Energy Reports 8 (2022) 15067–15080

Contents lists available at ScienceDirect

Energy Reports

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

Research paper

A double-deck deep reinforcement learning-based energy dispatch strategy for an integrated electricity and district heating system embedded with thermal inertial and operational flexibility

Bin Zhang [a], Amer M.Y.M. Ghias [b,*], Zhe Chen [a]

[a] Department of Energy Technology, Aalborg University, Aalborg, Denmark
[b] School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 11 July 2022
Received in revised form 30 October 2022
Accepted 1 November 2022
Available online xxxx

Keywords:
Integrated energy systems
Renewable energy
Machine learning
Deep reinforcement learning
Energy dispatch

ABSTRACT

With the high penetration of wind power connected to the integrated electricity and district heating systems (IEDHSs), wind power curtailment still inevitably occurs in the traditional IEDHS dispatch. Focusing on the flexibilities of the IEDHS is considered to be a beneficial solution to further promote the integration of wind power. In the district heating network, the thermal inertia is utilized to improve such flexibility. Therefore, an IEDHS dispatch model considering the thermal inertia of district heating network and operational flexibility of generators is proposed in this paper. In addition, to avoid the tendency of traditional reinforcement learning (RL) to fall into local optimality when solving high-dimensional problems, a double-deck deep RL (D3RL) framework is proposed in this study. D3RL combines with a deep deterministic policy gradient (DDPG) agent in the upper level and a conventional optimization solver in the lower level to simplify the action and reward design. In the simulation, the proposed model considering the transmission time delay characteristics of the district heating network and the operational flexibility of generators is verified in four scheduling scenarios. Besides, the superiority of the proposed D3RL method is validated in a larger IEDHS. Numerical results show that the considered scheduling model can use the heat storage characteristics of heating pipelines, reduce operating costs, improve the operational flexibility and encourage wind power utilization. Compared with traditional RL, the proposed optimization method can improve its training speed and convergence performance.

1. Introduction

Increasingly, serious environmental issues and energy crisis concerns caused by conventional fossil fuels are bringing great pressure on power system design and implementation (Obama, 2017). Globally, governments are responding with plans for clean energy development, aiming to promote a thriving renewable energy (RE) industry (Yang, 2021). However, the fluctuations of RE resources bring increasing burden to power grid regulation. Instead, due to the characteristics of internal autonomous operation and external friendly power, integrated energy systems (IES) have been recognized as an effective means of distributed RE consumption in the past several years (Nejabtkhah and Li, 2015). Especially, researches on integrated electricity and district heating systems (IEDHS) are receiving increasing attention. Moharram et al. (2022) assessed the capability of the district integrated electrical and heating systems to dissipate solar energy. With temperature feedback mechanism of district heating system (DHS) in mind, Xu et al. (2021) discussed the potential of the integrated electricity and heating system to further accommodate wind power. According to 100+ literature review, the barriers to reach flexible district electricity and heating system are analyzed and characterized into different technology types (Sneum, 2021). Instead of a centralized control solution, Chen et al. (2021) proposed a distributed coordinated operation strategy for a regional large-scale IEDHS to improve energy efficiency, which considers the thermal inertial of pipeline networks and different heating modes of buildings.

However, with the increasing intention of clean energy, the phenomenon of wind power curtailments have inevitably occurred, especially in the heating season, which has caused unnecessary economic losses. For example, under the conventional model of heat to electricity, combined heat and power plant (CHP) unit supplies the high heat load in winter and generates excess electricity, which limits the scheduling flexibility of the CHP and simultaneously causes the issue of large amounts of wind curtailments (Bagherian and Mehranzamir, 2020). As a result, a series of

* Corresponding author.
E-mail addresses: bzh@et.aau.dk (B. Zhang), amer.ghias@ntu.edu.sg (A.M.Y.M. Ghias), zch@et.aau.dk (Z. Chen).

https://doi.org/10.1016/j.egyr.2022.11.028

Nomenclature

AI	Artificial Intelligence
CHP	Combined heat and power unit
DDPG	Deep deterministic policy gradient
DHS	District heating system
D3RL	Double-deck deep reinforcement learning
DNN	Deep neural network
EB	Electric boiler
IES	Integrated energy system
IEDHS	Integrated electricity and district heating system
MDP	Markov decision process
ML	Machine Learning
MILP	Mixed-integer linear programming
RL	Reinforcement learning
RE	Renewable energy
ReLU	Rectified linear unit
WT	Wind turbine

researches are dedicated to the IEDHS, in which how to alleviate the issues of wind power curtailments in the heating season is being addressed. To cope with increasing wind capacity, Xia et al. (2022) established a bi-level planning-operation IEDHS framework, and the role of the participation of compressed air energy storage studied. Zhang et al. (2019b) proposed a novel integrated electricity and heat system covering local and national infrastructures, and assessed its performance on utilizing RE. Zhang et al. (2022a) integrated the reserve provision from large-scale heat pumps into the district electricity and heat networks to realize economic operation, wind power accommodation and load availability. Wang et al. (2020) used a district thermal energy storage to improve the energy utilization efficiency of the combined electricity and heat networks. While the above studies indeed improve the system's ability to accommodate RE to some extent, the majority of them rely on thermal storage devices such as storage tanks and heat pumps, which are not economically feasible. In fact, due to the transmission characteristics in the district heating pipelines, there is a time delay in the supply of energy from the heat source to the heat load. The network of pipelines in the heating system can be viewed as a thermal energy storage device. Zhang et al. (2021a) investigated the impact of heat energy storage in heat-supply net on the energy utilization efficiency of district heating system, which is described by the quantitative calculation model. However, the above literature neglects to promote RE integration by improving the operational flexibility of the units while considering the transmission delay.

Furthermore, despite sufficient thermal storage capacity in the heating pipelines and potential for thermal inertia to facilitate RE integration, RE utilization still faces significant challenges in some special occasions, and operational flexibility of the IEDHS itself should be emphasized. Ma et al. (2021) aims to mitigate the conflict between the inflexible operation of efficient CHP systems and the demand for grid flexibility improvement. A two-stage cogeneration dispatch model is proposed, in which lower stage quantifies the optimal trade-off between supply flexibility and conversion efficiency, and upper stage achieves global coordination by formulating a convex optimal power flow problem. Zhu and Li (2022) considered multiple modeling factors of thermal energy storages, such as heat transfer delay, mass flow rate and heat exchanger, to improve the flexibility of CHP units. Then, a decomposition coordination method is applied to solve the strongly complex non-linear problem. Daraei et al. (2021) integrated a hydrotreated pyrolysis oil production into the CHP plant for flexible energy supply. The integrated pyrolysis to CHP plants and onsite hydrogen use can improve flexibility and RE utilization. Therefore, a proper flexibility improvement method is worthy of investigation in the IEDHS researches.

In addition, traditional energy management ways (e.g., deterministic rules and abstract models) in the above existing literatures are mainly hampered by two dilemmas: (a) deterministic rules are difficult to cope with the time-varying parameters in non-stationary systems and may result in high costs; (b) the performance of the abstract model mostly relies on the experience of the modeler and may differs somewhat from the realistic model. Therefore, a model-free deep reinforcement learning (DRL) -based energy dispatch strategy is proposed in this paper. In recent years, with the rising enthusiasm of artificial intelligence (AI), machine learning (ML) algorithms are being widely adopted in smart grid, including voltage control (Sun and Qiu, 2021), operation costs optimization and electricity market bidding (Ye et al., 2020). The economic operation problem in an IES is formulated as a Markov decision process (MDP), and a DRL-based economic energy management strategy is proposed in Zhang et al. (2019a) and Yang et al. (2021). Indeed, these works aim to minimize economic operating costs with consideration of uncertainties of wind power, but flexibility of the CHP units is not fully guaranteed. Zhang et al. (2020) proposed a DRL-based dynamic energy management strategy for an IES to balance the flexibility of the unit while ensuring economic operation. However, the issue of how to effectively ensure the flexibility of the units while coping with the high dimensionality brought by multiple generators in the case of large-scale RE penetration has not been fully investigated.

In the view of above, a double-deck RL (D3RL) framework-based collaborative IEDHS dispatch model considering thermal inertia of heating pipelines and operational flexibility indicators is proposed in this paper. The dispatch objectives are to satisfy the intension of promoting integration of clean energy, reduce operating costs and improve operational flexibility of generators. Specifically, the main contributions of this paper are summarized below:

(1) A model-driven IEDHS energy management model is proposed to solve the non-convex multi-objective function without accurate thermal dynamic model and IEDHS topology information.
(2) Multiple uncertainties associated with electricity and heat loads, instant outputs of wind turbines, thermal plants, CHP, are considered.
(3) The proposed IEDHS dispatch model considers thermal inertia of heating pipeline network and operational flexibility of the combined heating and power units. Simulation results demonstrate that the proposed model has a positive effect on promoting wind power utilization and minimizing energy cost.
(4) To the best of our knowledge, it is the first study to construct a hybrid optimization framework that combines an upper-level DRL with a lower-level conventional optimization solver to accomplish the optimal IEDHS dispatch strategy. A large IEDHS is used to verify the proposed double-deck framework can effectively accelerate training speed and improve final convergence results with respect to original DRL.

The remaining structure of this work is introduced as follows. Section 2 describes the mathematical model of the IEDHS and the scheduling objective. The framework of the D3RL and the principles of DDPG algorithm are presented in Section 3. Section 4 studies the performance of the scheduling model based on four scenarios. Conclusion and future work are given in Section 5.

