



ARTICLE

Energy System Low-Carbon Transformation Operation Optimization Based on Deep Deterministic Policy Gradient Algorithm

Jing Shi¹, Zesen Li¹, Delv Zhu¹, Bingjie Li¹ and Lang Gao^{2,*}

¹Economic and Technological Research Institute of State Grid Jiangsu Electric Power Co., Ltd., Nanjing, China

²Sichuan Energy Internet Research Institute Tsinghua University, Chengdu, China

*Corresponding Author: Lang Gao. Email: langgao21@outlook.com

Received: 11 December 2025; Accepted: 02 February 2026; Published: 18 October 2026

ABSTRACT: In view of the multi-energy subject coupling and operation optimization problems faced by the integrated energy system in the low-carbon transformation, taking a certain city in Jiangsu Province as the experimental object, the research first constructs a city-level low-carbon integrated energy system with hydrogen energy storage as the core hub. Then, the mathematical models of key equipment such as gas turbines, gas boilers, and thermal storage tanks are established. Finally, the multi-energy subject system achieves autonomous optimization and collaborative control through multi-agent DDPG. This method showed good energy utilization efficiency and scheduling flexibility in different seasons. The comprehensive energy utilization rate in heating season, non-heating season and transition season reached 82%, 80% and 78%, respectively. As the photovoltaic coverage rate increased from 5% to 40%, the wind power proportion increased from 10% to 50%, the total system operating cost decreased from US\$73.87/MWh to US\$49.00/MWh, carbon emissions dropped by approximately 32.4%, and the power curtailment rate decreased by approximately 21.7%. Compared with the single-agent algorithm, the average operating cost was reduced by 9.6%, the carbon emissions were reduced by about 8%, the energy balance error was reduced by about 39.8%, and the convergence speed was reduced by about 28%. These results fully verify the efficient decision-making and dynamic optimization capabilities of the multi-agent DDPG in complex multi-energy interaction environments, providing a feasible technical path for the intelligent operation.

KEYWORDS: DDPG; multi-agent; energy system; low-carbon; operation optimization

1 Background

With the increasingly severe global climate change and the transformation of energy structure, low-carbon, efficient and sustainable energy systems have become an important issue in the energy field [1]. The Integrated Energy System (IES) integrates different energy forms (such as heat, cooling, gas, etc.) and various conversion and storage devices (such as gas turbines, boilers, heat pumps, energy storage batteries, etc.) to achieve cascade utilization and complementary coordination of energy. It can significantly optimize energy utilization and reduce carbon emissions [2,3]. Existing research has shown that IES systems have demonstrated significant economic and environmental benefits in different application scenarios [4]. For example, considering the meteorological sensitivity of agricultural production, Fu and Zhou established an IES system that combined agricultural production with photovoltaic power generation by integrating carbon energy, electric energy, and thermal energy. In a greenhouse with a photovoltaic coverage rate of 25% and an area of 3996 m², the system saved 15% of energy costs [5]. For the energy optimization problem on the

residential user side, Ali et al. proposed an intelligent multi-carrier IES system for residential users that considered the uncertainty of natural gas and electricity prices. A three-stage modeling optimization method was used to minimize operating costs. Even under fluctuating energy prices, the system's operating deviation and average rate were still reduced by 2.14% and 2.64%, respectively, verifying the operational optimization advantages of IES system in dealing with various uncertain factors [6]. In addition, Jarwar systematically examined the role of thermal energy storage and electrical energy storage in reducing the levelized energy cost of different IES. Experimental results showed that thermal energy storage could enhance load transfer, seasonal balancing and waste heat utilization, reducing the operating costs of district heating applications by 25%. Electrical energy storage can effectively reduce power curtailment and support grid stability. It is expected to reduce the levelized energy cost of solar battery systems to less than US\$60/MWh by 2030 [7]. The above research not only verified the optimization potential of the IES system in multiple scenarios, but also revealed its important value in promoting low-carbon transformation.

However, faced with the complexity of multi-agent environments, traditional operation optimization methods are hard to deal with the nonlinear, high-dimensional and dynamic characteristics of IES [8,9]. Artificial intelligence and reinforcement learning have provided new ideas for solving such complex optimization problems [10]. For example, Qiu et al. modeled the local integrated energy system as a Markov decision process that considered multiple types of uncertainties. An improved Twin Delayed Deep Deterministic Policy Gradient (TD3) was built. Compared with the Deep Deterministic Policy Gradient (DDPG), the system operating cost was reduced by 2.76%, and the energy supply imbalance was reduced by approximately 88%, fully verifying the effectiveness and advantages of Deep Reinforcement Learning (DRL) in the operation optimization [11]. Li et al. used individual attention mechanisms and deep dense neural networks to perform reinforcement learning and centrally train all agents to achieve coordinated control. After training, the proposed method could manage multiple energy sources within the IES [12]. To lower power loss and voltage violations, Hu et al. proposed a dual-time scale hybrid voltage control strategy relying on the multi-agent reinforcement learning. This control strategy had excellent sampling efficiency and robustness, and also could effectively suppress voltage fluctuations [13]. In addition, Zhou et al. utilized the multi-agent DRL method to explore energy management issues focusing on joint transactions in multi-microgrid systems. A joint energy and carbon trading market was established to formulate the dispatch optimization problem as an inexact modeled Markov decision process. The operating cost was reduced by 14.86% and carbon emissions were reduced by 19.04% [14].

In summary, although existing research has verified the economics and emission reduction benefits of IES in different scenarios, and explored the application of DRL in scheduling, control, and carbon market transactions, relevant research is still limited when considering multi-agent interaction, low-carbon target constraints, and dynamic changes in carbon market prices simultaneously. Therefore, the research innovatively conducts modeling and related theoretical analysis on the IES containing hydrogen energy storage. On this basis, to further coordinate the scheduling of multi-energy entities and low-carbon development, an IES low-carbon transformation operation optimization method based on multi-agent DDPG is proposed. The research aims to make up for the shortcomings of existing research on multi-agent collaborative optimization under low-carbon constraints, and provide new theoretical and methodological support for applying IES systems in energy transition and carbon emission reduction.

2 Methods

2.1 Low-Carbon IES System Modeling and Related Theoretical Analysis

Hydrogen energy storage can convert excess wind power and photovoltaic resources into hydrogen for storage through “green power hydrogen production” technology. When power demand rises or new energy

output is insufficient, the hydrogen can be converted back into electrical energy through fuel cells [15]. The waste heat generated during the conversion process can also provide heat for the city, effectively balancing the output fluctuations of new energy power generation [16,17]. Based on this background, the study first takes a certain city of Jiangsu Province as the experimental object. A low-carbon integrated energy system architecture with hydrogen energy storage as the core hub is proposed, as shown in Fig. 1.

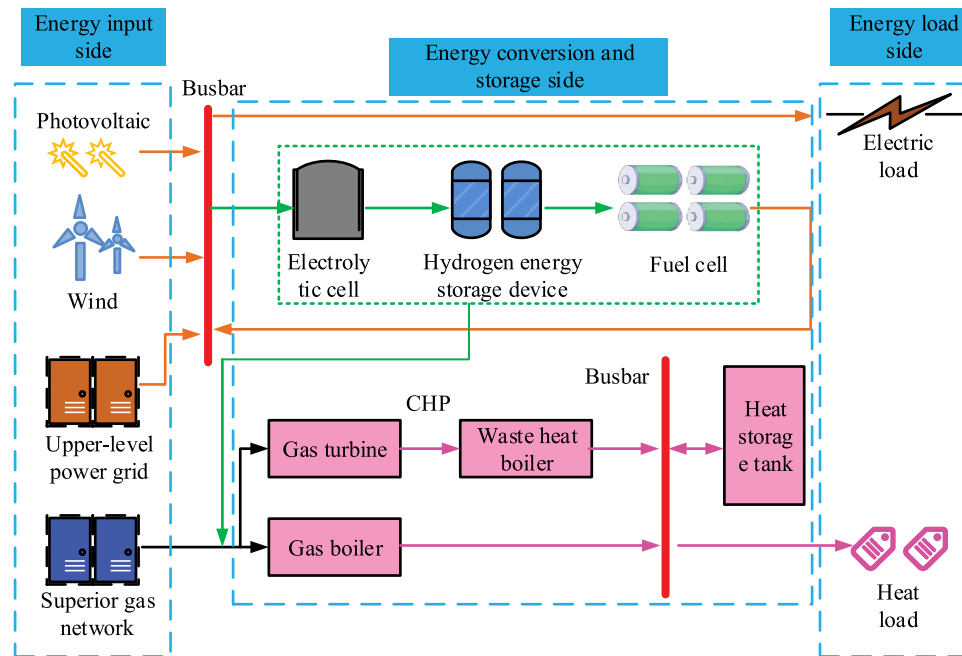


Figure 1: Low-carbon IES architecture.

From Fig. 1, the system mainly consists of the energy input side, the energy conversion and storage side, and the energy load side. On the energy input side, the system is linked to the upper-level power grid and upper-level gas network, and is equipped with renewable energy devices (photovoltaic and wind turbines) to realize on-site consumption of clean electricity. To improve the flexibility of energy utilization, an electrolytic water hydrogen production unit (electrolyser, fuel cell and hydrogen energy storage device) is set up to convert excess electricity into hydrogen energy. On the energy conversion and storage side, the system is equipped with gas turbines, gas boilers, waste heat boilers, heat storage tanks and Combined Heat and Power (CHP) equipment, which can realize complementary conversion such as electricity-heat, gas-heat, gas-electricity, etc. On the energy load side, the system covers multiple types of terminal needs such as electricity and heat. Electric energy is supplied by the power grid, renewable energy devices and hydrogen energy storage devices. Heat energy is provided by gas turbines, gas boilers, waste heat boilers, heat storage tanks and CHP units, thereby synergistically complementing different energy forms and dynamically balancing load demands. The hydrogen energy storage device not only assumes the buffer storage function of hydrogen, but can also inject hydrogen into the gas pipeline network and complete the deep coupling of electricity, heat and gas with the help of cogeneration equipment [18,19]. The principle of combined heat and electricity supply for hydrogen energy storage devices is shown in Fig. 2.

