ARTICLE

Low Carbon Economic Dispatch of Integrated Energy System Considering Power Supply Reliability and Integrated Demand Response

Jian Dong, Haixin Wang, Junyou Yang[*], Liu Gao, Kang Wang and Xiran Zhou

School of Electrical Engineering, Shenyang University of Technology, Shenyang, 110870, China
[*]Corresponding Author: Junyou Yang. Email: junyouyang@sut.edu.cn

ABSTRACT

Integrated energy system optimization scheduling can improve energy efficiency and low carbon economy. This paper studies an electric-gas-heat integrated energy system, including the carbon capture system, energy coupling equipment, and renewable energy. An energy scheduling strategy based on deep reinforcement learning is proposed to minimize operation cost, carbon emission and enhance the power supply reliability. Firstly, the low-carbon mathematical model of combined thermal and power unit, carbon capture system and power to gas unit (CCP) is established. Subsequently, we establish a low carbon multi-objective optimization model considering system operation cost, carbon emissions cost, integrated demand response, wind and photovoltaic curtailment, and load shedding costs. Furthermore, considering the intermittency of wind power generation and the flexibility of load demand, the low carbon economic dispatch problem is modeled as a Markov decision process. The twin delayed deep deterministic policy gradient (TD3) algorithm is used to solve the complex scheduling problem. The effectiveness of the proposed method is verified in the simulation case studies. Compared with TD3, SAC, A3C, DDPG and DQN algorithms, the operating cost is reduced by 8.6%, 4.3%, 6.1% and 8.0%.

KEYWORDS

Integrated energy system; twin delayed deep deterministic policy gradient; economic dispatch; power supply reliability; integrated demand response

1 Introduction

In recent years, with the development of the world economy and the increasing depletion of fossil fuels, the problem of insufficient energy supply has become increasingly prominent [1,2]. On 05 March, 2021, China pointed out in the government work report that CO_2 emissions would be at the peak by 2030 and achieve carbon neutrality by 2060. As an important carrier for the development of the energy internet, the integrated energy system (IES) can promote the coordination and complementation of various energy sources [3,4]. Meanwhile, the IES has made significant contributions to building clean, low-carbon and efficient energy systems. However,

with the deepening of energy coupling, the IES is confronted with enormous challenges due to fluctuating wind and PV power outputs and the uncertainty of multi-energy demands [5–7].

The operation economy and power supply reliability are two essential factors in energy management and optimization of the IES [8–10]. Power supply reliability is an important indicator to measure the stable operation of the power grid, and economic benefit is an important goal in the development of IES [11].

Regarding the low-carbon operation of IESs, Wang et al. [12] proposed an optimal scheduling model based on the carbon trading mechanism. IES operators can purchase or sell carbon quotas in the carbon trading market. The results show that considering carbon trading can reduce the operation cost of the IES. Zhai et al. [13] proposed an economic dispatch method for low-carbon power system considering the uncertainty of electric, thermal and cold loads. Yang et al. [14] proposed an optimal scheduling model for the combination of the microturbine (MT) and power to gas (P2G) units. The typical load scenarios are obtained by scenario generation and reduction techniques to improve wind power consumption and reduce carbon dioxide emissions. Although the above literature can realize the low-carbon operation of the system, the flexible resources of the demand side are not considered.

Considering integrated demand response (IDR) in IES can promote renewable energy consumption and reduce carbon emissions. Bahrami et al. [15] established a power system scheduling model considering demand response (DR) resources and carbon trading and verify its effectiveness in promoting wind power consumption and reducing carbon emissions. Zeng et al. [16] introduced the split-flow carbon capture power plant and DR into the IES to achieve low carbon.

Although remarkable results are achieved in the economic dispatch of the IES, the above models are solved by traditional methods, and the optimization effect is dependent on the prediction accuracy of sources and loads. With the development of the artificial intelligence (AI) technique, reinforcement learning (RL) has been paid more attention to the optimal control of power system [17–20]. The RL model can accumulate experience and improve policies by continuous interaction with the environment. In particular, the deep reinforcement learning algorithm combining deep neural network and reinforcement learning is of better adaptive learning ability and optimized decision-making ability for nonconvex and nonlinear problems [20–22]. In [23], microgrid (MG) real-time energy management based on deep reinforcement learning was proposed, and MG energy management is described as a Markov decision process (MDP) to minimize the daily operating cost. In [24], the energy management of IES was described as a constrained optimal control problem and solved by the asynchronous advantage actor critic algorithm.

The above studies provided the foundation for the application of the DRL approach in the IES. However, most of the above models only consider the economy and security of IES, without consideration of carbon dioxide emissions and other indicators in the system. In addition, in the face of collaborative optimization of multi-energy and energy storage, the model training may be time-consuming and prone to non-convergence.

This paper proposes a TD3-based integrated energy system source-load coordination optimization scheduling framework. Firstly, a combined optimization model of combined thermal and power (CHP), power to gas (P2G) and carbon capture system (CCS) is established, which can realize thermal-power decoupling and reduce carbon emissions. Secondly, the multi-objective optimization problem of IES is described as a Markov decision process, the environmental model of IES is established, and the action space, state space and reward function of the agent are designed. Finally, the low-carbon economic dispatch problem is solved by the twin delayed deep

deterministic policy gradient (TD3), and the convergence ability and stability of this method are analyzed. Finally, the effectiveness of this method in the low-carbon economic dispatch of the IES is validated.

2 IES Model and Problem Description

Low-carbon operation optimization of the IES aims to improve the economic and environmental benefits of the system with the constraint of safe operation of the system. This paper studies a multi-objective optimization problem with optimal economic cost, carbon emissions, and reliability of system operation. The structure of the IES studied in this paper is shown in Fig. 1. The power grid includes combined thermal and power (CHP), carbon capture system (CCS), wind power, photovoltaic, battery storage (BS) and electricity load. The gas network comprises the natural gas station, gas storage (GS) and gas load. The thermal supply network mainly consists of thermal storage (HS) and thermal load. The energy conversion equipment mainly includes gas turbine, P2G and MT.
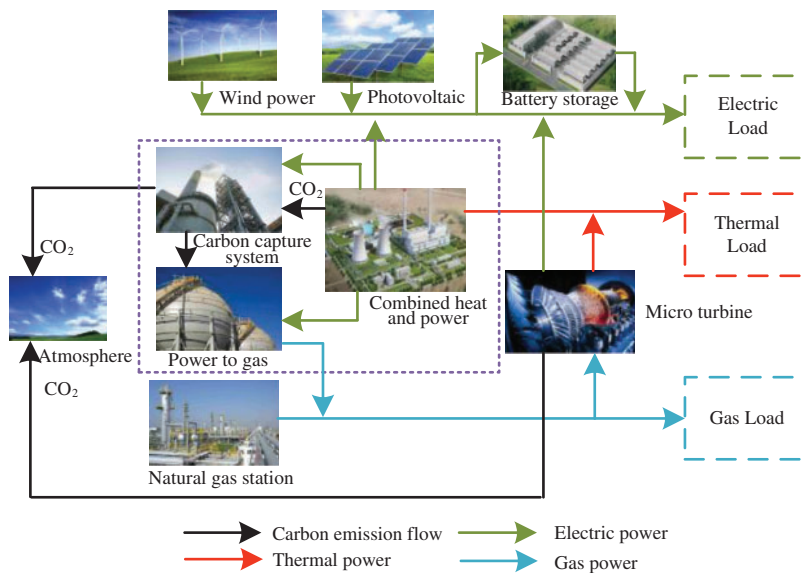

Figure 1: Structure diagram of the IES

2.1 IES Equipment

(1) CHP-CCS-P2G (CCP) Mathematical Model

The CHP units provide electric power and thermal power in the IES. The CHP has high carbon emissions and thermoelectric coupling characteristics, which causes severe environmental pollution. Therefore, according to the literature [25], the optimization model of CCP is established to reduce carbon emissions and improve the economic benefits of the IES. The principle of the CCP is shown in Fig. 2. P2G converts the electric power generated by CHP into natural gas, which strengthens the connection of the power-thermal-gas system and reduces the power output of CHP. The electric power consumed by P2G and CCS is directly taken from electric power generated by CHP. To reduce carbon emissions of CHP and carbon source cost of P2G, the CO_2 is captured by CCS, and transmitted to P2G for recycling.
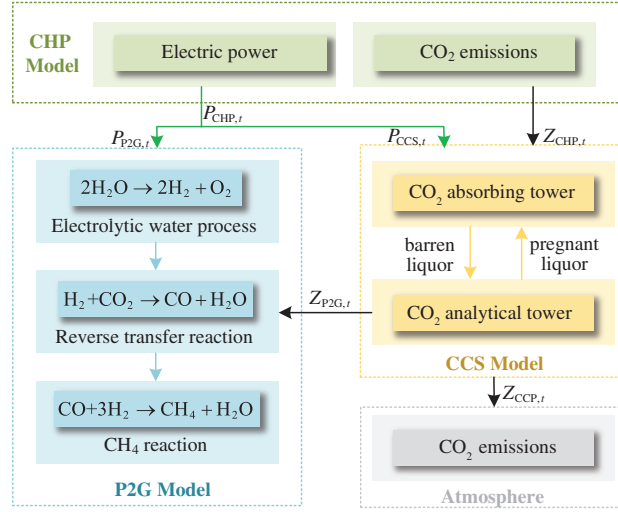
Figure 2: The principle of CCP combined optimization model

The electric power output of the CCP can be calculated by Eq. (1).

$$P_{CCP,t} = P_{CHP,t} - P_{P2G,t} - P_{CCS,t}, \tag{1}$$

where $P_{CCP,t}$ is the electric power output of the CCP at time t, $P_{CHP,t}$ is the electric power generated by CHP, $P_{P2G,t}$ is the electric power consumed by P2G, $P_{CCS,t}$ is the electric power of CCS.

