Renewable Energy 214 (2023) 216–232

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Coordinated energy management strategy for multi-energy hub with thermo-electrochemical effect based power-to-ammonia: A multi-agent deep reinforcement learning enabled approach

Kang Xiong [a], Weihao Hu [a,*], Di Cao [a], Sichen Li [a], Guozhou Zhang [a], Wen Liu [b], Qi Huang [c], Zhe Chen [d]

[a] School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China
[b] Copernicus Institute of Sustainable Development, Utrecht University, Princetonlaan 8a, 3584 CB, Utrecht, the Netherlands
[c] Southwest University of Science and Technology, Mianyang, 621010, China
[d] Department of Energy Technology, Aalborg University, Aalborg, Denmark

ARTICLE INFO

Keywords:
Power-to-ammonia
Renewable energy
Multi-energy hub
Multi-agent deep reinforcement learning

ABSTRACT

Power-to-ammonia (P2A) technology has attracted more and more attention since ammonia is recognized as a natural zero-carbon fuel. In this context, this paper constructs a renewable energy powered multi-energy hub (MEH) system which integrates with a thermo-electrochemical effect based P2A facility. Subsequently, the energy management of proposed MEH system is casted to a multi-agent coordinated optimization problem, which aims to minimize operating cost and carbon dioxide emissions while satisfying constraints. Then, a novel multi-agent deep reinforcement learning method called CommNet is applied to solve this problem to obtain the optimal coordinated energy management strategy of each energy hub by achieving the distributed computation of global information. Finally, the simulation results show that the proposed method can achieve better performance on reducing operating cost and carbon emissions than other benchmark methods.

1. Introduction

In recent decades, the penetration of renewable energy is continually increasing to reduce the consumption of fossil fuels and the carbon emissions. However, the utilization of renewable energy sources (RES) such as solar and wind energy also take the intermittency nature into power system [1,2]. To this end, the energy storage systems (ESS) such as battery and power-to-gas (P2G) technology are widely applied to mitigate the stochastic and intermittent generation of renewable energy [3]. Research in Refs. [4,5] proposed different hybrid energy systems structures which integrated battery with renewable energy to deal with the uncertainty of renewable energy generation and improved the reliability of the system. Nevertheless, the frequent charging and discharging behavior of battery will not only increase additional degradation cost, but also shorten the long-term lifetime [6–8].

In this context, P2G technology is an ideal substitute that can storage the surplus energy from renewable energy and product gaseous energy carriers such as hydrogen (H_2) and methane (CH_4). Tebibel [9] proposed a wind-hydrogen production system that utilized the wind turbines to produce electricity to supply the electrolyzer. Zhong et al. [10] designed a solar-driven power-to-methane system, and the simulation results have shown that the total energy efficiency and power-to-methane efficiency were obviously improved. However, the storage and transportation of hydrogen/methane at ambient condition requires the special process such as liquefaction and compression, which will increase the additional storage cost [11]. Besides, the combustion products of methane include carbon dioxide, which can aggravate the greenhouse effect [12]. Compared with these two P2G technologies, power-to-ammonia (P2A) is a better alternative, which is a technology to combine hydrogen (H_2) produced by electrolysis of water with nitrogen (N_2) separated from air to synthesis ammonia (NH_3) [13]. NH_3 can be liquefied just at −33.3 °C and atmospheric pressure [14], while H_2 requires −252.8 °C and 100 bar [15]. This characteristic allows ammonia more easily stored and transported without utilizing the special gas tanks, so as to efficiently decrease the cost [16]. Besides, considering the low cost of N_2 source, P2G is undoubtedly a more economical technology [12]. Moreover, NH_3 is a both zero-carbon and

* Corresponding author.
E-mail addresses: xiongkanguestc@163.com (K. Xiong), whu@uestc.edu.cn (W. Hu), caodi@uestc.edu.cn (D. Cao), sichenli@std.uestc.edu.cn (S. Li), zgz@std.uestc.edu.cn (G. Zhang), w.liu@uu.nl (W. Liu), hwong@uestc.edu.cn (Q. Huang), zch@et.aau.dk (Z. Chen).

https://doi.org/10.1016/j.renene.2023.05.067
Received 11 January 2023; Received in revised form 6 May 2023; Accepted 14 May 2023
Available online 9 June 2023

Nomenclature

Abbreviations

RES	renewable energy sources
ESS	energy storage systems
P2G	power-to-gas
P2A	power-to-ammonia
MCES	multi-carrier energy system
MEH	multi-energy hub
SO	stochastic optimization
RO	robust optimization
DRL	deep reinforcement learning
SADRL	single-agent deep reinforcement learning
MADRL	multi-agent deep reinforcement learning
PVT	photovoltaic thermal system
DES	district energy system

high energy density fuel whose combustion reaction products with oxygen are water and nitrogen and the energy density is 1.5 times higher than hydrogen, which indicates it is an environmentally friendly energy [17]. Conventionally, the route of P2A to synthesis ammonia is by the Haber-Bosch process based catalyst at 500 °C and 300 bar [18]. However, this method not only requires exacting reaction condition, but also has a very low yield of ammonia caused by the unfavorable chemical equilibrium [19]. Therefore, some researchers have explored a new reaction route which can synthesis ammonia from nitrogen at ambient condition and obviously improve the yield of ammonia by utilizing catalysts to promote the electrochemical synthesis process [20,21].

Hence, incorporating the renewable energy powered P2A with other energy systems such as gas and heat to construct a multi-carrier energy system (MCES), can better enhance the system flexibility, reliability and stability as well as reduce operating cost. Wen et al. [22] proposed an electricity-heat-ammonia coupled multi-generation system and evaluated the energy and exergy efficiencies of system. Wang et al. [23] developed an ammonia-water based combined heating and power (CHP) energy system that simultaneously supplied the electricity and hot air for consumers. However, the above studies do not consider thermo-electrochemical effect of the synthesis process of ammonia when construct the system model, which will influence the yield of ammonia. MCES always be constructed as an energy hub (EH) to better manage and optimize energy generation and consumption, which multiple energy carriers can be converted, conditioned and stored [24,25]. Furthermore, the complex MCES model can be constructed as several simple EHs which called multi-energy hub (MEH) so that the interconnection of each EH can be considered.

However, the energy management and operation optimization of MEH is a huge challenge as the unpredictable state changes and uncertainties of system can easily influence the performance [26–28]. To address this challenge, stochastic optimization (SO) method is widely applied to conduct energy management [29]. Thang et al. [30] developed a stochastic optimization (SO) method to simultaneously optimize the MEH and distribution networks considering the uncertainty of the system. Zou et al. [31] implemented an adaptive SO-based method to solve a Stackelberg game among a multi-energy microgrid. Heidari et al. [32] constructed a EH optimization model with combined cooling, heating, and power (CCHP) and ice storage, and applied the SO-based method to solve this model. Nevertheless, the SO-based methods always need to pre-sample various information of different scenarios according to the given probability distribution [33], which requires vast computational ability and even cannot be realized in practice. Therefore, the robust optimization (RO) method is utilized to solve the optimization model to achieve robust operation. Poursmaeil et al. [34] implemented RO-based method to obtain the optimal schedule of the constructed MEH system and considered the operation of electric vehicles (EV). Lu et al. [35] proposed a MEH system with electricity, gas, heat energy and applied RO method to solve the coordinated optimization model to reduce the total cost of MEH. However, the optimization results obtained by RO method are conservative in order to keep the safety of system operation, which makes the optimal operation strategy may not be obtained.

To this end, the machine learning (ML) based methods are applied to solve the optimization model of MEH, which can extract powerful operation knowledge from historical data to obtain the real-time operation strategy [33]. Specifically, the deep reinforcement learning (DRL) is a method that combine the perception ability of deep learning (DL) and the decision-making ability of reinforcement learning (RL), which is suitable to solve the stochastic decision-making problem. For instance, Zhou et al. [36] utilized an improved soft actor critic (SAC) algorithm to study the energy management strategy of MEH under the challenges of stochastic renewable supplies and energy demands. Tan et al. [37] developed a DRL approach to realize efficient energy management of hybrid electric vehicle. Nevertheless, these studies only apply single-agent based DRL (SADRL) algorithms for solving optimization problems, which is unable to achieve optimal operation when deal with the overlap of multiple adjustable parameters of multiple devices [38]. Besides, the SADRL methods focus on a centralized perspective and lack of cooperativity, which may lead to concerns about data privacy [38], communication delay [39] as well as single point failure issue [40]. In this context, it is necessary to find a more suitable and effective DRL method to solve the multi-objective cooperative energy management problems in a distributed manner with desirable performance and guaranteed scalability. To this end, multi-agent based DRL (MADRL) methods are naturally required, in which agents can be trained in a distributed manner to obtain a coordinated strategy [41]. Ahrarinouri et al. [42] implemented a multi-agent deep Q learning (MADQN) for residential multi-carrier energy management. Zhang et al. [43] proposed a multi-agent twin delayed deep deterministic policy gradient (MATD3) to learn the optimal operation strategy for bottom-layer microgrid. However, most MADRL approaches have several limitations: 1) the MADQN suffers from the curse of dimensionality when coping with continuous problems; 2) the MATD3 is so sensitive to hyper-parameter that cannot learn a suitable policy as the complexity of the environment grows exponentially [44].

Motivated by the above concerns, this paper develops a renewable energy powered MEH system which integrates thermo-electrochemical effect based P2A in each EH, and a novel multi-agent deep reinforcement learning (MADRL) algorithm is proposed to achieve the cooperative energy management of MEH. The main contributions of this paper are summarized as follows:

1). An integrated P2A technology multi-energy hub (MEH) framework which powered by renewable energy sources (wind and solar) is constructed in this paper. The consumption of nature gas can be reduced by utilizing the ammonia in each EH. Unlike the previous MEH studies [25,30], both the operation cost and carbon emissions of proposed MEH can be effectively reduced by cooperatively managing the coupled multi-energy in MEH.
2). A thermodynamic model of P2A which considers the thermo-electrochemical effects from renewable energy is integrated in each EH to enhance the yield of ammonia. Compared with the previous P2A models [22,23], the electrolytic temperature of electrolyzer when synthesis ammonia is controllable by adjusting the available renewable thermal energy for P2A reaction heating so that the yield of ammonia can be efficiently improved.
3). To achieve the cooperative optimization of MEH, the energy management scheme is formulated as a Markov game and solved by a novel MADRL algorithm called CommNet. Different from the existing centralized DRL based methods [36,37], the proposed approach can obtain the optimal coordinated energy

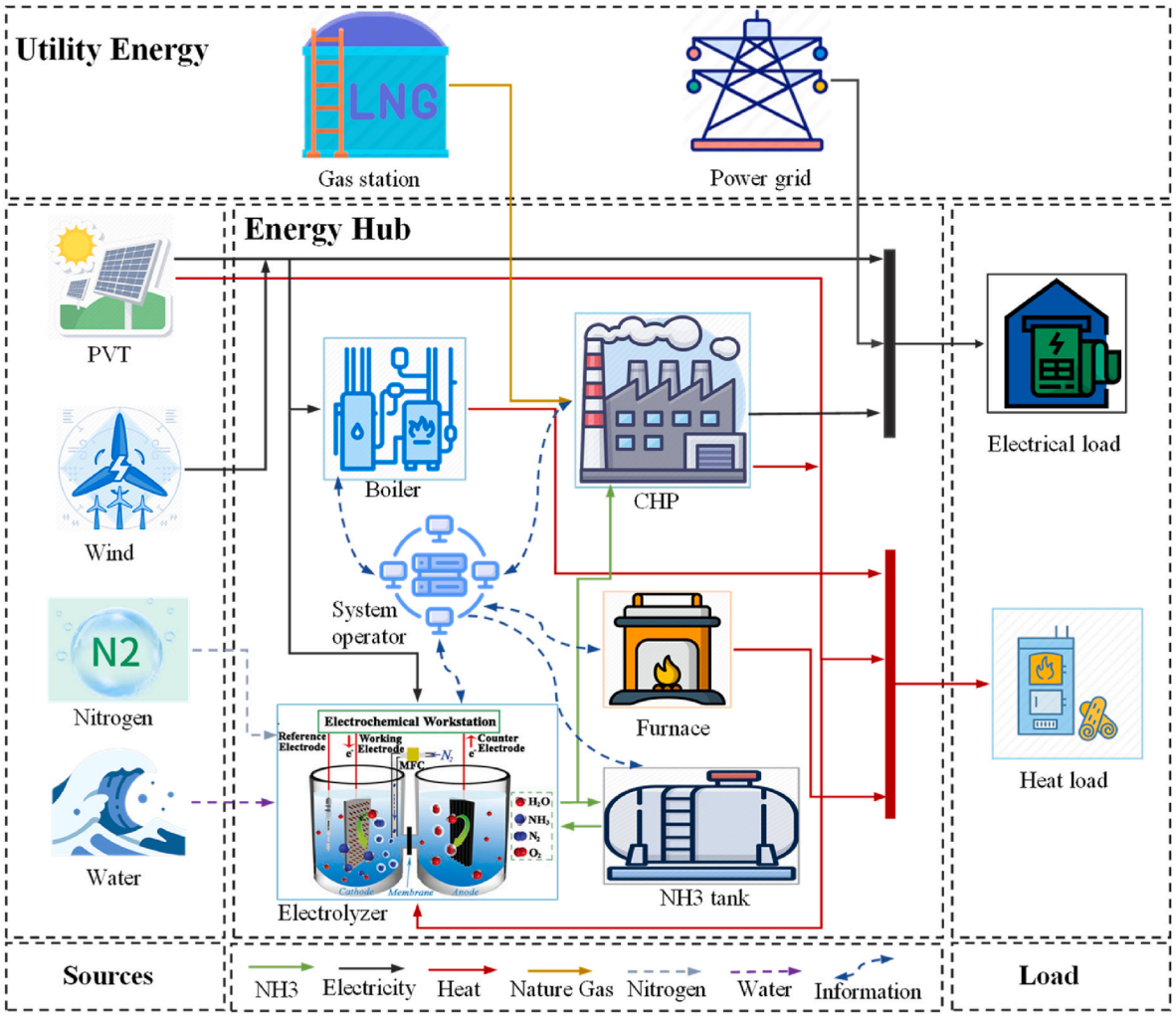

Fig. 1. Structure of the proposed single energy hub.