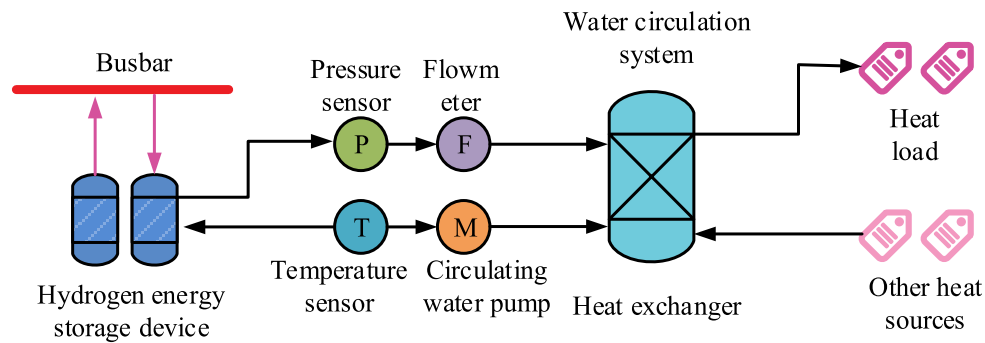


Figure 2: The principle of combined heat and electricity supply for hydrogen energy storage devices.

From Fig. 2, when the fuel cell is discharging, the stored hydrogen and oxygen react electrochemically to generate electrical energy and waste heat [20,21]. Electric energy is supplied to the power load through the electric bus. The waste heat is transferred to the circulating water system through the cooling circulating water and heat exchanger, and then transported to the heat load end of the city through the water pump, thereby realizing the heating needs for building heating or process heating. The heat exchanger can work together with other heat sources to improve the comprehensive utilization of system energy. The water circulation system is equipped with temperature sensors, pressure sensors, flow meters and circulating water pumps to adjust the operating status of circulating water in real-time to ensure the stability and efficiency of heat transfer. The hydrogen-doped transmission method for natural gas pipelines is presented in Fig. 3.

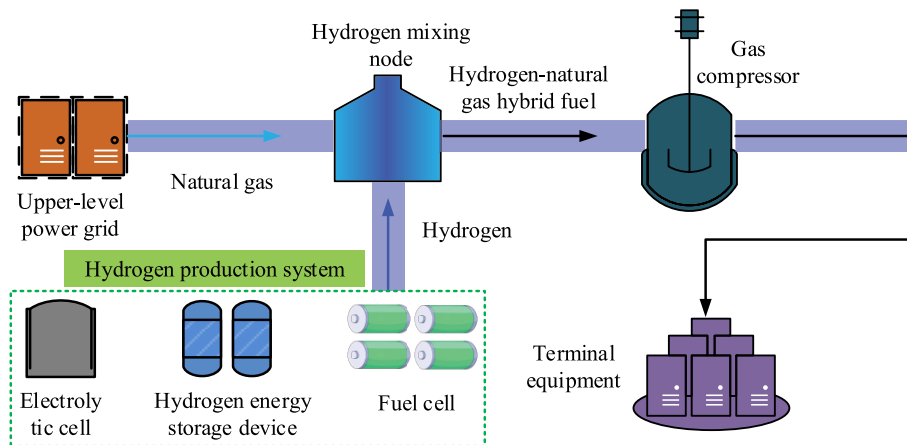


Figure 3: Hydrogen-doped transmission method for natural gas pipelines.

From Fig. 3, the electrolyzed water hydrogen production unit uses renewable energy as power input and produces hydrogen through electrolysis reaction to achieve localized production of green hydrogen energy. After pressure stabilization and purification, the produced hydrogen is transported to the hydrogen mixing node, and is fully mixed with natural gas from the upper-level gas network in a static mixer according to a set ratio to form a hydrogen-natural gas mixed fuel. After the mixed gas is pressurized by a gas compressor, it is transported to various terminal energy-consuming equipment through the natural gas transmission and distribution network in the city to provide fuel supply for industrial production or civil energy needs. In this process, the hydrogen doping ratio can be dynamically adjusted depending on system operation requirements and safety standards, utilizing renewable energy and reducing carbon emissions while ensuring

gas transmission safety and combustion stability, and achieving a clean and low-carbon transformation of the natural gas system. Subsequently, the research further establishes a mathematical model of the key equipment of the low-carbon IES. Combining the relationship between wind speed and fan output, the fan output is presented in Eq. (1).

$$P_t^{out} = \begin{cases} 0 & v_t < v_{in}^{cut}, v_t > v_{out}^{cut} \\ f(v) & v_{in}^{cut} < v_t < v_r \\ P & v_r < v_t < v_{out}^{cut} \end{cases} \quad (1)$$

In Eq. (1), P_t^{out} signifies the actual output power of wind turbine at time t . v_{in}^{cut} and v_{out}^{cut} signify the cut-in and cut-out wind speeds. v_r signifies the rated wind speed. v_t signifies the wind speed at time t . P signifies the rated output power of the wind turbine. $f(v)$ signifies the fitting polynomial. To reduce the complexity of photovoltaic output modeling, the study only selects two key factors, irradiance and ambient temperature, to construct a mathematical model of photovoltaic output, as shown in Eq. (2).

$$P_t^{PV} = P_{max} \frac{G_{c,t}}{G_{STC}} \{1 + k_p [T_{c,t} - T_{STC}]\} \quad (2)$$

In Eq. (2), P_t^{PV} signifies the photovoltaic output power at time t . P_{max} signifies the maximum output power. G_{STC} is the standard irradiance, which is 1000 W/m^2 . $G_{c,t}$ signifies the irradiance at time t . T_{STC} signifies the standard photovoltaic cell temperature, which is 25°C . $T_{c,t}$ is the photovoltaic cell temperature at time t . k_p signifies the power temperature coefficient. After the gas network is doped with hydrogen, the gas turbine is shown in Eq. (3).

$$P_t^{GT} = \eta_{GT}^e (H_t^{H_2} + H_t^{CH_4}) \quad (3)$$

In Eq. (3), P_t^{GT} is the output electric power of the gas turbine at time t . η_{GT}^e signifies the electrical conversion efficiency of the gas turbine. $H_t^{H_2}$ and $H_t^{CH_4}$ are the hydrogen and natural gas power consumed by the gas turbine, respectively. Under the hydrogen-natural gas mixed gas utilization, the mathematical model of the gas boiler is shown in Eq. (4).

$$Q_t^{GB} = \eta_{GT}^h (H_t^{H_2} + H_t^{CH_4}) \quad (4)$$

In Eq. (4), Q_t^{GB} signifies the output thermal power of the gas boiler at t , and η_{GT}^h signifies the heat conversion efficiency. $H_t^{H_2}$ and $H_t^{CH_4}$ satisfy the hydrogen and natural gas power consumed by the gas boiler, respectively. The heat storage tank is presented in Eq. (5).

$$Q_t^{TST} = (1 - \mu_{loss}) Q_{t-1}^{TST} + \eta_{in}^{TST} Q_{in,t}^{TST} - Q_{out,t}^{TST} / \eta_{out}^{TST} \quad (5)$$

In Eq. (5), Q_t^{TST} and Q_{t-1}^{TST} signify the heat storage capacity of the thermal storage tank at time t and $t - 1$. μ_{loss} signifies the heat loss coefficient of the thermal storage tank. η_{in}^{TST} and η_{out}^{TST} signify the charging and heat release efficiencies of the heat storage tank. $Q_{in,t}^{TST}$ and $Q_{out,t}^{TST}$ signify the input and output thermal power of the heat storage tank at time t . The relationship between the efficiency of the electrolyzer and its input power and hydrogen production is shown in Eq. (6).

$$P_{elec}(t) = \eta_{elec} \cdot \dot{m}_H(t) \cdot \Delta H_f \quad (6)$$

In Eq. (6), $P_{elec}(t)$ is the input power of the electrolyzer at time t . η_{elec} is the efficiency of the electrolyzer. $\dot{m}_H(t)$ is the hydrogen flow rate. ΔH_f is the enthalpy change of hydrogen. The energy balance

during hydrogen storage can be modeled by the relationship between the temperature, pressure and hydrogen storage capacity of the hydrogen storage tank, as shown in Eq. (7).

$$E_{\tan k}(t) = E_{\tan k}(t-1) + \eta_{storage} \cdot \dot{m}_H(t) \cdot \Delta H_f - \dot{m}_H(t) \cdot \Delta H_f \cdot \lambda_{loss} \quad (7)$$

In Eq. (7), $\eta_{storage}$ represents the energy efficiency of the hydrogen storage process. λ_{loss} represents the hydrogen loss coefficient during storage. The process of a fuel cell converting hydrogen into electrical energy can be modeled by its electrical conversion efficiency, as shown in Eq. (8).

$$P_{fuel}(t) = \eta_{fuel} \cdot \dot{m}_H(t) \cdot \Delta H_f \quad (8)$$

In Eq. (8), $P_{fuel}(t)$ represents the output power of the fuel cell at time t . η_{fuel} represents the efficiency of the fuel cell. When hydrogen is mixed with natural gas, the output power is shown in Eq. (9).

$$P_{gas}(t) = \eta_{gas} \cdot (m_{H_2}(t) \cdot m_{NG}(t)) + \Delta H_f \quad (9)$$

In Eq. (9), $P_{gas}(t)$ represents the gas turbine output power at time t . $\eta_{gas}(t)$ represents the gas turbine efficiency. $m_{H_2}(t)$ and $m_{NG}(t)$ represent the flow rates of hydrogen and natural gas, respectively. The hydrogen mass and energy balance of each device in the system are constrained by the integrated energy flow method, as shown in Eq. (10).

$$\dot{m}_H^{in}(t) - \dot{m}_H^{out}(t) = \dot{m}_H^{storage}(t) \quad (10)$$

In Eq. (10), $\dot{m}_H^{in}(t)$ represents the hydrogen input flow rate. $\dot{m}_H^{out}(t)$ represents the hydrogen output flow rate. $\dot{m}_H^{storage}(t)$ represents the amount of hydrogen stored in the hydrogen storage system.

