, and $price(t)$ represents the real-time electricity price. The action variable is defined as $a_t = [W_{ch}(t), W_{dis}(t), P_{ele}(t), P_{fc}(t), P_{buy}(t)]$, $P_{ele}(t)$ represents the power consumption of the electrolyzer, $P_{fc}(t)$ represents the power generation of the fuel cell, and $P_{buy}(t)$ represents the purchased power from the outside. The boundary constraints are $0 \leq \mathrm{SOC(t)} \leq 1$, $0 \leq Q_t \leq Q_{max}$, $0 \leq W_{ch} \leq W_{ch,max}$, $0 \leq W_{dis} \leq W_{dis,\max}$, $0 \leq P_{ele} \leq P_{ele,\max}$, $0 \leq P_{fc} \leq P_{fc,\max}$, and the charging and discharging are mutually exclusive with $W_{ch} \cdot W_{dis} = 0$. Q_{max} represents the upper limit of the hydrogen storage capacity. The power balance equation is shown in Equation (4) when ignoring line losses.

$$P_w(t) + P_{pv}(t) + P_{buy}(t) + P_{fc}(t) + W_{dis}(t) = L(t) + W_{ch}(t) + P_{ele}(t) \quad (4)$$

The terminal condition is the end of the scheduling period, with $\mathrm{SOC(t)} \geq 0.2$ and $Q_t \geq 0.1Q_{max}$. The current limiting processing is as follows: when the wind and solar power output exceeds the maximum charging power, if the hydrogen storage is full, then wind and solar power generation will be abandoned, and load shedding is only allowed in emergency situations. The load shedding amount $\leq 0.05L(t)$, and it is included in the penalty term. The penalty clause is as shown in Equation (5).

$$\begin{aligned} h_{t2} = {} & \lambda_1 \max(0, SOC_{low} - SOC)^2 + \lambda_2 \max(0, Q_{low} - Q_t)^2 + \lambda_3 L_{\mathrm{s}}^2 \\ & + \lambda_4 \max(0, P_{buy} - P_{buy,lim})^2 \end{aligned} \quad (5)$$

In Equation (5), L_{s} represents the amount of load shedding, SOC_{low} is set to 0.15, Q_{low} to $0.05Q_{max}$, $P_{buy,lim}$ to $0.8L_{max}$. For the cost coefficient, the purchase price of electricity is taken from the real-time data of PJM, the battery operation loss cost coefficient is 0.015 yuan/kWh, and the hydrogen system operation cost coefficient is 0.025 yuan/kWh. The penalty weights $\lambda_1 = 100$ yuan/kW^2, $\lambda_2 = 80$ yuan/kW^2, $\lambda_3 = 200$ yuan/kW^2, and $\lambda_4 = 50$ yuan/kW^2. To address the multidimensional constraints and uncertainties faced by this electric-hydrogen hybrid microgrid in real-time scheduling, a modified DDPG with decaying random noise is introduced. This modified

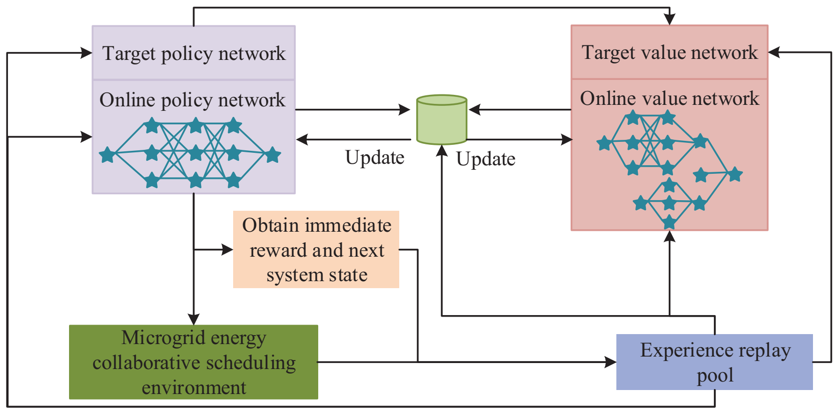


Figure 2 Electro-hydrogen coordinated scheduling DDPG algorithm framework diagram.

DDPG can directly handle the continuous action space. The architecture of the electric-hydrogen collaborative scheduling DDPG algorithm is shown in Figure 2.

As shown in Figure 2, the architecture achieves scheduling through online interaction and updates with the target policy network and the value network. The system state is normalized and input into the online network, combined with decaying random noise to generate actions. After interacting with the environment, a state-action-reward-next state sample is formed and stored in the experience replay pool. The reward function is shown in Equation (6).

$$h_t = \gamma \cdot h_t^1 - h_t^2 \tag{6}$$

In Equation (6), h_t is the instantaneous reward value obtained by the agent in the time slot t; γ represents the weight coefficient; h_t^1 represents the net profit of running within the time slot t; and h_t^2 represents the penalty term. During the update, random sampling is performed from the experience pool, the online value network evaluates the value of the state action, the target network calculates the target value based on the next state, and the target value is obtained by combining the reward. The minimized loss Z is shown in Equation (7).

$$Z = \frac{1}{P}\sum_i (f_i - U(m_i, n_i \mid \delta^U))^2 \tag{7}$$

In Equation (7), P represents the number of mini-batch samples; i represents the index of a single sample; f_i represents the target U value calculated for the i-th sample; $U(m_i, n_i \mid \delta^U)$ represents the U value predicted by

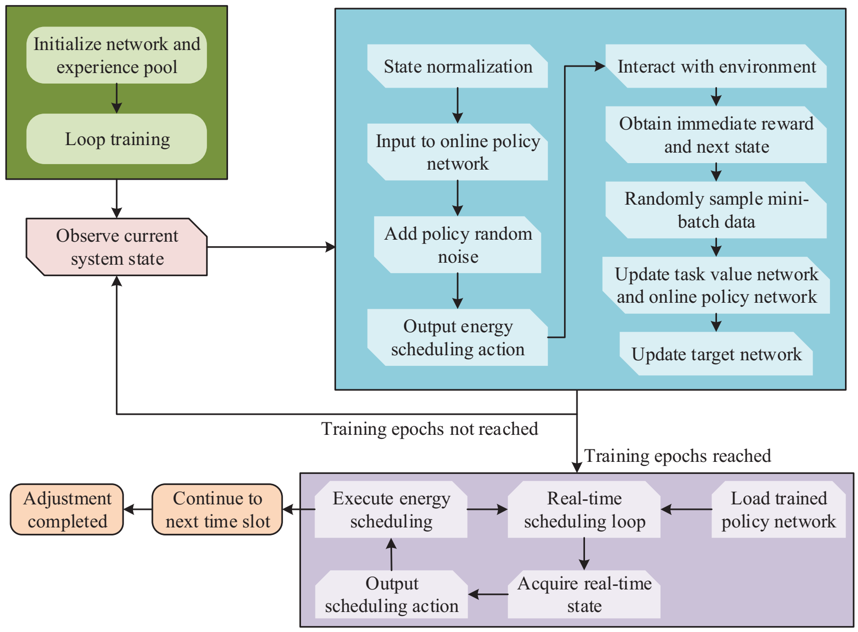


Figure 3 DDPG coordinated scheduling flowchart for electro-hydrogen hybrid energy.

