sustainability

MDPI

Article

Low-Carbon Economic Dispatch of Integrated Energy Systems for Electricity, Gas, and Heat Based on Deep Reinforcement Learning

Xiaojuan Lu [1], Yaohui Zhang [1], Duojin Fan [2,*], Jiawei Wei [1] and Xiaoying Yu [1]

[1] School of Automation Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; luxj@lmail.lzjtu.cn (X.L.)
[2] Research Institute of Photothermal Energy Storage, Lanzhou Jiaotong University, Lanzhou 730070, China
* Correspondence: fanduojin@lzdctc.com

Abstract

Under the background of "dual-carbon", the development of energy internet is an inevitable trend for China's low-carbon energy transition. This paper proposes a hydrogen-coupled electrothermal integrated energy system (HCEH-IES) operation mode and optimizes the source-side structure of the system from the level of carbon trading policy combined with low-carbon technology, taps the carbon reduction potential, and improves the renewable energy consumption rate and system decarbonization level; in addition, for the operation optimization problem of this electric–gas–heat integrated energy system, a flexible energy system based on electric–gas–heat is proposed. Furthermore, to address the operation optimization problem of the HCEH-IES, a deep reinforcement learning method based on Soft Actor–Critic (SAC) is proposed. This method can adaptively learn control strategies through interactions between the intelligent agent and the energy system, enabling continuous action control of the multi-energy flow system while solving the uncertainties associated with source-load fluctuations from wind power, photovoltaics, and multi-energy loads. Finally, historical data are used to train the intelligent body and compare the scheduling strategies obtained by SAC and DDPG algorithms. The results show that the SAC-based algorithm has better economics, is close to the CPLEX day-ahead optimal scheduling method, and is more suitable for solving the dynamic optimal scheduling problem of integrated energy systems in real scenarios.

Keywords: integrated energy systems; low-carbon economic dispatch; deep reinforcement learning; soft actor–critic; optimal energy management

Academic Editors: Ning Zhang, Yan Juan, Chao Ning and Yumeng Song

Received: 16 September 2025
Revised: 10 October 2025
Accepted: 11 October 2025
Published: 13 October 2025

Citation: Lu, X.; Zhang, Y.; Fan, D.; Wei, J.; Yu, X. Low-Carbon Economic Dispatch of Integrated Energy Systems for Electricity, Gas, and Heat Based on Deep Reinforcement Learning. *Sustainability* **2025**, *17*, 9040. https://doi.org/10.3390/su17209040

1. Introduction

With the depletion of traditional fossil fuels and the advancement of renewable energy technologies, countries worldwide are actively reshaping their energy mix to diminish their reliance on conventional fossil-based energy sources. The development of integrated energy systems (IES), enhancing energy efficiency, and boosting the capacity to integrate intermittent renewable sources via heterogeneous energy networks, represents pivotal pathways towards achieving a low-carbon, sustainable energy future [1,2].

Certain power grids in China's "Three North" regions feature a high penetration of distributed renewable energy and are coupled with intricate gas and heat networks, forming a typical multi-energy flow integrated energy system. Power-to-gas (P2G) technology converts surplus wind power or off-peak electricity into hydrogen, which can be further synthesized into methane. This enables a two-way interaction between the

power grid and the gas grid, offering a novel pathway for renewable energy integration [3]. Ref. [4] proposed a two-stage joint operation strategy involving P2G and Hydrogen fuel cells (HFCs) to promote wind power utilization while reducing energy losses and carbon emissions. Ref. [5] introduced P2G and carbon capture technologies, proposing an optimized scheduling strategy for an IES with carbon capture-electricity-to-gas coupling, which enhanced the renewable energy absorption rate. Addressing photovoltaic (PV) uncertainty, Ref. [6] demonstrated that the joint operation of hydrogen fuel cells and cogeneration units improves the rationality of PV consumption and equipment output. Ref. [7] utilized the thermal energy storage characteristics of heating pipelines to improve the operational flexibility of combined heat and power (CHP) systems and constructed a flexibility evaluation method for generalized thermal energy storage models to quantitatively analyze the flexibility of district heating networks.

The primary problem facing the integrated energy system is the coordinated optimization and scheduling of multi-energy flows. Ref. [8] established a comprehensive optimization model, which was solved by the non-dominated ranking genetic algorithm (NSGA-II) with the goals of operating cost, carbon emissions, and energy efficiency utilization, and the Pareto optimal frontier solution set was output. Ref. [9] established a steady-state energy flow and carbon flow calculation model for the integrated electricity–gas–hydrogen energy system and performed iterative calculations using the Newtonian method. Ref. [10] applied the fully distributed internal point conjugate gradient method to the problem of correcting equations in the distributed optimal scheduling of integrated electrical energy systems. Ref. [11] studied the scheduling strategy of the integrated electrical–gas–thermal energy system at multiple time scales and constructed a data-driven optimization model for the split-brud rod. Ref. [12] used an improved multi-objective optimization algorithm to enhance the operational economy and energy efficiency and to reduce the carbon emission level of the integrated electric–heat–hydrogen–cooling energy system. Ref. [13] addressed the uncertainty and correlation between wind power and electricity/gas loads by calculating the probabilistic optimal power flow model of the electric-gas interconnection system using the three-point estimation method of Nataf transformation. Ref. [14] considered the influence of wind and solar uncertainty and constructed a stochastic scheduling model for integrated energy virtual power plants. This model couples "coal-fired" power generation with electricity–carbon–hydrogen–chemical coupling, aiming to maximize benefits.

In the establishment of the optimal scheduling model, the stochastic programming method shows obvious advantages in reducing the operating cost of the system compared with the deterministic method. Ref. [15] considered adding hydrogen vehicle emission reductions to carbon trading and used the Monte Carlo algorithm to generate scenarios for wind and solar output uncertainty. Stochastic programming uses random sampling, chance constraint generation, and other methods to convert uncertainty problems into deterministic models and calculate the operating status of the system through multiple scenarios. However, the large number of scenarios increases both the computational burden and solving difficulty. Therefore, it is necessary to balance calculation accuracy with computational load. Robust optimization is mainly aimed at optimizing the operation of the system in extreme scenarios. Ref. [16] used stochastic optimization and robust optimization to deal with the uncertainty of load-side power generation measurement. It also added a coordination strategy to the second-stage optimization objectives, bringing the real-time optimization results closer to the global optimization value. However, the former faces a bottleneck in computational efficiency due to its heavy reliance on scenario generation, while the latter suffers from model complexity and a difficult trade-off between economic efficiency and conservatism. More importantly, these traditional methods belong to "static" optimization—once the model or parameters are determined, they struggle to adaptively

learn new uncertainty patterns online, demonstrating limited capability in coping with continuously dynamic real-world environments.

In recent years, artificial intelligence technology has flourished, with reinforcement learning (RL) as a model-free approach. It does not need to understand environmental changes in advance, has strong adaptability to many uncertainties and interferences, makes optimal decisions through continuous learning and interaction, and has good generalization ability, so it is more and more important in the optimal control of the power system. Ref. [17] used a deep Q-network (DQN) to adaptively respond to random fluctuations in power generation and demand, solving the energy management problem. Ref. [18] used the Nash equilibrium Q-learning algorithm to enable the coordinated scheduling of integrated energy microgrids. Ref. [19] used the double deep expected Q-network (DDEQN) algorithm to efficiently solve the real-time stochastic economic scheduling problem of microgrids. However, the above reinforcement learning methods often discretize actions, which not only reduces the accuracy of optimization decisions but also increases the number of discrete actions exponentially due to the increase in action dimensions, causing "dimensional disasters" that are difficult to solve. At present, some studies have begun to explore continuous control deep reinforcement learning models. Ref. [20] used the deep deterministic policy gradient (DDPG) algorithm to enable dynamic regulation of the integrated electrical–heat–gas energy system. Ref. [21] used the DDPG algorithm to solve the continuous control problem in the coordinated and optimized operation of active distribution networks. However, existing studies still exhibit two main limitations: Firstly, at the algorithmic level, the DDPG algorithm itself suffers from issues such as hypersensitivity to hyperparameters, limited exploration efficiency, and Q-value overestimation, which may lead to training instability and suboptimal policy performance; secondly, at the system modeling level, most existing works focus on traditional electricity–heat–gas-coupled systems, failing to deeply integrate hydrogen energy as a key low-carbon carrier with carbon capture, utilization, and trading mechanisms, thereby restricting the system's deep decarbonization and operational flexibility under the "dual-carbon" goals.