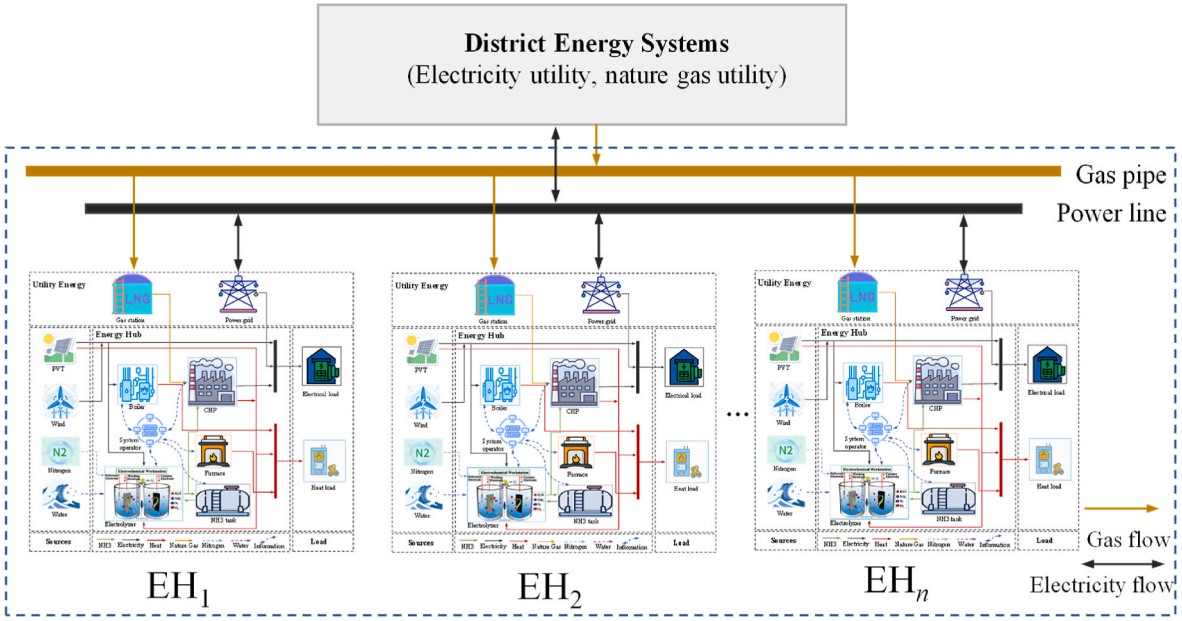

Fig. 2. Overview of the proposed multi-energy hub (MEH).

Table 1
Related parameters of PVT.

Parameter	value	Parameter	value
η_{opt}	0.85	T_{ref}	25 °C
η_{inv}	0.9	β_{ref}	0.004/°C
$\eta_{T_{ref}}$	0.12	$P_{PVT,rated}$	300 kW

management in a distributed perspective as well as guarantee the stable convergence of agent training.

The rest of this paper is organized as follows: the system model and problem formulation are presented in Section II and Section III, respectively. In Section IV, the proposed MADRL algorithm is introduced in detail. The numerical results are presented in Section V to verify the effectiveness of proposed method and the conclusions are presented in Section VI.

2. System model of multi-energy hub

The Fig. 1 shows the specific structure of each EH, which includes photovoltaic thermal system, wind turbine, CHP, boiler, furnace, P2A and NH_3 tank. The system operator of each EH aims to collect and dispatch information for each energy device to achieve stable and efficient operation of each EH. The modelling of each energy facility is described in detail in this section. Moreover, the framework of proposed MEH is shown in Fig. 2, which includes various single EH. The district energy system (DES) includes electricity utility and gas utility is backup energy supply for each EH as long as the EH cannot be self-sufficient. Moreover, the EH can sell their surplus energy to the electricity utility or other EHs through the power line.

2.1. Model of photovoltaic thermal system

The excessive increase of temperature of PV panels will lead to an obvious decrease of electrical efficiency. To this end, the photovoltaic thermal (PVT) system [45] has designed to overcome this challenge which consists of solar thermal collectors and photovoltaic panels to generate both electricity and heat energy. The output of photovoltaic panels is mainly related to solar irradiance intensity S_t, which can be formulated as follow [46,47]:

$$H_{PVT,t} = S_t C \eta_{opt} A \tag{1}$$

$$P_{PVT,t} = H_{PVT,t} \eta_{PV,t} \eta_{inv} \tag{2}$$

where C represents the concentration ratio; A denotes the module area of PV; η_{opt} and η_{inv} are the optical efficient and the efficiency of inverter, respectively; the efficiency of the PV module can be calculated in Eq. (3),

$$\eta_{PV,t} = \eta_{T_{ref}} \left[1 - \beta_{ref} \left(T_{C,t} - T_{ref}\right)\right] \tag{3}$$

where $\eta_{T_{ref}}$ denotes module's efficiency at the reference temperature; β_{ref} represents the temperature coefficient; T_{ref} refers to the reference temperature of cell; The related parameters of PVT are shown in Table 1 [47].

2.2. Model of wind turbine

A piecewise function is utilized to mathematically model the electricity output of WT, which the real time output $P_{WT,t}$ is closely related to the real time wind speed v_t. There is no output of WT when the wind speed lower than the cut-in wind speed v_{cin} or higher than the cut-out wind speed v_{cout}. The output can be calculated according to Eq. (4) when wind speed between cut-in speed and rated wind speed v_{rated}. Besides, the output of wind turbine set to be rated output $P_{WT,\text{rated}}$ when wind speed between rated wind speed and cut-out wind speed [48].

Table 2
Applied parameters of wind turbine.

Parameter	value	Parameter	value
v_{cin}	4 m/s	v_{cout}	25 m/s
v_{rated}	11.4 m/s	R	70 m
$P_{WT,\text{rated}}$	1 MW	–	–

$$P_{WT,t} = \begin{cases} 0, 0 \le v_t < v_{\text{cin}} \\ \frac{1}{2}\rho\pi R^2 v_t^3 C_p(\beta,\lambda) \Big/ 10^6, v_{\text{cin}} \le v_t < v_{\text{rated}} \\ P_{WT,\text{rated}}, v_{\text{rated}} \le v_t < v_{\text{cout}} \\ 0, v_t \ge v_{\text{cout}} \end{cases} \tag{4}$$

where ρ denotes the air density; R is the rotor diameter; C_p is utilization factor of wind energy; β and λ represent blade pitch angle and tip speed ratio, respectively. Table 2 specifically shows the applied parameters of WT.

2.3. Model of proposed power to ammonia

P2A is an emerging technology that utilizes the hydrogen from electrolysis of water and nitrogen separated from air to synthesis ammonia [7], which can be represented as:

$$\begin{aligned} &6H_2O + 2N_2 \rightarrow 4NH_3 + 3O_2 \\ &2H_2O \rightarrow O_2 + 4H^+ + 4e^- \\ &N_2 + 6H^+ + 6e^- \rightarrow 2NH_3 \end{aligned} \tag{5}$$

The reaction process of P2A and the yield of ammonia are mainly influenced by the type of catalysts [19], electron structure of active surfaces [49] and the PH value of reaction environment [50]. Without losing the generality as well as keep the yield of ammonia, the proposed P2A is conducted in electrochemical cells whose conducting electrolyte is NH_4^+/H^+ and electrodes are Pt/C based under ambient environment (room temperature and atmospheric pressure). As Eq. (5) shown, the reaction of electrolysis of water is conducted in anode, where the water is decomposed into oxygen O_2 and hydrogen ion H^+. After that, the hydrogen ion H^+ will transfer to the cathode and combine with the nitrogen N_2 separated from air to product NH_3. Generally, lots of repeating electrolysis cells are electrically connected in series to form a branch and several branches are combined in parallel way to finally form the electrolyzer module to ensure an ample yield of NH_3 [51].

2.3.1. Production model of NH_3

It can be obtained from the Faraday's law that the production of ammonia is approximately proportional to the current density. Besides, parasitic current and cross permeation when electrolyzer is operating also influence the yield of ammonia. Thus, the Faradaic efficiency $\eta_{F,t}$ is applied to mathematically construct the production model of ammonia [52]. The ammonia production of can be calculated as follow:

$$y_{t,n} = \frac{N_c \eta_{F,t} V_{t,n} I_{t,n} \Delta t}{N_e F} \tag{6}$$

$$Y_t = \sum_{n=1}^{N_b} y_{t,n} \tag{7}$$

where N_c represents the number of series electrolysis cell and N_b is the number of electrolysis cell branch; $V_{t,n}$ denotes the molar volume of cell branch n and F is Faraday constant.

For proposed P2A, environment pressure is the only negligible factor when consider the performance of electrolyzer so that which can be treated as a constant [17]. Thus, the $V_{t,n}$ is proportional to the reaction temperature according to the isobaric ideal gas law, which can be expressed as:

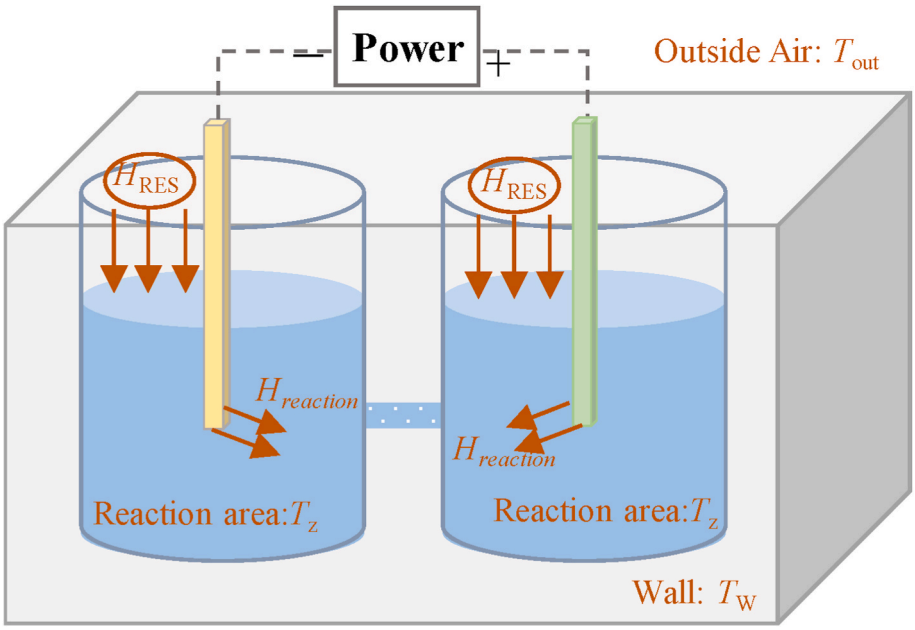

Fig. 3. Physical structure diagram of electrolysis device.

$$\frac{V_{t,n}}{T_{Z,t}}=\frac{V'_{t,n}}{T'_{Z,t}} \tag{8}$$

The electricity consumption of electrolyzer $P_{E,t,n}$ is closely related to electrolytic efficiency $\eta_{E,t}$, which can be described as follow:

$$\begin{aligned} P_{E,t,n} &= \frac{y_{t,n}Q_a}{\eta_{E,t}\Delta t} \\ P_{E,t} &= \sum_{n=1}^{N_b} P_{E,t,n} \end{aligned} \tag{9}$$

where Q_a is calorific value of ammonia that indicates the enthalpy change of reaction process under the standard condition; $P_{E,t}$ denotes the total electricity consumption of all electrolyzers. PSA is an efficient device to separate nitrogen from air for P2A reaction, which has low cost and longer lifetime as well as higher efficiency than the original nitrogen producer [53]. The cost of PSA will be specifically described in next section.