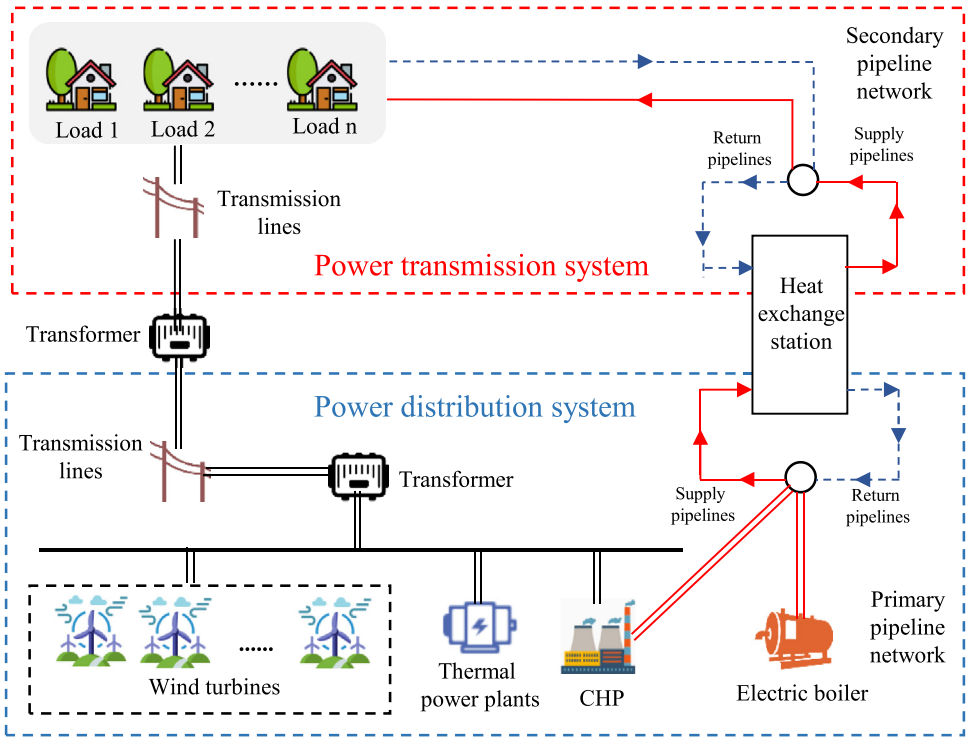

Fig. 1. Example architecture of the IEDHS with wind power.

2. System description

Fig. 1 depicts the example architecture of the IEDHS with wind power. Referring to Wang et al. (2021), a typical IEDHS includes wind turbines (WTs), thermal power unit, CHP unit, electric boiler (EB), and load demands. Electrical and heating networks are coupled through the CHP and EB units. Similar to the power system, the heat sources produce heat energy, the DHS transfers the heat energy (the form of hot water) by the primary pipeline network to the heat exchange station, and then distributes the heat energy to the consumers via the secondary pipeline network.

2.1. District heating system

A typical thermodynamic system consists of four components: heat source, heat network, heat exchange station and heat loads (Chen et al., 2015). As presented in Fig. 2, similar to power system, DHS can be divided into transmission system and distribution system. This paper mainly considers the transmission system. In the transmission system, the heat transfer medium transmits heat from each heat source to each heat exchange station through the water supply pipe network, and then returns to the heat source through the return pipe network, which continuously circulates in the heat network.

(1) Heat sources

As shown in Fig. 1, heat sources in the considered IEDHS are composed of the CHP and the EB. The output of CHP follows the principle of heat to power. According to Chen et al. (2015), the mathematic model of the feasible operational region of the CHP is expressed as:

$$P_{CHP_i}(t)=\sum_{k=1}^{N_i} y_i^k \alpha_i^k(t),\ Q_{CHP_i}(t)=\sum_{k=1}^{N_i} x_i^k \alpha_i^k(t),\ \forall i \in N_{CHP} \tag{1}$$

$$0 \le \alpha_i^k(t) \le 1,\ \sum_{k=1}^{N_i} \alpha_i^k(t)=1,\ \forall i \in N_{CHP},\ k \in \{1,2,\ldots,N_i\} \tag{2}$$

where $(Q_{CHP_i}(t), P_{CHP_i}(t))$ represent the heat and electricity generated by the CHP i at time slot t. (x_i^k, y_i^k) are the heat and electricity output of the extreme point i, and N_i is the total number of corner points of the CHP.

In Eq. (3), EB generates heat by consuming electricity:

$$Q_{EB}(t)=\eta_{EB} P_{EB}(t) \tag{3}$$

$$P_{EB}^{\min} \le P_{EB}(t) \le P_{EB}^{\max} \tag{4}$$

where $(Q_{EB}(t), P_{EB}(t))$ stand for the generated heat and consumed electricity at time slot t, respectively. η_{EB} is denoted as the conversion ratio between the input and output of the EB. $(P_{EB}^{\min}, P_{EB}^{\max})$ represent the allowable minimum and maximum input of the EB.

(2) Pipeline heating network

A distributed heating pipeline network is divided into the supply network and return network. The relationship between nodal thermal power Q and nodal temperature T can be described by Eq. (5) (Bin et al., 2019; Li et al., 2020).

$$Q = C_p m_q (T_s - T_r) \tag{5}$$

where C_p is the specific heat capacity of water, m_q is the mass flow of the pipeline, and (T_s, T_r) represent nodal temperature of the supply and return networks, respectively. Therefore, the thermal power ϕ_i^{CHP} of the CHP at the source node i can be calculated by:

$$\phi_i^{CHP} = C_p m_q^{HS}(T_s^{HS} - T_r^{HS}),\ \forall i \in N_{HS} \tag{6}$$

Similarly, the thermal power ϕ_j^{HE} of the heat exchange station located at node j can also be calculated, as presented in Eq. (7).

$$\phi_j^{HE} = C_p m_q^{HE}(T_s^{HE} - T_r^{HE}),\ \forall j \in N_{HE} \tag{7}$$

As shown in Fig. 2, the structure of pipeline heating network is presented, and the corresponding injected mass flow and nodal temperature can be calculated by Eqs. (8) and (7) (Shao et al., 2017).

$$\sum_{k\in SP}(T_{k,t}^{s,out} \cdot m_{k,t}^{s}) = T_{n,t}^{s} \cdot \sum_{k\in SP} m_{k,t}^{s} \tag{8}$$

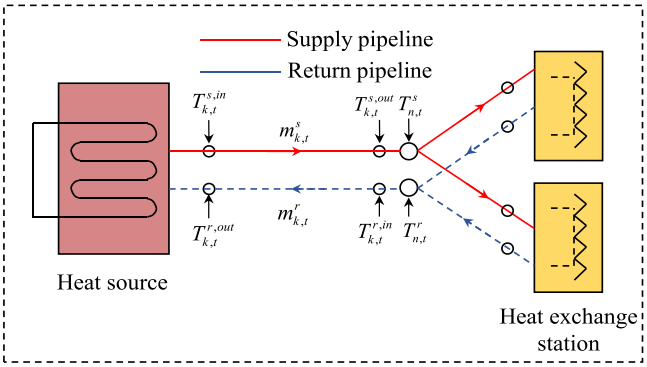

Fig. 2. Structure of pipeline heating network.

$$\sum_{k\in RP}(T^{r,out}_{k,t}\cdot m^{r}_{k,t}) = T^{r}_{n,t}\cdot\sum_{k\in RP} m^{r}_{k,t} \tag{9}$$

where (SP, RP) are the sets of pipelines in supply and return networks, respectively. $T^{s,out}_{k,t}$ is the outlet temperature of the pipeline k in the supply network at time slot t, $m^{s}_{k,t}$ is the mass flow of the pipeline k in the supply network, and $T^{s}_{n,t}$ is the temperature of the mixing node n in the supply network at time slot t. Similarly, $(T^{r,out}_{k,t}, m^{r}_{k,t}, T^{r}_{n,t})$ are the outlet temperature and mass flow of the pipeline k and temperature of the mixing node n in the return network at time slot t, respectively.

Besides, the nodal temperature of both the supply network and the return network should meet the following constraints:

$$T^{s}_{i,\min} \le T^{s}_{i,t} \le T^{s}_{i,\max} \tag{10}$$

$$T^{r}_{i,\min} \le T^{r}_{i,t} \le T^{r}_{i,\max} \tag{11}$$

where $(T^{s}_{i,t}, T^{r}_{i,t})$ are temperature of the node i in the supply and return networks at time slot t, respectively. $(T^{s}_{i,\max}, T^{r}_{i,\max})$ and $(T^{s}_{i,\min}, T^{r}_{i,\min})$ are the corresponding upper and lower limits of the nodal temperature in the supply and return networks, respectively.

In the distributed heating network, the heat medium is in the form of hot water (Gu et al., 2017). This leads to a certain thermal inertia for the process from the thermal power plant to each connected node. Because the hot water medium flows in the pipeline at a certain velocity ν, there is a certain time delay in the temperature change between the inlet and outlet of the pipeline. The temperature change delay T_{delay} of hot water medium in a pipe is related to the length L of the pipeline and flow rate ν, which can be expressed as:

$$T_{delay} = K_{delay}\frac{L}{\nu} \tag{12}$$

where K_{delay} is denoted as the thermal delay coefficient.

On the other hand, heat losses in the pipeline transmission system can be indirectly represented as the decrease of the nodal temperature in the pipeline. Heat loss of the pipeline can be calculated by Eq. (13) (Dai et al., 2018).

$$T_{end} = (T_{start} - T_a)e^{-\frac{\lambda L}{C_p m_q}} + T_a \tag{13}$$

where (T_{start}, T_{end}) are the inlet and outlet temperature of the pipeline, respectively. T_a is the ambient temperature, and m_q is the mass flow through the pipeline. (λ, L) describe the parameters of physical property, which are the transmission impedance and the length of the pipeline, respectively.