The gas power output of the CCP can be calculated by Eq. (2).

$$G_{CCP,t} = G_{P2G,t} = \eta_{P2G} P_{P2G,t}, \tag{2}$$

where $G_{CCP,t}$ is the gas power output of the CCP at time t, $G_{P2G,t}$ is the gas power output of the P2G, η_{P2G} is the conversion efficiency of P2G.

The electric power and ramp rate constraints of CCP can be described by Eqs. (3) and (4).

$$P_{CCP}^{min} = P_{CHP}^{min} - P_{P2G}^{max} - P_{CCS}^{max} \leq P_{CCP,t} \leq P_{CHP}^{max} - P_{P2G}^{min} - P_{CCS}^{min} = P_{CCP}^{max}, \tag{3}$$

$$\Delta P_{CCP}^{min} = \Delta P_{CHP}^{min} - \Delta P_{P2G}^{max} - \Delta P_{CCS}^{max} \leq P_{CCP,t+1} - P_{CCP,t} \leq \Delta P_{CHP}^{max} - \Delta P_{P2G}^{min} - \Delta P_{CCS}^{min} = \Delta P_{CCP}^{max}, \tag{4}$$

where $P_{CCP}^{min}, P_{CHP}^{min}, P_{P2G}^{min}$ and P_{CCS}^{min} are the lower limits of electric power of CCP, CHP, P2G and CCS, $P_{CCP}^{max}, P_{CHP}^{max}, P_{P2G}^{max}$ and P_{CCS}^{max} are the upper limits of electric power of CCP, CHP, P2G and CCS, $\Delta P_{CCP}^{min}, \Delta P_{CHP}^{min}, \Delta P_{P2G}^{min}$ and ΔP_{CCS}^{min} are the lower limits of electric power ramp rate of CCP, CHP, P2G and CCS, $\Delta P_{CCP}^{max}, \Delta P_{CHP}^{max}, \Delta P_{P2G}^{max}$ and ΔP_{CCS}^{max} are the upper limits of electric power ramp rate of CCP, CHP, P2G and CCS.

The thermal power and ramp rate constraints of CCP can be described by Eqs. (5) and (6).

$$H_{CCP}^{min} = H_{CHP}^{min} \leq H_{CCP,t} \leq H_{CHP}^{max} = H_{CCP}^{max}, \tag{5}$$

$$\Delta H_{CCP}^{min} = \Delta H_{CHP}^{min} \leq H_{CCP,t+1} - H_{CCP,t}, \leq \Delta H_{CHP}^{max} = \Delta H_{CCP}^{max}, \tag{6}$$

where $H_{CCP,t+1}$ and $H_{CCP,t}$ are the thermal power output of the CCP at time t and $t + 1$, H_{CCP}^{min} and H_{CHP}^{min} are the lower limits of thermal power of CCP and CHP, H_{CCP}^{max} and H_{CHP}^{max} are the upper limits of thermal power of CCP and CHP, H_{CCP}^{min} and ΔH_{CHP}^{min} are the lower limits of

thermal power ramp rate of CCP and CHP, $\Delta H_{CCP}^{\max}$ and $\Delta H_{CHP}^{\max}$ are limits of thermal power ramp rate of CCP and CHP.

The gas power and ramp rate constraints of CCP can be described by Eqs. (7) and (8).

$$G_{CCP}^{\min} = G_{P2G}^{\min} \leq G_{CCP,t} \leq G_{P2G}^{\max} = G_{CCP}^{\max}, \tag{7}$$

$$\Delta G_{CCP}^{\min} = \Delta G_{P2G}^{\min} \leq G_{CCP,t+1} - G_{CCP,t} \leq \Delta G_{P2G}^{\max} = \Delta G_{CCP}^{\max}, \tag{8}$$

where $G_{CCP,t+1}$ and $G_{P2G,t}$ are the gas power output of the CCP at time t and $t+1$, $G_{CCP}^{\min}$ and $G_{P2G}^{\min}$ are the lower limits of gas power of CCP and P2G, $G_{CCP}^{\max}$ and $G_{P2G}^{\max}$ are the upper limits of gas power of CCP and P2G, $\Delta G_{CCP}^{\min}$ and $\Delta G_{P2G}^{\min}$ are the lower limits of gas power ramp rate of CCP and P2G, $\Delta G_{CCP}^{\max}$ and $\Delta G_{P2G}^{\max}$ are the upper limits of gas power ramp rate of CCP and P2G.

(2) MT Mathematical Model

MT driven by natural gas is the essential equipment of the IES [25]. The mathematical model of MT is shown by Eqs. (9) and (10).

$$P_{MT,t} = G_{MT,t} \cdot \eta_{MT}, \tag{9}$$

$$H_{MT,t} = G_{MT,t} \cdot (1 - \eta_{MT} - \eta_{loss}), \tag{10}$$

where $P_{MT,t}$ and $H_{MT,t}$ are the electric and thermal power of MT at time t, $G_{MT,t}$ is the natural gas consumption power of MT, η_{MT} is the power generation efficiency of MT, η_{loss} is the energy loss coefficient of MT.

The power and ramp rate constraints of MT can be described by Eqs. (11)–(14).

$$P_{MT}^{\min} \leq P_{MT,t} \leq P_{MT}^{\max}, \tag{11}$$

$$H_{MT}^{\min} \leq H_{MT,t} \leq H_{MT}^{\max}, \tag{12}$$

$$\Delta P_{MT}^{\min} \leq P_{MT,t} - P_{MT,t-1} \leq \Delta P_{MT}^{\max}, \tag{13}$$

$$\Delta H_{MT}^{\min} \leq H_{MT,t} - H_{MT,t-1} \leq \Delta H_{MT}^{\max}, \tag{14}$$

where $P_{MT}^{\min}$ and $P_{MT}^{\max}$ are the lower limits and upper limit of electric power of the MT, $H_{MT}^{\min}$ and $H_{MT}^{\max}$ are the lower limits and upper limits of thermal power of the MT, $\Delta P_{MT}^{\min}$ and $\Delta P_{MT}^{\max}$ are the lower limits and upper limit of electric power ramp rate of the MT, $\Delta H_{MT}^{\min}$ and $H_{MT}^{\max}$ are the lower and upper limits of thermal power ramp rate of MT.

(3) Energy Storage Battery Model

The mathematical model of battery storage is expressed by Eqs. (15)–(17) [8].

$$SOC_{t+1} = SOC_t + a_{ch,t}\eta_{ch}\frac{P_{BS,ch,t}}{Q_{BS}}\Delta t + a_{x,disch,t}\frac{P_{BS,disch,t}}{\eta_{disch}}\Delta t, \tag{15}$$

$$a_{ch,t}, a_{disch,t} = \{0, 1\}, \forall t \in T, \tag{16}$$

$$a_{ch,t} \wedge a_{disch,t} = 0, \forall t \in N, \tag{17}$$

where SOC_{t+1} and SOC_t are the states of charge of the BS at time $t + 1$ and t, $P_{BS,ch,t}$ and $P_{BS,disch,}$ are the charging and discharging power of BS at time t ($P_{BS,ch,t} \geq 0$, $P_{BS,disch,t} \leq 0$). Q_{BS} is the capacity of BS, η_{ch} and η_{disch} are the charging and discharging efficiency of BS, Δt

is the time interval, $a_{ch,t}$ and $a_{disch,t}$ are the charging and discharging state parameters, $a_{ch,t} = 1$ represents the charging operation of the BS at time t, and similarly, $a_{disch,t} = 1$ represents the discharging operation of the energy BS at time t.

The energy storage equipment is limited by the following charging and discharging power constraints and capacity constraints in Eqs. (18)–(21).

$$SOC^{\min} \leq SOC_t \leq SOC^{\max}, \tag{18}$$

$$0 \leq P_{BS,ch,t} \leq P_{BS,ch}^{\max}, \tag{19}$$

$$0 \geq P_{BS,disch,t} \geq P_{BS,disch}^{\max}, \tag{20}$$

$$SOC_1 = SOC_T, \tag{21}$$

where $SOC^{\min}$ and $SOC^{\max}$ are maximum and minimum SOC values of the BS, $P_{BS,ch}^{\max}$ and $P_{BS,disch}^{\max}$ are maximum charging and discharging power of the BS, SOC_1 and SOC_T are the initial and final values of the BS.