Then, both the Faradaic efficiency and electrolytic efficiency are strongly influenced by the reaction temperature, the efficiencies increase when the reaction temperature rises and the relationship can be described as a piecewise function [53,54]:

$$\eta=\begin{cases} a\left(T_{Z,t}-T_{Z,\min}\right), T_{Z,\min}\le T_{Z,t}\le T_1 \\ a\left(T_1-T_{Z,\min}\right)+b\left(T_{Z,t}-T_1\right), T_1\le T_{Z,t}\le T_2 \\ a\left(T_1-T_{Z,\min}\right)+b\left(T_2-T_1\right)+c\left(T_{Z,t}-T_2\right), T_2\le T_{Z,t}\le T_{Z,\max} \end{cases} \tag{10}$$

where a, b, c are the coefficients of the efficiency function.

2.3.2. Thermo-electrochemical model

In proposed P2A, the thermo-electrochemical effect which caused by the injection of external renewable thermal energy is fully considered. The physical structure diagram of electrolysis device is described in Fig. 3. As Fig. 3 shown, the whole physical environment of utilized P2A electrolysis device consists of reaction area, surrounding wall as well as outside air, and these three parts can conduct the heat exchange. To mathematically model the thermo-electrochemical process, the thermal

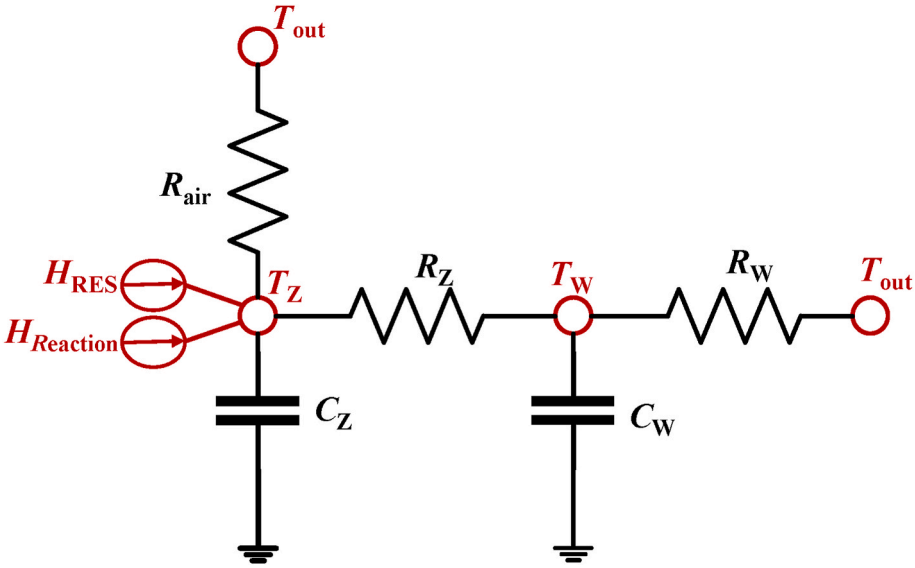

Fig. 4. Equivalent circuit diagram of electrolysis device [7].

Table 3
Applied parameters of proposed P2A.

Parameter	value	Parameter	value
F	96485 C/mol	$N_{W,n}$	3.09
$I_{t,\ n}$	1000 A/m^2	$\lambda_{Z,n}$	0.6 W/(m.°C)
N_c	5	$\lambda_{W,n}$	1.1 W/(m.°C)
N_b	9	$\lambda_{out,n}$	200 W/(m.°C)
$T_{Z,min}$	5 °C	y_{ng}	0.75 m^3
$T_{Z,max}$	65 °C	P_{ng}	0.5 kW
W_n	0.5 m	$C_{Z,n}$	0.2917 kWh/°C
L_n	1 m	$C_{W,n}$	0.0205 kWh/°C
$N_{Z,n}$	4	–	–

resistance and capacitance are utilized to measure the heat storage capacity and heat conversion ability of electrolysis device elements [7,55]. The resistor-capacitor based equivalent circuit diagram of electrolysis device is given in Fig. 4, which shows the temperature dynamic of electrolysis device. The heat storage ability and heat conversion ability of reaction area and surrounding wall are modeled as thermal resistors and capacitors as the figure shown.

It can be obtained from Fig. 4 that the thermal transferring process among reaction area, surrounding wall and outside air is represented as a resistor-capacitor circuit. The thermal resistances can be obtained by Eq. (11) under the laminar flow condition.

$$R_{Z,n}=\frac{l_{h,Z,n}}{N_{Z,n}\lambda_{Z,n}}, R_{W,n}=\frac{l_{h,W,n}}{N_{W,n}\lambda_{W,n}}, R_{\text{out},n}=\frac{1}{W_n L_n \lambda_{\text{out},n}} \tag{11}$$

where $N_{Z,n}$, $N_{W,n}$ represent Nusselt constant number; $\lambda_{Z,n}$, $\lambda_{W,n}$, $\lambda_{\text{out},n}$ denote the thermal conductivity of electrolyte, wall and outsider air; $l_{h,Z,n}$, $l_{h,W,n}$ are the geometric characteristic length of heat transfer surface of reaction zone and surrounding wall, which can be calculated as:

$$l_{h,Z,n}=\frac{2W_n h_{Z,n}}{W_n+h_{Z,n}} \tag{12}$$

$$l_{h,W,n}=\frac{2W_n h_{W,n}}{W_n+h_{W,n}} \tag{13}$$

where $h_{Z,n}$, $h_{W,n}$ are the height of reaction zone and surround wall.

Thus, the thermo-electrochemical process of P2A can be mathematically modeled as follow:

$$C_{Z,n}\frac{dT_{Z,t,n}}{dt}=\frac{T_{W,t,n}-T_{Z,t,n}}{R_{Z,n}}-\frac{T_{Z,t,n}-T_{out,t,n}}{R_{out,n}}+H_{\text{RES},t,n}+P_{E,t,n}\Delta t-y_{t,n}Q_a \tag{14}$$

$$C_{W,n}\frac{dT_{W,t,n}}{dt}=\frac{T_{Z,t,n}-T_{W,t,n}}{R_{Z,n}}+\frac{T_{out,t,n}-T_{W,t,n}}{R_{W,n}} \tag{15}$$

where the process of heat transfer and exchange among the reaction area, surrounding wall and outside air are described by the Eq. (14)-(15). The thermodynamic process of reaction area and wall are described in the left hand of Eq. (14)-(15). The first three terms of the right hand of Eq. (14)-(15) show the heat interaction among the inside electrolyte, surrounding wall and outside air, and the last three terms of the Eq. (14) are the heat energy injection from renewable energy, enthalpy change of reaction process and Joule heat from the current density. The applied parameters of proposed P2A are shown in Table 3 [16].

2.4. Model of combine heating and power system

There are two kinds of CHP system in the proposed EH framework, nature gas-fired CHP and ammonia-fired CHP system. CHP system can simultaneously generate heat and electricity, which consists of the power generation unit and heat recovery unit. The nature gas or ammonia is fired in power generation to product electricity, and the excessive power is captured by the heat recovery unit to product heat. The CHP system can be mathematically modeled as follow:

$$P_{CHP,t}=G_{i,t}Q_i\eta_e, i=\{\text{gas, ammonia}\} \tag{16}$$

$$H_{CHP,t}=\left(G_{i,t}Q_i-P_{CHP,t}\right)\eta_{rec}, i=\{\text{gas, ammonia}\} \tag{17}$$

where $G_{i,t}$ is the consumption of nature gas or ammonia at each time-step; the η_e and η_{rec} are the efficiency of power generation unit and the efficiency of heat recovery unit, which set to 0.38 and 0.8 [56], respectively; The Q_g and Q_a represent calorific value of nature gas and ammonia, which are 4.339 kWh/m^3 and 10.729 kWh/m^3.

2.5. Electrical boiler and gas furnace model

Electrical boiler and gas furnace are utilized to generate thermal energy to meet the heat demand. The energy conversion process can be described according to the following equations:

$$H_B=P_{B,t}\eta_B \tag{18}$$

$$H_{F,t}=Q_{gas}G_{gas,t}\eta_F \tag{19}$$

where the η_B and η_F are the thermal efficiency of electrical boiler and gas furnace, respectively. $P_{B,t}$ denotes the real time electricity input for boiler. $G_{gas,t}$ is the gas input for furnace at time t.

3. Problem formulation

After the MEH model is constructed, this section will form the energy management strategy of MEH. Thus, both the objective function and system constraints are explicitly discussed.

3.1. Objective function

This study aims to minimize the operating cost as well as reduce the carbon dioxide emission of MEH. The operating cost of MEH includes purchase cost $C_{p,t}$ and maintenance cost $C_{m,t}$ of all EHs. Besides, the reduction of carbon dioxide emission achieves by levied carton tax [57,58], which can be regarded as another cost $C_{c,t}$. Thus, the objective function can be represented as a total cost function as the Eq (20):

$$C_{total,t}=C_{p,t}+C_{m,t}+C_{c,t} \tag{20}$$

The total purchase cost of each EH is composed of the electricity purchase cost and the nature gas purchase cost as well as the cost of nitrogen production, which can be calculated as follow:

$$C_{p,t}=\sum_{i=1}^{k}\left(P_{grid,t,i}\times price_{p,t}+G_{gas,t,i}\times price_{g,t}+Y_{t,i}\times price_{N_2}\right) \tag{21}$$

where $P_{grid,t,i}$ and $G_{gas,t,i}$ represent the amount of electricity and nature gas import from power grid and gas station of EH i at time t; $price_{p,t}$ and $price_{g,t}$ are the real-time price of importing electricity and nature gas, respectively; $Y_{t,i}$ is the production of ammonia of EH i at time step t; $price_{N_2}$ denotes the required nitrogen cost for producing 1 m^3 ammonia, which can be chosen as 0.03125 $/m^3 [59].

The maintenance cost of each EH includes WT, PVT, CHP, boiler and furnace maintenance cost, which can be calculated as follow:

$$C_{m,t}=\sum_{i=1}^{k}\left(M_{WT}P_{WT,t,i}+M_{PVT}P_{PVT,t,i}+M_{CHP}P_{CHP,t,i}+M_B P_{B,t,i}+M_F P_{F,t,i}\right) \tag{22}$$

where $P_{WT,t,i}$, $P_{PVT,t,i}$, $P_{CHP,t,i}$, $P_{B,t,i}$, $P_{F,t,i}$ are the real-time electricity consumption of WT, PVT, CHP, boiler and furnace of EH i; M_{WT}, M_{PVT}, M_{CHP}, M_B, M_F denote the maintenance cost per unit of WT, PVT, CHP, boiler, furnace, which are specifically given in Table 4 [56,60].

Table 4
The applied maintenance cost factors.

Parameter	value	Parameter	value
M_{WT}	4.794 $/MW	M_B	4.644 $/MW
M_{PVT}	3.500 $/MW	M_F	3.870 $/MW
M_{CHP}	9.288 $/MW	–	–

The cost of carbon dioxide emission is related to the real-time carbon dioxide production and carbon tax. The main carbon dioxide production of proposed MEH comes from the importing electricity and natural gas since the importing electricity originate from the fossil energy and carbon dioxide would be produced during natural gas combustion. Thus, the cost of carbon dioxide emission can be calculated as follow:

$$carbon_{\ t} = \sum_{i=1}^{k} \left(\theta_{elec} \times P_{grid,t,i} + \theta_{gas} \times G_{gas,t,i}\right) \tag{23}$$

$$C_{c,t} = carbon_{\ t} \times \delta \tag{24}$$

where $carbon_{\ t}$ is the carbon dioxide emission of MEH at time t; θ_{elec}, θ_{gas} denote the carbon dioxide emission factor of importing electricity and natural gas, which can be chosen as 0.95 kg/kWh and 1.93122 kg/m^3 [56]; δ represents the carton tax that is set to 0.01527 $/kg [57,61].

3.2. System constraints

The system constraints are constructed to ensure the stable and efficient operation of MEH when solve the optimization problem, which will discuss in detail in this subsection. It is worth noting that the constraints are on hourly basis.

3.2.1. Constraints of power to ammonia

The yield of ammonia is limited by the maximum production of each electrolysis cell $y_{\max,n}$ to ensure the safe operation of reaction process, which can be described as:

$$0 \le y_{t,n} \le y_{\max,n} \tag{25}$$

Besides, influenced by the inertia of liquid electrolyte and velocity limits of gas streams [7], the electrolyzer need time to change the set-point so that the ramp capacity of ammonia production is required.

$$\left|y_{t,n} - y_{t-1,n}\right| \le y_{ramp,n} \tag{26}$$

3.2.2. Constraints of NH_3 tank

The constraints of NH_3 tank can be described as below:

$$G_{\text{tank,min}} \le G_{\text{tank},t} \le G_{\text{tank,max}} \tag{27}$$

$$SOC_{\text{tank,min}} \le SOC_{\text{tank},t} \le SOC_{\text{tank,max}} \tag{28}$$

$$SOC_{\text{tank},t} = SOC_{\text{tank},t-1} + \left(G_{c,t}\eta_c / cap - G_{d,t}\eta_d / cap\right)\Delta t \tag{29}$$

where $G_{\text{tank},t}$ denotes the stored NH_3 of tank; $G_{\text{tank,min}}$ and $G_{\text{tank,max}}$ are the limitation of NH_3 storage of gas tank; $SOC_{\text{tank},t}$ represents the state of charge of tank at time t; $SOC_{\text{tank,min}}$ and $SOC_{\text{tank,max}}$ are the upper and lower limit of SOC; $G_{c,t}$ and $G_{d,t}$ are the real-time amount of charging and discharging NH_3 of tank; η_c and η_d represent the charging and discharging efficiency. cap is the capacity of applied gas tank.