Based on the Soft Actor–Critic (SAC) framework, this paper constructs a deep reinforcement learning method for the operation of the hydrogen-coupled electrothermal integrated energy system (HCEH-IES). This method enables the algorithm to adaptively learn the characteristics of uncertain variations in wind power, photovoltaics, and various loads, thereby realizing optimal system scheduling under multiple scenarios.

(1) An HCEH-IES model is constructed, with the optimization objective being the minimization of the sum of the system's comprehensive operating costs, carbon capture and utilization costs, and carbon trading costs.

(2) The optimal scheduling of the integrated energy system is formulated as a Markov Decision Process (MDP), and the system's state space, action space, and reward function are defined.

(3) The SAC algorithm is utilized to optimize the dynamic energy scheduling of the system. The feasibility and effectiveness of the proposed optimal scheduling strategy and model are verified by comparing results obtained using different optimization algorithms and scenarios.

2. Hydrogen-Coupled Electro-Thermal Integrated Energy System Architecture

The system structure is shown in Figure 1 and is mainly composed of energy supply, energy conversion, energy storage, and load. The supply side mainly includes the upper gas grid, wind turbine photovoltaic, and upper gas network. The conversion side is mainly composed of an electrolyzer (EL), methane reactor (MR), gas turbine (GT), gas boiler (GB), hydrogen fuel cell (HFC), waste heat power generation device based on organic Rankine

cycle (ORC), and waste heat boiler (waste heat boiler, WHB); the energy storage end is mainly the electricity storage (ES), thermal storage tank (TST), and hydrogen storage tank (HST). On the energy load side, users are aggregated and uniformly characterized as electrical loads, gas loads, and heat loads.

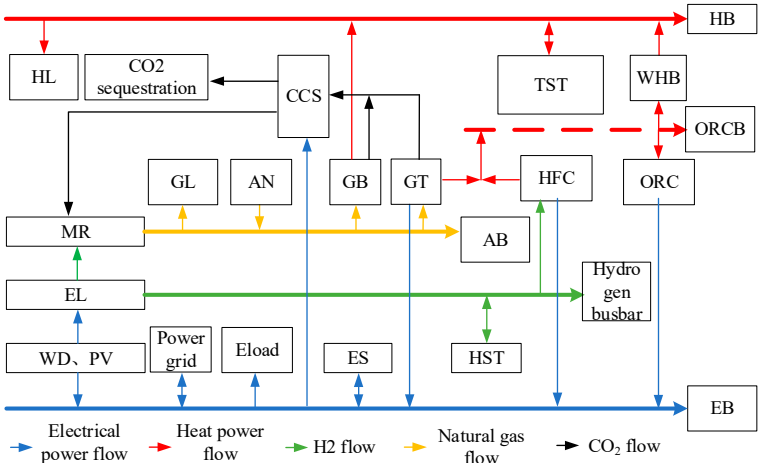

Figure 1. Structure of the hydrogen-coupled electrothermal integrated energy system (HCEH-IES).

Flow of energy and matter. We explicitly describe the core flows: electrical flow: from PV/WT/Grid, through converters (P2G, EL), to electrical loads and storage (ES). Gas flow: from the superior gas grid and the methane reactor (MR), through gas turbines (GT) and gas boilers (GB), to gas loads. Hydrogen flow: from electrolyzers (AWE, PEM) to storage (HST) and then to hydrogen fuel cells (HFC) or the methane reactor (MR). Heat flow: from cogeneration units (GT, HFC, GB), waste heat recovery (WHB, ORC), to heat loads and storage (TST). CO_2 flow: from the gas turbine's flue gas to the carbon capture system (CCS) and then to the methane reactor for utilization or to storage.

3. Modeling of Hydrogen-Coupled Electric–Thermal Integrated Energy Systems

3.1. Coupled Modeling of a Two-Stage Power-to-Gas and Carbon Capture (P2G-CCS) System

(1) Post-combustion carbon capture

In this paper, the installation of CCS in gas units for low-carbon flexible transformation is considered, and the mathematical model of carbon capture equipment is as follows:

$$\begin{cases} P_t^{CCS} = P_t^R + P^{TG} \\ P_t^R = \delta_{ccs} M_t^{CCS} \\ M_t^{CCS} = \kappa \eta_{ccs} M_G^{CO_2} (G_t^{GT} + G_t^{GB}) \end{cases} \quad (1)$$

This model captures the energy cost of carbon capture, which is crucial for economic dispatch. The variable P_t^{CCS} represents the operational energy consumed per unit of CO_2 captured, making carbon capture a dispatchable resource with an associated cost. The captured M_t^{CCS} becomes a feedstock for the P2G process, linking the carbon and gas networks.

P_t^R is the operating energy consumption power at time t; P^{TG} is to fix carbon capture energy consumption; δ_{ccs} is the energy consumption per unit of carbon capture; κ is the flue gas split ratio; η_{ccs} is the carbon capture efficiency; $M_G^{CO_2}$ is the CO_2 molar mass; G_t^{GT} and G_t^{GB} consume natural gas at t times, respectively.

(2) Two-stage P2G

P2G is a mathematical model for energy conversion efficiency. These equations model the core energy conversion process of P2G.

$$\begin{cases} H_t^{AWE} = \eta_{AWE} \frac{q_{eh} \cdot P_t^{AWE}}{h_{rz}} \\ H_t^{PEM} = \eta_{PEM} \frac{q_{eh} \cdot P_t^{PEM}}{h_{rz}} \\ G_t^{MR} = \eta_{MR} \frac{h_{rz} \cdot H_t^{MR}}{g_{rz}} \end{cases} \tag{2}$$

These equations model the core energy conversion process of P2G. η_{MR}: electrolyzer and methane reactor efficiencies. These are core parameters determining the economic viability of the electro-hydrogen–gas conversion pathway. H_t^{MR}, G_t^{MR}: hydrogen production rate and gas production rate. These decision variables serve as key control mechanisms for integrating renewable energy and producing green gas.

H_t^{AWE} and H_t^{PEM} are the hydrogen production volume of the AWE and PEM electrolyzers at t moment; P_t^{AWE} and P_t^{PEM} are the electrical power consumed by the AWE and PEM electrolyzers at the t moment; η_{AWE} and η_{PEM} are the energy conversion efficiency for ALK and PEM electrolyzers; q_{eh} is the heat energy that can be converted by a unit of electrical energy; h_{rz} is the calorific value of hydrogen per unit volume; g_{rz} is the calorific value per unit volume of natural gas.

Finally, methane reactors need to consume carbon dioxide in the process of synthesizing methane. Considering that the amount of CO_2 supplied by CCS may not be fully utilized, the excess is stored. At the same time, the amount of CO_2 consumed by the H2G process at some point may not be fully supplied, so CO_2 needs to be purchased. The CO_2 required for H2G is

$$\begin{cases} M_t^{MR} = \frac{q_{eh} \cdot \rho_{CO_2}}{1000 \cdot g_{rz}} G_t^{MR} \\ M_t^{CCS} = M_t^{H2G} + M_t^{CS} \\ M_t^{MR} = M_t^{\text{buy}} + M_t^{CCS} \end{cases} \tag{3}$$

where M_t^{MR} is the amount of CO_2 required for the H2G process at t moment; M_t^{CS} is the amount of CO_2 stored at the time t; g_{rz} is the calorific value of natural gas per unit volume; ρ_{CO_2} is the density of CO_2; M_t^{buy} and M_t^{CCS} are the amount of CO_2 purchased by the H2G process and the amount of CO_2 provided by CCS, respectively.