the online value network for the state-action pair (m_i, n_i) of the i-th sample under the current parameters δ^U; m_i and n_i are the system state and the action performed by the agent recorded in the i-th sample, respectively; and δ^U represents the trainable parameters of the online value network. Finally, a soft update mechanism is used to synchronize the parameters of the online network and the target network, and the algorithm is iterated until convergence is achieved, thereby optimizing the microgrid energy scheduling strategy. Based on the DDPG algorithm architecture concept of electric-hydrogen co-scheduling, the study clearly demonstrates the entire process from network initialization, experience collection and training optimization to final real-time scheduling. The DDPG co-scheduling process for electric-hydrogen hybrid energy is shown in Figure 3.

As shown in Figure 3, the algorithm completes system initialization upon startup, establishing the policy network, value network, and experience pool. During the training phase, the real-time state of the micronetwork is continuously observed and normalized, as shown in Equation (8).

$$\hat{s}_t = \frac{s_t - \mu_s}{\sigma_s} \tag{8}$$

In Equation (8), $\hat{s}_t$ represents the normalized system state vector at time t; s_t is the original system state vector observed at time t; μ_s is the vector composed of the mean values of each state dimension; and σ_s represents the vector composed of the standard deviations of each state dimension. Continuous scheduling instructions are then generated by the policy network, while attenuating random noise is introduced to balance exploration and utilization. The temporal attenuation of the attenuating random noise is shown in Equation (7).

$$\sigma_t = \sigma_{init} \cdot e^{-\frac{t}{T_{decay}}} \tag{9}$$

In Equation (9), σ_t represents the noise standard deviation at time t; σ_{init} represents the initial value of the noise standard deviation; T_{decay} represents the decay time constant; and e represents the natural constant. After the scheduling action interacts with the environment, it obtains a reward and a new state, and the interaction record is stored in the experience pool. During training, random samples are periodically drawn from the experience pool, and the parameters of the policy network and value network are updated synchronously through gradient calculation. The target network is updated incrementally to ensure stability. After reaching the preset training rounds, the real-time decision-making stage begins. The trained policy network can directly output the optimal scheduling instruction based on the current state, achieving a dynamic balance between efficient utilization of renewable energy and minimization of operating costs.

3.2 Energy Scheduling Method Based on Improved Multi-agent Proximity Strategy

With the advancement of the "dual carbon" goal and the construction of new power systems, regional multi-microgrid interconnection and coordination has become an inevitable trend to enhance energy resilience and promote the consumption of renewable energy [20]. However, the coordinated operation of multi-microgrid systems faces the prominent contradiction of difficulty in coordinating energy dispatch and carbon management. Existing studies often treat carbon capture and utilization and peer-to-peer trading separately, ignoring the deep coupling between the two in terms of resource sharing and cost allocation, resulting in the inability to balance the overall economic efficiency and low carbon performance of the system, thus restricting the synergistic efficiency potential of the regional energy internet [21, 22]. To this end, this study integrates carbon capture and utilization technology with

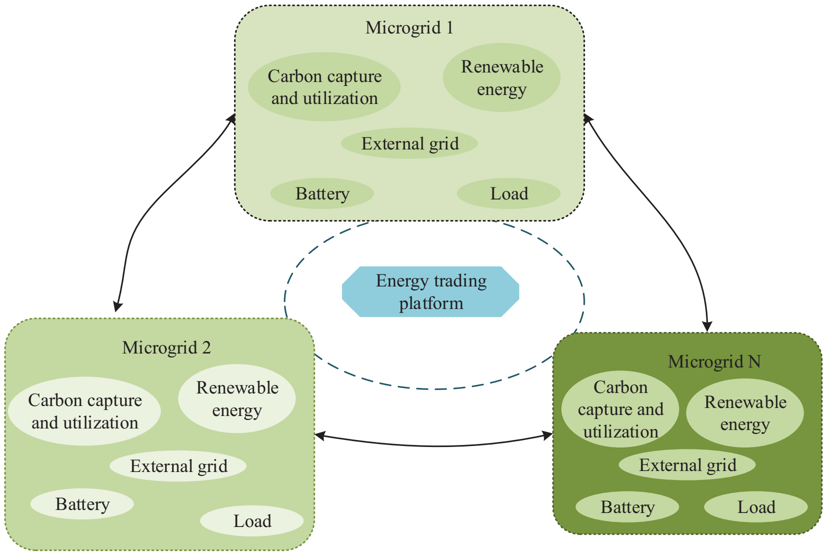


Figure 4 Multi-microgrid system architecture diagram.

peer-to-peer energy trading mechanism to construct a distributed dispatch framework that can coordinate multiple independent microgrids and achieve the dual goals of global economy and low carbon. The architecture of this multi-microgrid system is shown in Figure 4.

As shown in Figure 4, wind power and energy storage work together within each microgrid to maintain local power balance. Surplus or insufficient electricity is traded between microgrids through a point-to-point network, achieving regional energy mutual assistance. Each microgrid is equipped with a carbon capture system to capture carbon dioxide generated during power generation and convert it into synthetic methane using hydrogen produced by local water electrolysis, thus realizing waste resource utilization and the organic connection and flexible conversion of the electricity-gas network. The carbon capture rate is set to 85%, the energy consumption of the capture process is 0.30 kWh/kg CO_2, and the methane conversion efficiency is 70% (0.182kg H_2 is required per kilogram of CO_2 according to stoichiometry, and hydrogen is provided by electrolyzed water, with an electrolysis efficiency of 65%). The carbon emissions are calculated based on the purchased electricity emission factor of 0.85 kg CO_2/kWh and the thermal power portion, and the actual emissions after capture are the remaining uncaptured portion. The power consumption of the carbon capture system is included in the power balance as an additional load, and its operating constraint is $0 \leq \mathrm{P_{CCU}} \leq \mathrm{P_{CCUmax}}$, where $\mathrm{P_{CCU}}$ represents the actual power consumption of the carbon capture system during period t, and $\mathrm{P_{CCUmax}}$ represents the maximum rated power consumption of the carbon capture system. The energy consumption penalty has been included in the cost coefficient of the hydrogen system. A communication network covering the entire system collects real-time status and market information of each microgrid and accurately issues scheduling instructions generated by the collaborative optimization algorithm to each controllable unit, thereby integrating distributed devices into a smart energy system that can operate autonomously and collaborate globally. To address the environmental non-stationarity and dimensionality curse problems in multi-microgrid collaborative scheduling, the CTDE framework is introduced. This distributed execution mechanism mines globally optimal information through centralized training, achieving a balance between algorithm performance and scalability. The CTDE architecture is shown in Figure 5.

