A digital twin of multiple energy hub systems with peer-to-peer energy sharing

Shiyao Li [a,b], Yue Zhou [b,*], Jianzhong Wu [b], Yiqun Pan [c], Zhizhong Huang [d], Nan Zhou [e]

[a] School of Mechanical Engineering, Tongji University, Shanghai, China
[b] School of Engineering, Cardiff University, Cardiff, Wales, UK
[c] School of Architecture, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[d] Sino-German College of Applied Sciences, Tongji University, Shanghai, China
[e] Energy Technologies Area, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA

HIGHLIGHTS

- Extensive digital twin for multi-energy hub systems with P2P energy sharing.
- EXDT enables virtual testing & evaluations of MEHs under various scenarios.
- MARL-based model for predicting stochastic decision-making process of EHs.
- Economic and technical benefits of P2P sharing: increased renewables and financial gains.
- Evaluation of various decision-making and P2P strategies for energy system operations.

ARTICLE INFO

ABSTRACT

As climate change has become a global concern, the decarbonization of energy systems has become a prominent solution for CO_2 emission reduction. The recent emergence of multi-energy hub systems (MEHs), characterized by interconnected energy hubs (EHs) and facilitated by energy sharing, presents a promising solution for seamlessly integrating a significant share of renewable energy sources (RESs) and flexibility among EHs. Faced with the intricate interplay and uncertainty of future energy markets, an extensive digital twin (EXDT) is proposed to perform predictive testing and evaluate the performance of MESs. This EXDT provides energy system operators with insights into the coordinated behavior of interconnected EHs under various future scenarios, thus contributing to smarter decision-making processes. Specifically, an array of scenarios including different decision-making strategies and P2P energy sharing strategies were considered. For each of these scenarios,"what-if" tests were conducted using a multi-agent reinforcement learning (MARL)-based method to model the stochastic decision-making process of EHs belonging to different stakeholders with access to local information. Uncertainties during operation can be mitigated using Markov Game (MG) by capturing knowledge from historical energy data. Subsequently, the economic and technical performance were evaluated using multidimensional evaluation indexes. The proposed MARL-based EXDT was applied to a representative 4-EH multi-energy system in China. Simulation results indicate that P2P energy sharing facilitates the local consumption of renewable energy, providing additional financial benefits and self-sufficiency to each EH and offering peak shaving to the upstream grid. Additionally, system performance under various decision-making and P2P sharing strategies was tested and evaluated to identify the impact of these strategies on system operation.

---

* Corresponding author.
E-mail address: zhouy68@cardiff.ac.uk (Y. Zhou).

1. Introduction

1.1. Background

As the climate crisis has become a great global concern, it is a decisive moment for international efforts to tackle greenhouse gas (GHG) emissions. The number of countries pledging to reach net-zero emissions by mid-century or shortly thereafter continues to grow [1]. Given that the energy sector is a primary GHG emitter, its decarbonization is crucial for curtailing CO_2 emissions. Potential pathways for reducing carbon emissions in energy systems include substituting conventional fossil energy with various renewable energy sources, utilizing advanced energy conversion and storage technologies, and implementing demand response measures [2]. Over the past decade, the development of multi-energy systems (MESs) has been the trend. These systems provide a novel approach to accommodate increasing renewable energy integration and incorporate energy-saving technologies. By collectively managing multi-energy flows, which were previously dispatched independently, MESs harness complementary flexibility and provide a systematic solution for achieving peak emissions and carbon neutrality.

1.2. Literature review

An Energy Hub (EH) functions as a multi-energy management unit for MESs, enabling the district linking of various energy sources while regulating energy conversion and storage processes [3]. This modeling concept provides a feasible solution for enhancing the synergy and operational flexibility of integrated energy systems [4]. As a result, the modeling and operational optimization of EHs have garnered increasing research interest. References [5–7]focus on the mathematical models of EHs, developing energy coupling matrices for multi-energy converters, energy transmission lines, energy storage systems, and various load demands. Optimizing EH operations, which involves complex multi-energy and temporal couplings, presents a nonlinear optimization challenge. Extensive efforts using gradient descent [8] and heuristic methods [9] have been made to achieve optimal solutions. For example, in [10], a linearized model and the Karush-Kuhn-Tucker (KKT) conditions were adopted to obtain optimal coordination of flexible resources in an integrated gas-heat-electricity energy system. An improved version of particle swarm optimization (PSO) was used in [11] to solve the multi-energy coordinated optimization problem of community EHs.