2.2 IES Model

(1) Electric Network Model

The power flow constraints can be formulated by Eq. (22) [18].

$$\begin{cases} P_{i,t} = U_{i,t} \sum_{j \in i} U_{j,t} \left(G_{ij} \cos \theta_{ij,t} + B_{ij} \sin \theta_{ij,t} \right) \\ Q_{i,t} = U_{i,t} \sum_{j \in i} U_{j,t} \left(G_{ij} \sin \theta_{ij,t} - B_{ij} \cos \theta_{ij,t} \right) \end{cases}, \tag{22}$$

where $P_{i,t}$ is the active power of node i at time t, $Q_{i,t}$ is the reactive power of node i at time t, $U_{i,t}$ and $U_{j,t}$ are the voltage values of nodes i and j, G_{ij} is the electric conductance between nodes i and j, B_{ij} is the electrical susceptance between nodes i and j, $\theta_{ij,t}$ is the phase angle difference between nodes i and j.

(2) Natural Gas Network Model

The mathematical description of the natural gas network mainly includes pipe flow, compressor and gas flow models. The pipe flow equation of natural gas is obtained by Eqs. (23) and (24).

$$G_{mn} = Z_{mn} \sigma_{mn} \sqrt{\sigma_{mn} \left(\Phi_m^2 - \Phi_n^2 \right)}, \tag{23}$$

$$\sigma_{mn} = \begin{cases} 1, \ \Phi_m > \Phi_n \\ -1, \ \Phi_m \leq \Phi_n \end{cases}, \tag{24}$$

where G_{mn} is the steady-state flow rate of natural gas pipeline mn, Z_{mn} is the pipeline constant, which depends on various factors (e.g., inner diameter, length, friction coefficient and temperature of the pipeline), Φ_m and Φ_n are the pressure of nodes m and n, σ is the flow direction coefficient of natural gas.

The mathematical model of the compressor can be described by Eq. (25).

$$P_{com} = Z_{com} G_{com,mn} \left[(\Phi_n / \Phi_m)^{D_{com}} - 1 \right], \tag{25}$$

where P_{com} is the electric power of the compressor, Z_{com} is the compression ratio, $G_{com,mn}$ is the gas flow through the compressor, D_{com} is the compressor parameter.

The gas flow constraints of the natural gas network can be described by Eqs. (26) and (27).

$$Z_{com}^{\min} \leq Z_{com} = \Phi_n / \Phi_m \leq Z_{com}^{\max}, \tag{26}$$

$$\Phi_n^{\min} \leq \Phi_n \leq \Phi_n^{\max}, \tag{27}$$

where $Z_{com}^{\min}$ and $Z_{com}^{\max}$ are the lower and upper limits of the compression ratio of the compressor, $\Phi_n^{\min}$ and $\Phi_n^{\max}$ are the lower and upper limits of natural gas pressure at node n.

(3) Thermal Network Model

The thermal network model mainly includes the hydraulic model and thermodynamic model [26]. The hydraulic model is described by Eq. (28).

$$\begin{cases} A_k M = M_k \\ B H = 0 \end{cases}, \tag{28}$$

where A_k is the node-to-branch incidence matrix of the thermal network, M is the water flow matrix of the pipeline, and M_k is the water flow matrix of the node k, B is the loop-to-branch incidence matrix of the thermal network, H is the pipeline pressure loss matrix.

The thermodynamic model of the thermal supply network is expressed by Eq. (29).

$$\begin{cases} H_k = C_H m_k \left(T_{k,s} - T_{k,o} \right) \\ T_{k,out} \sum_{u \in E} m_u = \sum_{l \in Y} m_l T_{kl,in} \\ T_{end} = \left(T_{start} - T_g \right) e^{-\varepsilon L / c_H m} + T_g \end{cases}, \tag{29}$$

where H_k is the thermal power of node k of the thermal network, C_H is the thermal capacity of water, m_k is the water flow out of node k, $T_{k,s}$ and $T_{k,o}$ are the temperature of water flowing into and out of node k, E and Y are the upstream and downstream pipeline sets of node k, m_u and m_l are the water flow of pipelines u and l, $T_{k,out}$ is the initial temperature of the downstream pipeline, $T_{kl,out}$ is the terminal temperature of downstream pipeline l, T_{start} and T_{end} are the initial and terminal temperature of the pipeline, T_g is the ambient temperature, ε is the thermal transfer coefficient of the pipeline, and L is the pipeline length.

2.3 IDR Model

As a flexible resource, IDR is conducive to achieving low-carbon economic operation and improving the operation reliability of the IES. IDR can reduce energy demand during peak load periods by reducing, converting and shifting electric power demand. Renewable energy is used to replace high carbon emission units.

The consumption of transferable electric load in the peak demand period is shifted to other periods. The sum of all transferred load is within a certain little range to satisfy the load demand. The transferable load complies with the following operating constraints:

$$\begin{cases} P'_{Load,i,t} = P_{Load,i,t} - P_{Tra,i,t} \\ P'_{Load,i,t+\upsilon} = P_{Load,i,t+\upsilon} + P_{Tra,i,t} \end{cases}, \tag{30}$$

$$\delta_{Tra,i}^{\min} P_{Load,i,t} \leq P_{Tra,i,t} \leq \delta_{Tra,i}^{\max} P_{Load,i,t}, \tag{31}$$

where $P'_{Load,i,t}$ and $P'_{Load,i,t+\upsilon}$ are the electric load of node i after transfer at time t and $t+\upsilon$, $P_{Load,i,t}$ and $P_{Load,i,t+\upsilon}$ are the electric load of node i before transferring at time t and $t+\upsilon$, $P_{Tra,i,t}$ is the electric load of node i transferred from time t to time $t+\upsilon$, $\delta_{Tra,i}^{\min}$ and $\delta_{Tra,i}^{\max}$ are the maximum and minimum proportional coefficients of the transferable load of node i.

Interruptible loads comply with the following operating constraints in Eqs. (32) and (33).

$$P'_{Load,i,t} = P_{Load,i,t} - P_{Int,i,t}, \tag{32}$$

$$\delta_{Int,i}^{\min} P_{Load,i,t} \leq P_{Int,i,t} \leq \delta_{Int,i}^{\max} P_{Load,i,t}, \tag{33}$$

where $P_{Int,i,t}$ is the electric load of node i interrupted at time t, $\delta_{Int,i}^{\min}$ and $\delta_{Int,i}^{\max}$ are the maximum and minimum proportionality coefficients of the interruptible load of node i.

Users can convert peak electric load demand to other energy sources. The convertible load model is shown by Eqs. (34) and (35).

$$P_{Con,i,t} = \phi_{PG} G_{Con,m,t} + \phi_{PH} H_{Con,k,t}, \tag{34}$$

$$\delta_{Con,i}^{\min} P_{Load,i,t} \leq P_{Con,i,t} \leq \delta_{Con,i}^{\max} P_{Load,i,t}, \tag{35}$$

where P_{Con,i,t_a} is the reduced electric load of power grid node i at time t after conversion, G_{Con,m,t_a} is the increased gas load of gas network node m at time t after conversion, H_{Con,k,t_a} is the increased thermal load of thermal supply network node k, $\delta_{Con,i}^{\min}$ and $\delta_{Con,i}^{\max}$ are the maximum and minimum proportionality coefficients of the convertible load of node i.

2.4 Objective Function

In this paper, a low-carbon dispatching model of the IES is established to minimize the total cost of the system. The objective function is expressed by Eq. (36).

$$F = \min (C_O + C_C + C_{CWind} + C_{CPV} + C_{IDR} + C_{LS}), \tag{36}$$

where C_O is the system operating cost, C_C is the CO_2 emission cost, C_{QF} is the wind curtailment cost, C_{QG} is the PV curtailment cost, C_{IDR} is the IDR cost, and C_{LS} is the load shedding cost.

(1) System Operating Cost

The system operating cost includes CCP operating cost and MT operating cost, which can be described by Eq. (37).

$$\begin{cases} C_O = C_{CCP} + C_{MT} \\ C_{CCP} = \sum_{x=1}^{N} \left[\left(a_{CHP,x} P_{CHP,x,t}^2 + b_{CHP,x} P_{CHP,x,t} + c_{CHP,x} \right) + \left(a_{CCS,x} P_{CCS,x,t} + b_{CCS,x} Z_{CCS,x,t} \right) \right. \\ \qquad \left. + \left(a_{P2G,x} P_{P2G,x,t} + b_{P2G,x} Z_{P2G,x,t} \right) \right] \\ C_{MT} = \sum_{x=1}^{N} a_{MT,x} P_{MT,x,t} \end{cases},$$

(37)

where C_{CCP} and C_{MT} are the operating costs of CCP and MT, $a_{CHP,x}$, $b_{CHP,x}$ and $c_{CHP,x}$ are the operating cost coefficients of CHP, $a_{CCS,x}$ and $a_{P2G,x}$ are the operating cost coefficients of CCS and P2G, $Z_{CCS,x,t}$ is the carbon emissions captured by CCS at time t, $Z_{P2G,x,t}$ is the carbon consumed by P2G x at time t, $b_{CCS,x}$ is the cost coefficient of CO_2 capture of CCS, and $b_{P2G,x}$ is the cost coefficient of CO_2 purchase of P2G.