The mixture of hydrogen and natural gas is converted into hydrogen through a water electrolysis unit, converting surplus renewable electricity into hydrogen. The hydrogen is then transported through the natural gas pipeline network to the terminal equipment. To ensure system stability and safety, high-precision flow meters and gas mixers are used to adjust the hydrogen-to-natural gas flow ratio in real time, ensuring the hydrogen blending ratio remains within the range of 0%–30%. The Wobbe index is monitored to ensure the combustion characteristics of the mixed gas meet equipment requirements, preventing excessive hydrogen ratios from affecting equipment performance. To prevent hydrogen leakage, hydrogen leak sensors are installed at critical points (such as natural gas pipelines and storage tanks). These leak sensors, combined with an automatic alarm system, can quickly activate safety measures when a hydrogen leak is detected, such as closing valves or activating an automatic pressure relief system. Simultaneously, high-strength, leak-resistant materials (such as stainless steel and nickel alloys) are used in the natural gas pipelines and storage tanks to ensure system safety under high-pressure conditions. The system design process strictly adheres to relevant industry standards, such as ISO 14687 (hydrogen quality requirements) and EN 16723 (standard for blending hydrogen in natural gas systems), to ensure that gas mixing and transportation meet international standards and guarantee the long-term stable operation of the system. The overall control process of the low-carbon IES is shown in Fig. 4.

From Fig. 4, the system adopts a control method that combines “heating with electricity” and “optimal electricity price response”. Based on the actual electricity and heat load status and external electricity prices in a certain city in Jiangsu Province, the output of hydrogen energy storage devices, fuel cells, CHP units and gas boilers is dynamically adjusted. When the electric load is sufficient and the heat load is insufficient, priority is given to using the waste heat of the fuel cell for heating. If there is still a heat gap, it will be compensated by the gas boiler. When the electricity and heat loads are met and there is excess new energy, the system

starts electrolysis to produce hydrogen, converts the excess electricity into hydrogen storage, and stores the waste heat of the fuel cell into a heat storage tank. When the electricity load is insufficient and the heat load is sufficient, dynamic dispatch is performed based on the electricity price: If electricity prices are high, the fuel cell or gas turbine will give priority to power supply. If necessary, power will be purchased from outside. If electricity prices are low, electricity will be purchased directly from the grid. When the electricity and heat loads are insufficient, the system is optimized according to the electricity price zone: If electricity prices are high, the fuel cell and CHP operate jointly, and the heat shortage is compensated by the boiler. If electricity prices are low, the fuel cell and CHP are prioritized. If necessary, outsourced power or heat is purchased, thereby achieving dynamic optimal operation of economy and low-carbon. If the water electrolysis hydrogen production unit uses renewable energy, it is assumed that this part of the electricity does not generate carbon emissions. However, if grid electricity is used, the corresponding carbon emissions need to be calculated based on the marginal emission factor of the grid.

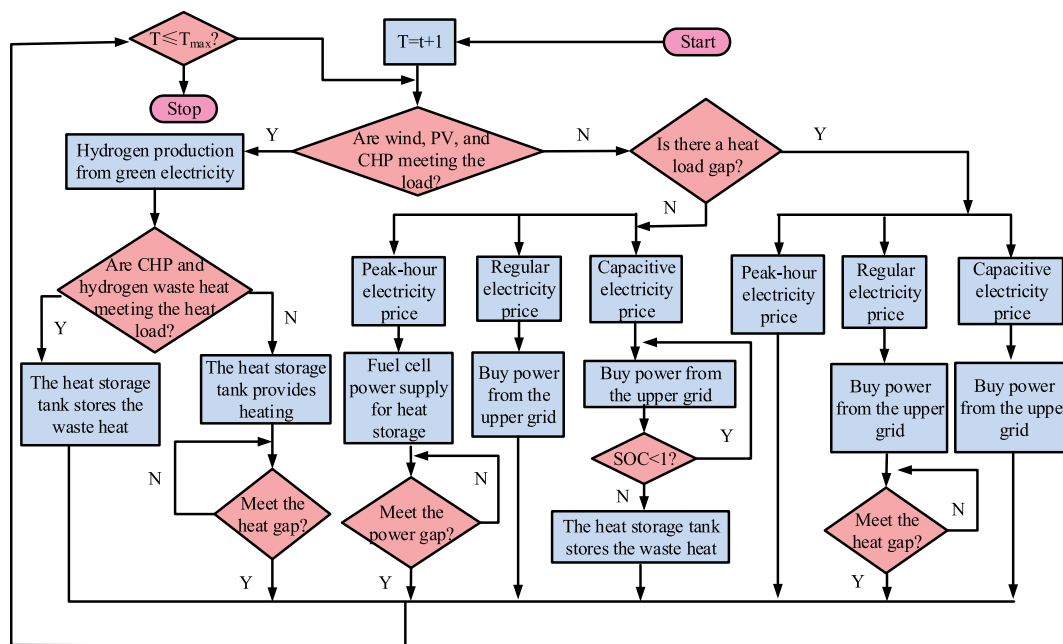


Figure 4: The overall control process of the low-carbon IES.

2.2 IES Low-Carbon Transformation Operation Optimization Method Based on Multi-Agent DDPG

The IES low-carbon system can coordinately output electric energy and thermal energy, which can not only meet the power and thermal needs of the city, but also optimize the energy utilization efficiency and overall economy. To deal with the operation optimization problem of multi-agent in the low-carbon IES, combined with the designed system architecture, an IES low-carbon transformation operation optimization method based on multi-agent DDPG is proposed. This method uses the overall idea of centralized training and decentralized execution, and divides the low-carbon IES into multiple intelligent entities according to different parks. While maintaining local independent perception and decision-making capabilities, each agent shares global information through centralized training to achieve the system operation optimization. In the model, each city park is treated as an independent agent, representing different energy operating companies or management units. During the training phase, all agents share global system state information (including electricity price, load, renewable energy output, hydrogen storage status, etc.), and the global Q-value function is uniformly evaluated by a central Critic network. In the execution phase, each campus

runs the Actor network independently and completes scheduling decisions based only on local observation data without the need to exchange detailed operational data in real time. This approach ensures both the data privacy and management autonomy of each operating entity, as well as the consistency and stability of the overall system scheduling. Furthermore, by embedding system operating costs, carbon emissions, and energy balance constraints into the reward function, each park can automatically meet city-level low-carbon operation goals while optimizing its own revenue, thereby achieving interest coordination and global collaborative optimization among multiple operating entities.

In the multi-agent DDPG algorithm, each agent adopts an actor-critic architecture. The core idea of the Actor-Critic architecture is to optimize through the Actor Network and Critic Network, improving the efficiency and performance of the decision-making process [22–24]. The Actor Network uses the local observation data as input and outputs the corresponding action decision. The Critic Network receives the global status information of the IES and the joint actions of all agents as input, which calculates the global Q-value function and evaluates the pros and cons of each strategy. Fig. 5 presents the Actor Network.

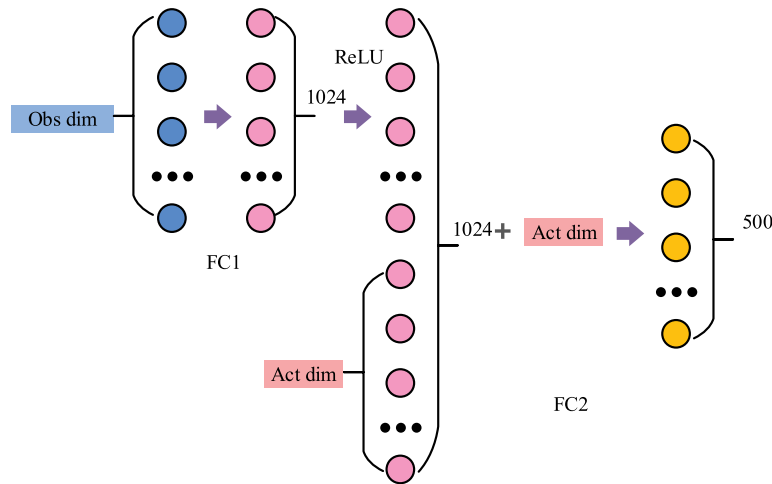


Figure 5: Actor network structure.

From Fig. 5, the Actor Network has two fully connected layers. The first layer is a fully connected layer with 1024 neuron nodes. ReLU is the activation function to enhance the nonlinear expression ability. The second layer is a fully connected layer with 500 neuron nodes, which is responsible for high-dimensional abstraction and mapping of features. The output layer of the network generates continuous action values based on the action dimensions of the corresponding agent, which are used to describe the optimal scheduling strategy, thereby completing the forward propagation process of action decisions. Fig. 6 presents the Critic Network.