Besides independent operation, by engaging in energy market transactions, an EH can interlink with other EHs to share energy surpluses or mitigate deficits. Peer-to-peer (P2P) energy sharing is gaining considerable research attention as an effective management scheme for prosumers with RESs within a local energy market (LEM) [12]. This approach facilitates direct energy and information exchanges among EHs. Technically, it provides access to energy resources in local grids, energy storage solutions and demand response measures implemented by end-users [13]. In this way, P2P energy sharing contributes to energy balancing and enhances the flexibility of EHs as system flexibility primarily stems from energy storage and demand response measures. Financially, the P2P LEM allows EHs to trade surplus or deficit renewable energy and benefit from more favorable pricing than traditional upstream grid transactions. Studies by Wei et al. [14] and Junhui et al. [15], have demonstrated the improved flexibility and economic gains from P2P sharing between EHs.

Existing research on P2P sharing frameworks generally focuses on two categories: hierarchical [16] and decentralized P2P paradigms [17–19]. A typical P2P energy sharing framework consists of physical models, business models, and control models. In terms of physical systems, studies have included various entities, ranging from individual buildings [20,21] and microgrids [16] to multi-microgrids [22], engaging as peers in the energy trading process. Regarding business models, extensive research has analyzed market designs for P2P trading [23], with a focus on pricing [24] and matching mechanisms [25] and

benefit allocation among participants [26]. Additionally, both cooperative [27,28] and non-cooperative [29] game theories have been applied to capture the conflicting interests of different participants in a computationally feasible way, optimizing bidding decisions to achieve either minimum cost [30] or maximum self-sufficiency [31] under the established market rules. For individual control, techniques such as reinforcement learning [32], alternating direction method of multipliers (ADMM)-based [33] and improved ADMM-based [34,35] optimization have been employed to achieve desirable outcomes, such as Nash equilibrium, through iterative processes.

Considering P2P energy sharing, coordinating multiple interlinked EHs (MEHs) offers a solution to integrating high shares of RESs and leveraging the complementary flexibility of energy systems across large-scale areas [36]. Despite the technical and financial advantages aforementioned, the integration of MEHs presents several challenges. 1) The decision space, which aggregates the operational behavior of each independent EH, is continuous and high-dimensional. When accounting for both P2P interactions among EHs and internal energy flow management within each EH, the complexity grows exponentially. 2) The multi-stakeholder nature of connected EHs raises privacy concerns, often resulting in a reluctance to share complete information between EHs. Consequently, the decision-making process of each EH becomes autonomous and stochastic, relying on incomplete information. 3) Additionally, the interactions and energy sharing among connected EHs further increase the uncertainties associated with future scenarios.

To address the complex nonlinearity and high dimensionality of decision-making processes while considering the privacy protection of autonomous EHs, simplifications have been widely employed in existing literature.

In response to the challenges of nonlinearity and high dimensionality, simplified models in literature [37,38] were used, where the operation of EHs was modeled using mixed-integer linear programming. Similarly, linear node power balance models were employed in [39,40] to determine the energy flow transmission and loss. Except for the linear model, independent operation hypotheses were adopted in [41], where a decentralized bi-level framework for multi-carrier system optimization was proposed.

Regarding privacy and safety concerns, the authors in [19,42] introduced a multi-agent system (MAS) structure for P2P energy sharing, demonstrating its effectiveness in protecting the privacy of multiple entities. Additionally, distributed ledger technology, such as blockchain, has been used to secure transactions in P2P sharing, with blockchain-based smart contracts being tailored for P2P energy sharing [43]. These approaches highlight the potential of P2P energy sharing to transform energy distribution and consumption within LEMs. However, the decision-making processes of EHs are typically formulated deterministically, aiming for an optimal solution based on perfect global information [44].

As a result of these simplifications, we often lose sight of the operational dynamics and interactions of EHs. For instance, linear processing may fail to capture intricate dynamics such as demand-side thermal dynamics, and fluctuations in energy markets. Simplified interaction models usually do not account for power exchange between EHs, leading to potential inefficiencies as the complementarity between different EHs is disregarded. The use of deterministic physical model-based [45] and programming-based [46] approaches to simplify the decision-making process of EHs overlooks the inherently stochastic nature of these processes. Furthermore, limited access to model parameters, global information, and significant computational demands [47] make these approaches impractical.