3.2. Cogeneration System Modeling

(1) SOFC-GT cogeneration

The system mathematical model is as follows:

$$\begin{cases} S_t^{GT} = P_t^{GT} + Q_t^{GT} \\ P_t^{GT} = \eta^{GT,e} G_t^{GT} \\ Q_t^{GT} = \eta^{GT,h} G_t^{GT} \\ S_t^{HFC} = P_t^{HFC} + Q_t^{HFC} \\ P_t^{HFC} = \eta^{HFC,e} H_t^{HFC} \\ Q_t^{HFC} = \eta^{HFC,h} H_t^{HFC} \\ Q_t^{GB} = \eta^{GB} G_t^{GB} \end{cases} \tag{4}$$

where S_t^{GT} is the total output of the gas turbine at the time of t; P_t^{GT} is the power supply to gas turbines at t time; Q_t^{GT} is the heating power for gas turbines at t time; $\eta^{GT,e}$ and $\eta^{GT,h}$ are the power and heating efficiency for gas turbines; P_t^{HFC} is the power supply to

hydrogen fuel cells; Q_t^{HFC} is the power for heating hydrogen fuel cells; S_t^{HFC} is the total output of hydrogen fuel cells; $\eta^{HFC,e}$ and $\eta^{HFC,h}$ are the electrical and thermal efficiency of hydrogen fuel cells; H_t^{HFC} is the hydrogen power consumed by the hydrogen fuel cell t at the moment; Q_t^{GB} is the heat generation power for the boiler at t moment; η^{GB} is the heat conversion efficiency of gas boilers; G_t^{GB} is the natural gas power consumed by the gas boiler at the t moment.

(2) Waste heat utilization

The heat energy of waste heat is supplied to the heat load through the waste heat boiler, and the excess part is generated through the ORC waste heat power generation device. The mathematical model of the waste heat utilization system is as follows:

$$\begin{cases} Q_t^{sum} = Q_t^{HFC} + Q_t^{GT} \\ \begin{cases} 0 \le \varphi_t^{ORC}, \varphi_t^{WHB} \le 1 \\ \varphi_t^{ORC} + \varphi_t^{WHB} = 1 \end{cases} \\ P_t^{ORC} = \eta^{ORC} \varphi_t^{ORC} Q_t^{sum} \\ Q_t^{WHB} = \eta^{WHB} \varphi_t^{WHB} Q_t^{sum} \end{cases} \tag{5}$$

where Q_t^{sum} is the heat energy collected by the waste heat bus of the system at time t; φ_t^{ORC}, φ_t^{WHB} are the proportion of waste heat utilized by the ORC waste heat power generation device and waste heat boiler input at t moment, respectively. P_t^{ORC} is the power generation of the ORC waste heat power generation device at t moment; η^{ORC} the power generation efficiency of the ORC waste heat power generation device; Q_t^{WHB} is the heat output of the waste heat boiler at the time of t; η^{WHB} is the efficiency of the waste heat boiler.

3.3. Carbon Trading Model

In this paper, the carbon quota is allocated free of charge, and the total carbon quota of the system is composed of the carbon quota of the power purchase unit, the carbon quota of the gas purchase unit, and the incentive quota of the green power unit:

$$\begin{cases} M_t^{quota} = M_t^{grid} + M_t^{gas} + M_t^{DG} \\ M_t^{grid} = \sum_{t=1}^{T} \theta_e P_t^{grid} \\ M_t^{gas} = \sum_{t=1}^{T} \theta_g \left(\chi P_t^{GT} + Q_t^{GT} + Q_t^{GB} \right) \\ M_t^{DG} = \sum_{t=1}^{T} \theta_{\mathrm{r}} (P_t^{PV} + P_t^{WD}) \end{cases} \tag{6}$$

where M_t^{quota} is the total carbon emission quota; M_t^{grid}, M_t^{gas}, and M_t^{DG}, respectively, are the carbon emission quotas for purchasing electricity from the power grid, purchasing gas from the higher-level gas grid, and the carbon quotas for green power incentives; θ_e, θ_g, and θ_{r} are $t$he carbon emission quota per unit of electricity and thermal power and the carbon quota of green power unit incentives, respectively. P_t^{grid} is the thermoelectric power conversion coefficient; P_t^{PV} and P_t^{WD}, respectively, represent the power generation of photovoltaics and wind turbines at time t.

Actual system carbon emissions consist of carbon emissions from equipment that interacts with the higher grid, purchases natural gas, and is a primary energy source.

$$\begin{cases} M_t^Z = M_t^{IN} + M_t^E + M_t^G \\ M_t^{IN} = M_t^{GT} + M_t^{GB} - M_t^{CS} - M_t^{MR} \\ M_t^E = \varepsilon_E \sum_{t=1}^{k} P_t^{grid} \\ M_t^G = \varepsilon_G \sum_{t=1}^{k} G_t^{Load} \end{cases} \tag{7}$$

where M_t^{IN} is the carbon emissions generated within the system at time t; M_t^E is carbon emissions generated by interaction with the superior power grid at the time of t; M_t^G is the carbon emission of the gas load; G_t^{Load} is the natural gas load at the time of t; ε_G is the carbon emission coefficient per unit of electricity in the regional power grid where the system is located; G_t^{buy} is the carbon emission coefficient of natural gas.

The net carbon emissions of the system are the difference between actual carbon emissions and carbon allowances and can be expressed as follows:

$$M_t^{net} = M_t^Z - M_t^{quota} \tag{8}$$

This carbon trading mechanism internalizes the cost of emissions into the economic objective. The actual emissions M_t^Z are calculated from interactions with the grid and gas consumption. The net emissions M_t^{net} directly translate into a cost (if positive) or revenue (if negative) through the carbon price $k_{CO_2}^{buy}$, creating a financial incentive for low-carbon operation.

4. Dynamic Optimization Scheduling Based on Deep Reinforcement Learning

The above hydrogen-coupled electrothermal integrated energy system model is transformed into a deep reinforcement learning model. This paper uses a deep reinforcement learning algorithm to solve the optimization decision-making problem with uncertainty factors, focusing on the dynamic optimization scheduling under the intermittent and random fluctuation of the user-side load of renewable energy generation in the integrated energy system.

4.1. Flexible Movement Evaluation Deep Reinforcement Learning

In the research field of integrated energy system optimization and regulation, this paper introduces the SAC algorithm to construct the regulation model of HCEH-IES and adopts the rolling optimization strategy to formulate the dynamic operation regulation scheme.

SAC is an offline learning algorithm, and its core innovation is to integrate the principle of maximum entropy into the strategy learning framework. The intelligences are motivated to not only maximize the long-term cumulative rewards but also to maintain the diversity of action choices during the learning process. The actual environment is often dynamically changing and full of uncertainty, and traditional algorithms may perform poorly when the environment changes because of over-reliance on specific optimal actions. The SAC algorithm, on the other hand, can better cope with these changes due to the diversity of its actions and can quickly adjust its strategy to maintain a better performance even when the environmental conditions change unexpectedly.

The discounted cumulative reward function J and the objective function π^* for SAC at any time slot t are

$$J = \sum_{t=0}^{T} E_{(s_t,a_t)\sim\rho_\pi}[r(s_t,a_t) + \alpha H(\pi_\phi(\cdot|s_t))] \tag{9}$$

$$\pi^* = \underset{\pi_\varphi}{\operatorname{argmax}} \sum_{t=0}^{T} E_{(s_t,a_t)\sim\rho_\pi}[r(s_t,a_t) + \alpha H(\pi_\varphi(\cdot|s_t))] \tag{10}$$

where π is the agent action strategy, which is essentially the probability distribution of the agent's action choice; s_t and a_t are $r(s_t,a_t)$, the current environmental state of the agent, the action output by the policy, and the reward value of the environment feedback to the agent, respectively; $(s_t, a_t)\sim\rho_\pi$ is the temperature coefficient of action entropy, where $H(\pi_\phi(\cdot|s_t))$ is used to characterize the influence of action entropy on reward; ϕ is for the strategy action trajectory; it is the action entropy of the policy in the state, which is the network parameter representing the policy.