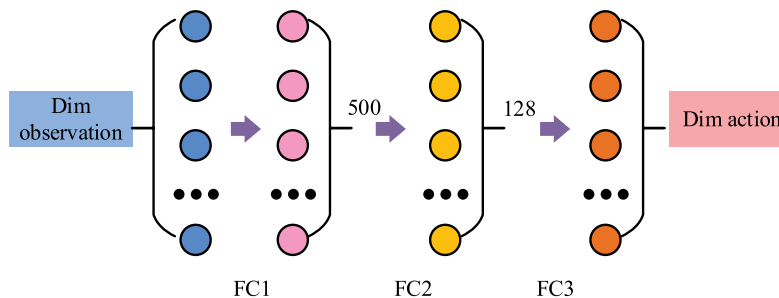


Figure 6: Critic network structure.

From Fig. 6, the Critic Network has three fully connected layers. The input layer is first mapped to a 500-dimensional feature representation through the first fully connected layer. Then, it is further compressed into a 128-dimensional feature space through the second fully connected layer. Finally, the Q-value estimate corresponding to the action space dimension is output through the third fully connected layer. Information is transferred layer by layer through forward connections between each layer, and optimal actions are found through gradient ascent. The parameters of the Actor Network and Critic Network of agent j are θ_j^μ and θ_j^Q , respectively, and the parameters of the target network are θ_j^Q and $\theta_j'^Q$, respectively. The output action A_t^j of the Actor Network is shown in Eq. (11).

$$A_t^j = \mu_j \left(O_t^j \mid \theta_j^\mu \right) + N_t \quad (11)$$

In Eq. (11), $\mu_j \left(O_t^j \mid \theta_j^\mu \right)$ is the action strategy function of agent j under parameter θ_j^μ . N_t is the noise term. The optimal scheduling strategy π_j^* is shown in Eq. (12).

$$\pi_j^* = \arg \max J_{\pi_j} \left(\theta_j^\mu \right) \quad (12)$$

In Eq. (12), $J_{\pi_j} \left(\theta_j^\mu \right)$ signifies the policy objective function of the Actor Network. π_j signifies the action strategy of j . The Q function of agent j at time t is $Q_j \left(O_t^l, \dots, O_t^j, \dots, O_t^U, A_t^l, \dots, A_t^j, \dots, A_t^U \right)$. In this partially observable Markov process S_t , the corresponding Q function is displayed in Eq. (13).

$$Q_j \left(O_t^l, \dots, O_t^j, \dots, O_t^U, A_t^l, \dots, A_t^j, \dots, A_t^U \right) = Q_j \left(S_t, A_t^l, \dots, A_t^j, \dots, A_t^U \right) \quad (13)$$

The optimal action $\nabla_{\theta^\mu} J_{\pi_j}$ is shown in Eq. (14).

$$\nabla_{\theta^\mu} J_{\pi_j} = E_{\theta_j^\mu} \left[\nabla_{a_i} Q_i \left(S_t^j, A_t^j \mid \theta_j^Q \right) \mid \nabla_{\theta_j^\mu} \mu_j \left(O_t^j \mid \theta_j^\mu \right) \right] \quad (14)$$

In Eq. (14), $E_{\theta_j^\mu}$ is the policy objective function of the Critic Network. $Q_i \left(S_t^j, A_t^j \mid \theta_j^Q \right)$ is the action state function of agent j . Based on the above design, the multi-agent DDPG algorithm operation optimization process is shown in Fig. 7.

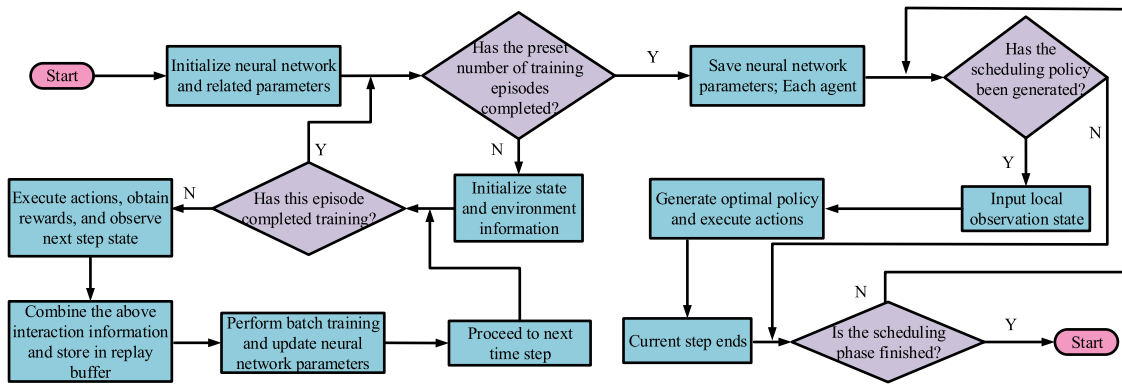


Figure 7: Run optimization process.

From Fig. 7, the neural network and related parameters are first initialized. At the beginning of each training cycle, the system determines whether the preset training rounds have been reached: If it has not

been reached, it will enter the centralized optimization stage. If the conditions are met, it will enter the decentralized training stage. In the centralized optimization stage, the low-carbon IES system first initializes the environmental state and detects the round progress. If the current training round has been completed, the corresponding action strategy is generated and executed based on network parameters and environmental information. The system calculates rewards and updates the environment status based on execution feedback, and stores interactive information such as status transitions, actions, and rewards into the experience pool for batch updates. After optimizing the network parameters through gradient descent, the process enters the next round of training. In the decentralized training stage, each agent loads the network parameters obtained by training and independently executes decisions, and determines whether to generate a scheduling strategy. If a scheduling strategy is generated, the optimal decision plan is output and executed based on the local observation state. When the current time step ends, the system determines whether the entire scheduling cycle has been completed. If it is completed, the training will be terminated. Otherwise, the policy update for the next period will continue.

The reasonable reward function is a key prerequisite for the multi-agent DRL algorithm to achieve long-term reward maximization [25–27]. The study uses the negative value of the objective function as the basic reward. A linear penalty term is introduced to constrain behaviors that do not meet the constraint conditions. The penalty term is not only related to overshoot (i.e., the extent to which the energy storage device exceeds its maximum charging capacity) and discharge (i.e., the extent to which it falls below the expected discharge level), but also includes the overcharge and discharge penalty coefficient of the energy storage device, which further strengthens the control over the dynamic behavior of the system. The reward function $R_{j,t}$ is presented in Eq. (15).

$$\begin{cases} R_{j,t} = -C_j - R_{j,t}^b \\ R_{j,t}^b = \varsigma_1 P_{j,t}^{ESS,oc} + \varsigma_2 P_{j,t}^{ESS,od} \\ P_{j,t}^{ESS,oc} = SOC_{j,t} - P_{j,c,max}^{ESS} \\ P_{j,t}^{ESS,od} = SOC_{j,t} - P_{j,d,max}^{ESS} \end{cases} \quad (15)$$

In Eq. (15), C_j and $R_{j,t}^b$ are the objective function and penalty term, respectively. $P_{j,t}^{ESS,oc}$ and $P_{j,t}^{ESS,od}$ are the overshoot and discharge amounts, respectively. ς_1 and ς_2 are the overcharge and discharge penalty coefficients of the energy storage device. SOC signifies the energy storage state of charge. The objective function optimizes the system's cost, emissions, and reduction targets through negative feedback, driving the system towards low-carbon and high-efficiency development. A linear penalty term ensures that the energy storage device is penalized for overcharging and over-discharging, thus preventing system instability caused by overcharging or discharging. The penalty coefficient is a dynamically adjusted coefficient whose value can be adjusted according to different operating conditions and objectives to balance the weights between rewards and penalties. The decision variables mainly include the output power of various energy devices (such as hydrogen storage devices, gas turbines, and gas boilers). The objective function for optimization is the system's operating cost, carbon emissions, energy balance error, and maximizing energy efficiency. Constraints include the operating constraints of each energy device (such as upper and lower power limits), load demand (such as electricity and heat demand), energy storage status (such as the charging and discharging limits of hydrogen storage devices), and policy constraints for low-carbon transition. The time frame is the entire operating cycle, with time steps discretized within each hour.

Rule-based control logic (such as “heat-to-electricity” and “optimal electricity price response”) primarily serves to guide and constrain, ensuring the system's basic operational stability under specific conditions. Meanwhile, the multi-agent DDPG strategy provides adaptive decision-making capabilities in dynamic

and complex environments. The combination of the two ensures that the system can flexibly respond to electricity price and load fluctuations while achieving optimal scheduling globally through reinforcement learning. Therefore, rule-based control logic is not completely replaced but rather serves as an auxiliary mechanism working in conjunction with the DDPG policy to guarantee optimal system operation under different circumstances.

3 Results

3.1 IES Low-Carbon System Performance Testing

To verify the low-carbon IES system under different seasons, different loads and renewable energy conditions, a suitable experimental environment was established. The annual load and meteorological information of a certain city of Jiangsu Province from April 2022 to March 2023 was selected. The sampling interval is 1 h, with a total of 8760 sets of data, including electrical load, heating load, natural gas consumption, photovoltaic output, wind power output, ambient temperature and electricity price information. The linear interpolation method and mean substitution method are used to repair missing data, the 3σ criterion is used to eliminate outliers, and the data are input into the system after normalization. With reference to the “Building Energy Saving Design Standard” (GB 50189-2015) and the typical climate characteristics of Jiangsu Province, the annual operation cycle is split into heating season (November-March of the following year), non-heating season (June-September), and transition season (April-May, and October). Multiple typical days (working days) are selected each quarter for comparative analysis, and rolling evaluations are conducted across multiple seasons and days based on these typical days. Each seasonal and typical day evaluation is an independent test set, ensuring that the system’s performance is verified under different environmental conditions to better assess its generalization capabilities. During the experiment, the system reads load, renewable energy output, meteorological, and electricity price information hourly, and dynamically generates optimal scheduling strategies based on equipment operating constraints, while simultaneously executing power allocation and energy dispatch. Furthermore, the traditional IES system is used as a benchmark. The system continuously records the consumption of electricity, heat, and gas, as well as related economic indicators, to evaluate the system’s performance under different configurations. The experimental environment software and hardware settings are presented in [Table 1](#).