Additionally, these approaches seldom address the uncertainties inherent in future scenarios, including EH interactions and the implementation of P2P energy sharing strategies.

To further investigate the energy behavior of EHs involving multiple stakeholders, it is essential to consider their stochastic, local information-based decision-making processes and the uncertainties

associated with future scenarios. Recognizing that obtaining practical, real-time machine-monitoring data under various operational conditions can be both costly and impractical, a high-resolution simulation approach is required to effectively model these complex systems.

Recently, the concept of the Digital Twin (DT) has been proposed, first used for the future generation of NASA and U.S. Air Force vehicles [48]. It provides a new paradigm to integrate ultra-high fidelity simulation (the virtual space) and the physical space [49]. Extensive research has focused on establishing the mapping between physical and virtual spaces. For instance, in [50], the connection between a robotic system and its digital twin was established using a data-driven approach to provide real-time decision-making. This cyber-physical mapping makes high-resolution mirroring of the previously mentioned MEH systems and their P2P energy sharing possible. However, In the energy sector, DTs find utility in real-time optimization [51], data monitoring [52], fault detection [53,54] and health management [55] of energy systems of various scales based on the real-time interaction of physical and virtual spaces. The historical operational data contained within DTs, capable of reflecting the operational logic or patterns of their physical systems, remain underutilized.

In summary, MESs composed of interconnected EHs and facilitated by P2P energy sharing represent a significant advancement in integrating substantial proportions of renewable energy and enhancing the flexibility of future energy systems.

However, due to the high-dimensional, nonlinear, and even non-convex nature of these systems, conventional centralized or decentralized methods, as reported in previous works, fail to capture the full dynamics of MEHs, accounting for both EH operation and interaction. These methods also overlook the stochastic decision-making processes and the consideration of future scenarios in MESs with P2P energy sharing. Moreover, such approaches, particularly model-based methods, rely on accurate models of and complete information, which may encounter challenges related to model construction, convergence, and computational demands [56].

To delve deeper into this novel type of energy system, The following research gaps require exploration:

1) Accurate modeling of the energy behavior of each EH, including P2P energy sharing and multi-energy flow management, while accounting for the stochastic nature of their decision-making processes. This stochastic element arises from imperfect information and the autonomy of decision-making.
2) Comprehensive testing and evaluation of the energy system across a range of future scenarios, including various decision-making and P2P energy sharing strategies. This is crucial for energy system operators to understand and adapt to potential future developments.
3) An extensive DT enables the testing and evaluation of energy system technologies (decision-making and P2P energy sharing strategies as discussed in this work). It can integrate the energy behavior model with future scenarios, and the corresponding evaluation system into a simulated environment, offering a practical alternative to physical system testing and providing valuable insights into the operation of MEHs with P2P for energy system managers.

2. Motivation and contribution

This work addresses the research gap and limited understanding of the emerging energy system paradigm, namely MEH. The novelty of this work primarily stems from:

1) Given that physical testing and commissioning for MEHs with P2P sharing are often unfeasible, and there is a need to explore potential future scenarios, we proposed an extensive digital twin (EXDT) approach. This approach aims to extend the function of DTs beyond mere historical data analysis and real-time data monitoring for energy systems, fully leveraging the high-resolution mirroring of physical systems and historical operational pattern data provided by CPSs.
2) Unlike existing centralized or model-based approaches, we employed a multi-agent reinforcement learning-based method to address the nonlinearity and high-dimensionality inherent in the EHs' decision-making process. Additionally, we account for the stochastic nature of each EH's operation under privacy concerns.
3) Multiple scenarios of P2P energy sharing are considered in this work, providing a comprehensive analysis of its impact on MEH systems.

By using this EXTD, we can predict and model the energy behavior of interconnected EHs under various scenarios, and subsequently evaluate the performance of the MEH system both economically and technically. Our approach empowers stakeholders to explore the full potential of MEH systems within a simulated environment, facilitating thorough testing of system functionalities, performance assessment across various scenarios, and validation of innovative energy management strategies. Ultimately, This work accelerates the transition from theoretical concepts to real-world implementations. This work contains prominent contributions as follows.