To sum up, after considering the transmission time delay in the heating network, the temperature change of the pipeline can be expressed as:

$$T_{end} - T_a = (T_{start}(t - T_{delay}) - T_a(t - T_{delay}))e^{-\frac{\lambda L}{C_p m_q}} \tag{14}$$

2.2. Electricity network

In this paper, DC power flow model (Yugeswar et al., 2022) is applied to analyze the power system. In addition, the node power balance model and the branch power flow model are shown in Eqs. (15)–(18).

$$\sum_{(i,j)\in\Omega_{pipe}} pf_{ij} - \sum_{(k,i)\in\Omega_{pipe}} pf_{ki} + \sum_{g\in i} P_g = P_{l,i},\ \forall i \in \Omega_{bus} \tag{15}$$

$$-PF^{\max}_{ij} \le pf_{ij} = \frac{\theta_{i,t} - \theta_{j,t}}{x_{ij}} \le PF^{\max}_{ij},\ \forall (i,j) \in \Omega_{line} \tag{16}$$

$$P^{\min}_{g} \le P_g \le P^{\max}_{g} \tag{17}$$

$$\theta^{\min}_{i} \le \theta_i \le \theta^{\max}_{i} \tag{18}$$

where the subscript ij indicates the line with i and j as nodes. (θ, x) are node phase angle and branch reactance, respectively. pf_{ij} is the power flow at line ij and $PF^{\max}_{ij,t}$ is the maximum transmission power of the line ij. P_g is the active power output of generators, including wind turbines and CHP units. $P_{l,i}$ is the total power load at bus i. $(P^{\min}_{g}, P^{\max}_{g})$ and $(\theta^{\min}_{i}, \theta^{\max}_{i})$ are minimum and maximum active power output of generator g and phase angle of node i, respectively.

2.3. Wind power generation

The consumption of wind power should not exceed its output, as expressed in Eqs. (19) and (20).

$$P^{c}_{WT}(t) = \eta_{WT}(t)P_{WT}(t) \tag{19}$$

$$0 \le \eta_{WT}(t) \le 1 \tag{20}$$

where $(P^{c}_{WT}(t), P_{WT}(t))$ are the wind power consumed and generated at time slot t. $\eta_{WT}(t)$ is the wind power conversion ratio at time slot t, which is regarded as one of the decision variables.

2.4. Flexible operation model

In this paper, the established scheduling model not only considers the economic operating costs, but also increases the flexibility of the electricity system to respond to sudden changes of wind power generations and load demands. Upward flexibility and downward flexibility should be considered simultaneously to improve the flexibility of the IEDHS, as described in Eqs. (21) and (22):

$$P^{d}(t) = \sum_{i=1}^{N}\min(P_i(t) - P^{\min}_{i}, \Delta t * r^{d}_{i}) \tag{21}$$

$$P^{u}(t) = \sum_{i=1}^{N}\min(P^{\max}_{i} - P_i(t), \Delta t * r^{u}_{i}) \tag{22}$$

where $(P^{d}(t), P^{u}(t))$ are separately downward flexibility and upward flexibility indexes for the CHP units, and (r^{d}_{i}, r^{u}_{i}) are downward and upward ramp rates of the generation unit i, respectively. $P_i(t)$ is the electric output of the generation unit i at time slot t. $(P^{\min}_{i}(t), P^{\max}_{i}(t))$ represent the minimum and maximum outputs of the unit i at time slot t, respectively. N is the amount of the generation units, and Δt is the adjacent time interval.

2.5. Objective function

In this paper, the optimization objective is to minimize the operation cost of the considered IEDHS in the dispatching cycle T by controlling the output of controllable equipment in real time. Specifically, the whole dispatch cycle T is set as one day, equal to

24 h. Besides, the operating costs are composed of the operating cost ($C_{CHP}(t)$, $C_{EB}(t)$, $C_G(t)$) of CHP, EB and thermal power units, and the penalty item $C_{WT}(t)$ of wind power curtailment.

$$\begin{cases} \min\{\sum_{t=1}^{T} C(t) - \varpi(\sum_{t=1}^{T_1} P^d(t) + \sum_{t=1}^{T_2} P^u(t))\} \\ C(t) = C_G(t) + C_{CHP}(t) + C_{EB}(t) + C_{WT}(t) \\ C_{CHP}(t) = a + b \cdot P_{CHP}(t) + c \cdot P_{CHP}(t)^2 + d \cdot Q_{CHP}(t) \\ \quad + e \cdot Q_{CHP}(t)^2 + f \cdot P_{CHP}(t) \cdot Q_{CHP}(t) \\ C_G(t) = g \cdot P_G(t)^2 + h \cdot P_G(t) + k \\ C_{EB}(t) = m \cdot P_{EB}(t) \\ C_{WT}(t) = \kappa \cdot (P_{WT}(t) - P_{WT}^c(t)) \end{cases} \tag{23}$$

where (T_1, T_2) are the number of peak and valley hours, and ϖ is denoted as the economic conversion coefficient used to project the target of scheduling flexibility into the economic dimension. $(a, b, c, d, e, f, g, h, k, m)$ are operating cost coefficients of CHP, EB and thermal power unit, respectively, which are constant values. κ is denoted as the penalty factor. Here, the first term in (23) denotes the operation cost of the IEDHS, and the second term represents the operation flexibility of the generators. It should be noteworthy that the symbol − in (23) indicates that the objective aims to avoid the unit output approaching the predefined maximum $P_i^{\max}$ and minimum $P_i^{\min}$ values. Therefore, the CHP unit can operate flexibly during the periods of the load peak and the low generation of wind power or the load valley and the high generation of wind power.

3. Double-deck scheduling model based on reinforcement learning

RL agent selects the action sequence by interacting with the environment by trial and error, in order to maximize the cumulative return (Cao et al., 2020). In RL, elements in action space are regarded as the decision variables of the optimal energy management problem of the IEDHS. However, the economic scheduling problem of the considered IEDHS includes a continuous action space with complicate constraints. Therefore, to solve the aforementioned characteristics, this paper aims to construct a D3RL decision model to improve the efficiency of learning process and realize real-time regulation of the IEDHS, and the framework is shown in Fig. 3. The lower level applies traditional solver to find the optimal wind power conversion ratio $a_t^{s^*}$ to maximize the immediate reward, when receiving the thermal power units, CHP and EB output a_t^l from the upper DRL layer. Since each search in the decision dimension of the actual wind power output is optimal, the learning efficiency of RL can be significantly improved, and the model training and convergence can be accelerated.

3.1. Upper-level DRL model

3.1.1. Markov decision process

The upper-level decision model is composed of a DRL-based agent, which controls the dispatch of the battery. In the framework of DRL, MDP (Zhang et al., 2022b) is used to formalize the interaction process between the agent and the environment. Six essential elements can be described the MDP, which is composed of the tuple $\langle s, a, r, \Gamma, \gamma, \pi \rangle$:

(1) State $s_t \in S$: s_t is the current state information, and S represents the state set. Sufficient state information should be provided for the DRL agent, including electricity loads $P_l(t)$, heat loads $Q_l(t)$, wind power generation $P_{WT}(t)$, and the output of thermal power unit, CHP and EB listed as:

$$s_t : \{P_l(t), Q_l(t), P_{WT}(t), P_G(t), P_{CHP}(t), P_{EB}(t)\} \tag{24}$$

(2) Action $a_t \in A$: a_t is the specific action, and A represents the action set. As shown in Eq. (25), the DRL agent regulates the amount of adjustment of thermal power unit, CHP and EB output.

$$a_t : \{\Delta P_G(t), \Delta P_{CHP}(t), \Delta P_{EB}(t)\} \tag{25}$$

(3) Reward $r_t \in R$: r_t indicates the immediate reward value, and R is the reward set. The immediate reward function r_t provided for the DRL agent is designed below:

$$r_t = -C(a_t^l; t) \tag{26}$$

in which, $C(a_t^l; t)$ is the operating costs obtained by the lower layer after receiving the action given by the upper layer.

(4) State transition function $\Gamma(s_{t+1}|s_t, a_t)$: the transition process can be divided into deterministic part and stochastic part. The deterministic part represents the effects of action a_t on state $\{P_G(t), P_{CHP}(t), P_{EB}(t)\}$, which means the relationship between $\{P_G(t), P_{CHP}(t), P_{EB}(t)\}$ and $\{P_G(t+1), P_{CHP}(t+1), P_{EB}(t+1)\}$ is only $\{P_G(t)+\Delta P_G(t), P_{CHP}(t)+\Delta P_{CHP}(t), P_{EB}(t)+\Delta P_{EB}(t)\}$. For the stochastic part, other state information with random and complex characteristics has unknown conditional probability $P(s_{t+1}|s_t, a_t)$.

(5) Discount factor γ: the function of the discount factor is to balance the immediate return and the future return, and the value is within the range [0, 1], as expressed in Eq. (27).

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \sum_{k=0}^{T} \gamma^k r_{t+k+1} \tag{27}$$

(6) Policy $\pi(a_t|s_t)$: policy π is used to map the state to the action. For the policy π, the expected cumulative return of an exploration process starting from state s_t and executing action a_t can be described by the action-value function $Q^\pi(s, a)$:

$$Q^\pi(s, a) = E^\pi[R_t|s = s_t, a = a_t] \tag{28}$$

3.1.2. Deep deterministic policy gradient

As discussed before, state transition process in the considered optimization model includes the stochastic part. Therefore, it is intractable for a model-based method to describe an environment that can represent the above randomness. In this paper, as the representative of the policy gradient-based algorithms, deep deterministic policy gradient (DDPG) (Timothy et al., 2015) is applied to train an agent that can provide the real-time optimal strategy for the IEDHS.