Action entropy is used to characterize the uncertainty of the policy π with respect to action selection. SAC maximizes the action entropy by introducing action entropy so that the actions output by its policy π during iterative training are as dispersed as possible, which allows the intelligent to consider more choice behaviors without omitting any potentially useful action choices. The action entropy is defined as

$$H(\pi_\phi(\cdot|s_t)) = E_{(s_t,a_t)\sim\rho_\pi}[-\ln \pi_\phi(\pi_\phi(a_t|s_t))] \tag{11}$$

As an Actor–Critic algorithm, the Actor network of SAC is responsible for modeling the action policy π_ϕ, and the Critic network evaluates the policy obtained by the Actor network using the value function $Q_\psi(a_t|s_t)$. The Q-value function and the state-value function of SAC are defined as follows, respectively:

$$Q_\psi(a_t|s_t) = \gamma E_{(s_t,a_t)\sim\rho_\pi}[V(s_{t+1})] + r(s_t,a_t) \triangleq \Gamma^\pi Q_\psi(a_t|s_t) \tag{12}$$

$$V(s_{t+1}) = E_{(s_t,a_t)\sim\rho_\pi}[Q_\psi(a_t|s_t) - \alpha \ln \pi_\phi(a_t|s_t) \tag{13}$$

where γ is the reward discount factor; $E_{(s_t,a_t)\sim\rho_\pi}[V(s_{t+1})]$ represents the sum of the expectations of all states ρ_π under trajectory s_{t+1}; Γ^π is the Bellman operator of strategy π.

SAC performs gradient backward updating of the parameters ϕ and ψ of the Actor and Critic networks, respectively, to obtain the optimal operation policy and Q-value function.

$$J_Q(\psi) = E_{(s_t,a_t)\sim D}[\frac{1}{2}Q_\psi(s_t,a_t) - ((r_t(s_t,a_t) + \gamma E_{s_{t+1}\sim\rho}[V_{\overline{\psi}}(s_{t+1})]))^2] \tag{14}$$

$$J_\pi(\psi) = E_{s_t\sim D}\left(D_{KL}\left([\pi_\phi(\cdot|s_t)]\middle\|\frac{\exp(\frac{1}{\alpha})(s_t\cdot)}{Z_\psi(s_t)}\right)\right) \tag{15}$$

where $V_{\overline{\psi}}(s_{t+1})$, the new state value function, is updated for the Critic network; $Z_\psi(s_t)$ is an allocation function for normalization; D is the updated sample set; $D_{KL}[||]$ represents the Kullback–Leibler (KL) divergence calculation to characterize the distance between two distributions.

4.2. State Space

The observed state of the system includes the electricity/gas/heat load demand, PV/wind turbine power, external energy price, and the charge state of the energy storage

equipment at the moment t − 1. For the low-carbon economic dispatch of the integrated energy system, the state can be expressed as follows:

$$s_t = \left(P_t^{Load}, G_t^{Load}, Q_t^{Load}, P_t^{PV}, P_t^{WD}, k_t^E, k_t^G, SOC_{t-1}^{ES}, SOC_{t-1}^{TST}, SOC_{t-1}^{HST}\right) \tag{16}$$

4.3. Action Space

The goal of low-carbon economic dispatch for integrated energy systems is to determine the optimal unit output profile. The actions in the low-carbon economic dispatch of the integrated energy system can be expressed as follows:

$$a_t = \left(P_t^{ES}, S_t^{GT}, S_t^{HFC}, P_t^{ES}, P_t^{TST}, P_t^{HST}\right) \tag{17}$$

where P_t^{ES}, P_t^{TST}, and P_t^{HST} are the outputs of electric energy storage, heat storage tank, and hydrogen storage tank t at the moment.

The SAC action space constraints are

$$\begin{cases} SOC_t^{ES} = (1-\mu_{ES})SOC_{t-1}^{ES} + (P_t^{ES,ch}\eta^{ES,ch} - P_t^{ES,dis}/\eta^{ES,dis})/E^{ES} \\ SOC^{ES,\min} \le SOC_t^{ES} \le SOC^{ES,\max} \\ 0 \le P_t^{ES,ch} \le f_t^{ES,ch} P^{ES,\max} \\ 0 \le P_t^{ES,dis} \le f_t^{ES,dis} P^{ES,\max} \\ f_t^{ES,ch} + f_t^{ES,dis} = 1 \\ P_t^{ES} = f_t^{ES,ch} P_t^{ES,ch} + f_t^{ES,dis} P_t^{ES,dis} \end{cases} \tag{18}$$

where SOC_t^{ES} is the state of charge of the power storage equipment at time t; μ_{ES} is the electrical loss coefficient of the power storage equipment; $P_t^{ES,ch}$ and $P_t^{ES,dis}$ are the charging and discharging power of the storage equipment, respectively; $\eta^{ES,ch}$ and $\eta^{ES,dis}$ charge and discharge efficiency, respectively; E^{ES} is the capacity of the power storage equipment; $SOC^{ES,\max}$ and $SOC^{ES,\min}$ are the upper and lower limits of the charging state of the power storage equipment, respectively; $P^{ES,\max}$ the maximum charging and discharging power of the power storage equipment; $f_t^{ES,ch}$ and $f_t^{ES,dis}$ are the charging and discharging state of the device at the time of t; P_t^{ES} is the net output of electric energy storage in the t period.

4.4. Reward Functions

Intelligent bodies maximize their cumulative returns through continuous learning during the scheduling cycle, and the setting of the reward function is generally related to the objective function of the system. The intelligent body reward mechanism designed in this paper includes the objective function F and the action-constrained penalty Fc, in which the action-constrained penalty Fc mainly includes the penalty for power overruns in interactions with the power-gas network, the penalty for overrunning the output of the unit, the penalty for overrunning the rate of change of output, and the penalty for overcharge and overdischarge of the energy storage equipment. The reward function guides the intelligent body to take actions to minimize the objective function while executing the constraints. This is accomplished by using the next reward function:

$$r(s_t, a_t) = -[\omega_1 F(s_t, a_t) + \omega_2 F_c(s_t, a_t)] \tag{19}$$

where ω_1 is the scaling factor of cost control; ω_2 is the scaling factor for the execution of the constraint penalty.

4.4.1. Objective Function

The total operating cost F of the system consists of two parts: the integrated operating cost F1 and the carbon control cost F2.

$$F = \min(F1 + F2) \tag{20}$$

(1) Comprehensive Running Costs

Comprehensive operating costs include the cost of higher-level grid interactions, the cost of gas purchased from the natural gas grid, and the operation and maintenance costs of each piece of equipment.

$$F1 = \min(C^{grid} + C^{gas} + C_t^{ma}) \tag{21}$$

$$\begin{cases} C^{grid} = \sum_{t=1}^{T} k_t^E P_t^{grid} \\ C^{gas} = \sum_{t=1}^{T} k_t^G G_t^{buy} \\ C^{ma} = \sum_{\mathrm{t}=1}^{T} \sum_{m=1}^{M} \mathrm{k}_m P_{\mathrm{t}}^m \end{cases} \tag{22}$$

where C^{grid} is the power purchase cost of the superior power grid; C^{gas} is the cost of purchasing gas for the gas network; k_t^E and k_t^G are the time-of-use electricity price of the t-period natural gas price; P_t^{grid} is the interaction power with the superior power grid at the time of t; G_t^{buy} is to purchase gas volume from the superior; C^{ma} is the cost of operation and maintenance of system equipment; k_m is the unit maintenance cost of the M type of equipment; P_{t}^m is the input or output power of the m type of device.

(2) Cost of Carbon Control

Carbon control costs mainly include carbon purchase costs incurred in the methanization process, carbon capture, storage, utilization equipment operation and maintenance costs, and carbon trading costs.

$$F2 = \min(C^{buy} + C^{ma} + C^{tra}) \tag{23}$$

$$\begin{cases} C_{CO_2}^{buy} = \sum_{t=1}^{T} k_{CO_2}^{buy} M_t^{buy} \\ C_{ccus}^{ma} = \sum_{\mathrm{t}=1}^{T} \sum_{\mathrm{n}=1}^{N} \mathrm{k_n} P_{\mathrm{t}}^{\mathrm{n}} \\ C^{tra} = \sum_{t=1}^{T} k_{CO_2}^{tra} M_t^{net} \end{cases} \tag{24}$$

where $C_{CO_2}^{buy}$ is the cost of carbon purchase; $k_{CO_2}^{buy}$ is the price of CO_2 per unit; M_t^{buy} is the purchase CO_2 volume for t moments; C_{ccus}^{ma} is the cost of operation and maintenance of carbon capture and utilization equipment; $\mathrm{k_n}$ is the unit price of natural gas; G_t^{buy} is the purchase gas power for t moment; C^{ma} is the cost of operation and maintenance of system equipment; $\mathrm{k_n}$ is the unit maintenance cost of the nth type of equipment; $P_{\mathrm{t}}^{\mathrm{n}}$ is the input or output power of the equipment for the nth type of carbon capture; C^{tra} is the carbon trading costs; $k_{CO_2}^{tra}$ is the carbon trading price; M_t^{net} is a net carbon emission for the system.