As shown in Figure 5, during the training phase, the system integrates all microgrid interaction data and the global state, centrally optimizing the value network parameters to alleviate multi-agent non-stationarity and ensure policy consistency and optimality. The policy network adjusts its actions based on feedback from the value network through the advantage function, while the value network continuously optimizes the evaluation accuracy based on

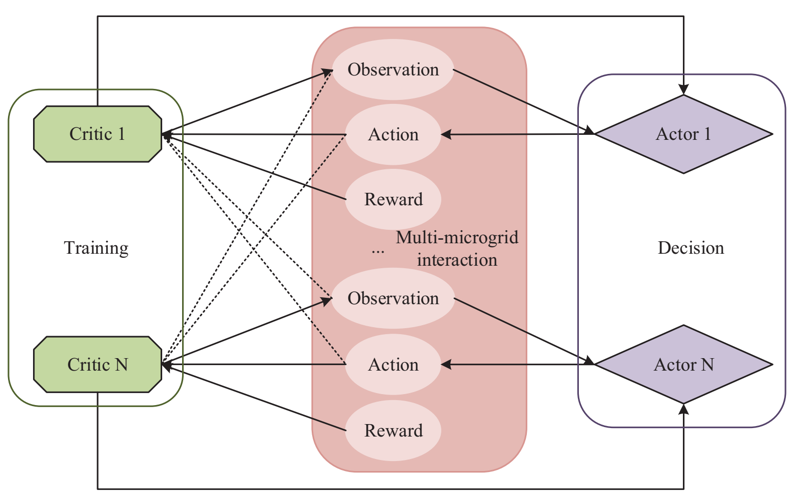


Figure 5 CTDE framework diagram.

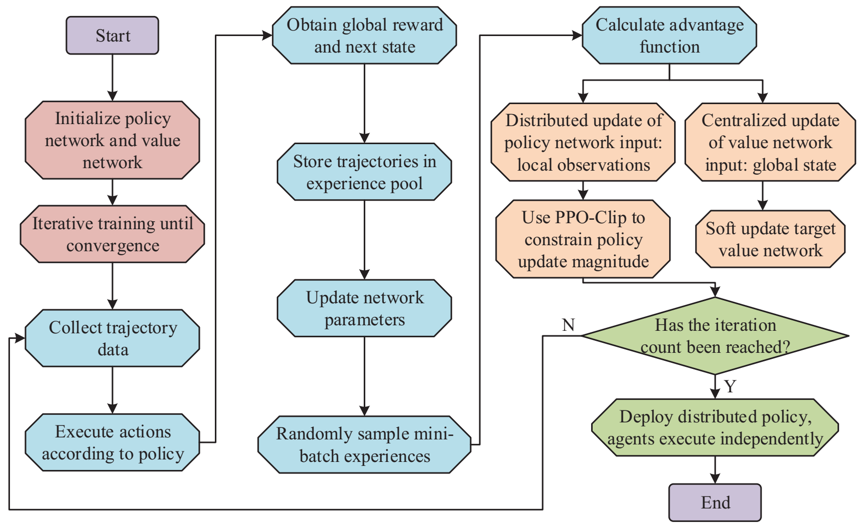


Figure 6 Scheduling flowchart based on C-MAPPO.

global data, thereby improving scheduling adaptability and overall operational efficiency. CTDE ensures both global optimization capabilities and local decision autonomy, perfectly meeting the improvement needs of multi-agent PPO algorithms. Therefore, this study proposes optimizing heterogeneous energy scheduling in multi-microgrids based on C-MAPPO, achieving a dual improvement in scheduling performance and system stability. The C-MAPPO-based scheduling process is shown in Figure 6.

As shown in Figure 6, the process first initializes the policy and value network parameters of all agents. After entering the loop training, each agent outputs an action based on the current policy and local observations, interacts with the environment to obtain global rewards and new states, and generates a complete batch of trajectories (on policy sampling). After accumulating sufficient trajectories, the algorithm uses these trajectories to calculate the advantage function and adopts a near end strategy to optimize the objective for multiple rounds (epochs $= 10$) of small batch gradient updates, reusing the current sampled data. The value network updates directly based on the global state, without using the target network or soft updates. This process is repeated until the strategy converges, and ultimately each agent relies solely on local observations for autonomous decision-making. The calculation of the multi-agent advantage function A_t is shown in Equation (10).

$$A_t = R_t + \eta V(s_{t+1}) - V(s_t) \tag{10}$$

In Equation (10), R_t is the immediate reward; η is the discount factor; $V(s_{t+1})$ is the value function of the next state; and $V(s_t)$ is the value function of the current state. The value network is updated accordingly to optimize the overall system value assessment. This process is repeated until policy convergence, and the final deployed distributed policy allows each agent to autonomously make optimal decisions based solely on local observations.