(2) Carbon Trading Cost

The carbon emissions of the system come from CHP and MT, and the part of them is absorbed by CCS. The carbon trading cost can be described by Eq. (38).

$$C_C = \lambda \sum_{x=1}^{N} \left(Z_{CHP,x,t} + Z_{MT,x,t} - Z_{CCS,x,t} - Z_{Q,t} \right),$$

(38)

where λ is the carbon trading cost coefficient, $Z_{CHP,x,t}$ is the carbon generated by CHP x at time t, $Z_{MT,x,t}$ is the carbon generated by MT, and $Z_{Q,t}$ is the carbon emission quota.

(3) Wind Curtailment Cost

The wind curtailment cost can be described as Eq. (39).

$$C_{CWind} = a_{CWind} P_{CWind,t},$$

(39)

where a_{CWind} is the penalty cost coefficient of wind power curtailment, $P_{CWind,t}$ is the wind power curtailment at time t.

(4) PV Curtailment Cost

The PV curtailment cost can be calculated by Eq. (40).

$$C_{CPV} = a_{CPV} P_{CPV,t},$$

(40)

where a_{CPV} is the penalty cost coefficient of PV curtailment, $P_{CPV,t}$ is the PV curtailment at time t.

(5) IDR Cost

The IDR cost is obtained by Eq. (41).

$$C_{IDR} = \sum_{i \in \Omega_{IDR}} \left(a_{Tra} P_{Tra,i,t_{ab}} + a_{Int} P_{Int,i,t_a} + a_{Con} P_{Con,i,t_a} \right)$$

(41)

where Ω_{IDR} is the set of nodes participating in demand response, a_{Tra}, a_{Int} and a_{Con} are the cost coefficients of transferable load, interruptible load and convertible load.

(6) Power Supply Reliability

To improve the power supply reliability of the IES, this paper uses the loss of power supply probability (LPSP) as the standard to measure the power supply reliability. The power supply reliability is calculated by Eq. (42).

$$LPSP_t = \frac{P_{LS,t}}{P_{Load,t}}, \tag{42}$$

where $P_{LS,t}$ is the load shedding power at time t.

The load shedding cost can be calculated by Eq. (43).

$$C_{LS} = a_{LS} P_{LS,t}, \tag{43}$$

where a_{LS} is the load shedding cost coefficient.

2.5 Objective Function

To meet the demand of electric-gas-heat load in each operation period, the system balance constraints are given by Eq. (44).

$$\begin{cases} P_{CCP,i,t} + P_{PV,i,t} + P_{Wind,i,t} + P_{BS,disch,i,t} + P_{BS,ch,i,t} = \\ P_{Load,i,t} - P_{Tra,i,t} - P_{Int,i,t} - P_{Con,i,t} - P_{LS,i,t} \\ G_{G,m,t} + G_{CCP,m,t} - G_{MT,m,t} = G_{Load,m,t} + G_{Con,m,t} \\ H_{CCP,k,t} + H_{MT,k,t} = H_{Load,k,t} + H_{Con,k,t} \end{cases}, \tag{44}$$

where $P_{PV,i,t}$ and $P_{Wind,i,t}$ are PV and wind energy output power of power grid node i at time t, $G_{G,t}$ is the purchased gas power of gas network node m, while $P_{Load,i,t}$, $G_{Load,m,t}$ and $H_{Load,k,t}$ are electric load, gas load and heat load.

3 Deep Reinforcement Learning Model for Low Carbon Economic Dispatch of the IES

In this paper, the multi-objective optimization problem of low-carbon economic dispatch of the IES is solved by the DRL method. In this section, the optimal problem of low-carbon economic dispatch of the IES is transformed into a DRL framework and solved by the TD3 algorithm.

3.1 Problem Transformation

Reinforcement learning is the process of interacting with the environment, obtaining feedback, updating strategy, iterative until learning the optimal strategy [23]. This environmental interaction is described by the Markov decision process (MDP), which consists of five elements: state space $\mathcal{S}$, action space $\mathcal{A}$, transfer function $\mathcal{T}$, reward function P and discount factor γ.

(1) State Space

The observed state of the IES is shown in Eq. (45), including the state of charge SOC_t of battery storage, wind power output $P_{wind,t}$, PV output $P_{PV,t}$, electric load $P_{Load,t}$, gas load G_{Load}, heat load $H_{Load,t}$ and time t.

$$s_t = \left(SOC_t, P_{Wind,t}, P_{Load,t}, G_{Load,t}, H_{Load,t}, t \right). \tag{45}$$

(2) Action Space

The action space of the IES is shown in Eq. (46), including electric power output $P_{CCP,t}$ and gas power output $G_{CCP,t}$ of the CCP, electric power $P_{MT,t}$ of gas turbine, output power $P_{BS,t}$ of energy storage battery, and electric power $P_{IDR,t}$ of the IDR.

$$a_t = \left(P_{CCP,t}, G_{CCP,t}, P_{MT,t}, P_{BS,t}, P_{IDR,t} \right). \tag{46}$$

(3) Reward

The reward function is set to guide the agent to acquire the maximum cumulative reward from the current action. Therefore, as the reinforcement learning agent generally employs maximizing cumulative reward, the reward is a negative value of the objective function. The reward function can be described by Eq. (47).

$$R_t = -\left(C_O + C_C + C_{CWind} + C_{CPV} + C_{IDR} + C_{LS} \right) + r_0, \tag{47}$$

where r_0 is a manually set constant, which is used to ensure that the cumulative reward of the agent changes from a negative value to a positive value in the learning process.

The punishment reward mainly includes two parts, agent action amplitude penalty and agent action change rate punishment, Which can be shown by Eq. (48).

$$\chi_t = \sum_{t=1}^{T} \sum_{a_i} \left[\kappa_{11,i} \max \left(a_{i,t+1} - a_{i,t} - a_{11,i,\max}, 0 \right) + \kappa_{12,i} \max \left(a_{i,t} - a_{i,t+1} - a_{i,12,\min}, 0 \right) \right.$$

$$\left. + \kappa_{21,i} \max \left(a_{i,t} - a_{i,21,\max}, 0 \right) + \kappa_{22,i} \max \left(a_{i,t+1} - a_{i,22,\min}, 0 \right) \right], \tag{48}$$

where $\kappa_{11,i}$, $\kappa_{12,i}$ are the upper and lower penalty coefficients of the agent's climbing action, $a_{11,i,\max}$, $a_{i,12,\min}$ are the upper and lower climbing limits of the action, $\kappa_{21,i}$, $\kappa_{22,i}$ are the upper and lower penalty coefficients of the agent's action, $a_{i,21,\max}$, $a_{i,22,\min}$ are the upper and lower limits of the action.

3.2 Problem Solving Based on TD3 Algorithm

In this paper, the multi-objective optimization problem of low-carbon economic dispatch of the IES is solved by the DRL method. In this section, the optimal problem of low-carbon economic dispatch of the IES is transformed into a DRL framework and solved by the TD3 algorithm. TD3 is a DRL algorithm based on the actor-critic framework and DDPG. The principle of the TD3 algorithm is shown in Fig. 3. The actor-critic framework adopts two neural networks. The actor generates action according to state, and critic inputs state and action to generate Q value and learns reward and punishment mechanism to evaluate the behavior selected by the actor. The agent updates relevant parameters in the continuous state by updating the strategy of the actor and achieves the single-step update effect.

In the DDPG algorithm, the actor generates deterministic actions according to the policy function $a_t = \pi_\theta (s_t)$. The parameter of the actor-network is θ, and the optimization objective can be described by Eqs. (49) and (50) [17].

$$J(\pi_\theta) = \mathbb{E}\left[Q\left(s_t, \pi_\theta(s_t) \right) \right], \tag{49}$$

$$\begin{aligned} \nabla_\theta J(\pi_\theta) &= \nabla_\theta \mathbb{E}\left[Q(s_t, a_t) \right] \\ &= \mathbb{E}\left[\nabla_\theta \pi_\theta(s_t) \nabla_a Q(s_t, a_t) \big|_{a_t = \pi_\theta(s_t)} \right], \end{aligned} \tag{50}$$

where $Q(s_t, a_t)$ is the action-value function, which indicates the expected reward value of the selected action a_t in the state s_t.