Table 1: Hardware and software environment settings.

Serial Number	Environment	Settings
1	CPU	Intel Xeon Gold 6338 2.0 GHz ×2
2	RAM	256 GB DDR4
3	GPU	NVIDIA A100 40 GB
4	Storage	4 TB NVMe SSD
5	Operating system	Ubuntu 22.04 LTS
6	Programming language	Python 3.11
7	Deep learning framework	PyTorch 2.1
8	Photovoltaic device	50~100 MW
9	Gas turbine	50~100 MW
10	Electric boiler	20~50 MW
11	Hydrogen energy storage device	50~200 MWh

According to the settings in Table 1, the study first conducted an energy utilization efficiency test on the low-carbon IES system. The results are shown in Fig. 8.

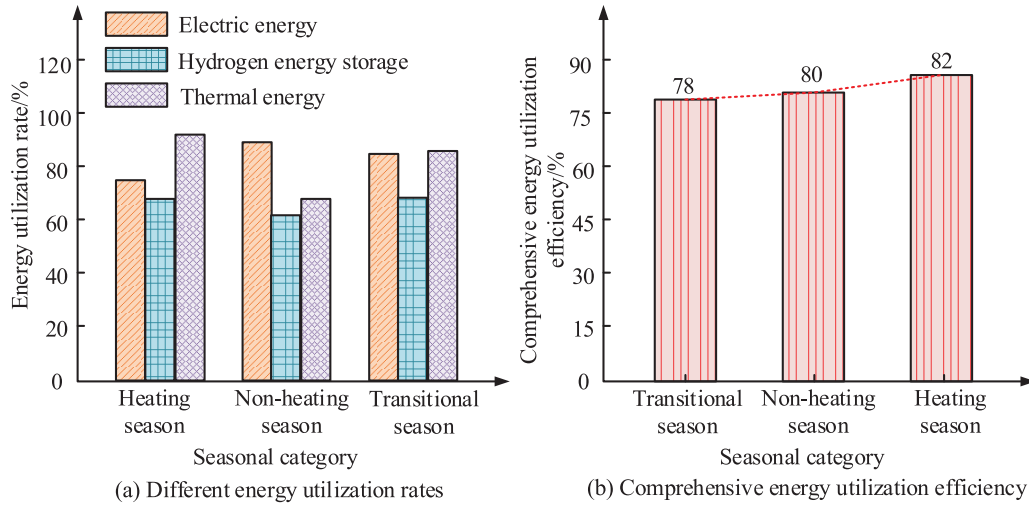


Figure 8: Energy utilization efficiency test.

Fig. 8a shows the electric energy, thermal energy and hydrogen energy storage utilization of the low-carbon IES in different seasons, and Fig. 8b shows the comprehensive energy utilization efficiency in each season. From Fig. 8a, during the heating season, the thermal energy utilization rate was the highest, reaching 92%, which could fully meet the city's heating needs. The electric energy utilization rate was 78%, and the hydrogen energy storage utilization rate was about 65%, indicating that the system could maintain reasonable electric energy and energy storage dispatching efficiency while ensuring the supply of heat loads. During the non-heating season, the electric energy utilization rate increased to 90%, the thermal energy utilization rate dropped to 65%, and the hydrogen energy storage utilization rate was about 60%, indicating that the thermal load of the city was low and the system scheduling was more biased towards electric energy utilization. During the transition season, each energy utilization rate was between the heating season and the non-heating season, and the overall energy utilization efficiency remained above 82%, indicating that the system could maintain high dispatch flexibility and energy utilization levels. From Fig. 8b, the comprehensive energy utilization efficiency was about 82% in the heating season, about 80% in the non-heating season, and about 78% in the transition season. The overall changes are stable, indicating that the low-carbon IES system can effectively coordinate the utilization in different seasons and achieve the stability of seasonal energy regulation. The photovoltaic coverage rates are set to 5%, 10%, 25%, and 40%, respectively, and the wind power capacity proportions are set to 10%, 20%, 35%, and 50%, respectively, and perform typical daily dispatching. The system economic and carbon emission performance test results under different renewable energy configuration proportions are shown in Fig. 9.

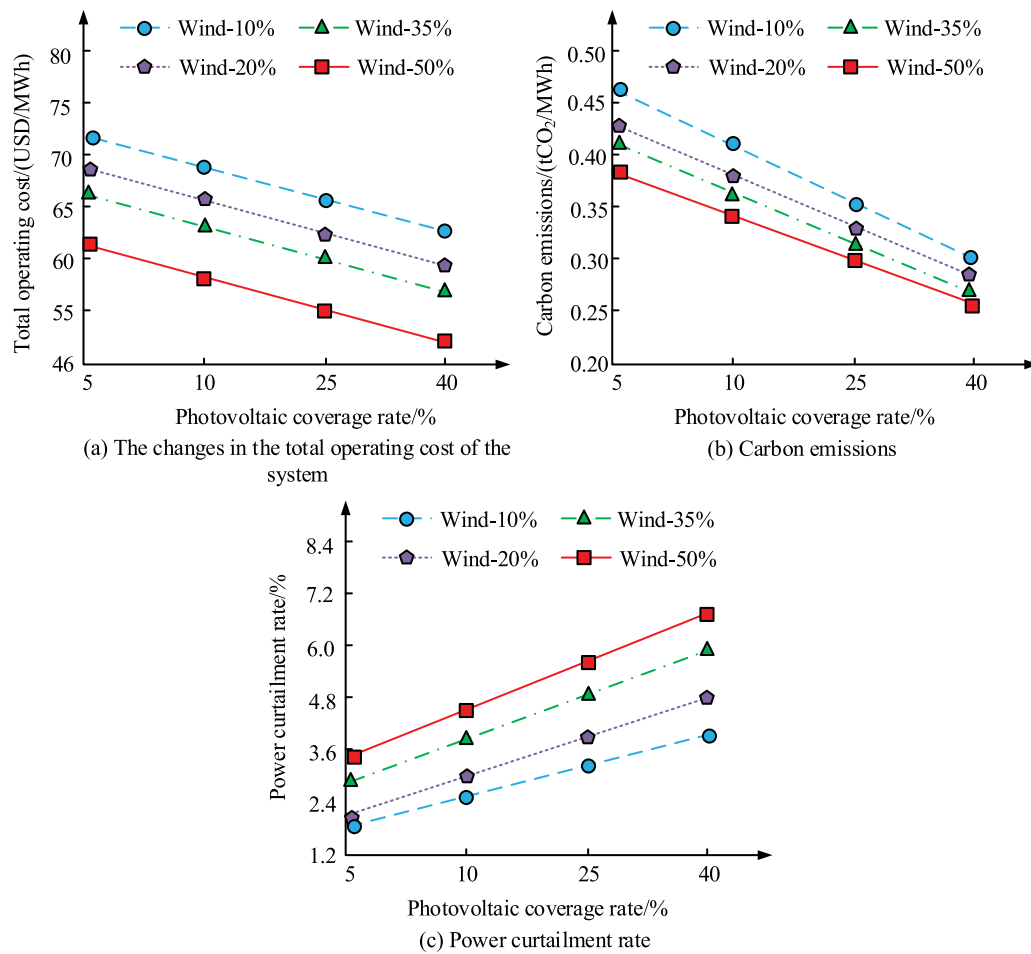


Figure 9: Test results of economic efficiency and carbon emission performance.

Fig. 9a shows the changes in the total operating cost under different proportions of photovoltaic and wind power capacity. Fig. 9b,c presents the corresponding carbon emissions and power curtailment rates, respectively. From Fig. 9a, the total operating cost of the system showed an obvious downward trend. The higher the wind power proportion, the lower the total cost. When the photovoltaic coverage rate increased from 5% to 40% and the wind power proportion increased from 10% to 50%, the total system cost decreased from US\$73.87/MWh to US\$49.00/MWh, a decrease of approximately 33.6%. From Fig. 9b,c, the system carbon emissions decreased significantly, from 0.46 to 0.28 t CO₂·MWh⁻¹, with an emission reduction rate of approximately 39.1%. However, as the proportion of new energy sources was too high, the power curtailment rate also increased from 1.8% to 6.7%. The increased volatility of wind and solar output has led to an imbalance in supply and demand during certain periods. Taking into account cost, carbon emissions and power abandonment rate, when the photovoltaic coverage rate is 25% and the wind power proportion is 35%, the system operation can reach an optimal state. Under this combination, the study further explored the impact of energy storage capacity on peak and valley adjustment capabilities and new energy absorption capabilities. The test results are shown in Fig. 10.