DDPG algorithm with the actor–critic architecture combines the core principles of the deep Q network (Volodymyr et al., 2013) and the DPG (David et al., 2014), and has achieved good performance in solving the RL problem with continuous action control. In DDPG algorithm, the instant reward (Eq. (25)) is obtained by the agent through interaction with the environment, and the cumulative reward (Eq. (27)) is maximized to obtain the optimal strategy π^*. The actor–critic architecture of the DDPG is composed of four fully connected layers, the corresponding inputs and outputs are listed in Table 1. Critic and actor functions are approximated by the critic and actor online networks, parameterized by θ^q and θ^μ, respectively. The targets network parameterized by θ^μ and θ^q are introduced to ensure the stability during training process. In addition, considering that the strategy provided by DDPG is deterministic, Gaussian noise N_t is added to the action to increase the random exploration, as expressed in Eq. (29). The updating process of the critic and action are introduced below.

$$\tilde{a}_t = a_t + N_t \tag{29}$$

The critic online network is updated by minimizing the loss function, as shown in Eq. (30):

$$L_{DDPG} = \frac{1}{m} \sum_{i=1}^{m} [y_i - Q(s_i, a_i; \theta^q)]^2 \tag{30}$$

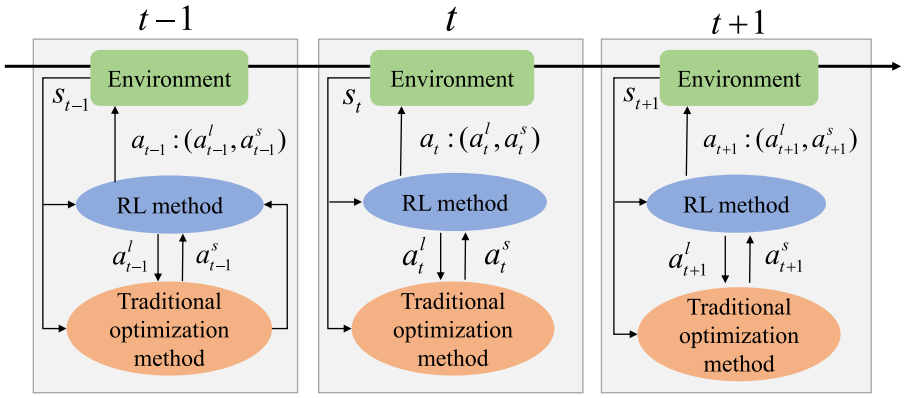

Fig. 3. Decision framework of the double-deck RL model.

Table 1
The neural network settings of DDPG algorithm.

Neural network	Input	Output
Actor online network μ	s_t	$a_t=\mu(s_t;\theta^\mu)$
Actor target network μ'	s_{t+1}	$a'_t=\mu'(s_{t+1};\theta^\mu)$
Critic online network q	$s_t, \widetilde{a}_t$	$q=Q(s_t,\widetilde{a}_t\|\theta^q)$
Critic target network q'	s_{t+1}, a'_t	$q'=Q'(s_{t+1},a'_t\|\theta^q)$

Table 2
The DDPG algorithm.

1. Initialize weights $(\theta^q,\theta^\mu,\theta^q,\theta^\mu)$ of online networks and their corresponding target networks.
2. Initialize replay buffer.
3. For each episode do
4. For each environment step do
5. Select an action using Eq. (29).
6. Execute action a_t, and return immediate reward r_t and next state s_{t+1}.
7. Store interaction information (s_t, a_t, s_{t+1}, r_t) in replay buffer.
8. Sample a random batch information $\{s_i, a_i, s_{i+1}, r_i\}$ with size K.
9. Update the weights of the critic online network by minimizing the loss using Eq. (30).
10. Update the weights of the actor online network by policy gradient using Eq. (32).
11. Update the weights of the corresponding target networks using Eq. (33).
12. End For
13. End For

$$y_i = r_i + \gamma Q'(s_{t+1}, \mu'(s_{t+1};\theta^\mu);\theta^q) \tag{31}$$

where m is defined as the batch size.

Similarly, the parameters of the actor online network are updated in the direction of increasing the $Q(s,a;\theta^q)$ by gradient descent method, which is given in Eq. (32).

$$\nabla_{\theta^\mu} J \approx \frac{1}{m}\sum_{t=1}^{m}[\nabla_a Q(s,a;\theta^q)|_{s_i,\mu(s_i)}\nabla_{\theta^\mu}\mu(s;\theta^\mu)|_{s_i}] \tag{32}$$

Target networks share the same parameters and architectures with the corresponding online network, and the updating of the target networks slowly tracks the online networks, which is called "soft update" mode:

$$\begin{cases}\theta^\mu \leftarrow \tau\theta^\mu + (1-\tau)\theta^\mu \\ \theta^q \leftarrow \tau\theta^q + (1-\tau)\theta^q\end{cases} \tag{33}$$

where τ is the soft update coefficient, and $0 < \tau \ll 1$.

The flow of the DDPG algorithm is provided in Table 2.

3.2. Lower-level solver model

The optimization objective of the lower-level solver model is to minimize the real-time operation cost after receiving the immediate action from the upper-level DRL model, as expressed in Eq. (34):

$$\begin{cases}\min C(t;a^s_t) \\ s.t.\ Equations\ (1)\text{–}(23)\end{cases} \tag{34}$$

Furthermore, Eq. (34) describes a mixed-integer linear programming (MILP) problem, which can be solved by SciPy toolkit. Then, the lower-level solver model provides the upper-level DRL model with the optimal instant operating cost and the optimal RE output values, which is used for immediate reward for the DRL model. Thus, it is a closed loop process.

3.3. Framework of the double-deck scheduling model based on DDPG algorithm

In Fig. 4, the workflow of the proposed D3RL scheduling model based on DDPG algorithm is concretely presented. Besides, the input, output and architecture of the actor and critic networks are also shown in detail. As for the critic network, it is composed of three fully connected layers, each of which has 300 neurons. As for the actor network, it includes two fully connected layers with the same neurons. Note that the rectified linear unit (ReLU) function and the tanh function are set as the activation functions of the DNNs.

The termination condition of the iteration is that the reward expectation converges to the maximum value, as expressed in Eq. (35):

$$|R_i - R_{i-1}| \le \delta \tag{35}$$

where (R_i, R_{i-1}) are the current future reward expectation and the previous future reward expectation, and δ is a constant, which is used as the maximum value of the gap. Finally, the optimal energy management strategy for the succeeding time slots is achieved.

4. Case studies

In this section, to illustrate the validity of the collaborative and flexible scheduling model and the proposed D3RL scheduling method, three scenarios are carried out on a IEDHS consisting of an IEEE-6 electricity system with an 8-node heat system. The potential benefits of considering thermal inertial and flexibility assessment are illustrated by comparing with the traditional economic dispatch model. Further, a comparison analysis with single DRL scheduling method is conducted on a IEDHS consisting of an IEEE-39 electricity system with a 16-node heat system to demonstrate the superiority of the DRL-based double-deck scheduling model. The implementation diagram of the case study is displayed in Fig. 5.

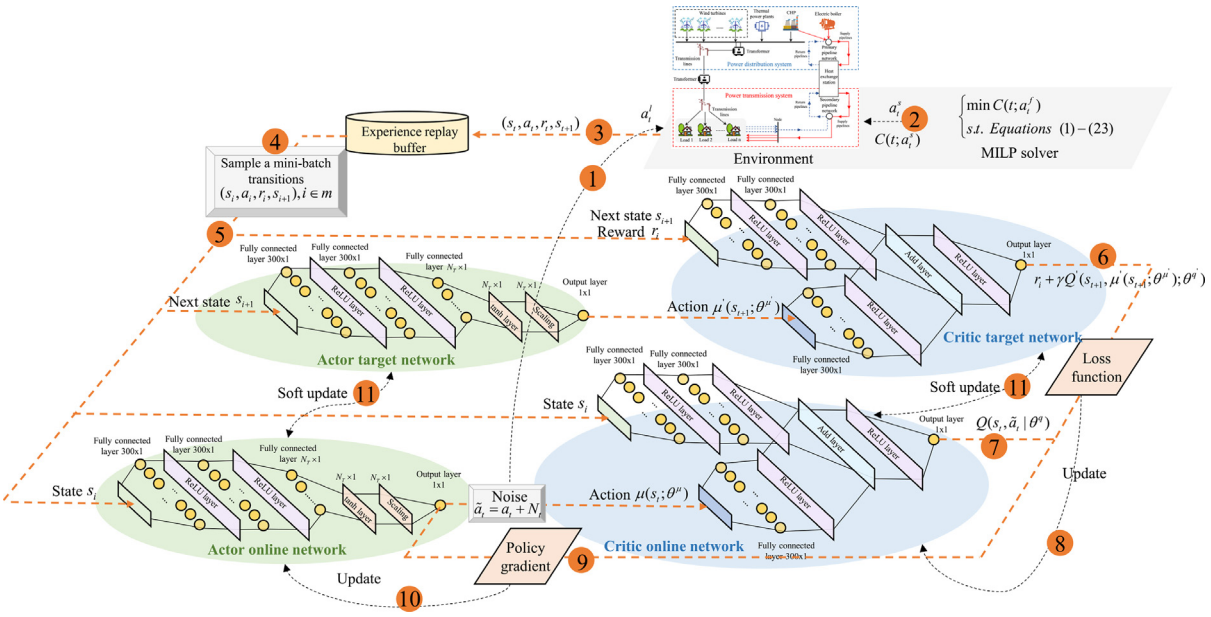

Fig. 4. Workflow of the double-deck scheduling model based on DDPG algorithm.

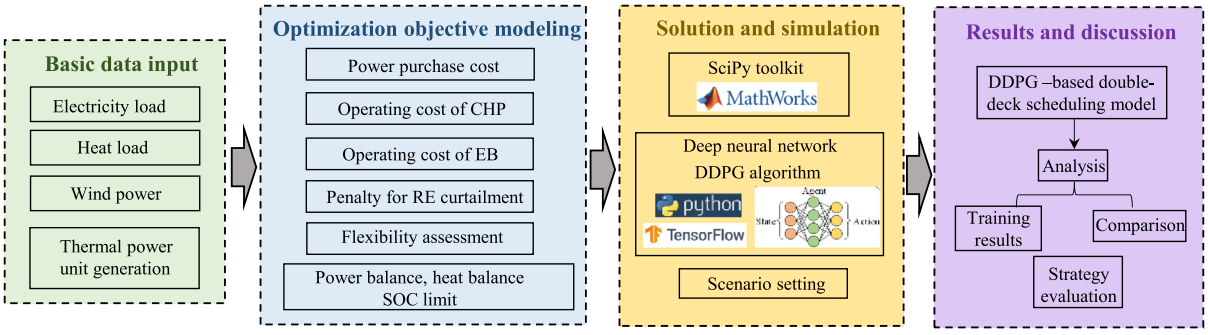

Fig. 5. Implementation diagram of the case system.