The critic is a Q function, which is used to fit the state-action value. The critic evaluates the action in the current state and provides gradient information for the actor. The target is calculated by Eq. (51).

$$y_t = r(s_t, a_t) + \gamma Q_{\varphi'}(s_{t+1}, \pi_\theta(s_{t+1})) \tag{51}$$

where $Q_{\varphi'}$ is the target Q network, y_t is the target value of the critic network.
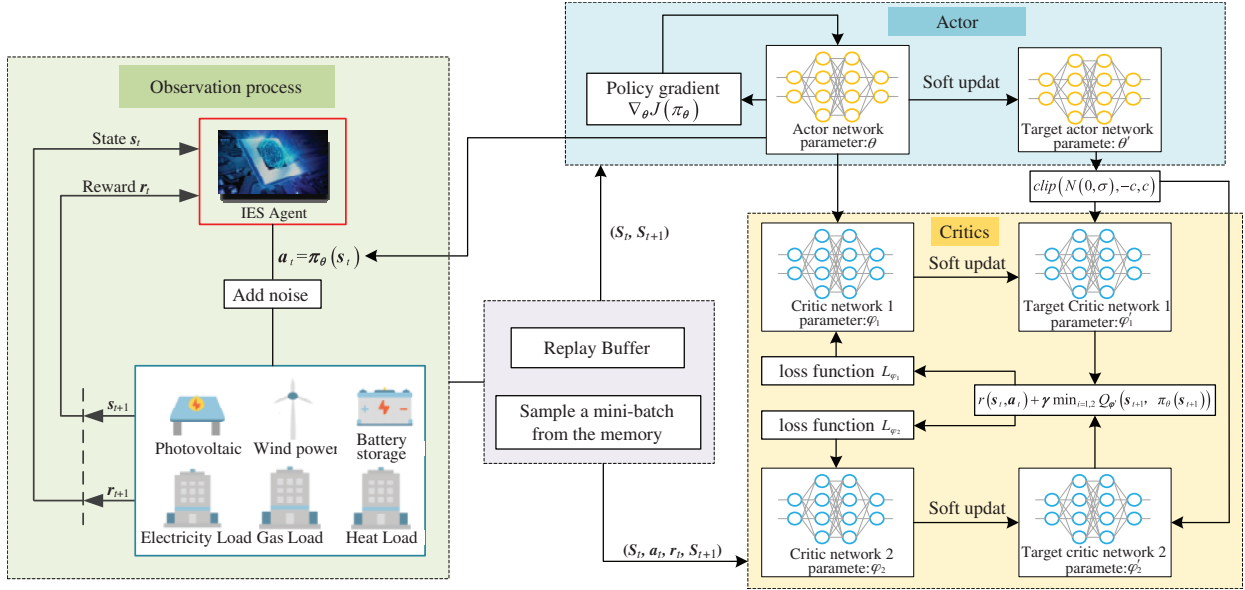

Figure 3: The principle of the TD3 algorithm

In DDPG, there is an overfitting phenomenon of the Q network, leading to the overestimation of the Q value. Under this circumstance, the policy network will affect the final performance by learning wrong information. TD3 algorithm can solve the overestimation problem of DDPG, restrict the overfitting of the Q network and reduce the deviation. Therefore, based on the DDPG framework, the improvements of TD3 are as follows:

(1) Clipped double-Q learning under actor-critic framework [27]. In the DDPG algorithm, both target actor-network and target critic network adopt the "soft update" method, which makes the actual network similar to the target network. It is difficult to separate the action selection and policy evaluation. Therefore, in the TD3 algorithm, the target value is obtained by clipping double-Q learning, and the Q value is constrained by two Q networks. Besides, corresponding minimum Q value in two Q networks are adopted to calculate the target Q value. According to the actor network π_θ, the target value y_t^2 of the critic network Q_{φ_2} is equal to y_t^1 as shown in Eq. (52).

$$y_t^1 = r(s_t, a_t) + \gamma \min_{i=1,2} Q_{\varphi'}(s_{t+1}, \pi_\theta(s_{t+1})). \tag{52}$$

(2) Smoothen the target action. In the continuous action space, it is usually expected that the same actions can have similar values [25]. Therefore, the action output of the agent is smoothened by adding random noise to the target action. The value function is updated by Eqs. (53) and (54).

$$y_t = r(\boldsymbol{s}_t, \boldsymbol{a}_t) + \gamma \min_{i=1,2} Q_{\varphi'}(\boldsymbol{s}_{t+1}, \pi_\theta(\boldsymbol{s}_{t+1}) + \varepsilon), \tag{53}$$

$$\varepsilon \sim clip(N(0,\sigma), -c, c), \tag{54}$$

where ε is the random noise.

(3) Delay the update of the policy network. Policy network cannot be trained according to poor performance Q network evaluation [28]. Therefore, in TD3, the actor is updated after the critic is updated for n times. The target network parameters are updated by Eq. (55).

$$\begin{aligned} \theta' &= (1-\tau) \cdot \theta' + \tau\theta, \\ \varphi' &= (1-\tau) \cdot \varphi' + \tau\varphi, \end{aligned} \tag{55}$$

where τ is the soft update rate.

4 Case Study

The simulation environment is established in the Gym toolkit of Open AI. The simulation structure of the IES in this paper is composed of IEEE 39-bus power system, 6-bus heating system and 20-bus natural gas system [1]. The natural gas price is 33 ¢/m. TD3 algorithm uses the neural network to fit actor and critic network. The number of hidden layers of the actor network is 3, and the number of neurons is 400, 200, 200. The hidden layer of the critic network is 3, and the number of neurons is 200, 100, 100. Simulation parameters of the IES are shown in Table 1 [5–7].

Table 1: Simulation parameters of the IES

Parameter	Value	Parameter	Value	Parameter	Value
$a_{CHP,x}$(USD/MW^2)	0.028	$b_{CHP,x}$(USD/MW)	24.85	$c_{CHP,x}$(USD/MW)	17.14
$a_{CCS,x}$(USD/MW)	0.95	$a_{P2G,x}$(USD/MW)	0.15	$b_{CCS,x}$(USD/MW)	18.6
$a_{P2G,x}$(USD/MW)	0.15	$b_{CCS,x}$(USD/MW)	18.6	$b_{P2G,x}$(USD/MW)	17.5
a_{CWind}(USD/MW)	36	a_{CPV}(USD/MW)	36	a_{Tra}(USD/MW)	45
a_{Int}(USD/MW)	64	a_{Con}(USD/MW)	28	a_{LS}(USD/MW)	28.8
$\delta_{Tra,i}^{min}$	0.02	$\delta_{Tra,i}^{max}$	0.15	$\delta_{Int,i}^{min}$	0.02
$\delta_{Int,i}^{max}$	0.18	$\delta_{Con,i}^{min}$	0.02	$\delta_{Con,i}^{max}$	0.18
λ(USD/MW)	34.5	P_{CHP}^{min}(MW)	5	P_{CHP}^{max}(MW)	60
P_{P2G}^{min}(MW)	0	P_{P2G}^{max}(MW)	15	P_{CCS}^{min}(MW)	1
P_{CCS}^{max}(MW)	0.2	$\Delta P_{CHP}^{min}(MW)$	0.1	$\Delta P_{CHP}^{max}(MW)$	25
ΔP_{P2G}^{min}(MW)	2	ΔP_{P2G}^{max}(MW)	5	ΔP_{CCS}^{min}(MW)	0
ΔP_{CCS}^{max}(MW)	0	P_{MT}^{min}(MW)	5	P_{MT}^{max}(MW)	30
ΔP_{MT}^{min}(MW)	2	ΔP_{MT}^{max}(MW)	10	SOC^{max}	0.95

(Continued)

Table 1 (Continued)

| Parameter	Value	Parameter	Value	Parameter	Value
$SOC^{\min}$	0.10	$P_{BS,ch}^{\max}$	10	$P_{BS,disch}^{\max}$	10
η_{ch}	0.90	η_{disch}	0.85	γ	0.95
$\kappa_{11,i}$	0.16	$\kappa_{12,i}$	0.28	$\kappa_{21,i}$	0.31
$\kappa_{22,i}$	0.22				

4.1 TD3 Algorithm Training Process

The data in this paper are taken from Liaoning province, China from 01 November 2020 to 28 February 2021. The training sets include the data from 01 November 2020 to 31 January 2021. The test sets from 01 February to 28th February are used to verify the optimized results after training. Training data is shown in Fig. 4.
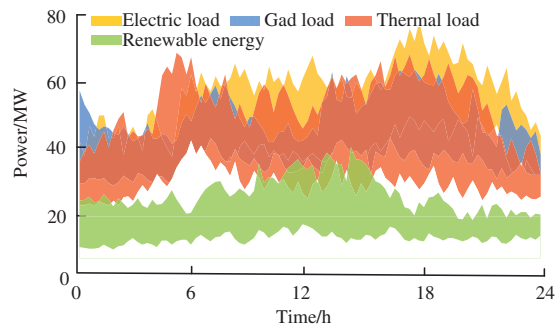
Figure 4: Historical sample data of IES

The training results of the TD3 algorithm are shown in Fig. 5. The reward value obtained by the agent at the initial stage of training is relatively low. The TD3 algorithm gets a stable optimal solution when the episodes approaching about 3000.
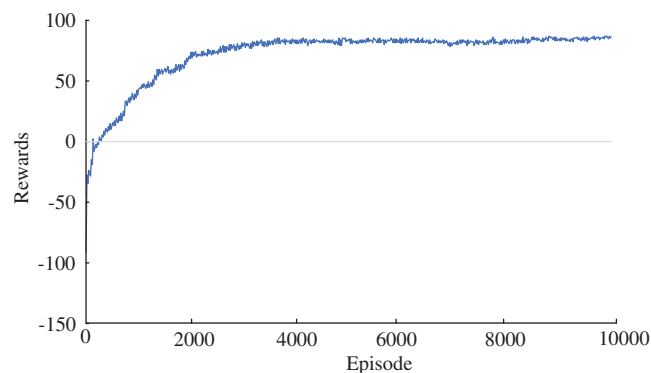
Figure 5: The cumulative reward value of agents

4.2 Scenario Analysis

To verify the effectiveness of the proposed multi-objective low-carbon scheduling method, the scheduling results of five different operating scenarios are compared.