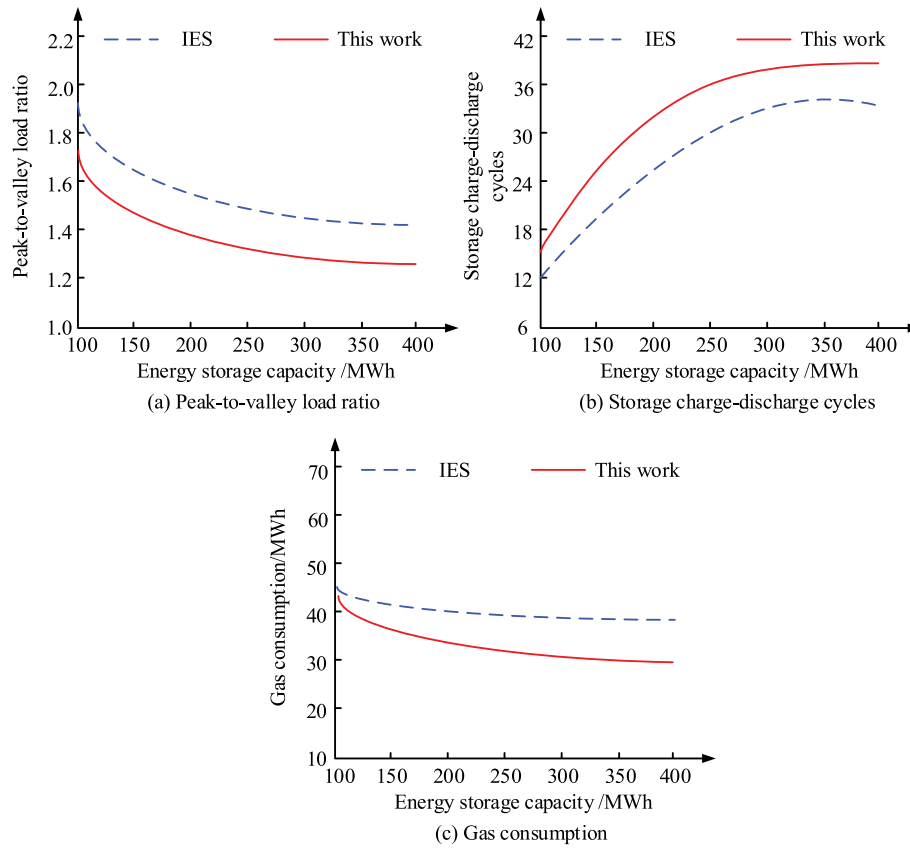


Figure 10: Peak-to-valley load ratio, charge-discharge cycle, and gas consumption change curves.

Fig. 10a–c respectively shows the peak-to-valley load ratio, storage charge-discharge cycles, and gas consumption change curves of the system under different energy storage capacities. From Fig. 10a, as the energy storage capacity increased from 100 to 400 MWh, the peak-to-valley load ratio of the low-carbon IES decreased from 1.75 to 1.25, with a decrease of approximately 28.6%. It shows that the energy storage device can effectively cut peaks and fill valleys. The load balancing is effectively achieved, the peak and valley load differences in the city are alleviated, and the overall scheduling flexibility is improved. From Fig. 10b, when the energy storage capacity increased, the storage charge-discharge cycles of the low-carbon IES showed an increasing trend, increasing from 15 times to 38 times. The increased energy storage capacity improves the adjustment capability, allowing energy storage to participate in power balancing more frequently during the peak time of new energy output or load fluctuations. From Fig. 10c, as the energy storage capacity increased, the gas consumption of the low-carbon IES gas turbine and gas boiler also decreased significantly, with a decrease of approximately 23.4%. The energy storage device uses new energy output and low-priced electricity to replace the operation of some gas equipment, reduce fossil energy consumption, reduce carbon emissions, and achieve dual optimization of economy and low-carbonity. Based on the above experimental results, the designed low-carbon IES can achieve efficient coordinated utilization in different seasons and different load conditions. After the renewable energy penetration rate and energy storage capacity are optimized, good results can be achieved in peak shaving and valley filling, reducing carbon emissions and controlling operating costs.

3.2 Analysis of Multi-Agent Operation Optimization Results

Based on the system environment constructed in Section 3.1, the study used a typical day in the heating season (10 January) and a typical day in the non-heating season (15 July) as the test period, and performed dynamic simulation with a time step of 1 h. The training step is 2×10^5 times, the learning rate is 1×10^{-4} , the discount factor is 0.99, the experience pool capacity is 10^6 , the batch size is 128, and the exploration noise obeys a Gaussian distribution with a mean of 0 and a variance of 0.2. The TD3 algorithm, single-agent DDPG algorithm, MILP, MIQP, MINLP and Gray Wolf Optimizer (GWO) algorithm were introduced for comparative testing. Under the same load, weather and electricity price conditions, the differences between different algorithms were compared in terms of system operating cost, carbon emissions, energy balance error and training convergence speed, as presented in Table 2.

Table 2: The system operation optimization results of different algorithms on typical days.

Season	Algorithms	Operating Cost (CN¥10,000/day)	Carbon Emissions (t CO ₂ /day)	Energy Balance Error (%)	Convergence Steps ($\times 10^4$)
Typical days of the heating season	TD3	24.86	82.37	2.94	10.1
	Single-agent DDPG algorithm	25.93	85.64	3.52	11.7
	GWO	28.46	92.73	4.81	15.6
	Multi-agent DDPG algorithm	23.42	79.18	2.12	8.4
	MILP	22.15	75.82	1.81	5.2
	MIQP	21.76	74.23	1.63	5.0
	MINLP	21.50	73.12	1.45	4.8
A typical day outside the heating season	TD3	16.88	57.06	2.45	8.7
	Single-agent DDPG algorithm	17.43	59.21	2.86	9.8
	GWO	19.02	63.97	3.72	13.8
	Multi-agent DDPG algorithm	15.91	54.52	1.83	7.2
	MILP	14.82	51.65	1.51	4.5
	MIQP	13.31	49.89	1.43	4.3
	MINLP	14.06	48.10	1.35	4.1

Note: Operating costs involve power purchase, gas purchase and carbon emission costs. Carbon emissions are converted based on gas consumption and grid power purchase carbon factor. Energy balance error is the root mean square error of supply and demand power deviation as a proportion of the total system load.

From Table 2, under typical days in different seasons, each algorithm showed consistent performance trends on operating costs, carbon emissions, and energy balance, but the difference varied due to seasonal load changes. On typical days during the heating season, the system heat load increased significantly and carbon emission levels generally rose. Although the GWO algorithm can quickly provide feasible solutions, it has a large energy balance error (4.81%) and the largest number of convergence steps (15.6×10^4), indicating that its optimization stability is limited in highly coupled nonlinear scenarios. The single-agent DDPG algorithm improved energy allocation decision-making through DRL strategies, reducing operating costs by about 8.9% and reducing carbon emissions by about 7.6% compared with the GWO algorithm. However,

because there is only a single strategy network, it is still easy to fall into local optimality. Influenced by delayed update and target smoothing mechanisms, the TD3 algorithm further improved convergence stability and global optimization capabilities, and the energy balance error dropped to 2.94%. The multi-agent DDPG algorithm establishes a distributed collaborative structure of electricity, heat, hydrogen, and other subsystems to maximize local benefits while achieving global energy coordination and carbon constraint consistency. Its operating cost was reduced to CN¥234,200/day during the heating season, and its energy balance error was only 2.12%, indicating its excellent stability and convergence efficiency under high load and complex coupling conditions.

On typical days during the non-heating season, the system load structure was dominated by power and cooling needs, the coupling intensity was low, and the overall operating cost and carbon emissions were significantly lowered. The multi-agent DDPG algorithm had the best performance. Its operating cost and carbon emissions were CN¥159,100/day and 54.52 t of CO₂/day, respectively, and the energy balance error was controlled at 1.83%. In addition, the number of convergence steps was the smallest (7.2×10^4), indicating that it had good generalization and transfer capabilities in different load scenarios. To further analyze the response characteristics of different algorithms in a dynamic electricity price environment, the study also selected the peak hours (17:00–22:00) and the trough hours (1:00–6:00) to evaluate the rationality of different algorithm scheduling strategies, as shown in Fig. 11.

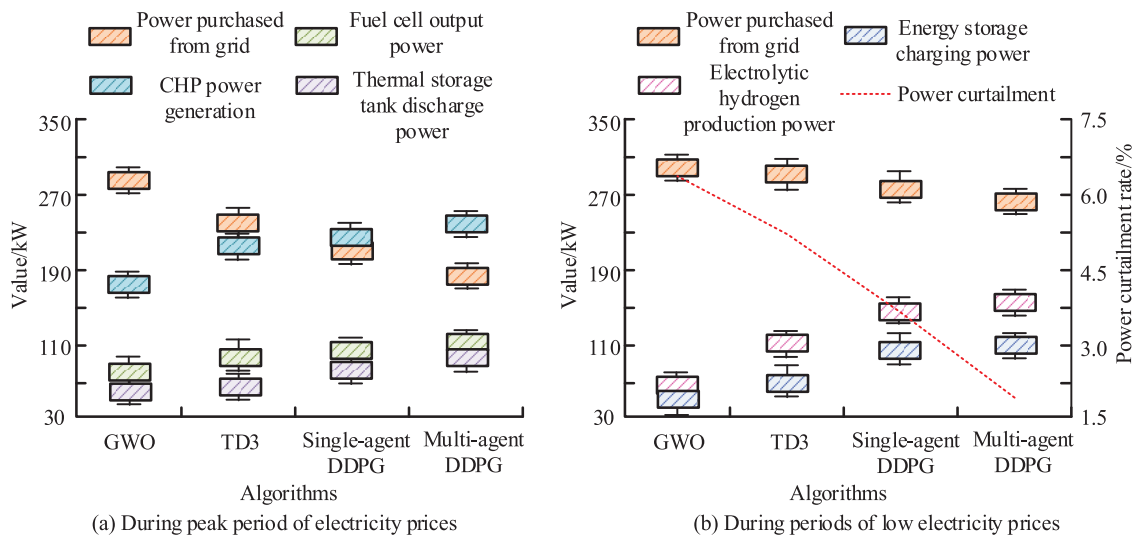


Figure 11: Comparison of system operation characteristics under different algorithms.

Fig. 11a shows the system power allocation and unit start-up and shutdown conditions of different algorithms during the peak period of electricity prices. Fig. 11b shows the energy flow direction and energy storage unit operating status of different algorithms during periods of low electricity prices. From Fig. 11a, the multi-agent DDPG algorithm could adaptively adjust the energy supply strategy based on real-time price signals and system status, prioritized the fuel cells and CHP units for power supply, and utilized the waste heat stored in the thermal storage tank to lower the frequency of starting and stopping the gas boiler, unifying economy and low-carbon. From Fig. 11b, during low electricity prices, the multi-agent DDPG algorithm could actively perform electrolytic hydrogen production operations and convert excess electricity into hydrogen energy storage. The system power abandonment rate declined from 6.2% in the GWO to 2.1%, effectively optimizing the renewable energy utilization. The change trend of the reward function of the multi-agent DDPG in different training stages is shown in Fig. 12.