3.2.3. Constraints of combine heating and power system

There are several constraints of CHP proposed to guarantee the safe operation, which can be expressed as follow:

$$P_{CHP,\min} \le P_{CHP,t} \le P_{CHP,\max} \tag{30}$$

$$P^{ramp}_{CHP,\min} \le P_{CHP,t} - P_{CHP,t-1} \le P^{ramp}_{CHP,\max} \tag{31}$$

where $P_{CHP,\min}$ and $P_{CHP,\max}$ are the maximum and minimum output of CHP; Eq. (31) is the ramp rate constraint of CHP; $P^{ramp}_{CHP,\min}$ and $P^{ramp}_{CHP,\max}$ denote the upper and lower limits of ramp rate output of CHP.

3.2.4. Constraints of boiler and furnace

The electrical boiler and gas furnace are only limited by their capacities, which can be described as:

$$H_{B,\min} \le H_{B,t} \le H_{B,\max} \tag{32}$$

$$H_{F,\min} \le H_{F,t} \le H_{F,\max} \tag{33}$$

where $H_{B,\min}$ and $H_{B,\max}$ are the maximum and minimum output of boiler; $H_{F,\min}$ and $H_{F,\max}$ are the maximum and minimum output of furnace.

3.2.5. Constraints of electricity, heat and ammonia balance

(a) **Electricity balance:** the generated electrical power must equal to the electrical consumption at each time-step, which can be described as:

$$\sum_{i=1}^{k}\left(P_{WT,t,i} + P_{PVT,t,i} + P^{NH_3}_{CHP,t,i} + P^{gas}_{CHP,t,i}\right) = \sum_{i=1}^{k}\left(P_{load,t,i} + P_{B,t,i} + P_{E,t,i}\right) \tag{34}$$

where $P_{load,t,i}$ is the real-time electrical load of each EH.

(b) **Heat balance:** the heat energy generated by CHP, boiler and furnace should equal to the real-time heat demand.

$$\sum_{i=1}^{k}\left(H^{NH_3}_{CHP,t,i} + H^{gas}_{CHP,t,i} + H_{B,t,i} + H_{F,t,i}\right) = \sum_{i=1}^{k} H_{load,t,i} \tag{35}$$

where $H_{load,t,i}$ is the real-time heat demand of each EH. Note that the thermal energy from the PVT module is utilized to heat the electrolyzer and enhance the yield of ammonia.

(c) **Ammonia balance:** the NH_3 consumption for CHP at each time-step is composed of the produced NH_3 by P2A and the charge/discharge amount of NH_3 tank $G^a_{\text{tank},t,i}$, which can be modeled as:

$$\sum_{i=1}^{k}\left(G^a_{\text{P2A},t,i} + G^a_{\text{tank},t,i}\right) = \sum_{i=1}^{k} G^a_{CHP,t,i} \tag{36}$$

where $G^a_{\text{P2A},t,i}$ denotes the real-time production of NH_3 each P2A unit.

4. Applied multi-agent deep reinforcement learning algorithm

In this section, to achieve the optimal operation of proposed MEH while satisfying the system constraints, the energy management of MEH is modeled as a Markov game first. Then, the formed Markov game solved by CommNet which is an improved MADRL algorithm based on communication architecture to realize the coordinated energy management for each EH in a distributed manner [62,63].

4.1. Modeling a Markov game

The energy management of proposed MEH is formulated as a Markov game [64], which is an extension of Markov decision process (MDP) for multi-agent system. The details of each component of Markov game are described as follow:

Agents: agents are the information centers which can receive the input states from environment and feedback suitable actions to maximize the corresponding reward. Note that three agents ($N = 3$) are applied to obtain the optimal energy management strategy for three EHs in this study.

State set: the state set is composed of the state of all agents at each

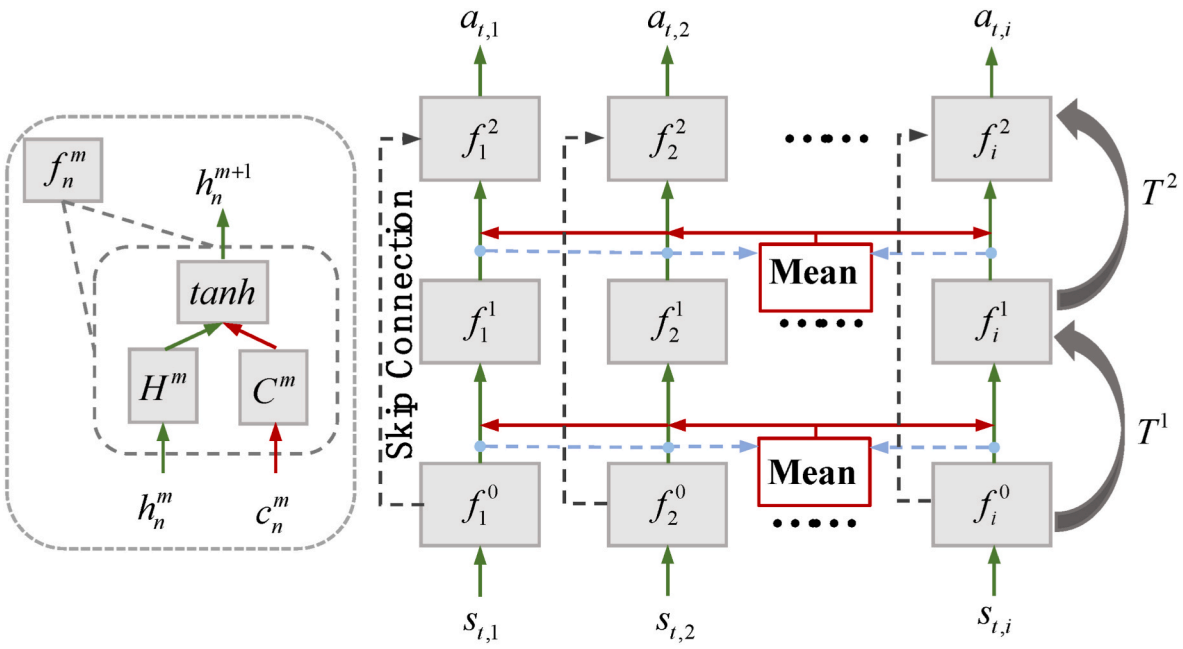

Fig. 5. The specific structure of CommNet [65].

time-step t, which can be denoted as $S_t = (s_{t,1},\ldots,s_{t,i},\ldots,s_{t,N})$. The state of agent i at time-step t $s_{t,i}$ can be describes as below:

$$s_{t,i} = \left(P_{WT,t,i}, P_{PVT,t,i}, H_{PVT,t,i}, P_{load,t,i}, H_{\text{load},t,i}, G^a_{\text{tank},t-1,i}, price_{p,t,i}, price^{\text{mean}}_{p,i}, price_{g,t,i}\right) \tag{37}$$

where $G^a_{\text{tank},t-1,i}$ represents the output of gas tank at last time-step and which will set to be 0 at the initial time-step; $price^{\text{mean}}_{p,i}$ denotes the daily average electricity price of EH i.

Action set: A_t denotes the action set at time-step t, which is composed of the actions of all agents $A_t = (a_{t,1},\ldots,a_{t,i},\ldots,a_{t,N})$. The actions of agent i at time-step t $a_{t,i}$ are the input electricity of electrical boiler at time-step t, the input gas of furnace at time-step t, and the charging/discharging amount of NH_3 of tank at time-step t, respectively, which can be described as follow:

$$a_{t,i} = \left(P_{B,t,i}, G_{gas,t,i}, G^a_{\text{tank},t,i}\right), i = 1, 2, \ldots, N \tag{38}$$

Note that the type of actions of each agent is same since they have the same system structure. However, the actual values of actions of each agent at every time-step are different as the states of each agent are different.

Reward function: the reward function at time-step t R_t includes the total cost function of MEH as show in Eq. (20) and the TSS reward function, which can be described as follow:

$$R_t = -\left(\alpha_1 * C_{total,t} + \alpha_2 * C_{TSS,t}\right) \tag{39}$$

where α_1, α_2 are the weight coefficient, $C_{TSS,t}$ represents the penalty cost of gas tank which is utilized to guide agent to explore the suitable actions for gas tank and the further study can be found in Ref. [42], which can be formulated as follow:

$$C_{TSS,t} = \sum_{i=1}^{N}\left(G^a_{\text{tank},t,i}\left(price_{p,t,i} - price^{mean}_{p,i}\right)\right) \tag{40}$$

The applied policy function of this study can be formulated as:

$$\pi_\theta(S, A) = P[A|S, \theta] \tag{41}$$

where P is the transition probability to map the state S to action An under the parameter θ. The expectation can be calculated by action-value function $Q^{\pi_\theta}(S_t, A_t)$ after policy is implemented, which is:

$$Q^{\pi_\theta}(S_t, A_t) = E_{\pi_\theta}\left[\sum_{i=t}^{T}\gamma^{(i-t)}R(s_i, a_i)\right] \tag{42}$$

At each time-step t, each agent obtains the current state $s_{t,i}$ from each EH i and the actions will be derived according to the policy π_θ to maximize the discounted cumulative reward, and then transfer to the next state.

4.2. Applying proposed MADRL method to solve the Markov game

To obtain the optimal energy management strategy of proposed MEH to minimize the operating cost and the carbon dioxide emission, a novel communication structure based MADRL algorithm called CommNet is utilized. In CommNet, each agent only sends its observed state from local environment to the communication channel and then all agents share the common average state by accessing the broadcasting communication structure. Then, the common average state is regarded as the input of next layer and the actions of agents are the output of the last layer.

The detailed structure of CommNet is shown in Fig. 5. From an overview perspective, each agent only input its observed information and the corresponding actions of its own are mapped by two communication steps T^1 and T^2. At each communication step, there are two input vectors of the module f_n^m (here m is the index of communication step and n is the index of agent): the hidden state vector h_n^m and the communication vector c_n^m, and one output vector h_n^{m+1}. The above vectors can be calculated as follow [65]:

$$h_n^{m+1} = f_n^m\left(h_n^m, c_n^m\right) = \tan h\left(H^m h_n^m + C^m c_n^m\right), m = 0, 1 \tag{43}$$

$$c_n^{m+1} = \frac{1}{N-1}\sum_{n' \neq n} h_{n'}^{m+1} \tag{44}$$

where H^m and C^m denote the coefficient matrices. Substituting the Eq. (43) into Eq. (42), we can obtain:

$$h_n^{m+1} = \tan h(T^m h^m) \tag{45}$$

here the related coefficients can be computed as follow:

$$h^m = \left[h_1^m, \ldots, h_N^m\right]^T \tag{46}$$

$$T^m = \begin{pmatrix} H^m & \overline{C}^m & \cdots & \overline{C}^m \\ \overline{C}^m & H^m & \cdots & \overline{C}^m \\ \vdots & \vdots & \ddots & \vdots \\ \overline{C}^m & \overline{C}^m & \cdots & H^m \end{pmatrix}, \overline{C}^m = C^m \Big/ (N-1) \tag{47}$$

The softmax activation function is placed at the final layer $output = \mathrm{softmax}(h_n^M)$. Thus, the output of softmax can be regarded as the probability of final output action a_t^n at the state s_t^n at time step t. The further study of convergence of CommNet model can be found in Refs. [39,66].

Then, the actor-critic based reinforcement learning model is utilized to train the network to obtain the optimal policy, the specific training process of proposed algorithm is described in Algorithm 1. Firstly, the weights of actor and critic networks θ, φ and the weights of target networks $\widehat{\theta}$, $\widehat{\varphi}$ are initialized at the beginning of training process (line 3). After that, the agents repeat the follow procedures (line 4-line 17) to obtain the optimal policy of energy management.

For each episode at time-step t, the action $a_{t,i}$ of each agent i are selected based on the policy $\pi_\theta(S_t, A_t)$ learned by actor network (line 7). Then, the obtained actions are executed in the MEH environment and calculate the corresponding reward R_t as well as observe the set of next state $S_t^{'}$ (line 8). The transition pair $(S_t, A_t, R_t, S_t^{'})$ is constructed and stored in replay buffer (line 9).