4 Results

4.1 Hybrid Energy Dispatch Test Based on DDPG

To test energy dispatching methods, this study utilized the NREL WIND Toolkit v3.0.0 and NSRDB v3.2.0 datasets from the National Renewable Energy Laboratory (NREL), which includes wind and solar power output time series (5-minute resolution), typical daily load curves, and PJM market electricity price data. A microgrid area in the central-western part of the United States, located at 41.98°N, 87.90°W, was selected. The annual data for the entire year of 2022 was used. The load data adopted the typical daily curve of the NREL Commercial Reference Building, the electricity price was obtained from the real-time node LMP of the PJM Northern Illinois Hub, and the time zone was UTC-6. The maximum wind power generation capacity was 3.2MW, the peak solar power generation capacity was 2.8MW, the peak-valley load ratio was 1:0.4, and the electricity price fluctuation range was \$0.02–0.35/kWh. Referring to the PJM node data for the whole year of 2022, all economic parameters denominated in US dollars were uniformly converted into Chinese yuan based on the average exchange rate of 1 US dollar to 6.73 yuan published by the People's Bank of China for that year. The cost coefficients for equipment operation loss, hydrogen system operation and maintenance already include depreciation and maintenance costs, and were not separately listed. Battery storage capacity was configured as 5 MWh/2.5 MW, hydrogen production efficiency was set to 65%. The original 5-minute data was aggregated and resampled according to the mean to a 15 minute scheduling interval, covering typical days and random fluctuations throughout the four seasons. The dataset was divided into training set (70%), validation set (15%), and testing set (15%) in chronological order to ensure that future information was not used during the training process. When making strategic decisions, this method only relied on the current observed wind and solar power output, load, electricity price, and energy storage status (i.e. battery state of charge and hydrogen storage), without relying

Table 1 Detailed information of the testing platform

–	Configure	Parameter
Hardware configuration	CPU	Intel Xeon W7-2495X
	GPU	NVIDIA RTX 4090
	Storage	2TB NVMe PCIe 5.0 SSD + 8TB NAS/RAID
	Operating system	Ubuntu Server22.04 LTS
	Programming language	Python 3.9.18
Software configuration	Deep learning framework	PyTorch 2.0.1+cu118
	Containerization tool	Docker Desktop4.25.0
	Distributed computing framework	Ray2.7.1
	Numerical computing tool	NumPy1.26.2

on any future predicted values. Therefore, this method achieved dynamic online scheduling without the need for accurate prediction. Hyperparameter settings were as follows: learning rate Actor $= 0.001$, Critic $= 0.002$, discount factor $\eta = 0.99$, soft update coefficient $\tau = 0.005$, experience replay pool capacity $= 10^6$, batch $\text{size} = 256$, and training epochs $\geq$2000. A simulation testing platform was built, and experiments were conducted on the Ubuntu Server 22.04 LTS operating system. The environment settings arepresented in Table 1.

To ensure reproducibility of the results, the random seeds for the three independent repeated experiments in this study were fixed at 42, 123, and 2024, respectively, and the global random seeds for the Ray distributed framework and NumPy were synchronously set during initialization to ensure consistency in all parallel environments and network initialization processes. Experimental results were obtained by taking the mean of three independent and repeated experiments, calculating 95% confidence intervals, and verifying the significance of the results using paired t-tests ($p < 0.05$). To explore the performance boundaries of the improved DDPG algorithm in a single electric-hydrogen hybrid microgrid, the experiment tested Renewable Energy Utilization (REU) rate, battery SOC effective volatility, and scheduling command tracking error rate under different renewable energy penetration rates and energy storage configurations, clarifying its advantages and bottlenecks as a single-agent method. The test results are shown in Figure 7.

In Figure 7(a), when the permeability was 30%, the REU of the improved DDPG algorithm was $96.5 \pm 1.0\%$, significantly higher than the $86.0 \pm 1.8\%$ when the permeability was 70% ($p < 0.05$). Due to the effective exploration

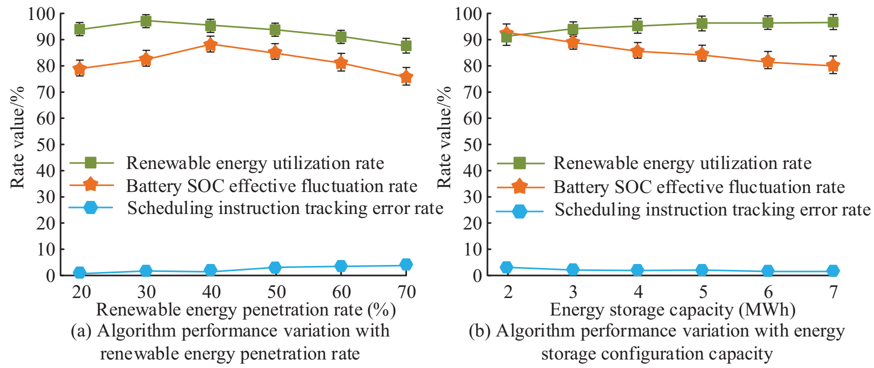


Figure 7 Performance analysis of improved DDPG algorithm in single microgrid scenario.

of the random noise attenuation mechanism, the optimal scheduling mode for the electric-hydrogen co-operation was found. Figure 7(b) shows that with a 4.0 MWh energy storage configuration, the improved DDPG algorithm had an REU of $96.8 \pm 1.1\%$ and an effective SOC fluctuation rate of $85.2 \pm 1.6\%$. Beyond this capacity, the marginal benefit of performance improvement decreased sharply, indicating that the algorithm's intelligence fully exploited the hardware's potential. To assess the effects of various microgrid configurations on REU, the experiment explored the role of electric-hydrogen bidirectional conversion facilities in improving energy absorption capacity. The experiment compared a microgrid system model, a model without battery storage, a model without electric hydrogen production facilities, and a baseline model (without both electric hydrogen production facilities and battery storage) to analyze the actual utilization level of renewable energy in random time periods. The test results are shown in Figure 8.

Figure 8(a) indicates that during the testing period, the average REU efficiency of the complete microgrid system model was 0.9877, which was higher than other models. Figure 8(b) shows that the efficiency of the battery-free energy storage model was 0.8955, close to the complete model, but still about 9.3% lower. Figure 8(c) shows that the efficiency of the model without an electric hydrogen production facility was only 0.5961. Figure 8(d) shows that the efficiency of the model missing both was even lower at 0.5110, a decrease of about 39.7% and 48.3% respectively compared to the complete model. Because the electricity-hydrogen conversion mechanism effectively avoided the waste of renewable energy by converting excess electricity into hydrogen for storage or reuse, especially during peak renewable energy generation periods, the electric hydrogen production facility could quickly