4.4.2. Constraints

(1) Power Balance Constraints

In order to satisfy the electric–hydrogen energy demand for each time period in the system, the power balance constraints as well as the external energy supply constraints in the system are shown as follows:

$$\begin{cases} P_t^{PV} + P_t^{WD} + P_t^{GT} + P_t^{HFC} + P_t^{ORC} + P_t^{ES} + \\ P_t^{grid} = P_t^{AWE} + P_t^{PEM} + P_t^{CCS} + P_t^{Load} \\ Q_t^{WHB} + Q_t^{GB} + Q_t^{TST} = Q_t^{Load} \\ H_t^{AWE} + H_t^{PEM} + H_t^{HST} = H_t^{MR} + H_t^{HFC} \\ G_t^{buy} + G_t^{MR} = G_t^{GT} + G_t^{GB} + G_t^{Load} \end{cases} \tag{25}$$

where P_t^{PV} is the output of the photovoltaic power station in the t period; P_t^{WD} is the output for the fan in the t period; P_t^{Load} is the electrical load of the t period; Q_t^{TST} is the net output for thermal energy storage in the t period; Q_t^{Load} is the heat load in the t period; G_t^{buy} is the purchasing natural gas for the t period; G_t^{Load} is the gas load in the t period; H_t^{HST} is the net output of the hydrogen storage tank in the t period.

(2) Interactive power constraints with higher-level electrical networks

$$\begin{cases} P^{grid,\min} \le P_t^{grid} \le P^{grid,\max} \\ G^{buy,\min} \le G_t^{buy} \le G^{buy,\max} \end{cases} \tag{26}$$

where $P^{grid,\max}$ and $P^{grid,\min}$ are the upper and lower limits of the interactive power between the system and the main power grid, respectively; $G^{buy,\max}$ and $G^{buy,\min}$ are the upper and lower limits of the interaction power between the system and the main gas network, respectively.

(3) Equipment output constraints

$$\begin{cases} P^{AWE,\min} \le P_t^{AWE} \le P^{AWE,\max} \\ P^{PEM,\min} \le P_t^{PEM} \le P^{PEM,\max} \\ G^{MR,\min} \le G_t^{MR} \le G^{MR,\max} \\ S^{GT,\min} \le S_t^{GT} \le S^{GT,\max} \\ S^{HFC,\min} \le S_t^{HFC} \le S^{HFC,\max} \\ Q^{GB,\min} \le Q_t^{GB} \le Q^{GB,\max} \end{cases} \tag{27}$$

where $P^{AWE,\max}$ and $P^{AWE,\min}$ are the upper and lower limits of the power consumed by electrolysis; $P^{PEM,\max}$ and $P^{PEM,\min}$ re the upper and lower limits of the electrical power consumed by electrolysis; $G^{MR,\max}$ and $G^{MR,\min}$ are the upper and lower limits of gas production power of methane reactors, respectively; $S^{GT,\max}$ and $S^{GT,\min}$ are the upper and lower limits of the total output of gas turbines; $S^{HFC,\max}$ and $S^{HFC,\min}$ are the upper and lower limits of the total output of hydrogen fuel cells; $Q^{GB,\max}$ and $Q^{GB,\min}$ are the upper and lower limits of the thermal power of gas boilers.

(4) Equipment output climbing constraints

$$\begin{cases} \Delta P^{EL,\min} \le P_{t+1}^{EL} - P_t^{EL} \le \Delta P^{EL,\max} \\ \Delta G^{MR,\min} \le G_{t+1}^{MR,CH_4} - G_t^{MR,CH_4} \le \Delta G^{MR,\max} \\ \Delta S^{GT,\min} \le S_{t+1}^{GT} - S_t^{GT} \le \Delta S^{GT,\max} \\ \Delta S^{HFC,\min} \le S_{t+1}^{HFC} - S_t^{HFC} \le \Delta S^{HFC,\max} \end{cases} \tag{28}$$

where $\Delta P^{EL,\max}$ and $\Delta P^{EL,\min}$ are the upper and lower limits of the electrolyzer power ramp, $\Delta G^{MR,\max}$ and $\Delta G^{MR,\min}$ are the upper and lower limits of methane reactor gas production power; $\Delta S^{GT,\max}$ and $\Delta S^{GT,\min}$ are the upper and lower limits of the total power of gas turbines; $\Delta S^{HFC,\max}$ and $\Delta S^{HFC,\min}$ are the upper and lower limits of the total output of hydrogen fuel cells.

5. Examples

5.1. Example Description

Simulation analysis is performed for the HCEH-IES built in Figure 1. This paper verifies the ability of reinforcement learning for offline training and online optimization of the model in this paper and conducts a comparative analysis of the three designed scenarios, as well as compares the ability of different optimization algorithms to solve the model. Under the premise of giving priority to meeting load demand, making full use of renewable energy sources, choosing appropriate optimal scheduling strategies, reducing the comprehensive operation cost and carbon control cost of the system, and finally planning the output of each unit. The parameters of each unit within HCEH-IES are shown in Appendix A.

5.2. Training Convergence Analysis

A total of 4000 cycles are trained in this simulation. In offline training, the SAC algorithm has the highest reward function value and the fastest convergence speed. The SAC reward curve gradually stabilizes in 3000 training cycles and converges to the reward value interval of -3.2×10^7–3.22×10^7, while the DDPG algorithm stabilizes only in 3500 training cycles, and the training results are poor. For detailed specifications, see Appendix B.

5.3. Analysis of Scheduling Results

After training the algorithmic network using historical data, the resulting network is saved and applied to the dynamic economic scheduling of the system. The results of scheduling actions are shown in Figures 2–5.

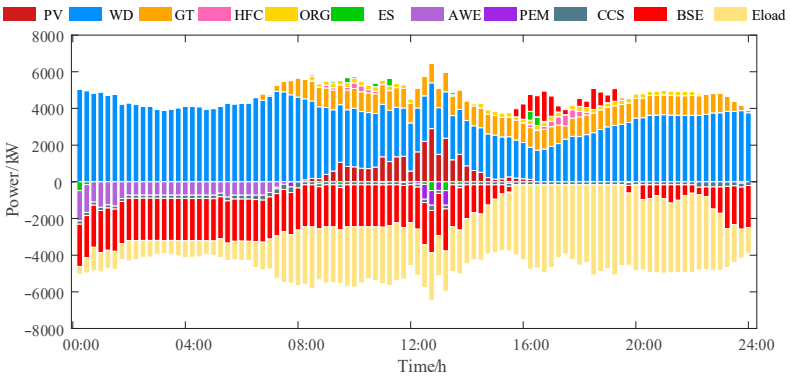

Figure 2. Results of optimal scheduling of electrical loads.

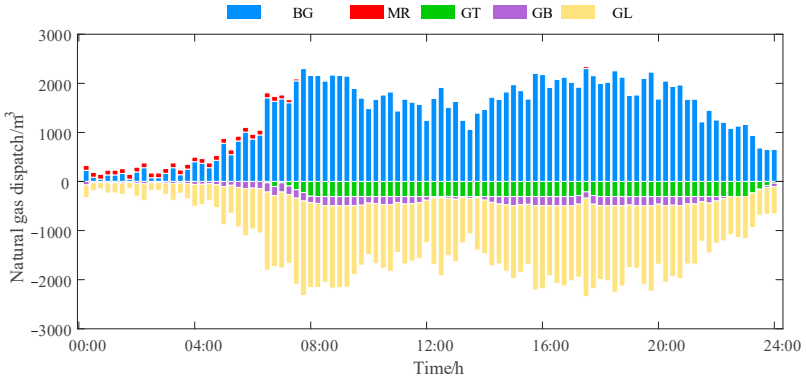

Figure 3. Results of gas load optimization scheduling.