Step (1) Based on Section 2, the input basic data set composed of historical data is established.

Step (2) Under the constraints of the power balance, heat balance and SOC limit, optimization objectives are developed, which is consist of comprehensive benefits and flexibility assessment.

Step (3) Three scenarios are used to verify the rationality of the collaborative and flexible scheduling model. Besides, the low-level MILP problem is solved by using SciPy toolkit on MathWorks. The upper-level DRL decision problem is solved by using DDPG algorithm in Python 3.6 on the TensorFlow platform.

Step (4) The dispatch results obtained from step 3 are evaluated from the perspective of the operation of each unit. In addition, the advantage of the proposed D3RL scheduling method is further illustrated by comparison analysis.

The case program is carried out on a 64-bit Windows-based laptop equipped with 4 GB of memory and Intel Core i7-4720HQ CPU. The algorithm is written in Python, and the system model is evaluated on the PandaPower.

4.1. Test system configuration

As displayed in Fig. 6, a modified system, an IEEE-6 power system combined with an 8-node heat network, is used for case study. There are three CHP units, which are located at Bus 1, Bus 4 and Bus 6, respectively. A WT is connected to Bus 1 in the electricity network. The thermal power unit are installed at Bus 1. Loads are divided into zone I, II, III and IV, where an EB is also equipped to supply thermal power. The return pipeline network has the same topology with the supply network. Each adjustable facility is dispatched every 15 min, and the whole dispatch period is a typical day (24 h). Detailed parameter settings of the test system can be referred to (Yao et al., 2019).

4.2. Scenario description

The scheduling results of the IEDHS from three simulation scenarios are analyzed to verify the effectiveness of the thermal inertial model and flexibility assessment model. Four scenarios are described below:

Scenario 1: Scheduling strategy with collaborative dispatch of the adjustable devices without involving the thermal inertial model and flexibility assessment model. In this scheduling, there is no transmission time delay characteristics in the heating pipeline network. Besides, the upward and downward flexibilities during peak and valley periods are not adopted in the objective functions.

Scenario 2: Collaborative dispatch strategy with the thermal inertial model. The transmission time delay is about 10 min. The impact of the heat storage characteristics in the heating pipeline network are investigated, but without the flexibility assessment model in the objective functions.

Scenario 3: Different from Scenario 2, dispatch strategy in Scenario 3 regards the operational flexibility model as one of the

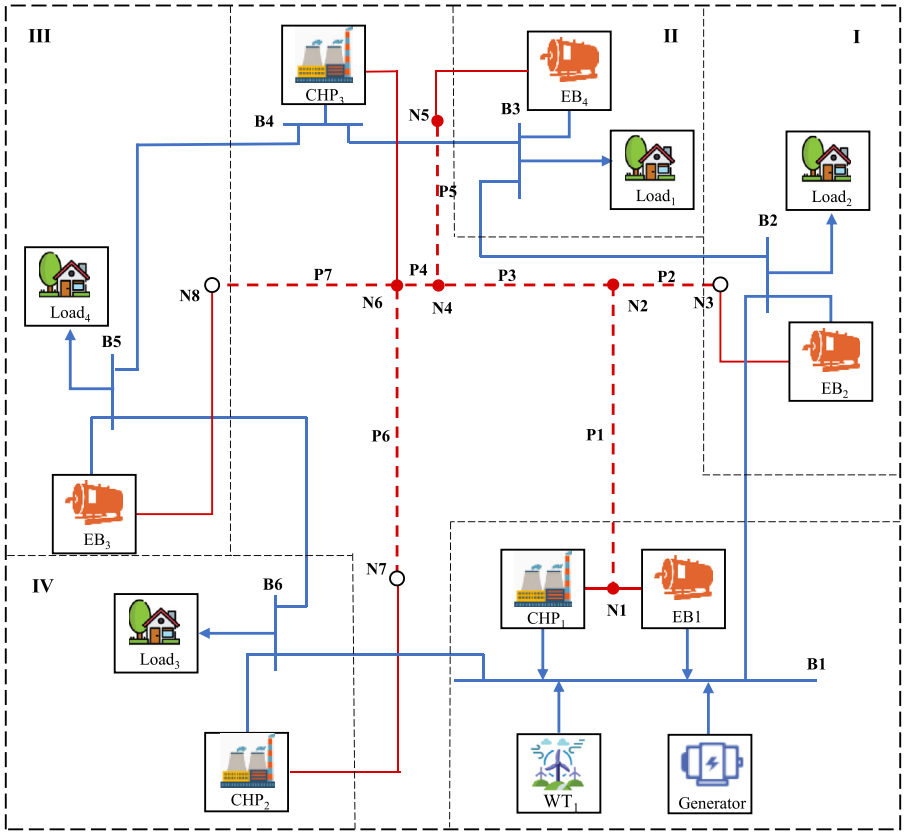

Fig. 6. Diagram of the case system with 6-bus electricity and 8-node district heating networks.

objective functions, instead of considering thermal inertial model in the distributed heating network.

Scenario 4: Compared with the above scenarios, the objective functions of this scheduling include minimizing operating costs and maximizing upward and downward flexibilities during peak and valley periods. Simultaneously, the thermal inertial model is included in the modeling process of the heating pipeline network.

Then the simulation results of the above four scenarios are performed and analyzed for the case system with 6-bus electricity and 8-node district heating networks.

4.3. Training process

In this sub-section, the parameter settings and training procedure are provided in detail. Continuous iterative simulations are required to successfully determine the optimal collaborative control strategy for the defined comparison scenarios. In one iteration, the instant energy information at the beginning of a day is gathered by the agent, such as wind power, electricity and heating loads, electricity price and initial SOC value; then, the agent provides current strategy (adjustable units output, i.e., ESS operation, EB and CHP generation and electricity purchased from main grid) and receives immediate reward and observation information of next time step. According to the algorithm presented in Fig. 4, the agent can adaptatively adjust the provided control strategy by updating the inner weights of DNNs to maximize the cumulative reward of the whole iteration. Thus, through a large amount of training episodes, the optimal collaborative scheduling strategy can be achieved.

The proposed D3RL scheduling method is a data-driven method, thus sufficient data is required to establish data sets for training. In this simulation, the historical data of one year (Conolly et al., 2015; Ashfaq and Ianakiev, 2018) for Aarhus, Denmark, including electricity and heating load profiles, wind power generations, are used to constitute the training set. Fig. 7 presents the concrete information of the historical data.

Table 3
The hyperparameters of DDPG algorithm.

Parameter	Value
Discount factor γ	0.9
Soft updating coefficient τ	0.01
Learning rate of the Actor α_0^{μ}	0.01
Learning rate of the Critic α_0^{q}	0.015
Batch size m	512
Experience buffer capacity D	20 000

The performance of the DRL algorithm mainly depends on settings of the DNNs, which has been discussed in Section 3. Additionally, the hyperparameters of the DRL DDPG algorithm are listed in Table 3. For the DDPG algorithm, the weights of the critic and actor networks are updated with the learning rates $\alpha_0^{\mu} = 0.01$ and $\alpha_0^{q} = 0.015$ according to Eqs. (30) and (32), respectively. Discount factor $\gamma = 0.9$ is provided for the critic to calculate the cumulative return mentioned in Eq. (27). The updating rates of the target networks are controlled by the given soft updating coefficient $\tau = 0.01$. The experience replay buffer with sufficient storage capacity $D = 20\,000$ is used to store historical interactive information. Batch information with size $m = 512$ is used as instant samples to update the weights of DNNs.

The change trend of cumulative reward corresponding to 24 time points of each training episode is given in Fig. 8. In the policy learning stage (before 4000 episodes), to increase the cumulative reward, the neural network parameters are modified and updated according to Eqs. (30), (32) and (33) to deal with the random changes of wind power, and load demands. After 4000 episodes,

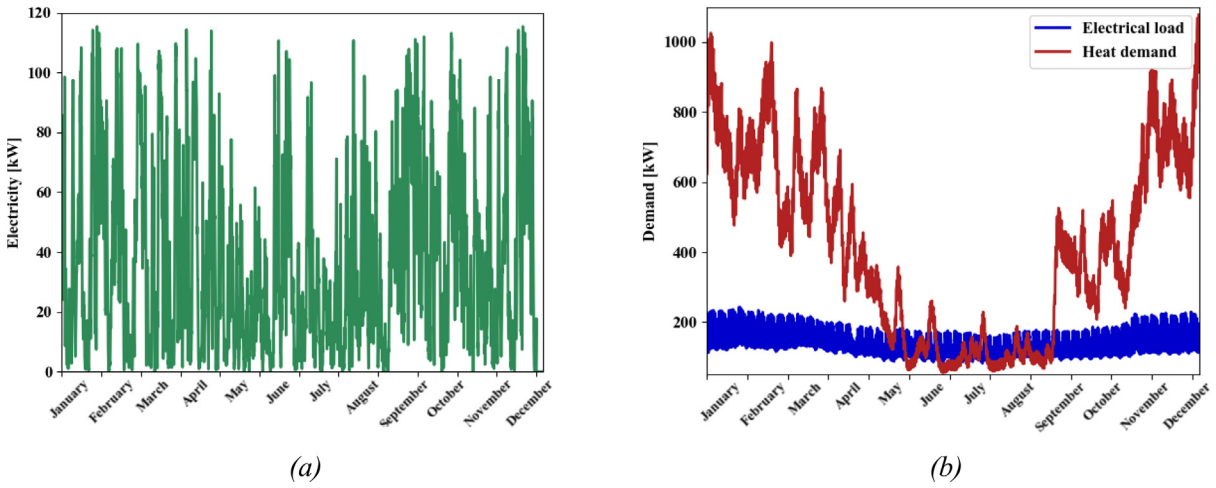

Fig. 7. Profiles of historical data based on one year: (a) Wind power generation, (b) Load demands.