1) We have broadened the scope of the current DT framework into a more versatile form, enabling predictive testing and evaluation of future energy systems. For the first time, an extensive DT supporting systematic testing and evaluation specific to MEH systems considering P2P energy sharing has been developed. This EXDT encompasses a range of hypothetical scenarios, energy behavior predictive modeling, multidimensional evaluation indexes, CPS, and data exchange among them. Using this EXDT, a comprehensive evaluation of MESs with P2P energy sharing under diverse future scenarios can be observed.
2) Considering the stochastic decision-making process of each EH, we introduced a novel multi-agent reinforcement learning-based method for their energy behavior predictive modeling. This is the first implementation of such a method, designed to extract energy behavior patterns from the historical data of CPS and to simulate the decision-making processes of each EH in various future scenarios. This multi-agent framework effectively models the stochastic nature and autonomous decision-making process of each EH's energy behavior.
3) In consideration of potential future scenarios, various decision-making strategies (both cooperative and non-cooperative) of EHs, as well as different P2P energy sharing strategies, were examined in this study. A comparative analysis of these scenarios was conducted using a multi-dimensional evaluation index system, identifying the strengths and weaknesses of the proposed strategies.

3. The framework of the proposed extensive digital twin

In this work, the function of DT for energy systems is broadened. Traditionally, DTs have focused on historical data analysis and real-time data monitoring. Few DTs for energy systems consider the CPS's function as a virtual testbed or simulation environment for conducting digital system testing and predicting and evaluating system performance under future scenarios. The EXDTs proposed in this work enable more comprehensive and forward-looking analyses, facilitating better decision-making and system design in the energy sector. The superstructure of the proposed EXDT for MEH systems is presented in Fig. 1.

In this work, potential scenarios in the future including decision-making models and P2P sharing are hypothesized. The energy behavior decision-making process of each energy hub across these scenarios is predicted and modeled by intelligent agents within the multi-agent system depicted in Fig. 1. The Scenario module (Hypothetic scenarios in Fig. 1) and Energy behavior model are interlinked in a "what-if" way indicating how each EH would behave under specific scenarios. By running "what-if" simulations, the operational pattern of energy systems
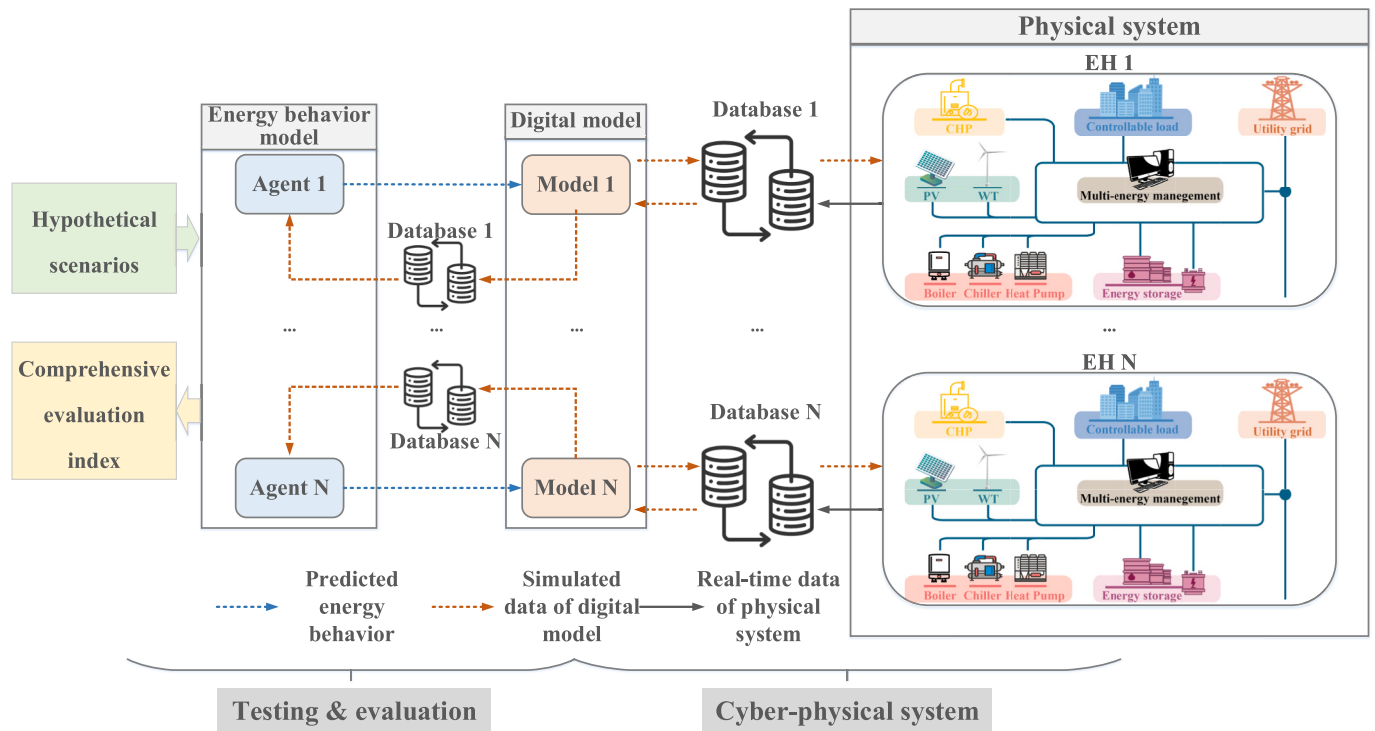
Fig. 1. The superstructure of the extensive digital twin proposed in this work.