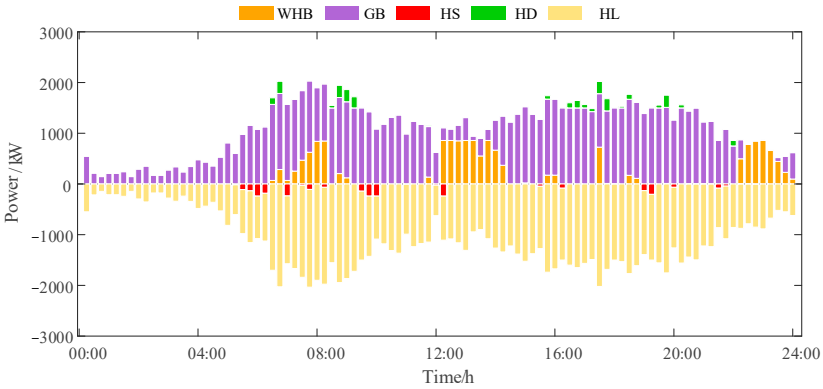

Figure 4. Heat load optimization scheduling results.

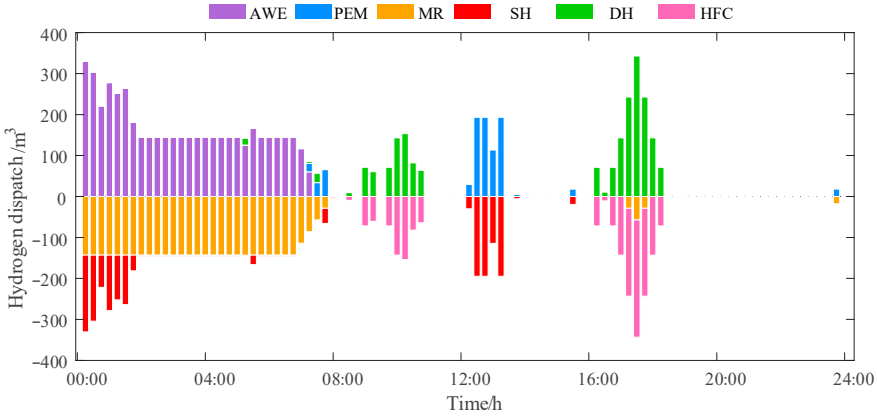

Figure 5. Hydrogen load optimization scheduling results.

As shown in Figure 2, power dispatch exhibits characteristics of multi-timescale collaborative optimization. During 00:00–04:00, there is PV plant shutdown, wind turbine as the main power supply of renewable energy, gas turbine low power operation to fill the gap between wind power and electric load, and hot standby maintained. Due to wind power overcapacity, the system sells power to the grid. During 04:00–06:00, wind turbine power drops and gas turbine power increases to ensure the stability of the electric load. From 06:00 onwards, the increase in light makes the power of photovoltaic power generation rise, and the system continues to optimize the dispatch to maintain electric balance. After 6:00, solar power generation increases but experiences a timing mismatch with peak electricity demand. At this time, the system coordinates gas turbines, hydrogen fuel cells, and low-temperature waste heat power generation units to form a diversified, complementary power supply structure. Notably, during the 18:00 to 20:00 peak load

period, the system achieved peak shaving and valley filling by preemptively discharging stored energy, adjusting P2G operation strategies, and promptly activating hydrogen fuel cells. This demonstrates that the SAC algorithm has mastered forward-looking decision-making capabilities in power dispatch, effectively mitigating renewable energy fluctuations through multi-energy flow conversion.

The thermal load scheduling in Figure 3 clearly demonstrates the system's full utilization of thermal inertia. From 00:00 to 04:00, during low-load periods, the system maintains baseline heating solely through fluctuating operation of gas boilers, while thermal storage units perform intermittent heat storage. This "valley-filling" operation reserves capacity for subsequent adjustments. From 06:00 onwards, as thermal load increases, the system synchronously boosts output from both gas and waste heat boilers while activating thermal storage units for heat release, establishing a "storage-supply" coordination mode. During the high-load period from 08:00 to 20:00, the gas boiler, waste heat boiler, and thermal storage units operate in a coordinated state. Through multi-heat-source control, this achieves a balance between heating reliability and economic efficiency. This optimized scheduling, based on the thermal system's spatiotemporal characteristics, demonstrates the unique advantages of integrated energy systems in thermal energy management.

Figure 4 illustrates the pivotal role of the gas network in multi-energy conversion. During 00:00–04:00, due to the low heat demand for production and life, the gas boiler maintains the basic heat supply in a fluctuating mode, the power of the waste heat boiler is small, and the heat storage device intermittently carries out heat storage with small power. From 06:00 onwards, the heat load starts to rise, the power of the gas boiler increases significantly, the power of the waste heat boiler increases synchronously, and the heat storage tank participates in heat storage during part of the time period. During 08:00–20:00, accompanied by a continuous growth or fluctuation of the heat from 08:00 to 20:00, along with the continuous growth or fluctuation of heat load, gas boilers, waste heat boilers, and heat storage tanks operate synergistically to increase the heat supply. The methane reactor increased its output around 20:00. This strategic arrangement responded to the anticipated growth in gas load while leveraging its time-varying operational characteristics to participate in system regulation. Throughout the entire scheduling process, the natural gas system ensured gas volume balance and operational safety for the multi-energy system through the dual safeguards of "purchased gas + P2G gas production".

Figure 5 illustrates the core value of hydrogen energy dispatch in energy conversion. Between 00:00 and 04:00, alkaline electrolyzers maximize low-cost wind power for large-scale hydrogen production while proton exchange membrane electrolyzers remain on standby. This differentiated operation demonstrates the system's precise control over hydrogen production economics. Simultaneously, methane reactors continuously consume hydrogen to synthesize natural gas, while hydrogen storage tanks handle surplus storage, forming a complete "production-storage-consumption" hydrogen management chain. Between 14:00 and 16:00, PEM electrolyzers significantly increase output—aligned with their rapid response characteristics—to smooth power fluctuations during this period. During the peak hydrogen consumption period from 18:00 to 20:00, the system achieved precise matching of hydrogen supply and demand by coordinating electrolyzer load reduction, hydrogen fuel cell power generation, and hydrogen tank release. This multi-timescale hydrogen management strategy fully demonstrates hydrogen's critical role in enhancing system flexibility and facilitating renewable energy integration.

5.4. Comparison of Methods

(1) Model Comparison

The following scenarios are set up to verify the superiority of this paper's model:

Scenario 1: Introducing gas-fired units and a carbon trading mechanism, without adding hydrogen fuel cells to form a cogeneration system, and without carbon capture utilization technology.

Scenario 2: Based on scenario 1, carbon capture technology is added, and electrolytic hydrogen is converted to natural gas for utilization.

Scenario 3: Based on scenario 2, two-stage P2G is used, and hydrogen fuel cells and ORC low-temperature power generators are introduced to form a cogeneration system.

From Table 1, it can be seen that scenario 1 only deploys gas units, does not build a cogeneration system, and lacks technical means such as electricity-to-gas, resulting in significantly high gas grid interaction costs. Scenario 2 integrates carbon capture technology on the basis of scenario 1 and directly converts electrolyzed hydrogen into natural gas for utilization, which effectively reduces natural gas procurement expenditure by supplementing gas sources, reducing the gas grid interaction cost to 108,204.95 CNY. However, due to the constraints of the carbon trading mechanism, the additional carbon purchase cost was 1576.83 CNY. Scenario 3 uses two-stage power-to-gas (P2G) technology to introduce hydrogen fuel cells and ORC cryogenic power generation devices to build a cogeneration system, although the complexity of the system increases and changes the natural gas demand structure, resulting in the gas grid interaction cost rising to 113,808.10 CNY, however, its grid interaction cost is −59,777.26 CNY, achieving significant benefits, which is attributed to the fact that the cogeneration system improves energy utilization efficiency and grid interaction benefits through energy allocation optimization and interactive collaboration.