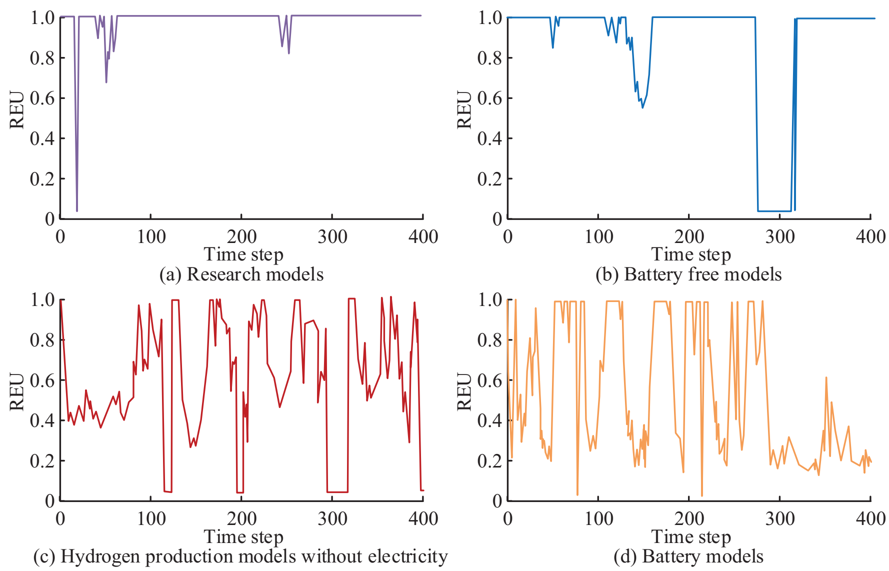


Figure 8 Comparison of REU for various models.

absorb excess electricity, while the energy storage system played a more significant role in regulating supply and demand and optimizing operational economics.

4.2 Energy Scheduling Test Based on Improved Multi-agent Proximal Strategy

The experiment aimed to evaluate the scalability and economic performance of the proposed CTDE-MAPPO optimization algorithm in multi-microgrid systems of different scales. Three scenarios with different numbers of microgrids were set up to compare the average system cost and average inter-microgrid electricity trading volume of C-MAPPO, the Single-agent PPO algorithm (SPPO), and the Independent PPO Algorithm Based on Shared Rewards (IPPO-S)under long-term operation, thus verifying the effectiveness of the algorithm in handling high-dimensional decision-making and promoting collaborative trading. The test results are shown in Figure 9.

As shown in Figure 9(a), when the number of microgrids increased from 3 to 5, the cost advantage of MAPPO-S compared to the fully centralized SPPO algorithm increased from 10.8% to 30.8%. Figure 9(b) shows that the MAPPO-S algorithm had the highest inter-microgrid electricity trading volume, which increased with the number of microgrids, reaching 23.8% higher

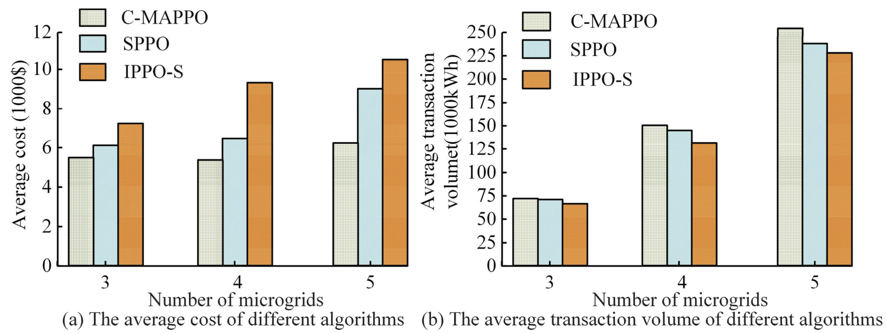
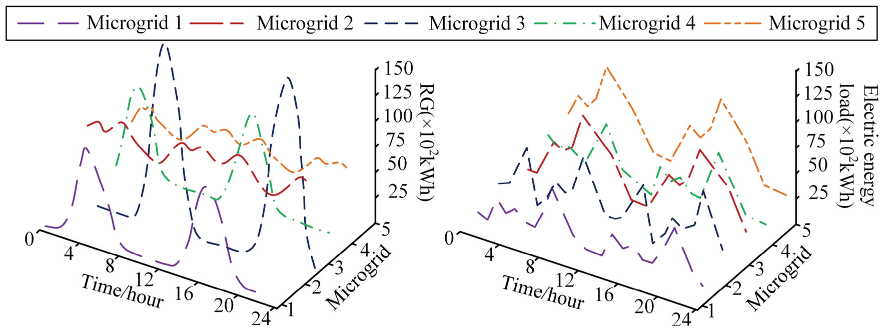


Figure 9 Comparison of average cost and average electricity trading volume for different algorithms.

than SPPO with 5 microgrids. Because MAPPO-S learned a global value function through centralized training to coordinate the behavior of multiple microgrids, its distributed execution decomposed decisions, thus overcoming the curse of dimensionality. This allowed it to accurately identify the spatiotemporal complementarity between microgrids, facilitating more P2P transactions, transforming high-priced externally purchased electricity into internal collaboration, and achieving the lowest operating cost. The experiment aimed to intuitively reveal the spatiotemporal differences in renewable energy generation and power load among microgrids in a multi-microgrid system. The experiment simulated 5 microgrids in different geographical locations, recording their operation over a continuous 24-hour period. The test results are shown in Figure 10.

As shown in Figure 10, the peak power generation of microgrid 1 occurred during hours 6–7 and 18–19, when its power load was low, resulting in a considerable power surplus. Microgrid 5, on the other hand, experienced extremely high load during hours 5–6 and 17–18 while power generation was at its lowest, facing a severe power shortage. The algorithm prompted microgrids to take more capture actions. This increased the total system cost due to the consumption of additional electrical energy and the increased operating costs of the Carbon Capture and Utilization (CCU) system. The experiment selected Mixed-Integer Linear Programming (MILP) for single microgrids and Distributed Alternating Direction Method of Multipliers (ADMM) for multi-microgrids as benchmark methods. The comprehensive performance of the proposed methods in single-microgrid and multi-microgrid scheduling scenarios was evaluated. The test results are presented in Table 2.



(a) Renewable energy generation of different microgrids

(b) The power load situation of different microgrids

Figure 10 Renewable energy generation and power load situation in single microgrid.

Table 2 Performance comparison of dispatch optimization methods

Method Type	Average Operating Cost (Yuan/kWh)	Renewable Energy Local Consumption Rate (%)	Daily Average Equivalent Cycle Degradation of Energy Storage (%)	Critical Section Power Limit Violation Time Proportion (%)
MILP	$0.436 \pm 0.012^{\#*}$	$93.2 \pm 1.2^{\mathrm{H}^*}$	$1.18 \pm 0.20^{**}$	$0.8 \pm 0.10^{**}$
Improved DDPG	0.381 ± 0.009	97.1 ± 0.9	0.92 ± 0.15	0.5 ± 0.08
ADMM	$0.480 \pm 0.015^{**}$	$88.5 \pm 1.5^{\#**}$	$1.05 \pm 0.18^{**}$	$1.2 \pm 0.18^{**}$
C- MAPPO	0.389 ± 0.007	95.8 ± 0.8	0.98 ± 0.12	0.7 ± 0.06

Note: "#′ indicates a significant difference compared to improved DDPG, and "*" indicates a significant difference compared to C-MAPPO, $p < 0.001$.