using P2P energy sharing in different scenarios can be analyzed.

Virtual digital models in Fig. 1 are connected with real physical systems by data exchange and databases. We extend the utility of DTs beyond physical systems by integrating digital models and energy behavior models through bidirectional data transmission and virtual databases. In the proposed EXDT framework, the energy behavior of each EH, determined by its corresponding agent through stochastic decision-making processes, is communicated to digital models for execution. Digital models then provide the system states and performance parameters after executing specific behaviors. System states refer to operational conditions such as meteorological parameters, renewable energy generation, storage status, and demand-side thermal conditions, while performance parameters include energy consumption and financial costs. Apart from serving as a testbed for providing system states feedback on executing given behavior signals, the virtual digital model also offers a digital database of virtual operational trajectory data for data-driven training of intelligent agents.

Subsequently, according to the performance parameter provided by the digital model, the operational performance of the entire MEH system can be evaluated using a multi-dimensional evaluation index, providing a comprehensive assessment.

To ensure high resolution in the virtual digital model, bidirectional data exchange and dynamic correction between the virtual model and the physical system are typically necessary. The receding horizon approach, extensively discussed in the literature, is commonly used for this correction. In this work, we focus on the virtual testing and evaluation aspects, specifically concentrating on the interaction of other modules (behavior model, scenarios, and evaluation index) with the CPS. This interaction aims to capture and predict the stochastic decision-making processes of multiple energy hubs in various scenarios, as well as to understand their impact on system operational states and performance. For simplification, the data exchange between the virtual and physical components of the CPS is omitted, and we use the virtual model solely as a high-resolution CPS, acting as a virtual testbed.

The rest of this paper is organized as follows: Multi-agent reinforcement learning is used to mirror the energy behavior of energy hubs.

Its mechanism is elaborated in Section 4. Section 5 provides a full overview of the hypothetical scenarios considered in this work and Section 6 interprets a comprehensive evaluation index used to assess system performance across various scenarios. A multiple energy hub system, which serves as the simulation case study, is introduced in Section 7, along with the presentation and discussion of the simulation results. Lastly, the conclusion and future work summarized in this work are discussed in Section 8.

4. Energy behavior modeling using multi-agent reinforcement learning

In prior research, centralized model-based optimization methods were commonly used to determine the energy behavior of EHs within a MEH system. This deterministic approach assumes that a central operator or each EH has access to global information, including private data. However, this assumption is often impractical, particularly when EHs are owned by different stakeholders, resulting in significant model inaccuracies. Given the nature of EHs' decision-making processes, several challenges arise: 1) The complexity of the model due to the nonlinearity or non-convexity introduced by the supply and demand dynamics of each EH, as well as the intricacies of P2P energy sharing market rules, is difficult to capture. 2) Private information, such as the renewable energy generation of each EH, is typically not shared.