Table 1. Optimize the comparison of scheduling results.

Scenario	Power Grid Interaction Cost/CNY	Gas Network Interaction Cost/CNY	Equipment Operation and Maintenance Cost/CNY	Carbon Purchase Cost/CNY	Carbon Capture and Storage Cost/CNY	Carbon Trading Cost/CNY	Total Cost/CNY
1	−49,686.20	110,188.20	17,895.48	0	0	21,339.10	99,990.24
2	−44,139.57	108,204.95	19,729.02	1576.83	866.22	18,116.87	104,511
3	−59,777.26	113,808.10	21,816.08	1246.84	836.31	17,551.39	95,601.14

Carbon control cost analysis: Scenario 1 does not involve carbon capture technology, so both carbon purchase costs and carbon capture and storage costs are zero. Scenario 2 introduces carbon capture technology, generating carbon purchase costs of 1576.83 CNY, carbon capture and storage costs of 866.22 CNY, and carbon trading costs of 18,116.87 CNY, initiating carbon emission control through technological and market-based measures. Scenario 3 continues relevant mechanisms and technologies, with carbon purchase costs of CNY 1246.84, carbon capture and storage costs of CNY 836.31, and carbon trading costs of CNY 17,551.39. Due to system optimization, some costs have been adjusted, but overall, carbon emission control and management continue.

(2) Algorithm comparison

To comprehensively evaluate the effectiveness of the proposed optimization scheduling strategy, this study employs both mathematical programming and deep reinforcement learning (DRL) methods for comparative analysis. Specifically, the commercial solver CPLEX (IBM ILOG CPLEX Optimization Studio) is introduced to solve the deterministic day-ahead optimization problem of the HCEH-IES model, providing a theoretical optimal solution as a benchmark under idealized forecasting conditions. Meanwhile, two representative DRL algorithms—deep Q-network (DQN) and deep deterministic policy gradient (DDPG)—are also optimized and applied to solve the same model.

From Table 2, the computational costs for the SAC, DDPG, and DQN algorithms are CNY 95,601.14, CNY 97,629.78, and CNY 99,287.60, respectively. Calculations show that

the operational cost increases relative to CPLEX for the SAC, DDPG, and DQN algorithms are approximately 2.76%, 4.94%, and 6.72%, respectively. Among these, the SAC algorithm exhibits a lower total operational cost than both DDPG and DQN, and its results are closer to those of CPLEX's current optimal scheduling method. It is important to emphasize that CPLEX achieves theoretical optimal solutions under ideal conditions where all source-load data is fully known. In contrast, the SAC algorithm develops scheduling strategies adaptable to uncertainty through interactive learning in dynamic environments. CPLEX's marginally superior economic performance in deterministic settings validates its theoretical advantage while highlighting SAC's effectiveness and robustness in scenarios closer to real-world operations. Compared to DDPG and DQN, SAC's operational costs were CNY 75,846.92, CNY 76,957.37, and CNY 77,633.45, respectively, demonstrating cost advantages that indicate greater efficiency in resource utilization or computational logic. Regarding carbon control costs, SAC incurred CNY 19,493.30, lower than DDPG's CNY 20,672.41 and DQN's CNY 21,654.15, demonstrating greater effectiveness in carbon emission control strategies.

Table 2. Comparative analysis results of different methods.

Algorithm	Running Cost/CNY	Carbon Control Cost/CNY	Total Cost/CNY
CPLEX	74,182.24	18,856.20	93,038.44
SAC	75,846.92	19,493.30	95,601.14
DDPG	76,957.37	20,672.41	97,629.78
DQN	77,633.45	21,654.15	99,287.60

This study adopts the SAC algorithm primarily based on the following considerations: First, this algorithm can adaptively learn the random fluctuation characteristics of wind and solar power generation and multi-energy loads through interaction with the environment without relying on precise predictive models; second, it possesses online learning and real-time adjustment capabilities, making it better suited for dynamic scheduling scenarios.

6. Conclusions

In this paper, an optimization method for the scheduling and operation of a hydrogen-coupled electrothermal integrated energy system is proposed. The source-side structure of the system is optimized by integrating carbon trading policies with low-carbon technology to improve the renewable energy consumption rate and system decarbonization level. Furthermore, to address the uncertainties in system source-load and the insufficient exploration in existing reinforcement learning algorithms, a deep reinforcement learning method based on Soft Actor–Critic (SAC) is proposed. The adaptive learning control strategy is obtained through interactions between agents and the energy system. The following conclusions are drawn:

(1) The proposed HCEH-IES framework and its optimization methodology, which synergizes carbon trading mechanisms with low-carbon technologies like P2G-CCS, increased the renewable energy consumption rate to over 85%. This architecture effectively matches the energy consumption of carbon capture and electrolytic hydrogen production with renewable generation profiles, resulting in a 12.7% reduction in total carbon emissions in scenario 3 compared to scenario 1, empirically demonstrating its significant effectiveness in enhancing the system's carbon reduction capability.

(2) The hybrid hydrogen production system, comprising AWE and PEM electrolyzers, operated in a complementary manner to meet hydrogen demand while effectively utilizing low-cost wind power and surplus PV generation, contributing approximately 15% to the system's flexibility regulation potential. The diversified utilization of hydrogen through

power generation in fuel cells, methanation, and direct storage fully unlocks its potential as a cross-seasonal storage medium and a coupling hub for multi-energy flows, proving crucial for the system's low-carbon and economic operation.

(3) Based on the deep reinforcement learning method of soft SAC, the adaptive optimization of control strategies is realized through the interaction learning between agents and energy systems. Compared with traditional reinforcement learning algorithms, this method can reduce the total cost of HCEH-IES and effectively improve the low-carbon and economic efficiency of the system.

Author Contributions: Data curation, X.L.; writing—original draft preparation, Y.Z.; writing—review and editing, D.F.; supervision, X.Y. and J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the National Natural Science Foundation of China (52567011); the Central Government Guidance Fund for Local Development (25ZYJA014); the Gansu Provincial Major Science and Technology Special Project (22ZD6GA063); and the Dunhuang Science and Technology Support Project (200501, 200502).

Data Availability Statement: Data cannot be shared publicly due to confidentiality agreements with the participants. Data are available upon reasonable request from the corresponding author (fanduojin@lzdctc.com) for researchers who meet the criteria for access to confidential data.

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

Abbreviations

The following abbreviations are used in this manuscript:

HCEH-IES	Hydrogen-Coupled Electrothermal Integrated Energy System
SAC	Soft Actor–Critic
P2G	Power to Gas
CCS	Carbon Capture and Storage
GT	Gas Turbine
GB	Gas-Fired Boiler
DDPG	Deep Deterministic Policy Gradient
SOFC	Solid Oxide Fuel Cell
DQN	Deep Q-Network

Appendix A

Equipment Parameters

Equipment	Parameter	Value
AEC	Maximum/Minimum Power Consumption/kW	1650/300
	Maximum/Minimum Power Consumption/kW	0.7
	Power Ramp-up Coefficient	0.25
	Maintenance Cost/CNY/kWh	0.022
PEM	Maximum/Minimum Power Consumption/kW	1000/0
	Energy Conversion Efficiency	0.85
	Maintenance Cost/CNY/kWh	0.056