After the replay buffer is fully stored, the agent come to the update period which the weight of target networks will be updated (line 11-line 16). A mini-batch of transition pairs are randomly selected from replay buffer and utilized to calculate the mean squared Bellman error to update the critic network. Here, the target action-value function of agent i δ_i can be formulated as:

$$\delta_i = R_i + \gamma Q_{\widehat{\omega}}(S_i^{'}, A_i^{'}) \tag{48}$$

The loss function of the critic network aims to optimize the minimum square error between the of target action-value function and the target value to update the weights of critic network φ, ich can be defined as:

$$L(\omega) = E_{(S,A,R,S')}\left[(\delta_i - Q_\omega(S_i, A_i))^2\right] = \frac{1}{\varphi}\sum_j \left[\delta_i - Q_\omega(S_i, A_i)\right]^2 \tag{49}$$

where φ represents the size of the sampled mini-batch. Thus, the parameter φ can be updated by minimizing the loss function (line 13), which is :

$$\omega^\pi = \underset{\omega}{\arg\min} L(\omega) \tag{50}$$

Then, the weights of actor network θ are updated based on the policy gradient theorem (line 14), which can be formulated as follow:

Algorithm 1. Learning procedures of proposed CommNet method

1: **Input:** states of all agent S_t.
2: **Output:** action for all EHs A_t.
3: **Initialize:** the weights of actor and critic networks θ, ϕ; the weights of target networks $\hat{\theta}$, $\hat{\phi}$.
4: **for** episode = 1 **to** max episode **do**
5: Initialize **MEH Environment**
6: **for** time step = 1 **to** max step **do**
7: Select actions $a_{t,i}$ for each agent i based on $\pi_\theta(S_t, A_t)$.
8: Execute the actions in MEH and obtain the reward R_t, observe the set of next state S_t'.
9: Store the transition pair (S_t, A_t, R_t, S_t') in the replay buffer.
10: **end for**
11: **If** time step >= update step **do**
12: Sample a mini-batch transition from the replay buffer.
13: Minimize the loss function to update the weights of critic network ϕ as Eq. (50) shows.
14: Update the weights of actor network θ by computed the policy gradient based on Eq. (51).
15: Update the weights of target networks based on Eq. (52).
16: **end if**
17: **end for**

Table 5
Hyper-parameters settings of the applied MADRL model.

Parameter	Value
Train episodes	8000
Learning rate of actor network	0.003
Learning rate of critic network	0.001
Batch size	256

$$\nabla_\theta J(\theta) \approx \frac{1}{\varphi}\sum_i \nabla_\theta \log \pi_\theta(S_i, A_i) Q_\omega(S_i, A_i) \tag{51}$$

Finally, the parameters of target networks can be updated to stabilize the training process (line 15), which is:

$$\begin{aligned} \widehat{\omega} &\leftarrow \eta\omega + (1-\eta)\widehat{\omega} \\ \widehat{\theta} &\leftarrow \eta\theta + (1-\eta)\widehat{\theta} \end{aligned} \tag{52}$$

where η denotes the learning rate. After that, the parameters of target networks are updated and which will use for solve the optimization problem after the training process is completed.

Algorithm 1. Learning procedures of proposed CommNet method

5. Case studies

To better illustrate the effectiveness of the proposed method, simulations are carried out based on the MEH system and specific numerical results are shown in this section.

5.1. Case settings

In this section, we construct a renewable energy powered MEH

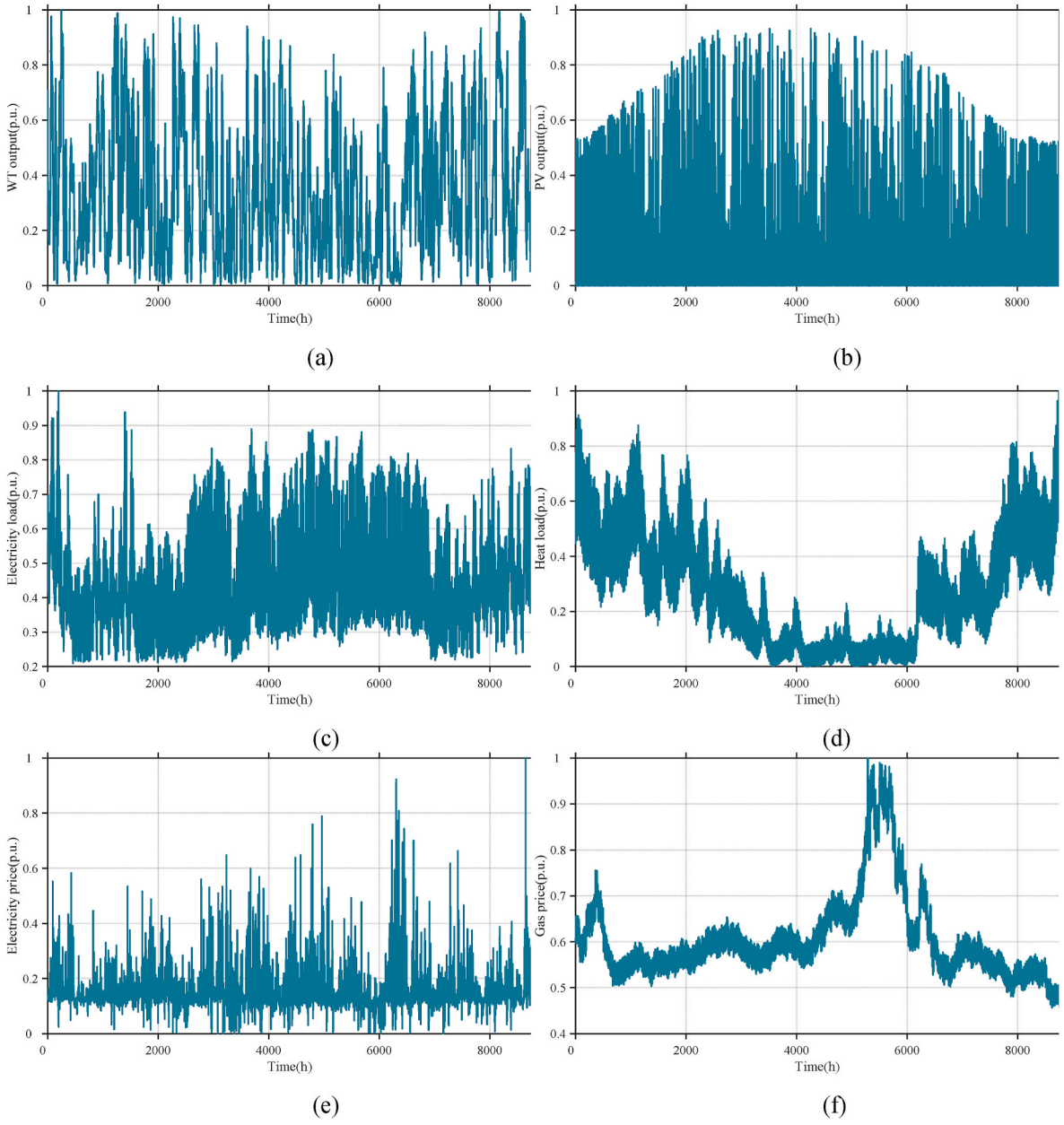

Fig. 6. The annual data for training the multi-agent: (a) WT output; (b) PV output; (c) Electricity load demand; (d) Heat load demand; (e) Electricity price; (f) Nature gas price.

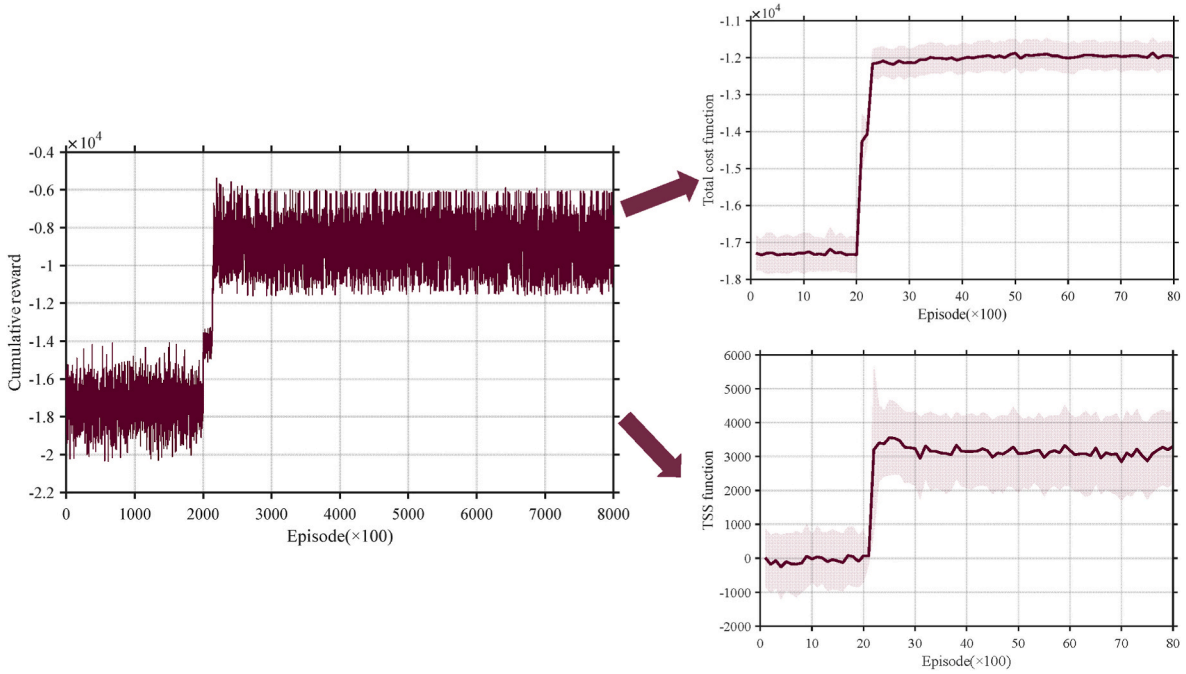

Fig. 7. The changes of cumulative reward of the MADRL method during training process.

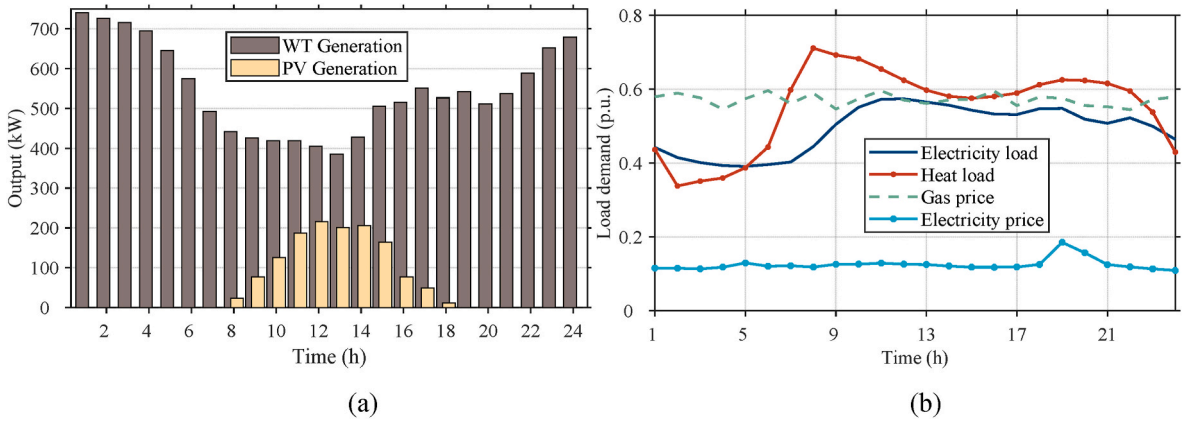

Fig. 8. The test data of selected day for the comparison experiments: (a) WT and PV generation; (b) load demand and price.

model which each EH includes PVT, WT, CHP, proposed P2A, NH_3 tank, boiler and furnace as the Fig. 1 shows to verify the effectiveness of proposed method. Note that there are three EHs in the MEH system. The installing capacities of PVT and WT are 300 kW and 1 MW, respectively. CHP can be powered by both nature gas and NH_3, the capacity of which is 500 kW. A 300 m^3 NH_3 tank is equipped to storage the produced NH_3 in a proper way to reduce the operating cost, whose storage level can only between 0.2 p.u. and 0.9 p.u. To ensure the heating capacity, an electrical boiler with 100 kW capacity and a nature gas furnace also with 100 kW capacity are built for supplying the heat energy.

Then, the energy management strategy of the MEH is achieved by the proposed MADRL method. The hyper-parameters of the applied MADRL method should be set before the training process, which are set as the Table 5 shown. To ensure the convergence of model, the total number of training episode is set to 8000, the learning rate of actor network and critic network are 0.003 and 0.001 and the mini-batch size is 256.

For training the agents, several data sets include the annual data of PVT output, the WT output, the electricity price, nature gas price, electricity load demand and heat load demand are applied, which are specifically shown in Fig. 6. It is worth noting that the training data is hourly based so that each data set includes 8760 (365 × 24) time points. The simulation model is constructed in MATLAB 2018b and the training procedure of MADRL method is conducted in Python based on a workstation computer with 32 GB RAM and Intel Core i9-10920X CPU.

5.2. *The training result of proposed MADRL method*

The CommNet based MADRL method has adopted to obtain the optimal coordinated energy management strategy of proposed MEH system. At each training episode, each agent observes the current state from the local EH and then transfers the information to the communication channel as the Section IV mentioned. After that, the agents choose and transfer the corresponding actions for environment models based on the obtained policy, then the reward can be calculated and feedbacked to the agents.