As presented in Table 2, the research method demonstrated superior performance across all metrics, significantly reducing the average cost per kilowatt-hour. In single-microgrid scenarios, the improved DDPG cost was 0.381 ± 0.009 yuan/kWh, a 12.6% reduction compared to MILP. In multi-microgrid scenarios, C-MAPPO costs 0.389 ± 0.007 yuan/kWh, an 18.9% reduction compared to ADMM. The absorption rates reached $97.1 \pm 0.9\%$ and $95.8 \pm 0.8\%$, respectively. The algorithm implicitly learned the system's safety boundary during training, thereby automatically avoiding risks. To verify the convergence of the algorithm training, the changes in reward values during the training process were tracked. The convergence curves of the improved DDPG algorithm and C-MAPPO algorithm are shown in Figure 11.

As shown in Figure 11(a), during the early stages of training ($0 \sim 300$ episodes), the strategy was in the strong exploration stage due to the influence of attenuated random noise. The average reward of the improved DDPG algorithm fluctuated sharply within the range of -380 yuan to -220 yuan.

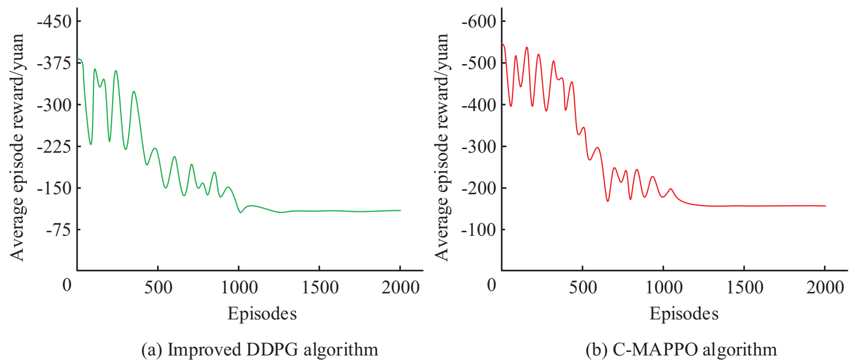


Figure 11 Convergence curves of improved DDPG algorithm and C-MAPPO algorithm.

During the mid training period (300∼800 episodes), the fluctuation narrowed to −200 yuan to −140 yuan. After 800 rounds, it converged and eventually stabilized at around −110 yuan. As shown in Figure 11(b), in the early stages of training (0 ∼ 400 episodes), it was necessary to coordinate the global interests of multiple agents, and the average reward calculated by C-MAPPO fluctuated within the range of −550 yuan to −380 yuan. During the mid training period (400˜1200 episodes), the price increased to −250 yuan to −170 yuan. After 1200 rounds, it converged and eventually stabilized at around −160 yuan. The smooth transition of the global reward curve without divergence or strategy collapse proved that both algorithms had good training stability. To evaluate the comprehensive impact of carbon capture on multi microgrid scheduling, this study used the C-MAPPO algorithm in 5 microgrid scenarios to compare the performance of the complete model (including carbon capture and P2P trading) with the CCU free model (removing carbon capture devices and methane reactors). The two models shared the same wind and solar power output curve, load data, energy storage configuration, P2P trading mechanism, and electricity price data. Carbon emissions only accounted for indirect emissions corresponding to purchased electricity (emission factor 0.85 kg CO_2/kWh) and the thermal power part within the microgrid (0.78 kg CO_2/kWh), and renewable energy generation did not account for emissions. The results of the ablation experiment are shown in Table 3.

As shown in Table 3, the carbon emissions of the complete model decreased by 74.5% compared to the model without CCU, but the operating costs increased by 7.5%. The cost increase was mainly due to the additional power consumption and maintenance costs of carbon capture systems. The

Table 3 Results of ablation experiment

Indicator	Complete Model	Model Without CCU
Carbon emissions (kg CO_2/kWh)	0.120	0.470
Operating cost (yuan/kWh)	0.389	0.362
Electricity purchase cost (yuan/kWh)	0.301	0.285

results indicated that carbon capture modules could effectively reduce carbon emissions. To verify the robustness and generalization ability of the proposed improved DDPG algorithm and C-MAPPO algorithm, this study selected the Mendeley Data dataset for testing. The original time series data of wind power, photovoltaic power and load in this dataset originated from the public records of NEMWEB of the Australian Energy Market Operator and were scaled and resampled according to the capacity ratio of microgrids, with a time interval of 15 minutes, covering 8760 points throughout the year and including complete seasonal variations. Since this dataset records only a single microgrid, to simulate different microgrid configurations, the original wind and solar power load time series were scaled by different capacity coefficients (randomly taken values from 0.7 to 1.3) and subjected to ± 2 hours of random time offset to construct 5 sub-microgrids with temporal and spatial differences. The proposed improved DDPG algorithm was compared with Twin Delayed Deep Deterministic Policy Gradient (TD3) and Soft Actor-Critic (SAC) in the single microgrid scenario. The C-MAPPO algorithm was compared with Multi-Agent Deep Deterministic Policy Gradient (MADDPG) and Q-value mixing network (QMIX) in the multi-microgrid scenario. All experiments were repeated 5 times (random seed 0–4), and the results were expressed as the mean $\pm$ standard deviation. The evaluation indicators were REU, Levelized Cost of Electricity (LCOE), and peak-time external power purchase ratio. The performance comparison results of each algorithm on the Mendeley dataset are shown in Table 4.