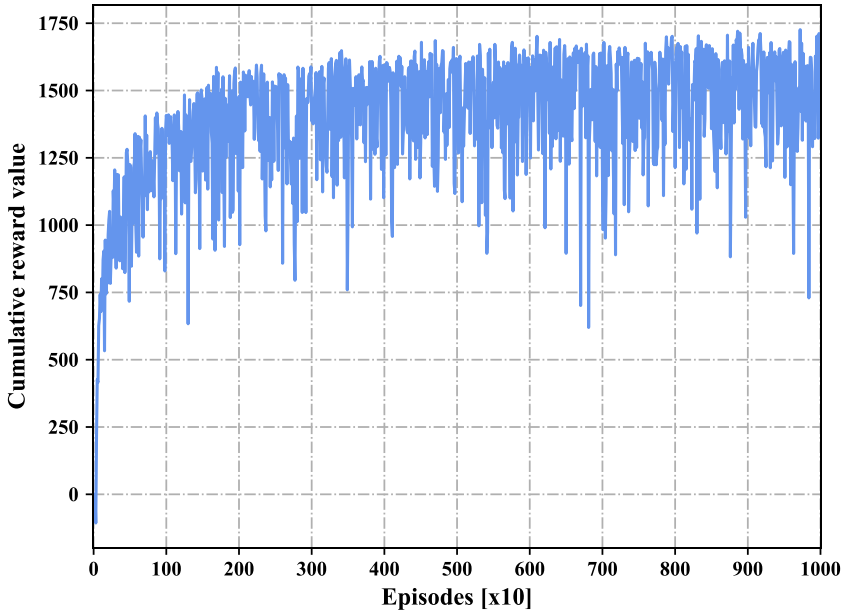

Fig. 8. Cumulative reward values change with episodes during the training process.

Table 4
Optimization results of economic costs and RE utilization under four scenarios.

Scenarios	Total economic costs (DKK)	Total RE utilization (kW)
Scenario 1	1884.4	4102.1
Scenario 2	1636.5	4717.3
Scenario 3	1589.8	4653.2
Scenario 4	1421.3	4952.9

the gap between the current future reward expectation and the previous future reward expectation is less than the predefined threshold (Zhang et al., 2019a). It means the agent learned how to deal with the random environment and maintain the cumulative reward, indicating that the optimal energy management strategy has been achieved.

4.4. Analysis of optimal scheduling results based on four scenarios

Based on the above training process, the corresponding optimal scheduling results of economic costs and RE utilization under four scenarios are listed in Table 4.

4.4.1. Impact of transmission time delay in the heating pipeline network

As displayed in Table 4, compared to Scenario 1, the daily economic operating costs of Scenario 2 decreased by 13.15%, that is from 1884.4 DKK to 1636.5 DKK. Furthermore, the amount of the RE utilization in Scenario 2 increased from 4102.1 kW to 4717.3 kW. In Scenario 2, the transmission time delay considered in the heating pipeline modeling indicates that heat storage capacity in the pipeline network can be utilized to supply heat power. Especially during the peak periods of wind power, CHP can reduce its own output by utilizing the residual heat in the pipeline network to supply heat loads, so as to improve the utilization rate of RE. Therefore, to reduce economic costs and promote wind power accommodation, it is generally reasonable to take the transmission time delay characteristic into account when modeling the district heating network.

4.4.2. Impact of flexibility assessment model

The total economic costs of Scenario 3 decrease from 1884.4 DKK to 1589.8 DKK, a decrease of 15.63% compared to Scenario 1. Additionally, the wind power utilization in Scenario 3 increase

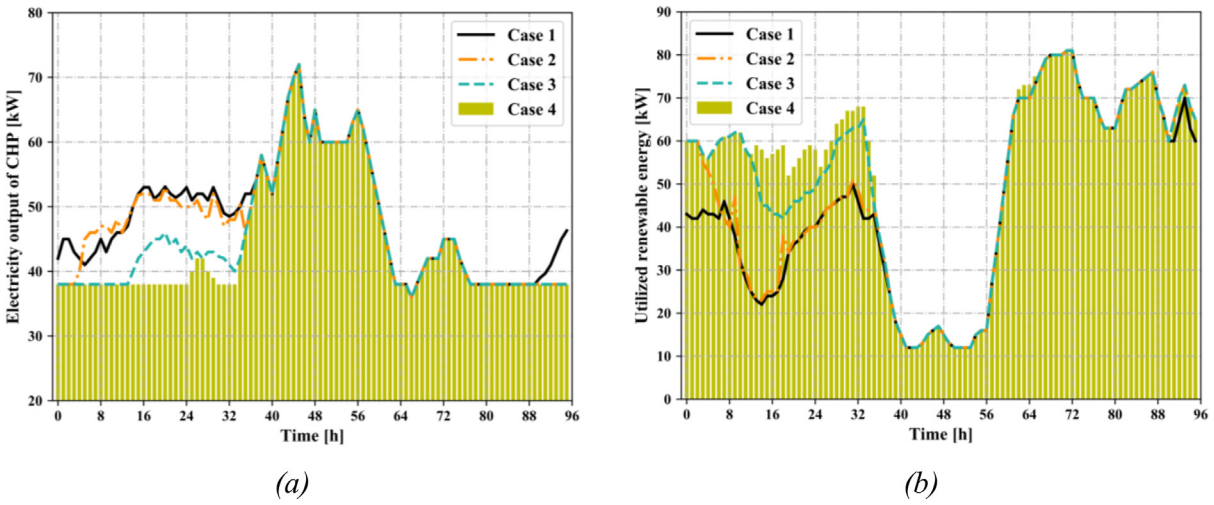

Fig. 9. Comparison results of scheduled electricity under four scenarios: (a) Electricity output of the CHP unit, (b) Utilized wind power.

from 4102.1 kW to 4653.2 kW, compared to Scenario 1. The optimization target of Scenario 3 is not only the economic operating costs, but also the flexibility of the IEDHS. By optimizing the flexibility during load peak period and valley period, the operational flexibility model considered in Scenario 3 can effectively cope with the increased variability of RE generations and load profiles, thereby the integration of wind power can be promoted with less efforts. The results shows that the flexibility assessment model can efficiently help to improve RE utilization and reduce operating costs.

4.4.3. Impact of combination of two above characteristics

In the Scenario 4, the above characteristics, including transmission time delay and flexibility assessment model, are considered. The total economic costs are reduced by 24.57% compared with Scenario 1, from 1884.4 DKK to 1421.3 DKK. The amount of wind power utilization increases 20.74% from 4102.1 kW to 4952.9 kW compared with Scenario 1. The simulation results demonstrate that the combination of thermal inertial model and flexibility assessment model can further reduce the economical operating costs and RE curtailment. The results can be attributed to two reasons: due to the thermal inertial existing in the heating pipelines, surplus heat energy can be stored in the distributed heating network; to ensure flexibility of the IEDHS, more RE resources are utilized to supply electricity loads and EB units for satisfying heating loads, instead of increasing the output of the CHP units. Hence, the total heat output of the CHP is relatively lower than that in other three scenarios, resulting in lower operating costs. Besides, more wind power is tended to integrated into the IEDHS.

In Fig. 9(a) and (b), a comparison analysis of the electricity output of the CHP unit and utilized RE under four scenarios is displayed. Due to heat storage characteristic and more EB participation, the electricity outputs of the CHP in Scenarios 2, 3 and 4 are lower than those in Scenario 1 in most scheduling time slots. However, during some time periods of hours 0–9, it can be observed that the electricity output of the CHP in Scenario 2 is higher than that in Scenario 1. The reason is that the CHP is needed to generate more heat energy and store it in the heating pipelined in advance, resulting in more electricity being produced. Hence, the least electricity output of the CHP is required in Scenario 4 because it takes heat storage characteristics and flexibility assessment into account at the same time. In Fig. 9(b), Scenario 4 has the greatest RE utilization among four scenarios.

Fig. 10 displayed the scheduled heat energy of Scenarios 1, 2, 3 and 4. Due to the influence factor of transmission time delay characteristic, compared with Scenario 1, more heat energy are scheduled in Scenario 2 during hours 3–4, hours 8–10 and hours 18–20, indicating that thermal energy is stored into the heating pipelines to minimize operating costs. In Scenario 3, without heat capacity of pipeline network, thermal energy at the time of hours 14–20 are actively scheduled to improve the operational flexibility of the CHP unit. Considering influence factors of thermal inertial and operational flexibility, Scenario 4 has the least thermal energy of CHP among four scenarios, which reduces the output of the CHP. At the time of hours 8–9, hours 14–20, more thermal energy is still required to be stored in advance. The results show that, after introducing these two kinds of factors, wind power integration is promoted and the operational flexibility of the CHP is improved by reducing the output of the CHP. The reason why the electricity output of CHP and utilized RE for cases 1–4 are the same during 35–89 h is that cases 1–4 have the same electricity balance constraints after storing sufficient thermal energy in the heat network. Besides, operation cost including reducing wind curtailment is the common objective in cases 1–4, and using the CHP units is more economical than the thermal plants.

Table 5
Flexibility comparison results of thermal power units under four scenarios.

Scenarios	Downward flexibility (kW)	Upward flexibility (kW)	Total flexibility (kW)
Scenario 1	62.3	112.5	174.8
Scenario 2	70.2	113.8	184
Scenario 3	82.3	114.2	196.5
Scenario 4	85.6	116.5	202.1

Furthermore, the downward flexibility and upward flexibility of the CHP unit are displayed in Table 5. It can be observed that flexibilities in Scenarios 2, 3 and 4 are significantly better than that in Scenario 1, which illustrates that thermal inertial has a certain function on improving flexibility. Besides, the downward flexibility of Scenario 4 increases from 62.3 kW to 85.6 kW, an increase of 37.40% compared to Scenario 1, resulting in flexibility improvement of the IEDHS. Since there are almost valley periods of electric load in the test day, the improvement of the upward flexibility is not obvious.