The changes of cumulative reward of proposed MADRL method during the training process is shown in Fig. 7. It can be seen that the left side of this figure is the cumulative reward function during the training process, which is combined with two parts: the total cost function (upper right side) and the TSS reward function (lower right side). As the figure

Table 6
System performance of EH 1 under difference schemes.

	Scheme 1	Scheme 2	Scheme 3	Scheme 4
Carbon emission (kg)	4890.7	5003.7	4985.4	5087.4
NH_3 yield (m^3)	1703.1	1480.8	1695.5	1693.3
Operation cost ($)	1157.9	1177.1	1185.3	1202.0
Maintenance cost ($)	176.6	177.1.	181.5	183.8
Total operating cost ($)	1334.5	1354.2	1366.8	1385.8

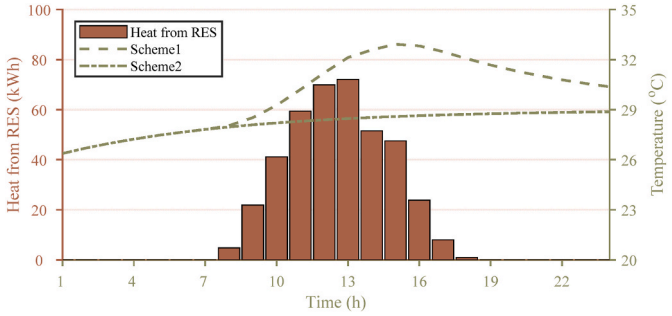

Fig. 9. The daily change of reaction temperature under different heating conditions.

shows, the cumulative reward stays in a low level with an average value −18000 in the initial exploration stage which means the agents cannot obtain the optimal energy management of MEH to reduce the operating cost and the carbon emissions. However, the cumulative reward gradually increases to −8500 after the ceaseless interaction between agents and environment, which indicates the better energy management of MEH has been obtained by the agents. Besides, it can be obtained that both the total cost function and the TSS reward function have also increased gradually during the training process, which means the agents have learned an energy management policy to minimize the total cost of MEH system and guided the gas tank to take suitable actions at each time-step. The reason is that the experiences learned from the interaction can facilitate the agents to update the parameters of networks and obtain the energy management policy. Finally, the training process of proposed MADRL method is convergent as the Fig. 7 shows, which indicates the agents have obtained the optimal or close-to-optimal energy management policy of proposed MEH.

5.3. Optimization results of the MEH based on proposed MADRL approach

To investigate the superiority of the constructed MEH system and verify the effectiveness of proposed MADRL method, four schemes are analyzed and listed:

Scheme 1this scheme constructs a renewable energy powered MEH model with P2A and gas tank as described in Section II. The objective is to reduce the operating cost and the carbon dioxide emissions as shown in Section III.
Scheme 2different from the scheme 1, this scheme removes the external heating for P2A devices as the Eq. (14) shown so that the yield of NH_3 will decrease.
Scheme 3this scheme does not equip with NH_3 tank, which the operating cost will increase as the influence of varying price of nature gas is not fully considered.
Scheme 4the system configuration of this scheme is same as the scheme 1, but the objective is only to minimize the operating cost without considering the influence of carbon tax so that the carbon dioxide emissions will increase.

After that, the well-trained agents are utilized to generate the energy management strategy of MEH for energy dispatching of each device based on the policy. The renewable energy generation data and load demand as well as price data of the selected test day as shown in Fig. 8

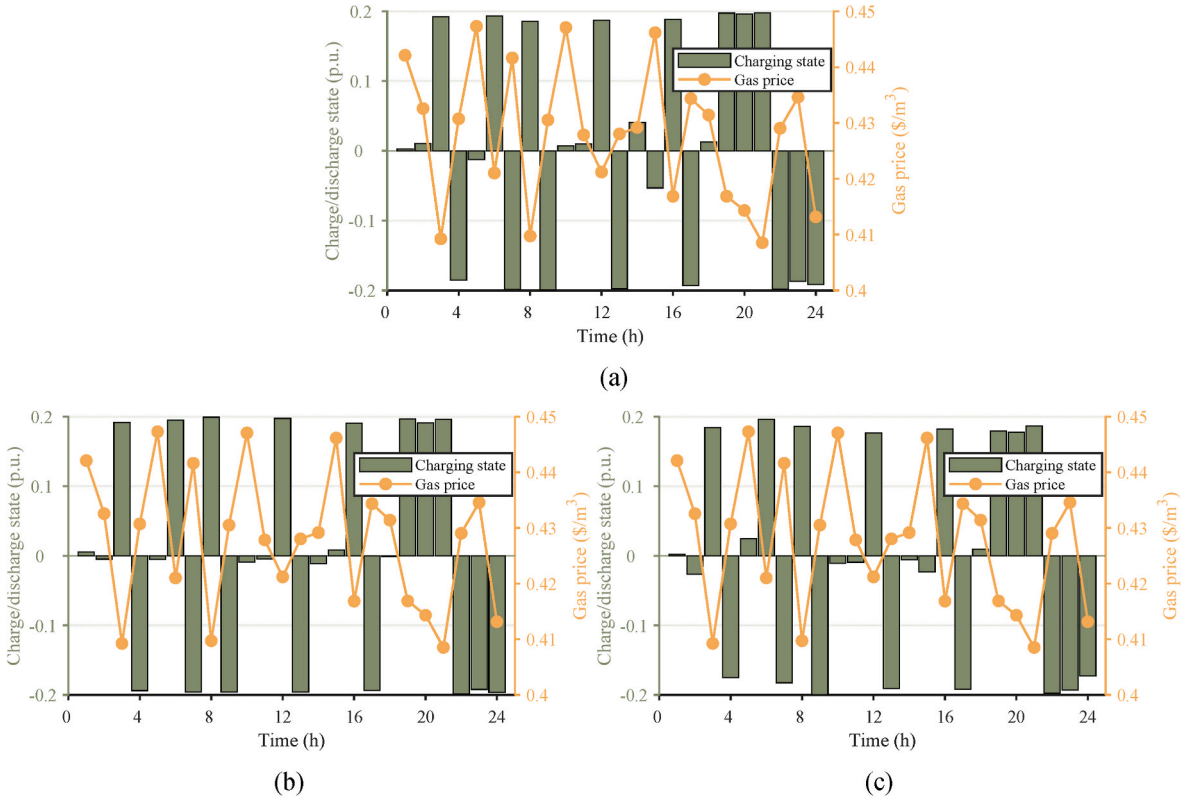

Fig. 10. State change of NH_3 tank under different schemes: (a) scheme 1; (b) scheme 2; (c)scheme 4.

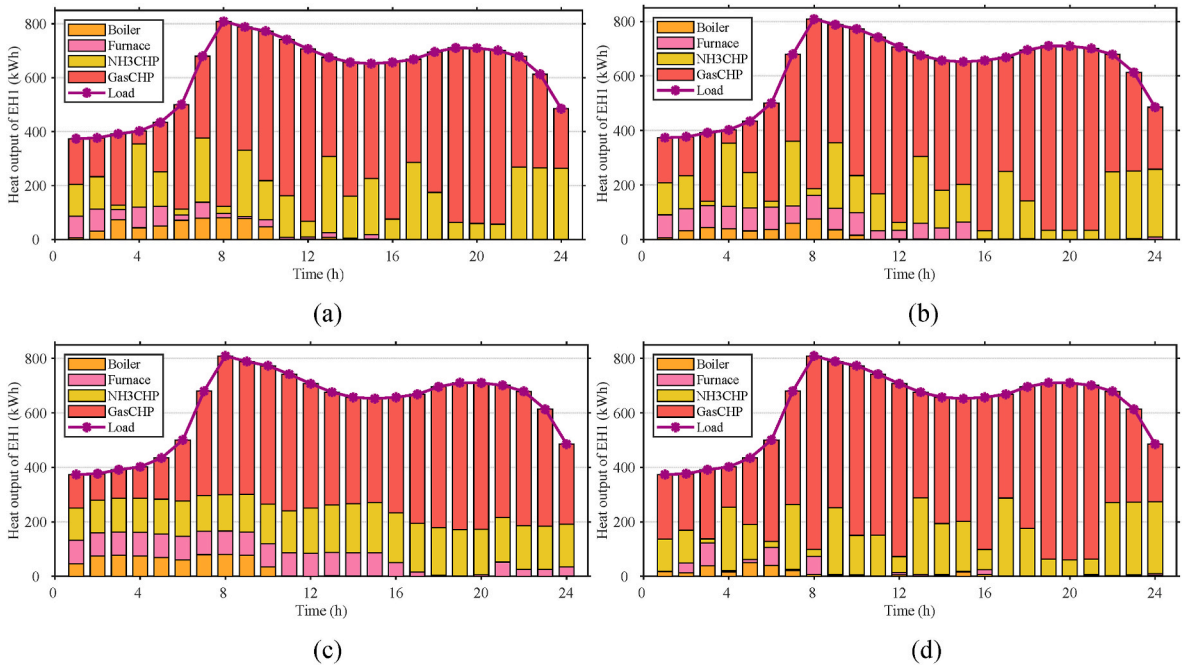

Fig. 11. The optimized energy management of thermal dispatch of EH 1: (a) scheme 1; (b) scheme 2; (c)scheme 3; (d)scheme 4.

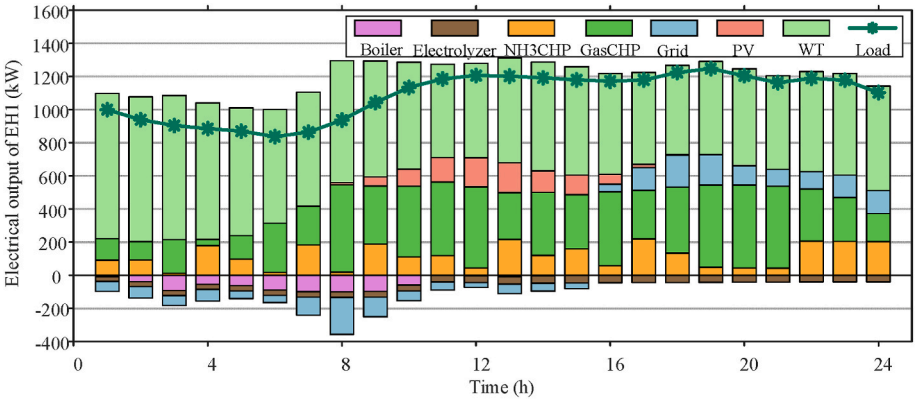

Fig. 12. The optimized energy management of electricity dispatch of EH 1 in scheme 1.

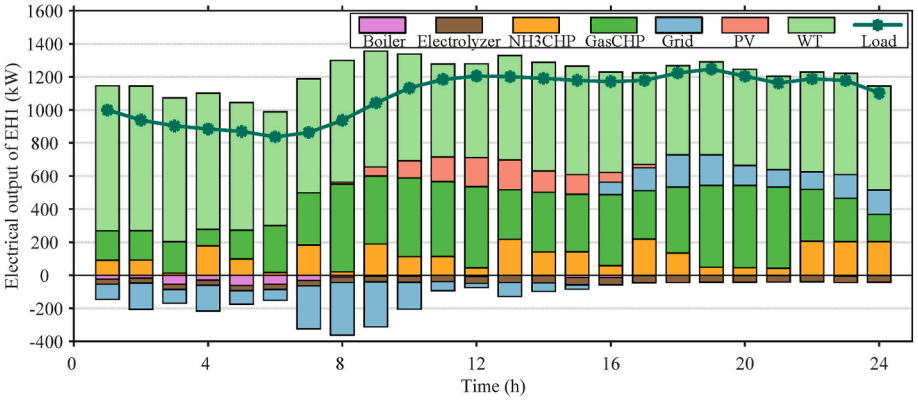

Fig. 13. The optimized energy management of electricity dispatching of EH 1 in scheme 4.

are utilized to further evaluate the performance of the well-trained agents. The optimization results are illustrated in Figs. 8–13 and Table 6. To avoid repetition as well as guarantee the effectiveness of analysis, only the simulation results of EH 1 are illustrated and analyzed in detail.

To evaluate the influence of external injected heating from renewable energy, the daily change of reaction temperature under scheme 1 and scheme 2 are shown in Fig. 9. It can be seen that the reaction

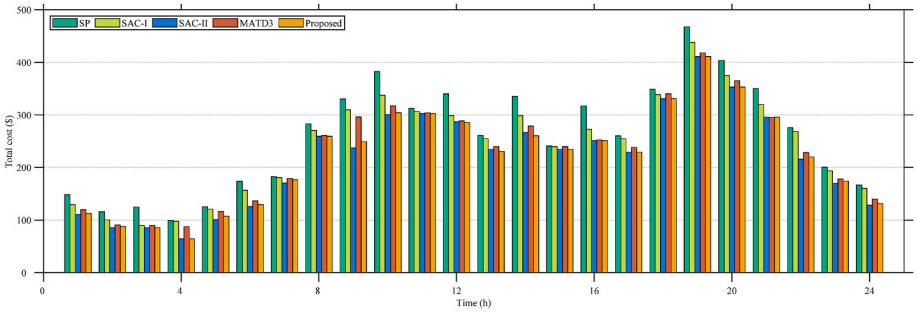

Fig. 14. The hourly cost of MEH system for different methods on the test day.