As shown in Table 4, in the single microgrid scenario, the REU of the proposed improved DDPG algorithm was $96.5 \pm 1.3\%$, the LCOE was 0.379 ± 0.010 yuan/kWh, and the proportion of external purchase during peak hours was $11.5 \pm 1.5\%$. This was significantly better than TD3 and SAC ($p < 0.05$). In the 5-subnet multi-microgrid scenario, the REU of the C-MAPPO algorithm was $95.0 \pm 1.2\%$, the LCOE was 0.385 ± 0.010 yuan/kWh, and the proportion of external purchase during peak hours was $9.4 \pm 1.1\%$. Both were significantly better than MADDPG and QMIX ($p < 0.05$). The results indicated that the proposed method was superior to the comparison algorithms in both single microgrid and multi-microgrid

Table 4 Performance comparison of various algorithms on the Mendeley dataset

Scene	Algorithm	REU (%)	Average Operating Cost (yuan/kWh)	Proportion of External Electricity Purchases During Peak Hours (%)
Single microgrid	Improved DDPG	96.5 ± 1.3	0.379 ± 0.010	11.5 ± 1.5
	TD3	$94.8 \pm 1.9^{\#}$	$0.398 \pm 0.016^{\#}$	$14.0 \pm 2.2^{\#}$
	SAC	$95.2 \pm 1.5^{\#}$	$0.390 \pm 0.013^{\#}$	$12.8 \pm 1.8^{\#}$
Multi microgrid	C-MAPPO	95.0 ± 1.2	0.385 ± 0.010	9.4 ± 1.1
	MADDPG	$92.5 \pm 1.8^{*}$	0.413 ± 0.019*	$12.0 \pm 2.0^{*}$
	QMIX	$93.3 \pm 1.4^{*}$	0.404 ± 0.015*	$10.9 \pm 1.6^{*}$

Note: "#" indicates a significant difference compared to improved DDPG, and "*" indicates a significant difference compared to C-MAPPO, $p < 0.05$.

scenarios, verifying the robustness and generalization ability of the method to seasonal variations and configuration differences.

5 Conclusion

This research aimed to improve energy utilization efficiency and reduce operating costs. For individual independent microgrids, a hybrid electric-hydrogen architecture and an improved DDPG algorithm were adopted. Relying on the centralized training and distributed execution mechanism of the C-MAPPO algorithm, energy trading and low-carbon coordination across microgrids were coordinated. Experiments showed that the improved DDPG achieved a utilization rate of 96.5% with a renewable energy penetration rate of 30%–40%, and an average cost per kilowatt-hour of 0.381 yuan/kWh, a 12.6% reduction compared to MILP. In a 5-microgrid scenario, C-MAPPO's cost advantage expanded to 30.8% compared to SPPO, and its electricity trading volume increased by 23.8% compared to SPPO. The research approach could improve the operating performance of different microgrid systems. Nevertheless, it did not fully consider extreme weather conditions and complex market trading rules. In the future, the algorithm can be extended to adapt to scenarios with deep coupling of multiple energy sources, and introduce uncertainty aware RL or robust optimization methods. By explicitly modeling the distribution deviation of wind and solar power caused by extreme weather, as well as price spikes and limit constraints in market trading rules, decision risks can be further quantified, thereby improving the robustness and safety of scheduling strategies in extreme scenarios.

Data Availability

The NREL dataset (Wind Toolkit v3.0.0 and NSRDB v3.2.0), PJM electricity price data, and Mendeley Data used in this study are all from publicly available resources. The pseudocode logic of the core algorithm refers to the improved DDPG (including attenuation noise detection mechanism) and C-MAPPO (including CTDE centralized training program) in Section 3. The network parameters, experience pool size, and training epochs are consistent with the experimental description in Section 4. With permission, the complete implementation code and trained model parameters can be obtained from the corresponding author.

References

[1] Yin Z, Zhang X, Luo W, Ruan S. Scalable Power Dispatch Automation Methods Based on Artificial Intelligence in Distributed Energy Systems. Distributed Generation & Alternative Energy Journal, 2026, 41(3): 791–814. DOI: 10.13052/dgaej2156-3306.41310.

[2] Zhong L S, Zeng Z L, Huang Z K, Shi X W, Bie Y M. Joint optimization of electric bus charging and energy storage system scheduling. Frontiers of Engineering Management, 2024, 11(4): 676–696. DOI: 10.1007/s42524-024-3102-2.

[3] Chen J, Hu N. Distributed Optimization Model for Economic Dispatch of Smart Grid. Distributed Generation & Alternative Energy Journal, 2025, 40(3): 457–480. DOI: 10.13052/dgaej2156-3306.4031.

[4] Zhou Z, Bian J Y, Yu Z Y. Incremental cost analysis model of distribution network based on economic dispatch of distributed new-energy storage system. Clean Energy, 2024, 8(2): 89–103. DOI: 10.1093/ce/zkae007.

[5] Huang Y, Pan L, Chen J, Pang Y, Shi F, Chen L, et al. Robust optimal operation scheduling method for integrated energy system considering flexible energy storage and mixed hydrogen natural gas. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2024, 46(1): 3476–3498. DOI: 10.1080/15567036.2024.2319723.

[6] Lu Z X, Xu X Y, Yan Z, Shahidehpour M, Sun W Q, Han D. Distributionally robust chance constrained optimization method for risk-based

routing and scheduling of shared mobile energy storage system with variable renewable energy. IEEE Transactions on Sustainable Energy, 2024, 15(4): 2594–2608. DOI: 10.1109/TSTE.2024.3429310.

[7] Dong L, Huang Y, Xu X, Zhang Z Y, Liu J Y, Pan L, Hu W H. Research on priority scheduling strategy for smoothing power fluctuations of microgrid tie-lines based on PER-DDPG algorithm. IET Generation, Transmission & Distribution, 2024, 18(20): 3221–3233. DOI: 10.104 9/gtd2.13267.

[8] Krishna R, Hemamalini S. Long short-term memory-based forecasting of uncertain parameters in an islanded hybrid microgrid and its energy management using improved grey wolf optimization algorithm. IET Renewable Power Generation, 2024, 18(16): 3640–3658. DOI: 10.1049/rpg2.13115.

[9] Todakar K M, Gupta P P, Kalkhambkar V, Sharma K C. Optimal scheduling of battery energy storage train and renewable power generation. Electrical Engineering, 2024, 106(5): 6477–6493. DOI: 10.100 7/s00202-024-02385-w.

[10] Yan D M, Wang H K, Gao Y J, Tian S J, Zhang H. Based on improved crayfish optimization algorithm cooperative optimal scheduling of multi-microgrid system. Scientific Reports, 2024, 14(1): 1–18. DOI: 10.1038/s41598-024-76041-5.

[11] Sui Q, Zhang J Y, Sun L, Liang J, Wei F R, Lin X N. Optimal scheduling of mobile energy storage capable of variable speed energy transmission. IEEE Transactions on Smart Grid, 2024, 15(3): 2710–2722. DOI: 10.1 109/TSG.2023.3329294.

[12] Sun H N, Zhang B C, Liu N. Imitation learning-based online optimal scheduling for microgrids: An approach integrating input clustering and output classification. IET Renewable Power Generation, 2024, 18(15): 3161–3172. DOI: 10.1049/rpg2.12909.