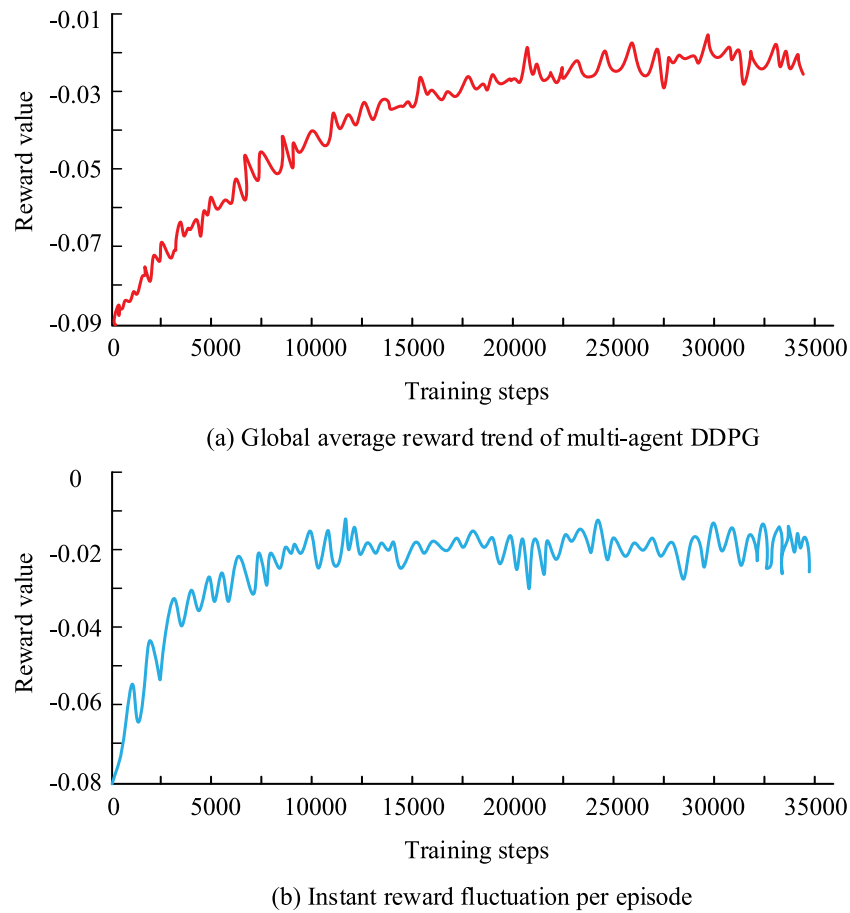


Figure 12: The changing trend of the reward function of the multi-agent DDPG at different training stages.

Fig. 12a shows the global average reward value, and Fig. 12b shows the change of the instant reward within a single round. From Fig. 12, in the initial stage of training ($0 \sim 1 \times 10^4$ steps), because each agent has not fully learned the environmental feedback and dynamic electricity price characteristics, the reward value shows large fluctuations, and negative rewards dominate, and the system scheduling strategy is not yet stable. As the number of training steps increased to about 2×10^4 steps, the agent gradually mastered the rules of energy distribution and load response, the average reward value began to rise significantly, and the fluctuation amplitude decreased significantly. When the number of training steps exceeded 3×10^4 steps, the reward curve was stable and finally stabilized in the range of $-0.015 \sim -0.020$, indicating that the algorithm achieved stable convergence. The multi-agent DDPG algorithm has good learning ability and strategy stability in a complex energy system environment, and can complete the convergence process from random exploration to the optimal scheduling strategy within limited training rounds. Finally, to further verify the robustness and dynamic adaptability of the multi-agent DDPG in complex operating scenarios, the study introduced random disturbance factors (electricity price fluctuation $\pm 15\%$, wind speed fluctuation $\pm 20\%$, and random load change $\pm 10\%$), and tested the system response performance and operational stability under disturbance conditions, as presented in Table 3.

Table 3: Operating performance under different disturbance conditions.

Disturbance Type	Algorithms	The Fluctuation Range of Operating Costs (%)	Mean Cost (CNY/h)	Energy Balance Error (%)	Scheduling Response Delay (s)	Stability Index (Standard Deviation)
Electricity price fluctuation	Single-agent DDPG algorithm	± 7.4	438.6	4.8	3.8	0.082
	Multi-agent DDPG algorithm	± 3.1	421.3	2.9	2.4	0.046
Wind speed fluctuation	Single-agent DDPG algorithm	± 6.8	445.2	5.1	3.5	0.071
	Multi-agent DDPG algorithm	± 2.5	427.5	2.5	2.2	0.039
Random variation of load	Single-agent DDPG algorithm	± 8.2	452.8	4.6	3.9	0.069
	Multi-agent DDPG algorithm	± 3.4	429.1	2.7	2.1	0.038
Comprehensive disturbance scenario	Single-agent DDPG algorithm	± 9.5	460.3	5.4	4.0	0.082
	Multi-agent DDPG algorithm	± 4.2	432.7	2.4	2.3	0.037

From Table 3, under random fluctuations in electricity prices, the system operating cost of the multi-agent DDPG algorithm fluctuated only $\pm 3.1\%$, indicating that the multi-agent algorithm had stronger dynamic adaptability under external economic disturbances. When wind speed disturbance caused uncertainty in renewable energy output, the multi-agent DDPG algorithm controlled the system energy balance error within 2.5% through collaborative compensation of energy storage and fuel cells in each region, showing significant robust control advantages. Under random load fluctuations, the scheduling response delay of the multi-agent DDPG algorithm was 2.1 s on average, which was nearly 45% shorter than that of the single-agent DDPG algorithm, indicating that it had higher real-time performance and scheduling agility under multi-energy dynamic fluctuations. Under a comprehensive disturbance scenario (the three factors of electricity price, wind speed, and load fluctuate simultaneously), the system stability index (measured by the standard deviation of energy balance) of the multi-agent DDPG algorithm was 0.037, which was about 55% lower than that of the single-agent DDPG algorithm, verifying its adaptive and robust optimization capabilities in complex energy environments.

4 Discussion

To verify the IES low-carbon transformation operation optimization method, the performance of typical energy optimization and scheduling methods in the existing literature was comparatively analyzed. First of all, in terms of economic optimization, the cogeneration system optimization method based on the improved PSO algorithm proposed by Li et al. reduced operating costs by approximately 11.4% under typical operating conditions [28]. Xu and Xiang utilized the DDPG to optimize micro-energy networks, with a cost reduction of 9.8% [29]. The proposed model achieved a cost optimization rate of 15.3% under the same test

conditions, which was better than that of the above-mentioned literature methods. The results show that the multi-agent structure has higher search efficiency and global optimal capabilities when dealing with multi-energy flow collaborative optimization problems, and can effectively reduce the operating costs of complex systems. Secondly, in terms of low-carbon performance, the CHP scheduling method based on reinforcement learning proposed by Tan et al. achieved a carbon emission reduction rate of approximately 17.2% [30]. The proposed model achieved a 21.6% reduction in carbon emissions while considering the multi-energy coupling constraints of heat, electricity, and hydrogen. The multi-agent collaboration mechanism and dynamic reward function design can further enhance the model's energy-saving and emission-reduction capabilities. In addition, the research improved peak and valley load reduction and energy storage utilization by 22.8% and 18.5%, respectively, further verifying the advantages of reinforcement learning algorithms in dynamic scheduling of complex energy systems.

In summary, the IES low-carbon transformation operation optimization method relying on multi-agent DDPG is superior to the optimization methods in the current mainstream literature in terms of economy, carbon emission reduction capability, convergence stability, and comprehensive performance. Its innovation is mainly reflected in three aspects: (1) At the algorithmic mechanism level, the search efficiency of reinforcement learning in high-dimensional state space is significantly improved through multi-agent parallel collaboration and shared experience mechanisms; (2) At the energy coupling level, the algorithm utilizes the complementary characteristics of heat, electricity, hydrogen and other multi-energy to achieve global optimal scheduling under complex constraints; (3) At the engineering application level, the research has verified the versatility and robustness in typical daily scenarios, providing a generalizable theoretical basis and practical path for intelligent control of multi-energy systems.

5 Conclusion

As energy demand continues to grow and awareness of environmental protection increases, the IES operation optimization method based on reinforcement learning has become the key to solving energy efficiency. This study proposed an IES low-carbon transformation operation optimization method based on multi-agent DDPG. This method built a low-carbon IES system with hydrogen energy storage as its core, forming a comprehensive energy supply architecture with complementary electricity, heat, and gas. By introducing the dual-channel energy conversion mechanism of “electricity-hydrogen-electricity” and “gas-hydrogen-gas”, deep coupling and dynamic balance between multiple energy sources were achieved, providing a flexible and efficient operation basis for low-carbon cities. Experimental results showed that as the energy storage capacity increased from 100 to 400 MWh, the peak-to-valley load ratio of the low-carbon IES decreased from 1.75 to 1.25, with a decrease of approximately 28.6%. The gas consumption of gas turbines and gas boilers also dropped significantly, at 23.4%. The designed low-carbon IES could achieve efficient coordinated utilization in different seasons and different load conditions, fully verifying the rationality and feasibility of the system. The multi-agent DDPG realized collaborative optimization among multiple energy entities. This algorithm completed strategy evaluation and update through the Actor-Critic structure in the continuous action space, effectively improving the learning efficiency and global convergence of complex systems. The operating cost of the multi-agent DDPG was reduced to CN¥234,200/day during the heating season, and the energy balance error was only 2.12%. On typical days during the non-heating season, its operating costs and carbon emissions were CN¥159,100/day and 54.52 t of CO₂/day, respectively, and the energy balance error was controlled at 1.83%, verifying its adaptive and robust optimization capabilities in complex energy environments. However, the current experimental data are mainly based on typical daily scenarios. In the future, multi-period, seasonal and sudden load data can be introduced to improve the time series generalization ability.