Table 7
Monthly cost of the MEH system under different methods.

Methods	Total cost ($)	Decline ratio
SP	194835.7	–
SAC-I	180490.3	7.36%
SAC-II	166331.4	14.62%
MATD3	176504.8	9.41%
Proposed	167016.3	14.28%

temperature of scheme 2 stays in a low level (below 29 °C) without extra heating, but the overall temperature change shows an upward trend due to the reaction entropy. However, the reaction temperature of scheme 1 can rise to 32.92 °C as the injected thermal power form renewable energy can heat the electrolyzer during the noon time. Note that the reaction temperature is closely related to the Faradaic efficiency, and which will further influence the yield of electrolyzer. Thus, the scheme 1 has better performance for NH_3 yield enhancement compares with scheme 2. Besides, the hourly nature gas price and the state change of NH_3 gas tank under different schemes are shown in Fig. 10. It can be found that the tank charging when gas price is low and discharging when the price is high, which can effectively reduce the cost of importing nature gas. Hence, the above results effectively demonstrate that the proposed approach can efficiently schedule the NH_3 tank in real-time to reduce the operating cost after the agents are trained.

Based on the well-trained agents, the optimized energy management strategy of thermal energy dispatching in EH 1 is derived, which is shown in Fig. 11. It can be found that all heat generating equipment are cooperatively dispatched to meet the heat demand at each time-step. Compared with scheme 1, the electrical boiler and gas furnace of EH 1 in scheme 2 generate more heat energy to satisfy the heat demand as the lower production of NH_3. The NH_3 produced in real-time is all transferring to the NH_3 CHP to generate electrical and thermal energy since no gas tank equipped in scheme 3. In scheme 4, the heat form gas furnace is less than that in scheme 1 and more heat is provided by gas CHP as the carbon tax is not considered. Furthermore, the optimized energy management strategy of electricity dispatching of EH 1 in scheme 1 and scheme 4 are illustrated in Figs. 12 and 13, respectively. It can be seen that the electricity productions from both renewable energy and CHP equipment are utilized to meet load demand and redundant electricity is sold to the upper grid. Because the carbon tax is not considered in scheme 4, the production of gas CHP is much higher than that in scheme 1 and more redundant electricity has been traded to the electricity utility, which means more nature gas is need and the more carbon dioxide is produced in this scheme.

Furthermore, the system performance of EH 1 under four schemes are listed in Table 6. It can be found that the scheme 1 has the lowest carbon emissions and the scheme 4 has the most carbon emissions, which 196.7 kg carbon dioxide are reduced when the carbon tax is considered. The NH3 yield of scheme 1, 3,4 are similar due to the external heating injected, which are much higher than scheme 2. The NH_3 yield of scheme 1 increases by 222.3 m^3 compared with that in scheme 2 due to the external heating. Besides, the proposed approach can coordinately schedule devices of EH 1 to reduce operating cost (includes purchase cost and maintenance cost but no carbon cost) in scheme 1, so which has the lowest operating cost. In conclusion, the proposed scheme (scheme 1) has the superiority in minimizing the operating cost and the carbon dioxide emissions when compared with other three schemes.

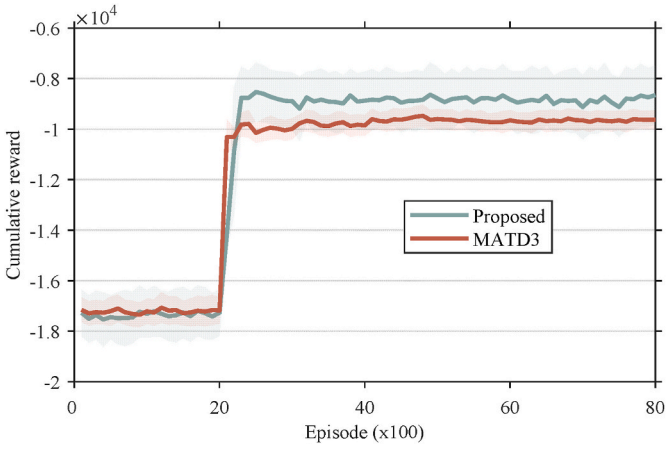

Fig. 15. The changes of cumulative reward of the propose and MATD3 methods.

5.4. Comparison results with other benchmark methods

To demonstrate the effectiveness of the proposed MADRL approach, the stochastic optimization (SP) approach [67] and MATD3 method [68] are utilized for a comparison analysis. Note that the SP method is a common optimization approach that is utilized to deal with the uncertainties and MATD3 is a traditional MADRL method. Besides, to further demonstrate the necessity of applying MADRL, a SADRL method soft actor-critic (SAC) is utilized for conducting the comparison experiment. Note that the SAC-based method is trained in a centralized manner with global state. Since the SADRL method easily faces to single-point failure caused by interruption of communication [40], a single-point failure is set on the communication line between the agent and EH 1 to obtain the performance of SAC-based method under this condition. Thus, two SAC-based methods: SAC-I (with single-point failure) and SAC-II (without single-point failure) are applied for conducting the experiment. The comparison results are specifically shown in Figs. 14–16 and the Table 7.

The total cost of MEH (includes operation cost, maintenance cost as well as carbon tax cost) at each time step on the test day with the three methods are shown in Fig. 14. It can be found that the DRL-based methods can achieve better performance on total cost reduction compared with SP method at every time-step, which indicates that the DRL-based method can obtain better energy management strategy to

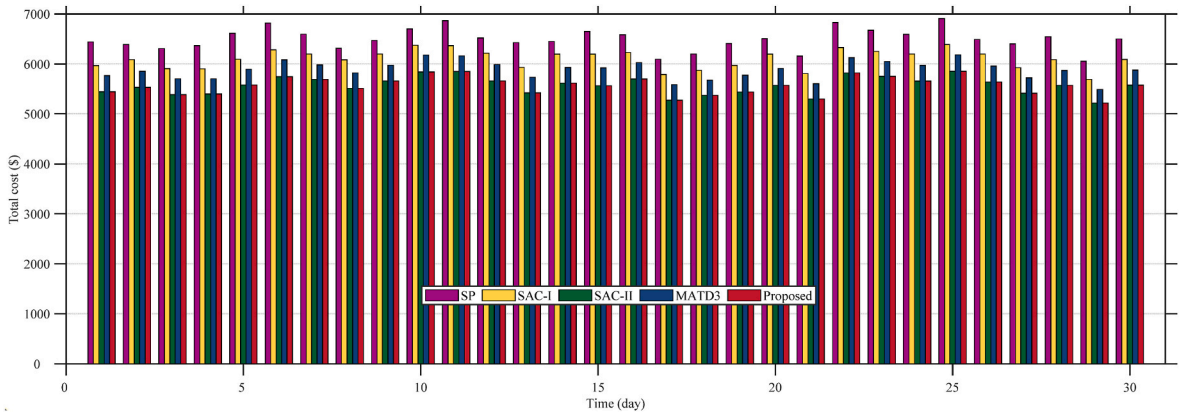

Fig. 16. Daily cost of MEH system with different control methods.

minimize the total cost of MEH. Moreover, it can be observed that the SAC-II can achieve better performance on total cost minimizing since which is trained in a centralized manner with global information. But the centralized methods always lead to concern about data privacy and single-point failure. Once the single-point failure occurs, the control performance of centralized method (SAC-I) will be severely affected. Thus, the MADRL methods which conduct optimization in a decentralized manner with local information are ideal substitutes.

To further verify the performance of the MADRL methods, the changes of cumulative reward of the proposed and MATD3 methods during the training process are illustrated in Fig. 15. It can be observed that the proposed method can achieve higher cumulative reward than MATD3 algorithm, which indicates that the proposed MADRL method can learn a better energy management strategy of MEH system. The reason is that the proposed MADRL can learn a better policy by collaboration based on CommNet. Before making a decision, each agent can observe the policies change of other agents by common communication channel and then accordingly change its own policy. Thus, agents in CommNet are able to obtain the optimal policies in a cooperative way.

Furthermore, to better demonstrate the superiority of the proposed method on a long-term operation, a further comparison is conducted on 30-day data selected from the dataset. The daily cost of MEH system under different control methods is illustrated in Fig. 16. The total monthly cost of MEH system of the three methods is listed in Table 7. It can be found that the MEH system controlled by the proposed method and SAC-II method can achieve the similar daily cost for each day, which indicates the proposed method can effectively reduce the total cost. From a quantitative perspective, the monthly cost of MEH system under the proposed method can be reduced up to 14.28% compared with SP based method. The monthly cost of MEH system can be reduced 9488.5 $ compared with MATD3 method, which means the proposed method can achieve better performance on cost minimizing. Besides, it can be observed that the monthly cost of MEH system under the SAC-I method is 14158.9 $ higher compared with SAC-II method, which indicates that the control performance can be highly reduced once the single-point failure occurs. Thus, the proposed method can achieve better performance on energy management when considering the robustness.

6. Conclusion

In this paper, a thermo-electrochemical effect based P2A technology is utilized to construct a MEH system to reduce the operating cost and carbon dioxide emissions of MEH. Then, the energy management problem of MEH is formed as a Markov game and solved by a novel MADRL algorithm called CommNet. The related conclusions can be summarized as follow:

1) The proposed renewable energy powered MEH system integrates with the thermo-electrochemical effect based P2A facility has shown superiority than other three schemes, which the operating cost can be reduced 19.7 $ at least and the carbon emissions can be reduced by 3.87% of one EH.
2) The energy management problem of proposed MEH is formed as a Markov game and a novel MADRL algorithm called CommNet is utilized to solve this problem. After 8000 episodes training, the agents have steadily converged and the results show that the proposed MADRL method has obtained better policy than MATD3 method and SP-based method, which the total monthly cost of MEH can be reduced by 14.28%.
3) In future work, more researches are still needed to improve the convergence speed of proposed MADRL method and more attention should be paid to enhance the robustness and generalization of the agents when applied in real-world scenarios.

CRediT authorship contribution statement

Kang Xiong: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Supervision. **Weihao Hu:** Methodology, Software, Validation. **Di Cao:** Investigation, Software, Validation. **Sichen Li:** Resources, Software, Validation. **Guozhou Zhang:** Supervision, Validation, Software. **Wen Liu:** Resources, Writing – review & editing. **Qi Huang:** Writing – review & editing, Supervision. **Zhe Chen:** Writing – review & editing, Supervision, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Sichuan Science and Technology Program under Grant 2023YFG0108.

References

[1] J.C.Y. Lee, C. Draxl, L.K. Berg, Evaluating wind speed and power forecasts for wind energy applications using an open-source and systematic validation framework, Renew. Energy 200 (2022) 457–475.
[2] E. Magadley, R. Kabha, I. Yehia, Outdoor comparison of two organic photovoltaic panels: the effect of solar incidence angles and incident irradiance, Renew. Energy 173 (2021) 721–732.
[3] S. Lv, H. Wang, X. Meng, C. Yang, M. Wang, Optimal capacity configuration model of power-to-gas equipment in wind-solar sustainable energy systems based on a novel spatiotemporal clustering algorithm: a pathway towards sustainable development, Renew. Energy 201 (2022) 240–255.