In this study, we utilize multi-agent reinforcement learning to model the decision-making processes of autonomous EHs in various scenarios. On one hand, the Markov Game in reinforcement learning algorithms allows the modeling of the P2P energy market matching and multi-energy dispatch of each EH, which are challenging to express explicitly in mathematical terms. By constructing a one-to-one mapping between EHs and agents, the decision-making process of each EH is formulated as the sequential decision-making problem of its corresponding agent. The data-driven training process extracts knowledge from historical energy data to address uncertainties. On the other hand, the multi-agent structure of multi-agent reinforcement learning aligns well with the MEH system. It reflects the interactions (whether

cooperative or non-cooperative) among EHs and captures the stochasticity of each EH's decision-making, particularly under imperfect information scenarios that consider privacy concerns.

In conclusion, multi-agent reinforcement learning is an effective solution to complex sequential decision-making problems involving nonlinearity and non-convexity among multiple entities with privacy concerns. It aids in protecting the privacy of each entity while reducing the iterations and communication requirements.

This section delves deeper into the energy behavior model shown in Fig. 1 and its linkage with other parts. As outlined in the scenario module (see Section 5), which encompasses decision-making and P2P energy sharing strategies, agents operating within the multi-agent energy behavior model engage in a "what-if" simulation. In this "what-if" simulation, agents take action (*a*), such as P2P energy sharing and multi-energy dispatch within the EH, according to state (*s*) provided by the cyber-physical system (CPS) and the specific scenario. The CPS then functions as a virtual environment providing feedback on the execution outcomes of the determined energy behavior, which, in turn, influences state (*s'*) and reward (*r*) updating. The closed-loop interaction between the multi-agent decision-making model and the virtual environment generates virtual operational data trajectories (*s, a, s', r*), which are then stored in a virtual database (*D*) for agent training. This training process significantly enhances the individual reasonability of agents, ultimately contributing to the formation of swarm intelligence among them. Following execution, the performance of these "what-if" cases, specifically the performance of the MEH system across various scenarios, is assessed through a comprehensive evaluation index.

4.1. Multi-agent deep deterministic policy gradient reinforcement learning

Recently, multi-agent reinforcement learning (MARL) has been reported in many prominent sequential decision-making problems, such as playing the game of Alpha Go [57,58], playing real-time strategy games, like StarCraft II [59] and Dota2 [60], and card games [61], robotic control [62], as well as autonomous driving [63].

There exist numerous MARL algorithms, with one notable example being the multi-agent deep deterministic policy gradient (MADDPG) algorithm. MADDPG is categorized as a multi-agent policy gradient algorithm, wherein agents learn policies solely based on local information during execution. A key feature of MADDPG is its decentralized framework, enabling it to effectively address continuous action spaces and adapt to either cooperative or competitive interactions among agents. As a model-free method without any assumptions about the mathematical model of the environment or the communication structure between agents, this algorithm finds application in the realm of decision modeling for complex multi-agent systems.

Extensive literature highlights the robustness of MADDPG, showcasing its superior convergence and cumulative rewards when compared to existing methods. These advantages are particularly evident in both cooperative and competitive scenarios, underscoring the algorithm's effectiveness in various contexts [64].

In this work, agents in Fig. 2 make energy behavior decisions according to local observations and undergo training to enhance their intelligence through a historical data replay buffer. This decision and training process employs MADDPG, and is elaborated as below.