Acknowledgement: Not applicable.

Funding Statement: The research is supported by the Science Program of State Grid Jiangsu Electric Power Co., LTD (No. J2025019).

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Jing Shi, Zesen Li and Delv Zhu; methodology, Jing Shi; software, Zesen Li; validation, Delv Zhu, Bingjie Li and Lang Gao; formal analysis, Jing Shi; investigation, Zesen Li; resources, Jing Shi; data curation, Zesen Li; writing—original draft preparation, Jing Shi, Zesen Li and Delv Zhu; writing—review and editing, Bingjie Li and Lang Gao; visualization, Bingjie Li and Lang Gao; supervision, Bingjie Li and Lang Gao; funding acquisition, Lang Gao. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Lang Gao, upon reasonable request.

Ethics Approval: This study strictly followed academic ethics and moral standards. No human participants or animals were involved in this research; therefore, ethical approval was not required. All publicly available data and materials used in this study have been properly cited in accordance with the journal's requirements.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wu X, Zhang Y, Fei X, Hu H, Sun X, Gong D, et al. Low-carbon economic operation optimization of mine integrated energy system considering energy supply uncertainty. *Complex Syst Model Simul*. 2024;4(3):258–73. doi:10.23919/csms.2024.0012.
2. Sun D, Liu Z, Shao J, Lin Z. Review on low carbon planning and operation of integrated energy systems. *Energy Sci Eng*. 2022;10(8):3201–15. doi:10.1002/ese3.1167.
3. Jin J, Wen Q, Cheng S, Qiu Y, Zhang X, Guo X. Optimization of carbon emission reduction paths in the low-carbon power dispatching process. *Renew Energy*. 2022;188:425–36. doi:10.1016/j.renene.2022.02.054.
4. Yan N, Ma G, Li X, Guerrero JM. Low-carbon economic dispatch method for integrated energy system considering seasonal carbon flow dynamic balance. *IEEE Trans Sustain Energy*. 2023;14(1):576–86. doi:10.1109/TSTE.2022.3220797.
5. Fu X, Zhou Y. Collaborative optimization of PV greenhouses and clean energy systems in rural areas. *IEEE Trans Sustain Energy*. 2023;14(1):642–56. doi:10.1109/TSTE.2022.3223684.
6. Ali A, Morsli A, Al-Zoubi OH, Nuñez-Alvarez JR, Khan MA, Hlail SH, et al. A multi-stage interval optimization approach for operation of the smart multi-carries energy system considering energy prices uncertainty. *Electr Eng*. 2024;106(6):6709–29. doi:10.1007/s00202-024-02397-6.
7. Jarwar AR. Advancing multi-energy systems with thermal and electrical storage: a literature review on leveled cost reduction. *Int J Sustain Appl Sci*. 2025;3(8):541–52. doi:10.59890/ijssas.v3i8.141.
8. Zhong J, Li Y, Wu Y, Cao Y, Li Z, Peng Y, et al. Optimal operation of energy hub: an integrated model combined distributionally robust optimization method with Stackelberg game. *IEEE Trans Sustain Energy*. 2023;14(3):1835–48. doi:10.1109/tste.2023.3252519.
9. Ye Y, Papadaskalopoulos D, Yuan Q, Tang Y, Strbac G. Multi-agent deep reinforcement learning for coordinated energy trading and flexibility services provision in local electricity markets. *IEEE Trans Smart Grid*. 2023;14(2):1541–54. doi:10.1109/TSG.2022.3149266.
10. Ukoba K, Olatunji KO, Adeoye E, Jen TC, Madyira DM. Optimizing renewable energy systems through artificial intelligence: review and future prospects. *Energy Environ*. 2024;35(7):3833–79. doi:10.1177/0958305x241256293.
11. Qiu Y, Zhou S, Xia D, Gu W, Sun K, Han G, et al. Local integrated energy system operational optimization considering multi-type uncertainties: a reinforcement learning approach based on improved TD3 algorithm. *IET Renew Power Gener*. 2023;17(9):2236–56. doi:10.1049/rpg2.12725.

12. Li S, Cao D, Hu W, Huang Q, Chen Z, Blaabjerg F. Multi-energy management of interconnected multi-microgrid system using multi-agent deep reinforcement learning. *J Mod Power Syst Clean Energy*. 2023;11(4):1606–17. doi:10.35833/mpce.2022.000473.
13. Hu D, Ye Z, Gao Y, Ye Z, Peng Y, Yu N. Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization. *IEEE Trans Smart Grid*. 2022;13(6):4873–86. doi:10.1109/TSG.2022.3185975.
14. Zhou Y, Ma Z, Wang T, Zhang J, Shi X, Zou S. Joint energy and carbon trading for multi-microgrid system based on multi-agent deep reinforcement learning. *IEEE Trans Power Syst*. 2024;39(6):7376–88. doi:10.1109/tpwrs.2024.3380070.
15. Ding T, Jia W, Shahidehpour M, Han O, Sun Y, Zhang Z. Review of optimization methods for energy hub planning, operation, trading, and control. *IEEE Trans Sustain Energy*. 2022;13(3):1802–18. doi:10.1109/tste.2022.3172004.
16. Zhu C, Zhang Y, Wang M, Deng J, Cai Y, Wei W, et al. Optimization, validation and analyses of a hybrid PV-battery-diesel power system using enhanced electromagnetic field optimization algorithm and ϵ -constraint. *Energy Rep*. 2024;11:5335–49. doi:10.1016/j.egyr.2024.04.043.
17. Liao H, Zhou Z, Liu N, Zhang Y, Xu G, Wang Z, et al. Cloud-edge-device collaborative reliable and communication-efficient digital twin for low-carbon electrical equipment management. *IEEE Trans Ind Inf*. 2023;19(2):1715–24. doi:10.1109/tii.2022.3194840.
18. Xu J, Gao H, Wang R, Liu J. Real-time operation optimization in active distribution networks based on multi-agent deep reinforcement learning. *J Mod Power Syst Clean Energy*. 2024;12(3):886–99. doi:10.35833/mpce.2023.000213.
19. Hady MA, Hu S, Pratama M, Cao Z, Kowalczyk R. Multi-agent reinforcement learning for resources allocation optimization: a survey. *Artif Intell Rev*. 2025;58(11):354. doi:10.1007/s10462-025-11340-5.
20. Chen P, Liu S, Chen B, Yu L. Multi-agent reinforcement learning for decentralized resilient secondary control of energy storage systems against DoS attacks. *IEEE Trans Smart Grid*. 2022;13(3):1739–50. doi:10.1109/TSG.2022.3142087.
21. Safiri S, Nikoofard A, Khosravy M, Senjyu T. Multi-agent distributed reinforcement learning algorithm for free-model economic-environmental power and CHP dispatch problems. *IEEE Trans Power Syst*. 2023;38(5):4489–500. doi:10.1109/TPWRS.2022.3217905.
22. Wang Y, Qiu D, Teng F, Strbac G. Towards microgrid resilience enhancement via mobile power sources and repair crews: a multi-agent reinforcement learning approach. *IEEE Trans Power Syst*. 2024;39(1):1329–45. doi:10.1109/TPWRS.2023.3240479.
23. Vu L, Vu T, Vu TL, Srivastava A. Multi-agent deep reinforcement learning for distributed load restoration. *IEEE Trans Smart Grid*. 2024;15(2):1749–60. doi:10.1109/tsg.2023.3310893.
24. Fan H, Lu E, Yu W, Du L, Wang H, Wang D. Multi-agent deep reinforced co-dispatch of energy and hydrogen storage in low-carbon building clusters. *IEEE Trans Netw Sci Eng*. 2024;11(6):5449–62. doi:10.1109/tnse.2023.3243202.
25. Qiu D, Wang J, Dong Z, Wang Y, Strbac G. Mean-field multi-agent reinforcement learning for peer-to-peer multi-energy trading. *IEEE Trans Power Syst*. 2023;38(5):4853–66. doi:10.1109/tpwrs.2022.3217922.
26. Khalatbarisoltani A, Kandidayeni M, Boulon L, Hu X. A decentralized multi-agent energy management strategy based on a look-ahead reinforcement learning approach. *SAE Int J Elec Veh*. 2022;11(2):151–64. doi:10.4271/14-11-02-0012.
27. Haruna AA, Muhammad LJ, Abubakar M. Novel thermal-aware green scheduling in grid environment. *Artif Intell Appl*. 2022;1(4):244–51. doi:10.47852/bonviewAIA2202332.
28. Li M, Yang S, Zhang M. Power supply system scheduling and clean energy application based on adaptive chaotic particle swarm optimization. *Alex Eng J*. 2022;61(3):2074–87. doi:10.1016/j.aej.2021.08.008.
29. Xu B, Xiang Y. Optimal operation of regional integrated energy system based on multi-agent deep deterministic policy gradient algorithm. *Energy Rep*. 2022;8:932–9. doi:10.1016/j.egyr.2022.08.066.
30. Tan Y, Shen Y, Yu X, Lu X. Low-carbon economic dispatch of the combined heat and power-virtual power plants: a improved deep reinforcement learning-based approach. *IET Renew Power Gener*. 2023;17(4):982–1007. doi:10.1049/rpg2.12650.