Scenario 1: The optimization objective is the operation cost, and the economy of system operation is considered in the optimization process.

Scenario 2: CCP combined optimization model is utilized.

Scenario 3: Power supply reliability is considered based on scenario 2.

Scenario 4: IDR is considered based on scenario 2.

Scenario 5: IDR and power supply reliability are considered based on scenario 2.

The optimal scheduling results in different scenarios are shown in Table 2. Compared with scenario 1, the system operation and total costs in scenario 2 are reduced by \$150 and \$1416. CCS captures CO_2 emitted by CHP, and transmits it to P2G, thus saving the cost of P2G transmission and purchasing CO_2. Compared with scenario 2, the total cost of the system in scenario 3 is increased by \$996 to meet load requirements after considering system reliability. Compared with scenario 2, the power load demand in the peak period and the starting capacity of CHP units in scenario 4 are reduced. Compared with scenario 2, the load rejection of the system in scenario 5 is reduced to 0, and the wind curtailment cost is significantly reduced. At the same time, the PV curtailment costs are reduced to 0, and the total cost of the IES is reduced by \$2201. Therefore, the method proposed can improve the reliability of the power supply and economic benefits.

**Table 1:** The optimal scheduling results in different scenarios

Scenario	1	2	3	4	5
C_O/\$	190257	190107	191468	188551	190425
C_C/\$	−15458	−16784	−16871	−16101	−17812
C_{QF}/\$	1024	987	1012	654	621
C_{QG}/\$	278	277	284	0	0
C_{IDR}/\$	0	0	0	951	987
C_{LS}/\$	321	319	109	201	0
Total cost/\$	176422	175006	176002	174256	174221

CO_2 emissions in different scenarios are shown in Fig. 6. In the load peak period (20:00–22:00), scenario 2 emits less CO_2 than scenario 1 as the requirements of CO_2 of P2G come from CCS. Scenario 4 has the lowest CO_2 emissions due to the less CHP unit output during peak load periods. Scenario 3 considers the power supply reliability, improves the output of CHP units, resulting in the highest CO_2 emissions. Considering power supply reliability and IDR, scenario 5 has slightly higher carbon emissions than scenario 4, but lower than scenario 2.
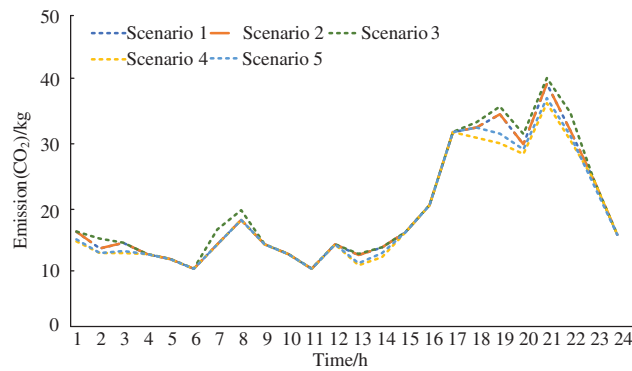
Figure 6: CO_2 emissions in different scenarios
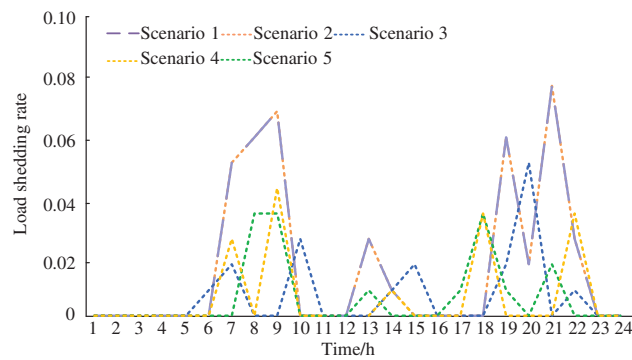
Figure 7: The load shedding rate in different scenarios

The load shedding rate in different scenarios is shown in Fig. 7. The load shedding rates in scenario 1 are the same as scenario 2, and the CCP combined optimization model does not affect power supply stability. In CCP combined optimization model, the electric power consumed by P2G and CCS is directly taken from electric power generated by CHP. The integrated energy system generates and consumes no additional power. Compared with scenario 1, the operation stability of the IES in scenario 3 is improved by considering the power supply reliability. Compared with scenario 1, the load demand in scenario 4 is less during the period (7:00–8:00, 19:00–22:00), and the power supply reliability is improved after considering the demand response.

The scheduling results of the IES in different scenarios are shown in Fig. 8. In scenario 3, CHP unit output is significantly higher than other scenarios thanks to the consideration of power supply reliability. In scenario 4, the outputs of CHP units during peak load period are lower than those of other scenarios, and the outputs of renewable energy are higher than those of other scenarios during under load period due to the considerations of power supply reliability and IDR. In scenario 5, during the underestimation period (0:00–06:00) of load, the peak load demand is transferred to the underestimation period after demand response, and renewable energy outputs are significantly higher than those in other scenarios. During the peak load period (19:00–21:00), CO_2 emissions are reduced while maintaining system reliability. The optimization method based on TD3 can achieve a better control effect in different scenarios.
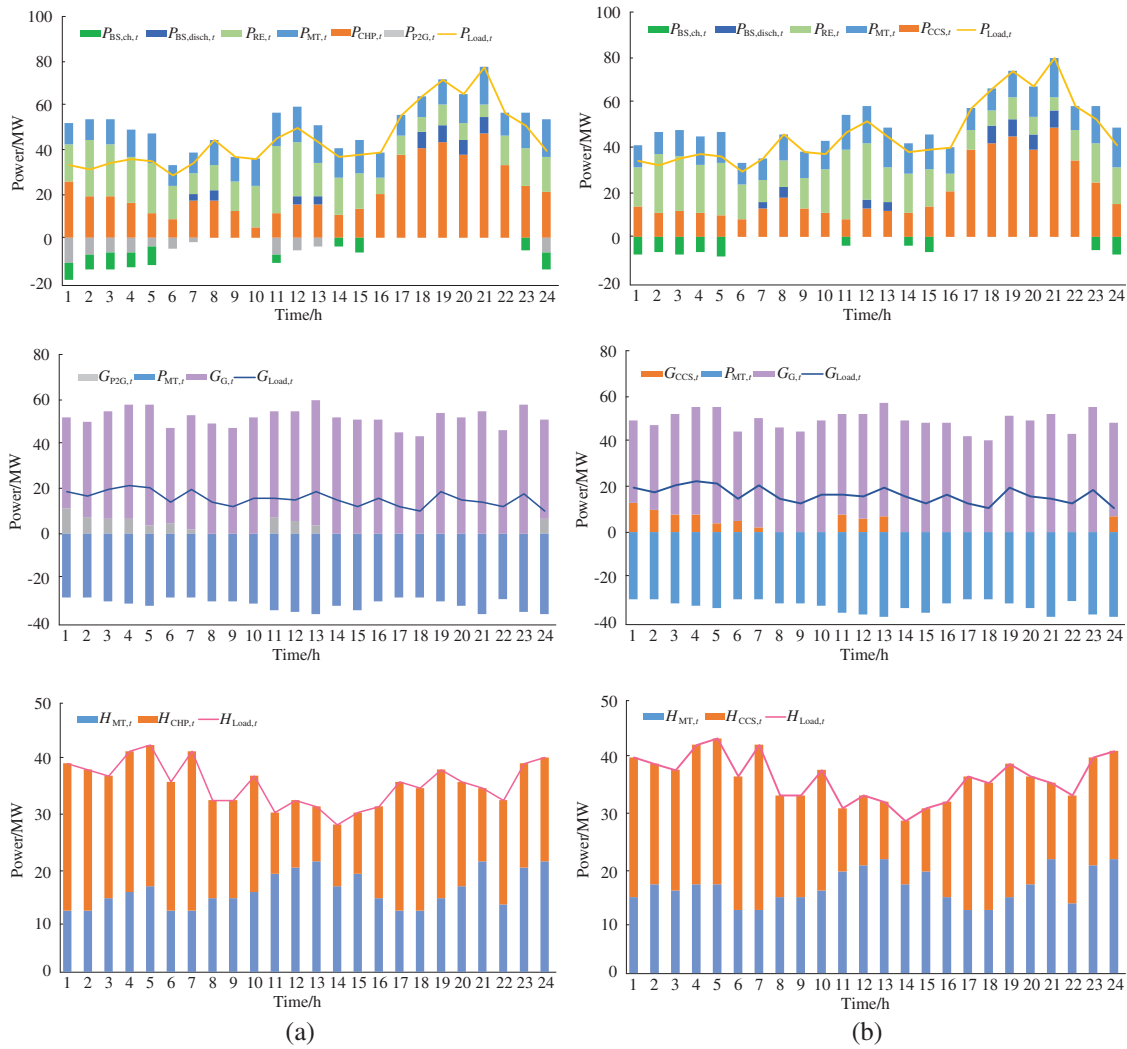
Figure 8: (Continued)