[13] Tian J J, Qian Y, Zhao F, Mo S L, Xiao H X, Zhu X T, et al. Energy-saving optimal scheduling under multi-mode “source-network-load-storage” combined system in metro station based on modified Gray Wolf Algorithm. Archives of Electrical Engineering, 2024, 73(1): 121–143. DOI: 10.24425/aee.2024.148861.

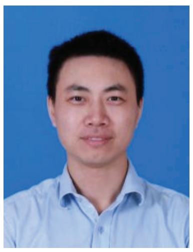
[14] Xue L, Niu T, Ge H, Zhang J, Xue Y, Fang S, et al. A joint distributed optimization framework for voltage control and emergency energy storage vehicle scheduling in community distribution networks. IEEE Transactions on Industry Applications, 2024, 60(4): 5317–5330. DOI: 10.1109/TIA.2024.3384474.

[15] Qian T, Ming W L, Shao C C, Hu Q R, Wang X L, Wu J Z, et al. An edge intelligence-based framework for online scheduling of soft open points with energy storage. IEEE Transactions on Smart Grid, 2024, 15(3): 2934–2945. DOI: 10.1109/TSG.2023.3330990.
[16] Ghaffarpour R, Zamanian S. A robust bi-level programming approach for optimal planning an off-grid zero-energy complex. IET Renewable Power Generation, 2024, 18(14): 2394–2415. DOI: 10.1049/rpg2.13083.
[17] Zhang H X. Application of day-ahead optimal scheduling model based on multi energy micro-grids with uncertainty in wind and solar energy and energy storage station. International Journal of Renewable Energy Development, 2024, 13(5): 873–883. DOI: 10.61435/ijred.2024.60218.
[18] Tang Y, Zhai Q Z, Zhao J X. Multi-stage robust economic dispatch with virtual energy storage and renewables based on a single level model. IEEE Transactions on Automation Science and Engineering, 2024, 21(4): 5490–5502. DOI: 10.1109/TASE.2023.3312379.
[19] Kang K, Zhang Y L, Miu Y J, Gao Q, Chen K W, Zeng Z H. Study on master-slave game optimization operation of integrated energy microgrid considering PV output uncertainty and shared energy storage. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2024, 28(3): 528–540. DOI: 10.20965/jaciii.2024.p0528.
[20] Shokri M, Niknam T, Mohammadi M, Dehghani M, Siano P, Ouahada K, et al. A novel stochastic framework for optimal scheduling of smart cities as an energy hub. IET Generation, Transmission & Distribution, 2024, 18(14): 2421–2434. DOI: 10.1049/gtd2.13202.
[21] Waqas M, Naseem A. Artificial intelligence in sustainable industrial transformation: A comparative study of Industry 4.0 and Industry 5.0. FinTech Sustain. Innov, 2025, 1: A2. DOI: 10.47852/bonviewFSI52025321.
[22] Kumar A, Dhillon J S. Emended Harris Hawk optimizer for mixed energy generation scheduling problem. Electric Power Components and Systems, 2024, 52(8): 1235–1268. DOI: 10.1080/15325008.2023.2239224.

Biographies

Yueli Zhou graduated from Huazhong University of Science and Technology in 2005 with a bachelor's degree in Water Resources and Hydropower Engineering; Assist engineers; With over 10 years of experience in the operation and management of pumped storage power stations, we specialize in centralized control and management of pumped storage power stations, as well as the construction and management of new energy storage power stations.

Shaohua Zhao, male, from Nanchong, Sichuan Province, graduated from the School of Electric Power of South China University of Technology in 2021 with a Bachelor's degree in Electrical Engineering and Automation. He is an assistant engineer and currently works as an automation operation and maintenance engineer at the Operation Center of Southern Power Grid Peak shaving and Frequency Regulation (Guangdong) Energy Storage Technology Co., Ltd. He has been engaged in the operation and maintenance management of electrochemical energy storage stations since he started my career, serving as the person in charge of Yaogu Energy Storage Station and possessing rich experience in the operation and management of electrochemical energy storage stations. Since 2021, as the project leader, he has organized multiple technical renovation projects and have some experience in the transformation of energy storage systems.



Jiasheng Wu graduated from South China University of Technology with a Master's degree in Fluid Machinery and Engineering in 2009, as a Senior Engineer. From August 2018 to April 2020, served as the Production Comprehensive Management Supervisor (Level 3) in the Production Technology Department of Peak shaving and Frequency Modulation Company. From April 2020 to July 2022, served as the Production Planning and Indicator Management Supervisor (Level 2) in the Production Technology Department of Peak shaving and Frequency Modulation Company. In July 2016, he won the second prize of China Electric Power Science and Technology Progress Award, in January 2018, he won the model worker of China Southern Power Grid, in June 2021, he won the Outstanding Communist Party Member, and in February 2021, he won the exemplary individual of Safe Production.

Qihua Lin (1999.11); he graduated from Wuhan University in 2022 with a Bachelor's degree in Electrical Engineering and Automation. From 2022 to

2023, he worked as a technician in the Electrical Department at the Technology Company Construction Center. Focused on the electrochemical energy storage infrastructure industry, served as the leader of the progress team and a member of the quality team for the independent battery energy storage project on the Nanhai power grid side in Foshan, Guangdong. Participated in the pre construction preparation work and construction management of the project.

Xiaodong Zheng, male, from Shanwei, Guangdong, graduated from the School of Electric Power, South China University of Technology in 2022 with a master's degree. Currently, he serves as the assistant project construction manager of the Construction Center of Southern Power Grid Peak shaving and Frequency Regulation (Guangdong) Energy Storage Technology Co., Ltd. I have been engaged in electrochemical energy storage related work since I started my career and have rich experience in electrochemical energy storage construction. Since 2023, he has been continuously involved in the development and construction of energy storage cloud platforms and centralized control management centers, and have accumulated certain experience in the fields of energy storage participation in electricity market trading and algorithm design.

Hanfeng Bai graduated as a Master of Engineering in Control Science and Engineering from Beijing University of Information Science and Technology in 2023. During his studies, he focused on research in the areas of network security and deep learning, and he have a certain understanding of fields such as information encryption. In 2019, he won the second prize in the Sichuan Division of the Electronic Design Contest. For my undergraduate degree, he graduated from the Automation program and possess solid skills in embedded development. During my graduate studies, he published an academic paper and accumulated rich experience in scientific research. Currently, he works as an Energy Storage Operations Engineer, primarily dealing with business related to electrochemical energy storage.