4.5. Performance evaluation on IEEE 39-bus system

To further validate the effectiveness and superiority of the proposed D3RL dispatch model, an IEEE 39-bus system combined with a 16-node distributed heating network is applied, the

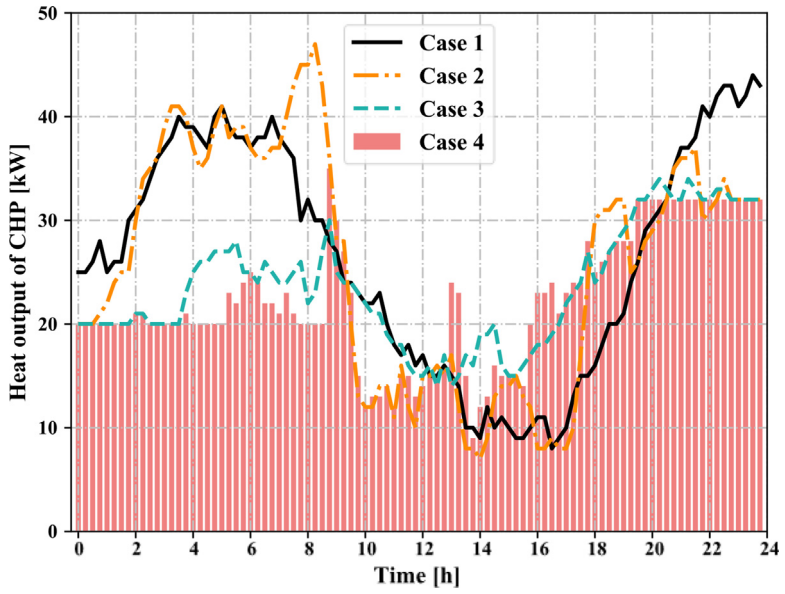

Fig. 10. Comparison results of scheduled heat of the CHP under four scenarios.

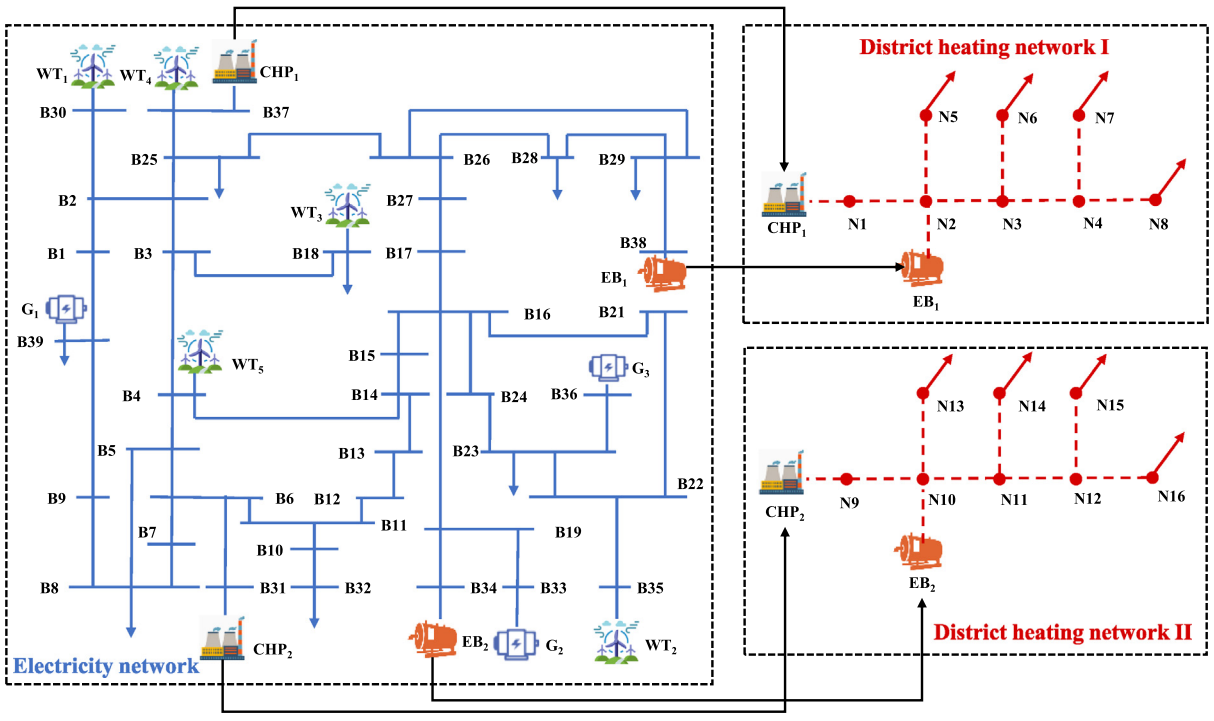

Fig. 11. Diagram of the case system with 39-bus electricity and 16-node district heating networks.

structure of which is shown in Fig. 11. Electricity network and heating network are coupled by the CHP and EB units. The 16-node distributed heating network is composed of two 8-node heating networks, and model parameters can be obtained in Li et al. (2015). Five wind farms are located at bus 30, bus 35, bus 18, bus 37 and bus 4, respectively. Two CHP units are placed at bus 31 and bus 37, which are connected with node 1 and node 9 of the district heating network to supply heat energy, respectively. EBs located at bus 38 and bus 34 are connected with node 2 and node 10, respectively, which are regarded as heat sources. Thermal power generators are placed at bus 33, bus 36 and bus 39 to supply electric loads.

In Fig. 12, the cumulative reward change of the proposed method and DDPG algorithm is displayed. The final convergence result of the proposed D3RL scheduling method is significantly better than that of the DDPG algorithm. Due to the volatility associated with the increasing share of wind power, the DDPG algorithm, which performs the optimization search by exploring the action space, causes significant oscillations compared to the proposed method. The training process shows that the proposed model can effectively reduce the invalid exploration and obtain better results because the conventional optimization algorithm can provide the optimal RE generators output to the DRL model in advance.

Furthermore, to demonstrate the superiority of the well-trained D3RL agent, comparison simulation with other benchmark methods on consecutive 30 days test data, e.g., DDPG and PSO-based stochastic optimization (Zhang et al., 2021b), are conducted. The

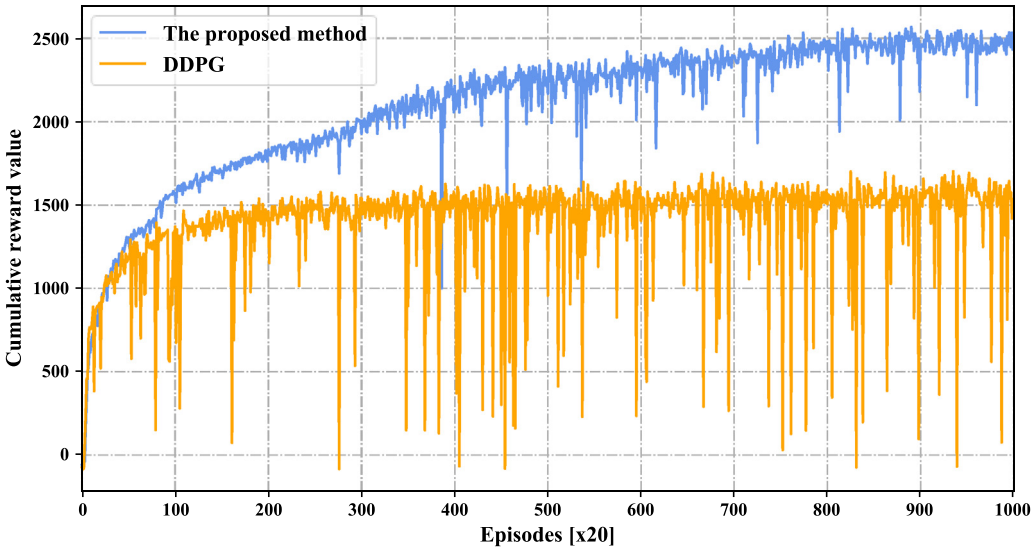

Fig. 12. Cumulative reward comparison of the proposed method and DDPG in IEEE 39-bus system during the training process.

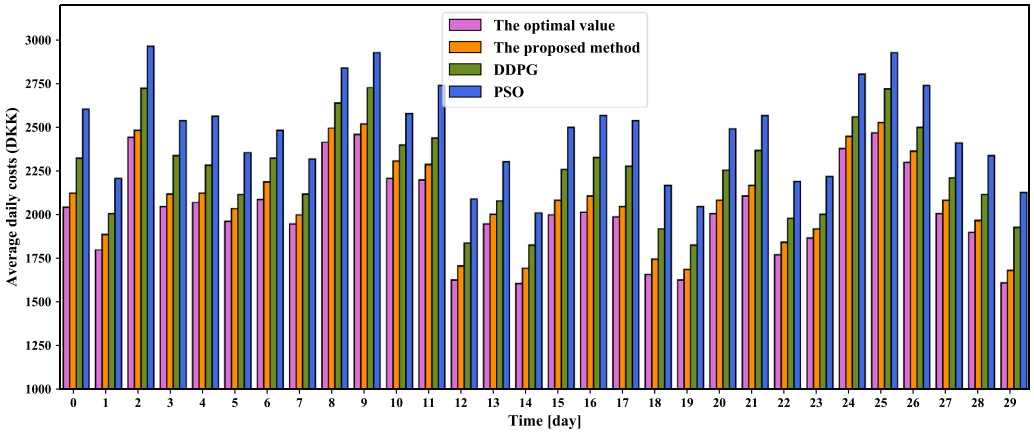

Fig. 13. Daily operating costs of the IEDHS under 30-day test data using the proposed approach, DDPG and PSO.

optimal value is obtained by formalizing the problem as a mixed-integer linear problem and using the Gurobi Optimizer to solve it, which is regarded as a theoretical benchmark. The control strategy of the PSO-based stochastic is formalized by averaging the optimal control strategy over a sample set of 500 historical days test data. It can be observed that the proposed D3RL energy management strategy achieves the least average daily costs among three methods. Besides, Table 6 presents the comparison results of daily operation costs under different wind power forecast accuracy. DRL-based algorithms are decision processes that address only current and historical information, eliminating the need for source/load prediction, and are able to respond adaptively to random dynamic changes in the environment. In contrast, the PSO stochastic algorithm is influenced by the prediction accuracy during the optimal scheduling process(see Fig. 13).