4.3 Comparison Analysis

SAC, A3C, DDPG and DQN algorithms are selected to verify the effectiveness of the proposed low carbon optimization model. The parameters of SAC, A3C, DDPG and DQN are taken from the literature [29–32].
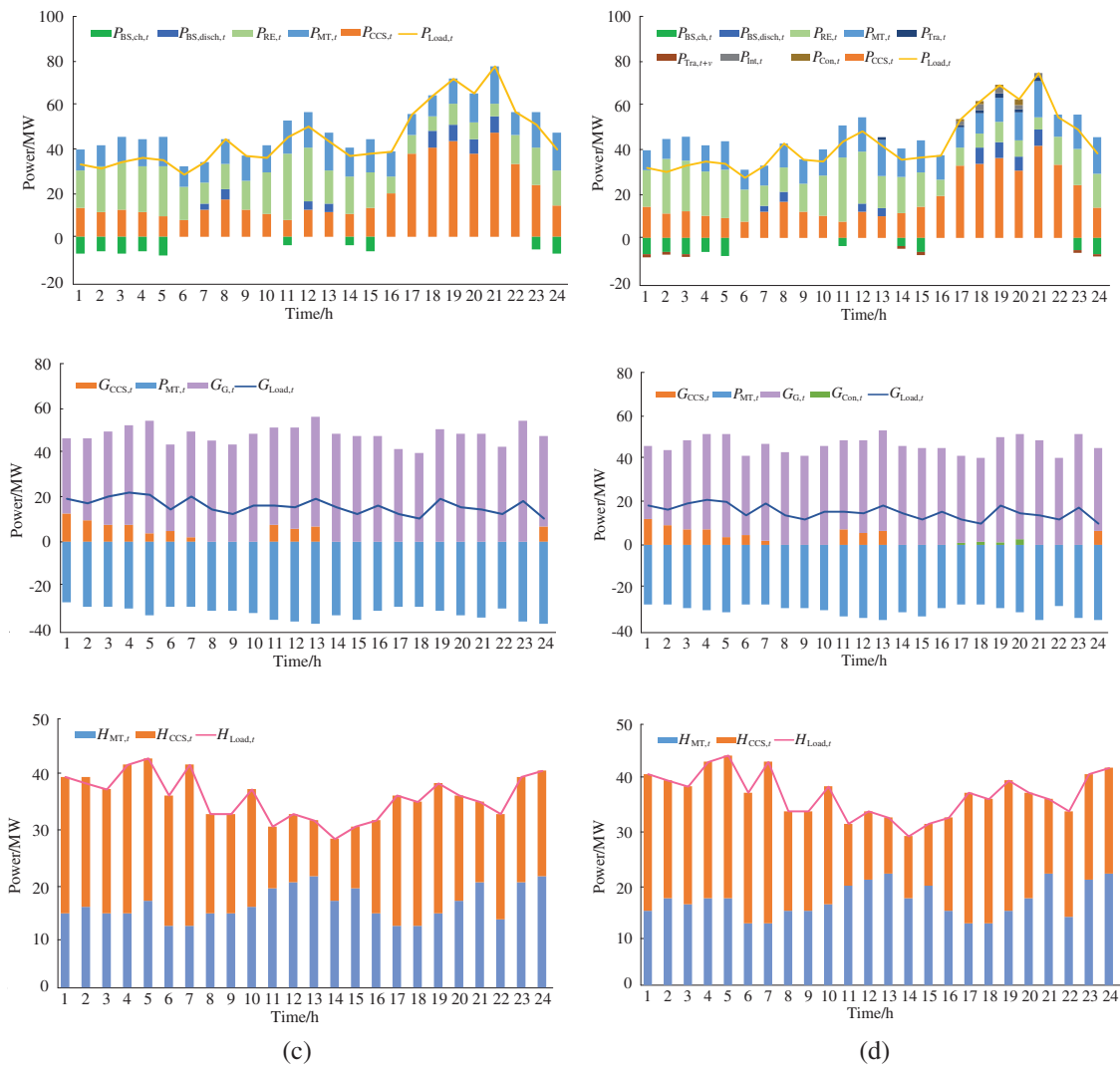
Figure 8: (Continued)

The reward curves of different comparison methods are shown in Fig. 9. The reward values obtained by DDPG and SAC algorithms are similar in convergence. The convergence speed of the A3C algorithm is higher than other methods due to the asynchronous architecture. The proposed optimization method based on TD3 has the best comprehensive performance and can obtain higher reward values.
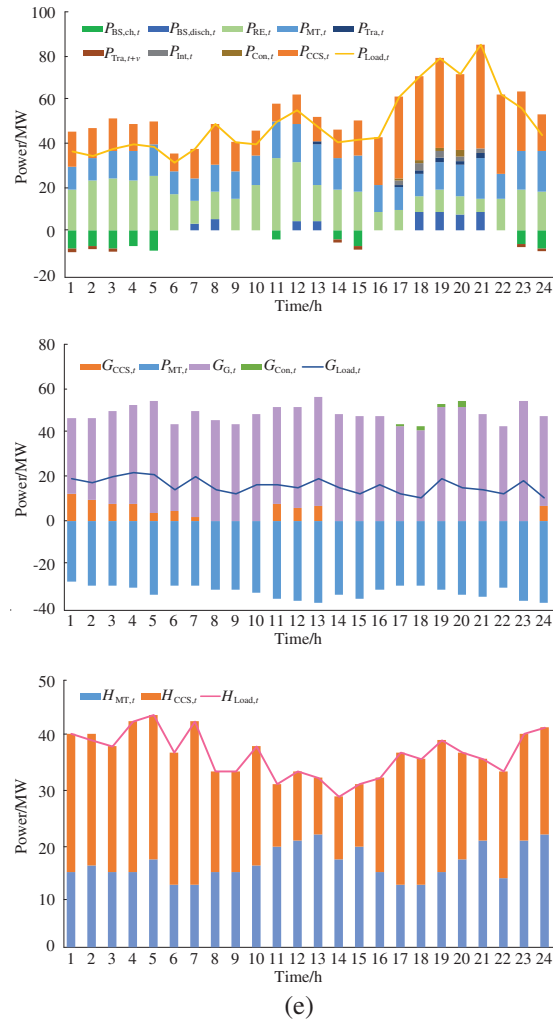
Figure 8: The power balances of the IES in different scenarios. (a) The power balances of the IES in scenario 1, (b) the power balances of the IES in scenario 2, (c) the power balances of the IES in scenario 3, (d) the power balances of the IES in scenario 4, (e) the power balances of the IES in scenario 5

The optimal scheduling results in different algorithms are shown in Table 3. DQN algorithm has the highest operating cost due to the discretization of the agent action. The optimization results of the SAC algorithm and DDPG algorithm are similar, and the wind power curtailment cost of DDPG is higher than the SAC algorithm. A3C algorithm uses asynchronous mechanism, so the operating cost is lower than DQN, DDPG, and SAC algorithms. TD3 algorithm has the lowest cost and, obtains higher carbon emission benefits.
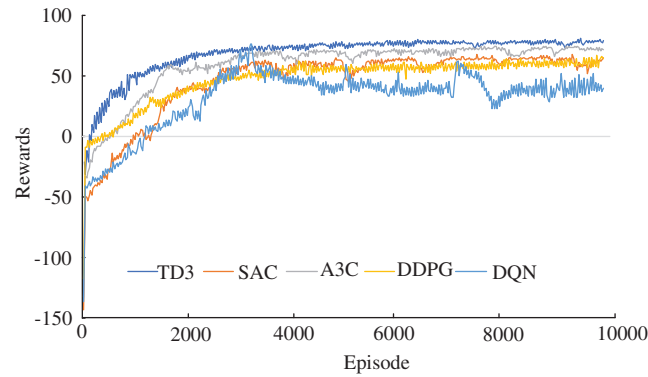
Figure 9: Reward curve in the training process of different algorithms

Table 3: The optimal scheduling results in different algorithms

Scenario	TD3	SAC	A3C	DDPG	DQN
C_O/\$	190425	192465	190897	193412	203512
C_C/\$	-17812	-17697	-17433	-17621	-16985
C_{QF}/\$	621	688	652	712	874
C_{QG}/\$	0	0	0	0	23
C_{IDR}/\$	987	1011	981	1001	1023
C_{LS}/\$	0	0	0	0	0
Total cost/\$	174221	176467	175097	177504	188447

5 Conclusions

In this paper, a multi-objective optimization method based on TD3 is proposed for the low-carbon scheduling problem of the IES. On the power generation side, we develop the CCP combined optimization model. IDR and power supply reliability are considered on the power supply side. Moreover, we describe the multi-objective optimization problem as MDP and use the deep reinforcement learning method based on TD3 to solve it. The proposed method can achieve a better control effect in different scenarios. The results show that the proposed method saves the operation cost of the system and effectively reduces the CO_2 emission of the IES. Compared with TD3, SAC, A3C, DDPG and DQN algorithms, the operating costs are reduced by 8.6%, 4.3%, 6.1% and 8.0%, respectively.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Zhou, Y., Hu, W., Min, Y., Dai, Y. (2019). Integrated power and heat dispatch considering available reserve of combined heat and power units. *IEEE Transactions on Sustainable Energy, 10(3),* 1300–1310. DOI 10.1109/TSTE.2018.2865562.