Equipment	Parameter	Value
Methane reactor	Maximum/Minimum Gas Production/m^3	100/0
	Energy Conversion Efficiency	0.7
	Power Ramp-up Coefficient	0.2
	Maintenance Cost/CNY/m^3	0.04
SOFC	Maximum/Minimum Total Power/kW	800/0
	Electrical Conversion Efficiency	0.6
	Thermal Conversion Efficiency	0.2
	Power Ramp-up Coefficient	0.25
	Maintenance Cost/CNY/kWh	0.04
GT	Maximum/Minimum Total Power Output/kW	2300/0
	Electrical Conversion Efficiency	0.35
	Thermal Conversion Efficiency	0.4
	Power Ramp-up Coefficient	0.25
	Maintenance Cost/CNY/kWh	0.04
GB	Maximum/Minimum Total Power Output/kW	1200/0
	Energy Conversion Efficiency	0.8
	Power Ramp-up Coefficient	0.25
	Maintenance Cost/CNY/kWh	0.04
Equipment	Parameter	Value
ORC	Energy Conversion Efficiency	0.2
	Maintenance Fee/CNY/kWh	0.04
WHB	Energy Conversion Efficiency	0.7
	Maintenance Cost/CNY/kWh	0.05
CCS	Flue gas diversion ratio	0.8
	Level of carbon capture	0.9
	Fixed Carbon Capture Energy Consumption/kWh	160
	Energy consumption per unit of carbon capture (kWh/t)	270
	Maintenance Cost/CNY/kWh	0.04
BES	Capacity/kWh	1500
	Maximum Input/Output (kW)	480
	Self-dispersion coefficient	0.005
	Energy Storage Efficiency	0.95
	Maximum/Minimum Charge State	0.9/0.2
	Maintenance Cost/CNY/kWh	0.026

Equipment	Parameter	Value
TES	Capacity/kWh	1500
	Maximum Input/Output (kW)	0.2
	Self-dispersive coefficient	0.01
	Energy Storage Efficiency	0.95
	Maximum/Minimum Thermal Load State	0.9/0.2
	Maintenance Cost/CNY/kWh	0.016
HES	Capacity/m^3	2000
	Maximum Input/Output per m^3	0.25
	Self-dispersion coefficient	0.006
	Energy Storage Efficiency	0.98
	Maximum/Minimum Hydrogen Loading State	0.9/0.2
	Maintenance Cost/CNY/kWh	0.032

Appendix B

Hyperparameters and other parameters

Parameter	Value
q_{eh}	3600 (kJ/kwh)
h_{rz}	12,586 (kJ/m^3)
g_{rz}	3600 (kJ/m^3)
ρ_g	0.71428 (kg/m^3)
M_g	16
M_{CO_2}	44
k^{CCS}	30 CNY/t
$k_{CO_2}^{tra}$	400 CNY/t
$\theta_e, \theta_g, \theta_r$	0.799, 0.324,0.5 (t/kwh)
$\varepsilon_E, \varepsilon_G$	0.867, 0.367 (t/kwh)
Reward Discount Factor	0.96
Actor Network Learning Rate	1×10^{-4}
Critic Network Learning Rate	3×10^{-4}
Q-value Network Learning Rate	4×10^{-4}
Soft Update Coefficient	4×10^{-4}
Number of Hidden Layers in Network	4
Number of Neurons in Hidden Layer	256
Experience Buffer Capacity	80,000

References

1. Yue, X.; Cai, H.; Gu, C.; Shen, X. Cost-benefit analysis of integrated energy system planning considering demand response. *Energy* **2020**, *192*, 116632. [CrossRef]
2. Zhang, H.; Yuan, T.; Tan, J.; Kai, S.; Zhou, Z. Hydrogen energy system planning framework for unified energy system. *Proc. CSEE* **2022**, *42*, 83–94.
3. Chen, D.; Liu, F.; Liu, S. Optimization of virtual power plant scheduling coupling with P2G-CCS and doped with gas hydrogen based on stepped carbon trading. *Power Syst. Technol.* **2022**, *46*, 2042–2053.
4. Cui, Y.; Yan, S.; Zhong, W.; Wang, Z.; Zhang, P.; Zhao, Y.P. Optimal thermoelectric dispatching of regional integrated energy system with power-to-gas. *Power Syst. Technol.* **2020**, *44*, 4254–4263.
5. Meng, M.; Ma, S.; Zhao, H.; Tao, X. Bi-level optimal operation strategy of integrated energy system with concentrating solar power plant and CCS-P2G[J/OL]. *J. N. China Electr. Power Univ. (Nat. Sci. Ed.)* **2023**, 1–10. Available online: http://kns.cnki.net/kcms/detail/13.1212.TM.20231016.1007.002.html (accessed on 11 December 2023).
6. Han, Z.; Li, Z.; Zhang, W.; Liu, K.; Dong, H.; Yuan, T. Economic operation strategy of hydrogen integrated energy system considering uncertainty of photovoltaic output power. *Electr. Power Autom. Equip.* **2021**, *41*, 99–106.
7. Jiang, Y.; WAN, C.; Botterud, A.; Song, Y.; Xia, S. Exploiting Flexibility of District Heating Networks in Combined Heat and Power Dispatch. *IEEE Trans. Sustain. Energy* **2020**, *11*, 2174–2188. [CrossRef]
8. Zhang, T.; Guo, Y.; Li, Y.; Yu, L.; Zhang, J. Optimization scheduling of regional integrated energy systems based on electric-thermal-gas integrated demand response. *Power Syst. Prot. Control* **2021**, *49*, 52–61.
9. Liu, H.; Wang, D.; Jia, H.; Dou, Z.; Zhang, C.; Wang, S. Construction and Analysis of Energy Carbon Security Region Model of Electric GasHydrogen Integrated Energy System. *Power Syst. Technol.* **2025**, *49*, 73–83.
10. Luo, Q.; Zhu, J.; Zhu, H.; Li, H.; Guo, T. A Fully Distributed Optimal Dispatch Method for Integrated Electricity and Gas Systems with Superlinear Convergence. *Power Syst. Technol.* **2025**, *49*, 1816–1825.
11. Yang, M.; Zhu, Y.; Yu, X. Distributionally robust low-carbon scheduling of integrated energy system considering source—load collaborative carbon reduction under multiple time scales. *Electr. Power Autom. Equip.* **2025**, *45*, 34–42.
12. Chen, R.; Tsay, Y.S.; Zhang, T. A multi-objective optimization strategy for building carbon emission from the whole life cycle perspective. *Energy* **2023**, *262*, 125373. [CrossRef]
13. Sun, G.; Chen, S.; Wei, Z.; Chen, S.; Li, Y. Probabilistic optimal power flow of combined natural gas and electric system considering correlation. *Autom. Electr. Power Syst.* **2015**, *39*, 11–17.
14. Cui, Y.; Sun, X.; Cheng, D.; Xu, Y.; Zhu, H.; Zhao, Y. Stochastic Low-carbon Scheduling of Integrated Energy Virtual Power Plant Considering"Coal—fired+" Coupling Power Generation and the Coupling of Electricity-carbon-hydrogen-chemical. *Power Syst. Technol.* **2025**, *49*, 2388–2397.
15. Li, J.; Cheng, R.; Zhou, B.; Liu, J.; Mao, T.; Zhao, W.; Wang, T.; Huang, G.; Xu, Y. Stochastic Optimal of Integrated Energy System in Low-Carbon ParksConsidering Carbon Capture Storage and Power to Hydrogen. *Electr. Power* **2024**, *57*, 149–156.
16. Liu, C.; Li, R.; Yin, Y. Two-stage optimization for community integrated energy system based on robust stochastic model predictive control. *Electr. Power Autom. Equip.* **2022**, *42*, 1–7.
17. Wang, X.; Zhao, Q.; Zhao, L.; Yang, T. Energy management approach for integrated electricity-heat energy system based on deep Q-learning network. *Electr. Power Constr.* **2021**, *42*, 10–18.
18. Liu, H.; Li, J.; Ge, S.; Zhang, P.; Chen, X. Coordinated scheduling of grid-connected integrated energy microgrid based on multi-agent game and reinforcement learning. *Autom. Electr. Power Syst.* **2019**, *43*, 40–50.
19. Feng, C.; Zang, Y.; Wen, F.; Ye, C.; Zhang, Y. Energy management strategy for microgrids based on deep expectation Q-network algorithm. *Autom. Electr. Power Syst.* **2022**, *46*, 14–22.
20. Yang, T.; Zhao, L.; Liu, Y.; Feng, S.; Pen, H. Dynamic economic dispatch for integrated energy system based on deep reinforcement learning. *Autom. Electr. Power Syst.* **2021**, *45*, 39–47.
21. Gong, J.; Liu, Y. Active distribution network coordination optimization based on deep determination strategy gradient algorithm. *Autom. Electr. Power Syst.* **2020**, *44*, 113–120.