[4] A. Merabet, A. Al-Durra, E.F. El-Saadany, Energy management system for optimal cost and storage utilization of renewable hybrid energy microgrid, Energy Convers. Manag. 252 (2022), 115116.
[5] K. Rahbar, J. Xu, R. Zhang, Real-time energy storage management for renewable integration in microgrid: an off-line optimization approach, IEEE Trans. Smart Grid 6 (1) (2015) 124–134.
[6] U.G.K. Mulleriyawage, W.X. Shen, Impact of demand side management on optimal sizing of residential battery energy storage system, Renew. Energy 172 (2021) 1250–1266.
[7] D. Xu, B. Zhou, Q. Wu, C.Y. Chung, C. Li, S. Huang, S. Chen, Integrated modelling and enhanced utilization of power-to-ammonia for high renewable penetrated multi-energy systems, IEEE Trans. Power Syst. 35 (6) (2020) 4769–4780.
[8] S.C. Li, W.H. Hu, D. Cao, T. Dragicevic, Q. Huang, Z. Chen, F. Blaabjerg, Electric vehicle charging management based on deep reinforcement learning, JOURNAL OF MODERN POWER SYSTEMS AND CLEAN ENERGY 10 (3) (2022) 719–730.
[9] H. Tebibel, Methodology for multi-objective optimization of wind turbine/battery/electrolyzer system for decentralized clean hydrogen production using an adapted power management strategy for low wind speed conditions, Energy Convers. Manag. 238 (2021), 114125.
[10] L. Zhong, E. Yao, H. Zou, G. Xi, Thermodynamic and economic analysis of a directly solar-driven power-to-methane system by detailed distributed parameter method, Appl. Energy 312 (2022), 118670.
[11] C. Tarhan, M.A. Çil, A study on hydrogen, the clean energy of the future: hydrogen storage methods, J. Energy Storage 40 (2021), 102676.
[12] M. Fasihi, R. Weiss, J. Savolainen, C. Breyer, Global potential of green ammonia based on hybrid PV-wind power plants, Appl. Energy 294 (2021), 116170.
[13] S. Giddey, S.P.S. Badwal, A. Kulkarni, Review of electrochemical ammonia production technologies and materials, Int. J. Hydrogen Energy 38 (34) (2013) 14576–14594.
[14] W.H. Avery, A role for ammonia in the hydrogen economy, Int. J. Hydrogen Energy 13 (12) (1988) 761–773.
[15] D. Mei, X. Qiu, H. Liu, Q. Wu, S. Yu, L. Xu, T. Zuo, Y. Wang, Progress on methanol reforming technologies for highly efficient hydrogen production and applications, Int. J. Hydrogen Energy (2022).
[16] J. Ikäheimo, J. Kiviluoma, R. Weiss, H. Holttinen, Power-to-ammonia in future North European 100 % renewable power and heat system, Int. J. Hydrogen Energy 43 (36) (2018) 17295–17308.
[17] Z. Wan, Y. Tao, J. Shao, Y. Zhang, H. You, Ammonia as an effective hydrogen carrier and a clean fuel for solid oxide fuel cells, Energy Convers. Manag. 228 (2021), 113729.
[18] J.M. Modak, Haber process for ammonia synthesis, Resonance 16 (12) (2011) 1159–1167.
[19] M. Nazemi, S.R. Panikkanvalappil, M.A. El-Sayed, Enhancing the rate of electrochemical nitrogen reduction reaction for ammonia synthesis under ambient conditions using hollow gold nanocages, Nano Energy 49 (2018) 316–323.
[20] J. Chatt, A.J. Pearman, R.L. Richards, The reduction of mono-coordinated molecular nitrogen to ammonia in a protic environment, Nature 253 (5486) (1975) 39–40.
[21] R. Lan, J.T. Irvine, S. Tao, Synthesis of ammonia directly from air and water at ambient temperature and pressure, Sci. Rep. 3 (2013) 1145.
[22] D. Wen, M. Aziz, Design and analysis of biomass-to-ammonia-to-power as an energy storage method in a renewable multi-generation system, Energy Convers. Manag. 261 (2022), 115611.
[23] J. Wang, J. Wang, P. Zhao, Y. Dai, Proposal and thermodynamic assessment of a new ammonia-water based combined heating and power (CHP) system, Energy Convers. Manag. 184 (2019) 277–289.
[24] A.A.M. Aljabery, H. Mehrjerdi, S. Mahdavi, R. Hemmati, Multi carrier energy systems and energy hubs: comprehensive review, survey and recommendations, Int. J. Hydrogen Energy 46 (46) (2021) 23795–23814.
[25] T. Liu, D. Zhang, T. Wu, Standardised modelling and optimisation of a system of interconnected energy hubs considering multiple energies—electricity, gas, heating, and cooling, Energy Convers. Manag. 205 (2020), 112410.
[26] D. Zhu, B. Yang, Q. Liu, K. Ma, S. Zhu, C. Ma, X. Guan, Energy trading in microgrids for synergies among electricity, hydrogen and heat networks, Appl. Energy 272 (2020).
[27] Y. Zou, Y. Xu, F. Xue, N. R.T, S. Boon-Hee, Transactive energy system in active distribution networks: a comprehensive review, CSEE J. Power Energy Syst. (2022).
[28] D. Zhu, B. Yang, C. Ma, Z. Wang, S. Zhu, K. Ma, X. Guan, Stochastic gradient-based fast distributed multi-energy management for an industrial park with temporally-coupled constraints, Appl. Energy 317 (2022), 119107.
[29] A. Zakaria, F.B. Ismail, M.S.H. Lipu, M.A. Hannan, Uncertainty models for stochastic optimization in renewable energy applications, Renew. Energy 145 (2020) 1543–1571.
[30] V.V. Thang, T. Ha, Q. Li, Y. Zhang, Stochastic optimization in multi-energy hub system operation considering solar energy resource and demand response, Int. J. Electr. Power Energy Syst. 141 (2022), 108132.
[31] Y. Zou, Y. Xu, C. Zhang, A Risk-averse adaptive stochastic optimization method for transactive energy management of a multi-energy microgrid, IEEE Trans. Sustain. Energy (2023) 1–12.
[32] A. Heidari, S.S. Mortazavi, R.C. Bansal, Stochastic effects of ice storage on improvement of an energy hub optimal operation including demand response and renewable energies, Appl. Energy 261 (2020), 114393.
[33] D. Cao, J. Zhao, W. Hu, F. Ding, N. Yu, Q. Huang, Z. Chen, Model-free voltage control of active distribution system with PVs using surrogate model-based deep reinforcement learning, Appl. Energy 306 (2022).
[34] B. Poursmaeil, P. Hosseinpour Najmi, S. Najafi Ravadanegh, Interconnected-energy hubs robust energy management and scheduling in the presence of electric vehicles considering uncertainties, J. Clean. Prod. 316 (2021), 128167.
[35] X. Lu, Z. Liu, L. Ma, L. Wang, K. Zhou, S. Yang, A robust optimization approach for coordinated operation of multiple energy hubs, Energy 197 (2020), 117171.
[36] Y. Zhou, Z. Ma, J. Zhang, S. Zou, Data-driven stochastic energy management of multi energy system using deep reinforcement learning, Energy 261 (2022), 125187.
[37] H. Tan, H. Zhang, J. Peng, Z. Jiang, Y. Wu, Energy management of hybrid electric bus based on deep reinforcement learning in continuous state and action space, Energy Convers. Manag. 195 (2019) 548–560.
[38] D. Cao, W. Hu, J. Zhao, Q. Huang, Z. Chen, F. Blaabjerg, A multi-agent deep reinforcement learning based voltage Regulation using coordinated PV inverters, IEEE Trans. Power Syst. 35 (5) (2020) 4120–4123.
[39] Z. Zhang, Y. Wan, J. Qin, W. Fu, Y. Kang, A deep RL-based algorithm for coordinated charging of electric vehicles, IEEE Trans. Intell. Transport. Syst. (2022) 1–11.
[40] G. Zhang, W. Hu, D. Cao, Z. Zhang, Q. Huang, Z. Chen, F. Blaabjerg, A multi-agent deep reinforcement learning approach enabled distributed energy management schedule for the coordinate control of multi-energy hub with gas, electricity, and freshwater, Energy Convers. Manag. 255 (2022).
[41] D. Zhu, B. Yang, Y. Liu, Z. Wang, K. Ma, X. Guan, Energy management based on multi-agent deep reinforcement learning for a multi-energy industrial park, Appl. Energy 311 (2022).
[42] M. Ahrarinouri, M. Rastegar, A.R. Seifi, Multiagent reinforcement learning for energy management in residential buildings, IEEE Trans. Ind. Inf. 17 (1) (2021) 659–666.
[43] B. Zhang, W. Hu, A.M.Y.M. Ghias, X. Xu, Z. Chen, Multi-agent deep reinforcement learning based distributed control architecture for interconnected multi-energy microgrid energy management and optimization, Energy Convers. Manag. 277 (2023).
[44] K. Xiong, D. Cao, G. Zhang, Z. Chen, W. Hu, Coordinated volt/VAR control for photovoltaic inverters: a soft actor-critic enhanced droop control approach, Int. J. Electr. Power Energy Syst. 149 (2023).
[45] E. Bisengimana, J. Zhou, M. Binama, Y. Yuan, Numerical investigation on the factors influencing the temperature distribution of photovoltaic/thermal (PVT) evaporator/condenser for heat pump systems, Renew. Energy 194 (2022) 885–901.
[46] A. Behzadi, A. Habibollahzade, P. Ahmadi, E. Gholamian, E. Houshfar, Multi-objective design optimization of a solar based system for electricity, cooling, and hydrogen production, Energy 169 (2019) 696–709.
[47] G. Kosmadakis, D. Manolakos, G. Papadakis, Simulation and economic analysis of a CPV/thermal system coupled with an organic Rankine cycle for increased power generation, Sol. Energy 85 (2) (2011) 308–324.
[48] G. Zhang, W. Hu, D. Cao, W. Liu, R. Huang, Q. Huang, Z. Chen, F. Blaabjerg, Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach, Energy Convers. Manag. 227 (2021), 113608.
[49] J. Hou, M. Yang, J. Zhang, Recent advances in catalysts, electrolytes and electrode engineering for the nitrogen reduction reaction under ambient conditions, Nanoscale 12 (13) (2020) 6900–6920.
[50] Y. Ren, C. Yu, L. Wang, X. Tan, Z. Wang, Q. Wei, Y. Zhang, J. Qiu, Microscopic-level insights into the mechanism of enhanced NH3 synthesis in plasma-enabled cascade N2 oxidation-electroreduction system, J. Am. Chem. Soc. 144 (23) (2022) 10193–10200.
[51] A. Ursua, L.M. Gandia, P. Sanchis, Hydrogen production from water electrolysis: current status and future trends, Proc. IEEE 100 (2) (2012) 410–426.
[52] J. Hou, M. Yang, J.J.N. Zhang, Recent Advances in Catalysts, Electrolytes and Electrode Engineering for the Nitrogen Reduction Reaction under Ambient Conditions, vol. 12, 2020.
[53] K. Karimi, S. Fatemi, Methane capture and nitrogen purification from a nitrogen rich reservoir by pressure swing adsorption; experimental and simulation study, J. Environ. Chem. Eng. 9 (5) (2021), 106210.
[54] D. Bao, Q. Zhang, F.L. Meng, H.X. Zhong, M.M. Shi, Y. Zhang, J.M. Yan, Q. Jiang, X.B. Zhang, Electrochemical reduction of N2 under ambient conditions for artificial N2 fixation and renewable energy storage using N2/NH3 cycle, Adv. Mater. 29 (3) (2017).
[55] B. Zhou, D. Xu, C. Li, C.Y. Chung, Y. Cao, K.W. Chan, Q. Wu, Optimal scheduling of biogas–solar–wind renewable portfolio for multicarrier energy supplies, IEEE Trans. Power Syst. 33 (6) (2018) 6229–6239.
[56] L. Li, S. Zhang, Techno-economic and environmental assessment of multiple distributed energy systems coordination under centralized and decentralized framework, Sustain. Cities Soc. 72 (2021).
[57] Y. Chen, X. Hu, W. Xu, Q. Xu, J. Wang, P.D. Lund, Multi-objective optimization of a solar-driven trigeneration system considering power-to-heat storage and carbon tax, Energy 250 (2022).
[58] R. Zeng, X. Zhang, Y. Deng, H. Li, G. Zhang, An off-design model to optimize CCHP-GSHP system considering carbon tax, Energy Convers. Manag. 189 (2019) 105–117.
[59] H. Zhang, L. Wang, J. Van herle, F. Maréchal, U. Desideri, Techno-economic comparison of green ammonia production processes, Appl. Energy 259 (2020), 114135.
[60] G. Zhang, W. Hu, D. Cao, Z. Zhang, Q. Huang, Z. Chen, F. Blaabjerg, A multi-agent deep reinforcement learning approach enabled distributed energy management schedule for the coordinate control of multi-energy hub with gas, electricity, and freshwater, Energy Convers. Manag. 255 (2022), 115340.

[61] S. Zhang, W. Hu, J. Du, C. Bai, W. Liu, Z. Chen, Low-carbon optimal operation of distributed energy systems in the context of electricity supply restriction and carbon tax policy: a fully decentralized energy dispatch strategy, J. Clean. Prod. 396 (2023).
[62] B. Zhang, W. Hu, X. Xu, T. Li, Z. Zhang, Z. Chen, Physical-model-free intelligent energy management for a grid-connected hybrid wind-microturbine-PV-EV energy system via deep reinforcement learning approach, Renew. Energy 200 (2022) 433–448.
[63] D. Cao, J. Zhao, W. Hu, J. Hu, Physics-informed Graphical Representation-Enabled Deep Reinforcement Learning for Robust Distribution System Voltage Control, IEEE Transations on Smart Grid, 2023 (early access).
[64] D. Cao, W. Hu, J. Zhao, G. Zhang, B. Zhang, Z. Liu, Z. Chen, F. Blaabjerg, Reinforcement learning and its applications in modern power and energy systems: a review, J. Modern Power Syst. Clean Energy 8 (6) (2020) 1029–1042.
[65] S. Sukhbaatar, A. Szlam, R. Fergus, Learning Multiagent Communication with Backpropagation, Neural Information Processing Systems, NIPS 2016, 2016.
[66] M.A. Khamsi, W.A. Kirk, An introduction to metric spaces and fixed point theory (Khamsi/An introduction) || appendix, 10.1002/9781118033074, Set Theory (2001) 273–287.
[67] M. Tostado-Véliz, P. Arévalo, F. Jurado, A comprehensive electrical-gas-hydrogen Microgrid model for energy management applications, Energy Convers. Manag. 228 (2021), 113726.
[68] Z.-c. Qiu, Y. Yang, X.-m. Zhang, Reinforcement learning vibration control of a multi-flexible beam coupling system, Aero. Sci. Technol. 129 (2022), 107801.