2. Liu, C., Li, Q., Wang, K. (2021). State-of-charge estimation and remaining useful life prediction of super-capacitors. *Renewable and Sustainable Energy Reviews, 150(5),* 111408. DOI 10.1016/j.rser.2021.111408.

3. Yang, J., Zhang, N., Cheng, Y., Kang, C., Xia, Q. (2019). Modeling the operation mechanism of combined P2G and gas-fired plant with CO_2 recycling. *IEEE Transactions on Smart Grid, 10(1),* 1111–1121. DOI 10.1109/TSG.2018.2849619.

4. Zhang, Z., Wang, D., Gao, J. (2021). Learning automata-based multiagent reinforcement learning for optimization of cooperative tasks. *IEEE Transactions on Neural Networks and Learning Systems, 32(10),* 4639–4652. DOI 10.1109/TNNLS.2020.3025711.

5. Gong, H., Rallabandi, V., Ionel, D. M., Colliver, D., Duerr, S. et al. (2020). Dynamic modeling and optimal design for net zero energy houses including hybrid electric and thermal energy storage. *IEEE Transactions on Industry Applications, 56(4),* 4102–4113. DOI 10.1109/TIA.2020.2986325.

6. Wang, H. X., Yang, J., Chen, Z., Li, G., Liang, J. et al. (2020). Optimal dispatch based on prediction of distributed electric heating storages in combined electricity and heat networks. *Applied Energy, 267(3),* 1–10. DOI 10.1016/j.apenergy.2020.114879.

7. Clegg, S., Mancarella, P. (2016). Integrated electrical and gas network flexibility assessment in low-carbon multi-energy systems. *IEEE Transactions on Sustainable Energy, 7(2),* 718–731. DOI 10.1109/TSTE.2015.2497329.

8. Wang, D., Liu, L., Jia, H. J., Wang, W. L., Zhi, Y. Q. et al. (2018). Review of key problems related to integrated energy distribution systems. *CSEE Journal of Power and Energy Systems, 4(2),* 130–145. DOI 10.17775/CSEEJPES.2018.00570.

9. Feng, X., Li, Q., Wang, K. (2020). Waste plastic triboelectric nanogenerators using recycled plastic bags for power generation. *ACS Applied Materials & Interfaces, 13(1),* 400–410. DOI 10.1021/acsami.0c16489.

10. Zhao, J., Li, F., Wang, Z., Dong, P., Wang, K. (2021). Flexible PVDF nanogenerator-driven motion sensors for human body motion energy tracking and monitoring. *Journal of Materials Science: Materials in Electronics, 32(11),* 14715–14727. DOI 10.1007/s10854-021-06027-w.

11. Lu, S., Gu, W., Meng, K., Dong, Z. (2021). Economic dispatch of integrated energy systems with robust thermal comfort management. *IEEE Transactions on Sustainable Energy, 12(1),* 222–233. DOI 10.1109/TSTE.2020.2989793.

12. Wang, Y., Qiu, J., Tao, Y., Zhao, J. (2020). Carbon-oriented operational planning in coupled electricity and emission trading markets. *IEEE Transactions on Power Systems, 35(4),* 3145–3157. DOI 10.1109/TPWRS.2020.2966663.

13. Zhai, J., Wu, X., Zhu, S., Liu, H. (2019). Low carbon economic dispatch of regional integrated energy system considering load uncertainty. *34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC),* pp. 642–647. Bohai University, China.

14. Yang, D., Zhou, X., Yang, Z., Guo, Y., Niu, Q. (2020). Low carbon multi-objective unit commitment integrating renewable generations. *IEEE Access, 8,* 207768–207778. DOI 10.1109/ACCESS.2020.3022245.

15. Bahrami, S., Sheikhi, A. (2016). From demand response in smart grid toward integrated demand response in smart energy hub. *IEEE Transactions on Smart Grid, 7(2),* 650–658. DOI 10.1109/TSG.2015.2464374.

16. Zeng, B., Liu, Y., Luo, Y., Xu, H., Ahmed, L. (2021). Optimal planning of IDR-integrated electricity-heat system considering techno-economic and carbon emission issues. *CSEE Journal of Power and Energy Systems, 3(4),* 1–12. DOI 10.17775/CSEEJPES.2020.02940.

17. Zhou, J., Xue, S., Xue, Y., Liao, Y., Liu, J. et al. (2021). A novel energy management strategy of hybrid electric vehicle via an improved TD3 deep reinforcement learning. *Energy, 224(5),* 1–16. DOI 10.1016/j.energy.2021.120118.

18. Wu, Y., Tan, H., Peng, J., Zhang, H., He, H. (2019). Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus. *Applied Energy, 247(6),* 454–466. DOI 10.1016/j.apenergy.2019.04.021.

19. Yan, Z. M., Xu, Y. (2019). Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search. *IEEE Transactions on Power Systems, 34(2),* 1653–1656. DOI 10.1109/TPWRS.2018.2881359.

20. Petrollese, M., Valverde, L., Cocco, D., Cau, G., Guerra, J. (2016). Real-time integration of optimal generation scheduling with MPC for the energy management of a renewable hydrogen-based microgrid. *IEEE Transactions on Power Systems, 166,* 96–106. DOI 10.1016/j.apenergy.2016.01.014.

21. Wu, S., Hu, W., Lu, Z., Gu, Y., Tian, B. et al. (2020). Power system flow adjustment and sample generation based on deep reinforcement learning. *Journal of Modern Power Systems and Clean Energy, 8(6),* 1115–1127. DOI 10.35833/MPCE.2020.000240.

22. Yan, Z. M., Xu, Y. (2020). A multi-agent deep reinforcement learning method for cooperative load frequency control of a multi-area power system. *IEEE Transactions on Power Systems, 35(6),* 4599–4608. DOI 10.1109/TPWRS.2020.2999890.

23. Ji, Y., Wang, J. H., Xu, J. C., Fang, X. K., Zhang, H. G. (2019). Real-time energy management of a microgrid using deep reinforcement learning. *Energies, 12(12),* 2291. DOI 10.3390/en12122291.

24. Zhang, B., Hu, W., Cao, D., Huang, Q., Chen, Z. (2021). Asynchronous advantage actor-critic based approach for economic optimization in the integrated energy system with energy hub. *2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES)*, pp. 1170–1176. Chengdu, China.

25. Ma, Y., Wang, H., Hong, F., Yang, J., Chen, Z. et al. (2021). Modeling and optimization of combined heat and power with power-to-gas and carbon capture system in integrated energy system. *Energy, 236(5),* 121392. DOI 10.1016/j.energy.2021.121392.

26. Rigo-Mariani, R., Zhang, C., Romagnoli, A., Kraft, M. K., Ling, V. et al. (2020). A combined cycle gas turbine model for heat and power dispatch subject to grid constraints. *IEEE Transactions on Sustainable Energy, 11(1),* 448–456. DOI 10.1109/TSTE.2019.2894793.

27. Zhang, F. J., Li, J., Li, Z. (2020). A TD3-based multi-agent deep reinforcement learning method in mixed cooperation-competition environment. *Neurocomputing, 411(3–4),* 206–215. DOI 10.1016/j.neucom.2020.05.097.

28. Wu, D., Dong, X., Shen, J., Hoi, S. C. H. (2020). Reducing estimation bias via triplet-average deep deterministic policy gradient. *IEEE Transactions on Neural Networks and Learning Systems, 31(11),* 4933–4945. DOI 10.1109/TNNLS.2019.2959129.

29. Wu, J., Wei, Z., Li, W., Wang, Y., Li, Y. et al. (2021). Battery thermal and health-constrained energy management for hybrid electric bus based on soft actor-critic DRL algorithm. *IEEE Transactions on Industrial Informatics, 17(6),* 3751–3761. DOI 10.1109/TII.2020.3014599.

30. He, Y., Wang, Y., Qiu, C., Lin, Q., Li, J. et al. (2021). Blockchain-based edge computing resource allocation in IoT: A deep reinforcement learning approach. *IEEE Internet of Things Journal, 8(4),* 2226–2237. DOI 10.1109/JIOT.2020.3035437.

31. Qiu, C., Hu, Y., Chen, Y., Zeng, B. (2019). Deep deterministic policy gradient (DDPG)-based energy harvesting wireless communications. *IEEE Internet of Things Journal, 6(5),* 8577–8588. DOI 10.1109/JIOT.2019.2921159.

32. Lin, T., Su, Z., Xu, Q., Xing, R., Fang, D. (2020). Deep Q-network based energy scheduling in retail energy market. *IEEE Access, 8,* 69284–69295. DOI 10.1109/ACCESS.2020.2983606.