5. Conclusion

In this paper, a collaborative IEDHS energy dispatch model considering the thermal inertia of the DHS and operational flexibility is established. Based on the transmission time delay characteristics of heating pipeline networks, thermal inertia of the DHS is constructed. Then, operational flexibility is included in the multi-objective optimization function, which also considers operating costs and wind power utilization. Instead of directly applying DRL methods, a D3RL optimization framework is proposed to cope with high-dimensional action space brought by large-scale wind turbines integrated into IEDHS. The D3RL optimization framework is composed of upper-level DRL model and lower-level traditional solver model. When receiving the instant action from the upper-level DRL model, the lower-level solver model determines the optimal wind power conversion ratio and the optimization objective, which is used as the reward value for DRL model. Four scenarios with are applied to analyze the influence of the different models. Simulation results show that due to the heat capacity of heating pipeline network, thermal inertia has positive potential in promote operation flexibility of CHPs, reducing operating costs and improving wind power utilization. To demonstrated the superiority of the proposed D3RL framework, a large infrastructure with 39-bus electricity and 16-node district heating networks is used. Compared with traditional RL, the proposed D3RL optimization framework can improve its training speed and convergence performance. In addition, it still has the

Table 6
Comparison analysis of the daily operation costs under different wind power forecast accuracy.

Method	Wind power forecast accuracy (Tawn and Browell, 2022)			
	100	95	90	85
PSO-based stochastic method	2512.3	2732.5	2962.1	3032.4
DDPG	2280.6	–	–	–
The proposed method	2019.3	–	–	–

advantages on dealing with optimization problem affected by wind power forecast accuracy.

In future work, the rescheduling problem between different IEDHS will be further investigated. Because IEDHS rescheduling in practical application gets into trouble with long training time, poor scenario transability and waste of domain knowledge, a transfer-RL based rescheduling method will be applied.

CRediT authorship contribution statement

Bin Zhang: Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing. **Amer M.Y.M. Ghias:** Writing – original draft, Writing – review & editing. **Zhe Chen:** Writing – review & editing, Supervision.

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.

Data availability

The data that has been used is confidential.

Acknowledgments

This work was supported by the School of Electrical and Electronic Engineering at Nanyang Technological University, Ministry of Education, Singapore, under Grant AcRF TIER 1 RG50/21.

References

Ashfaq, A., Ianakiev, A., 2018. Cost-minimized design of a highly renewable heating network for fossil-free future. Energy 152, 613–626.

Bagherian, M., Mehranzamir, K., 2020. A comprehensive review on renewable energy integration for combined heat and power production. Energy Convers. Manage. 224 (113454).

Bin, Z., et al., 2019. Deep reinforcement learning-based approach for optimizing energy conversion in integrated electrical and heating system with renewable energy. Energy Convers. Manage. 202 (112199).

Cao, D., et al., 2020. Reinforcement learning and its applications in modern power and energy systems: a review. J. Mod. Power Syst. Clean Energy 8 (6), 1029–1042.

Chen, X., Kang, C., Malley, M., et al., 2015. Increasing the flexibility of combined heat and power for wind power integration in China: Modeling and implications. IEEE Trans. Power Syst. 30 (4), 1848–1857.

Chen, H., Zhang, Y., Zhang, R., et al., 2021. Privacy-preserving distributed optimal scheduling of regional integrated energy system considering different heating modes of buildings. Energy Convers. Manage. 237 (114096).

Conolly, D., et al., 2015. Enhanced Heating and Cooling Plans to Quantify the Impact of Increased Energy Efficiency in EU Member States. Tech. rep., Aalborg University, Denmark.

Dai, Y., et al., 2018. Dispatch model for CHP with pipeline and building thermal energy storage considering heat transfer process. IEEE Trans. Sustain. Energy 10 (1), 192–203.

Daraei, M., Campana, P., Avelin, A., et al., 2021. Impacts of integrating pyrolysis with existing CHP plants and onsite renewable -based hydrogen supply on the system flexibility. Energy Convers. Manage. 243 (114407).

David, S., et al., 2014. Deterministic policy gradient algorithms. In: Proceedings of the 31st International Conference on Machine Learning.

Gu, W., et al., 2017. Optimal operation for integrated energy system considering thermal inertial of district heating network and buildings. Appl. Energy 199, 234–246.

Li, Z., Wu, W., Shahidehpour, M., et al., 2015. Combined heat and power dispatch considering pipeline energy storage of district heating network. IEEE Trans. Sustain. Energy 7 (1), 12–22.

Li, X., et al., 2020. Collaborative scheduling and flexibility assessment of integrated electricity and distributed heating systems utilizing thermal inertia of distributed heating network and aggregated buildings. Appl. Energy 258 (114021).

Ma, H., Chen, Q., Hu, B., et al., 2021. A compact model to coordinate flexibility and efficiency for decomposed scheduling of integrated energy system. Appl. Energy 285 (116474).

Moharram, N., Tarek, A., Gaber, M., et al., 2022. Brief review on Egypt's renewable energy current status and future vision. Energy Rep. 8, 165–172.

Nejabtkhah, F., Li, Y., 2015. Overview of power management strategies of hybrid AC/DC microgrid. IEEE Trans. Smart Grid 30 (12), 7072–7089.

Obama, B., 2017. The irreversible momentum of clean energy. Science 355 (6321), 126–129.

Shao, C., et al., 2017. Modeling and integration of flexile demand in heat and electricity integrated energy system. IEEE Trans. Sustain. Energy 9 (1), 361–370.

Sneum, D., 2021. Barriers to flexibility in the district energy-electricity system interface – A taxonomy. Renew. Sustain. Energy Rev. 145 (111007).

Sun, X., Qiu, J., 2021. A customized voltage control strategy for electric vehicles in distribution networks with reinforcement learning method. IEEE Trans. Ind. Inform. 17 (10), 6852–6863.

Tawn, R., Browell, J., 2022. A review of very short-term wind and solar power forecasting. Renew. Sustain. Energy Rev. 153 (111758).

Timothy, P., et al., 2015. Continuous control with deep reinforcement learning. arXiv:1509.02971.

Volodymyr, M., et al., 2013. Playing atari with deep reinforcement learning. ArXiv ID: 1312.5602.

Wang, X., Bie, Z., Liu, F., et al., 2021. Co-optimization planning of integrated electricity and district heating systems based on improved quadratic convex relaxation. Appl. Energy 285 (116439).

Wang, H., Yang, J., Chen, Z., et al., 2020. Optimal dispatch based on prediction of distributed electric heating storages in combined electricity and heat networks. Appl. Energy 267 (114879).

Xia, T., Li, Y., Zhang, N., et al., 2022. Role of compressed air energy storage in urban integrated energy systems with increasing wind penetration. Renew. Sustain. Energy Rev. 160 (112203).

Xu, F., Hao, L., Chen, L., et al., 2021. Discussion on the real potential of district heating networks in improving wind power accommodation with temperature feedback as one consideration. Energy Convers. Manage. 250 (114907).

Yang, Q., 2021. Prospective contributions of biomass pyrolysis to China's 2050 carbon reduction and renewable energy goals. Nature Commun. 12 (1698).

Yang, T., Zhao, L., Li, W., et al., 2021. Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning. Energy 235 (121377).

Yao, S., et al., 2019. Testdata for the case he-ies. http://dx.doi.org/10.21227/hznh-m758, [Online]. Available:.

Ye, Y., Qiu, D., Sun, M., et al., 2020. Deep reinforcement learning for strategic bidding in electricity markets. IEEE Trans. Smart Grid 11 (2), 1343–1355.

Yugeswar, R., et al., 2022. Stochastic optimal power flow in islanded DC microgrids with correlated load and solar PV uncertainties. Appl. Energy 307 (118090).

Zhang, Y., Cai, J., Liu, R., 2021a. Calculation and analysis of energy storage in heat supply nets of distributed energy. Energy Convers. Manage. 229 (113776).

Zhang, B., Hu, W., Cao, D., et al., 2019a. Deep reinforcement learning-based approach for optimizing energy conversion in integrated electrical and heating system with renewable energy. Energy Convers. Manage. 202 (112199).

Zhang, B., Hu, W., Cao, D., et al., 2020. Dynamic energy conversion and management strategy for an integrated electricity and natural gas system with renewable energy: Deep reinforcement learning dispatch. Energy Convers. Manage. 220 (113063).

Zhang, G., Hu, W., Cao, D., et al., 2021b. Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach. Energy Convers. Manage. 227 (113608).

Zhang, X., Strbac, G., Shah, N., et al., 2019b. Whole-system assessment of the benefits of integrated electricity and heat system. IEEE Trans. Smart Grid 10 (1), 1132–1145.

Zhang, M., Wu, Q., Wen, J., et al., 2022a. Day-ahead stochastic scheduling of integrated electricity and heat system considering reserve provision by large-scale heat pumps. Appl. Energy 307 (118143).

Zhang, G., et al., 2022b. 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. Manage. 255 (115340).

Zhu, M., Li, J., 2022. Integrated dispatch for combined heat and power with thermal energy storage considering heat transfer delay. Energy 244 (part. B, no. 123230).