4.1.1. Markov game
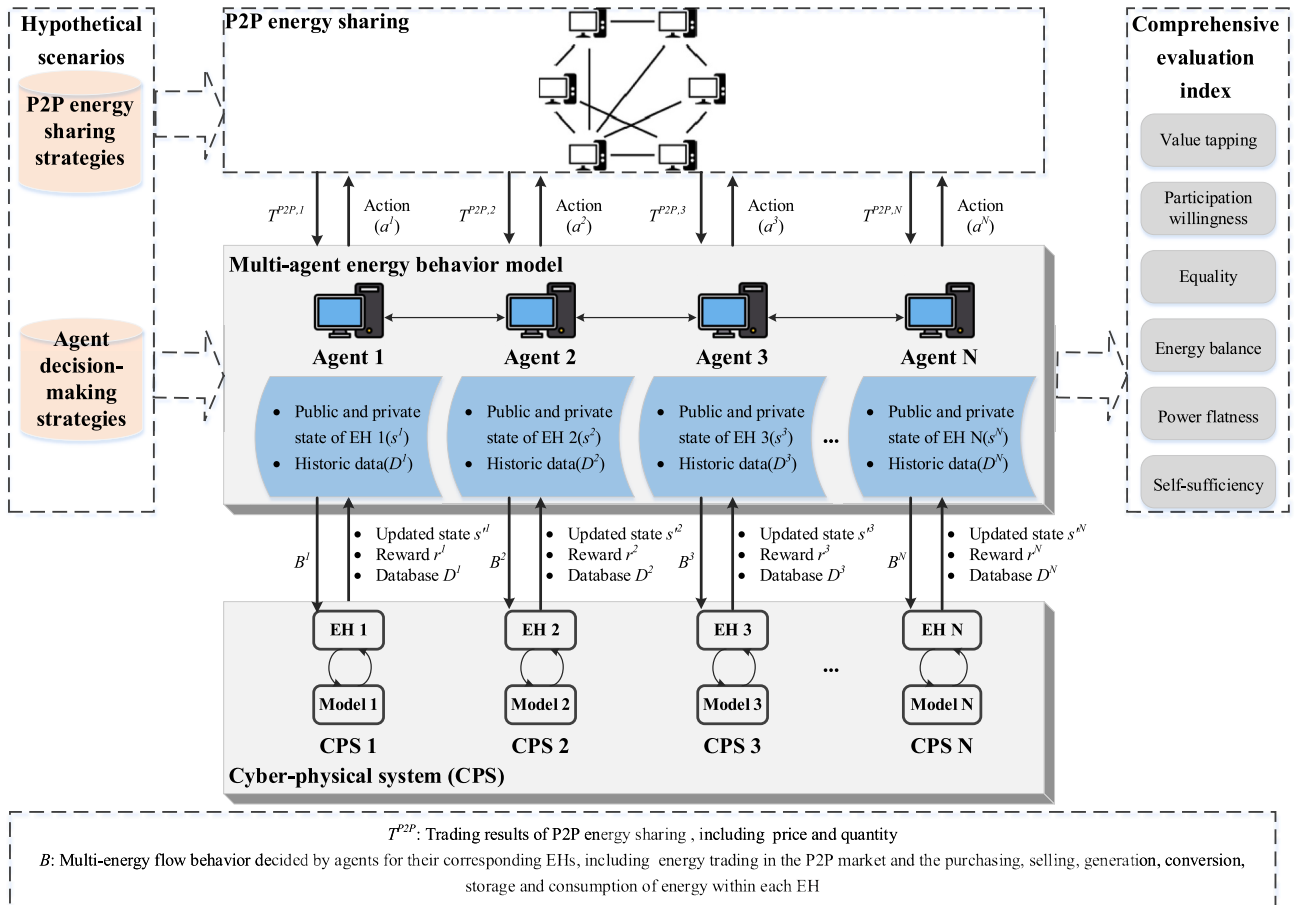In this work, we use a multi-agent extension of Markov decision

Fig. 2. Multi-agent energy behavior model in the proposed extensive digital twin.

processes (MDPs) called partially observable Markov games to formulate the behavior trajectory of a set of agents $\boldsymbol{N}$ [65]. A Markov game is defined by a tuple $\left(\boldsymbol{N}, \{S^i\}_{i\in\boldsymbol{N}}, \{A^i\}_{i\in\boldsymbol{N}}, \rho, \{R^i\}_{i\in\boldsymbol{N}}, \gamma\right)$, where $\boldsymbol{N} = \{1, 2, 3, ..., N\}$ denotes the set of agents, S^i denotes the state space of agent i, and A^i denotes the action space of agent i. To choose actions, each agent i uses a stochastic policy $\boldsymbol{\mu}_{\varphi^i}: S^i \rightarrow A^i$, which conducts the state transition to the next state, according to the transition function T. Unlike single-agent systems, multi-agent systems are not a mere simple aggregation of individual intelligent agents; there exist interactions and mutual influences among these agents. The state transitions in a multi-agent system environment and the rewards for each intelligent agent are influenced by the joint action set of all agents $T: S^i \times A^1 \times A^2 \times A^3 \times ... \times A^N \rightarrow S^{\cdot i}$. ρ denotes the transition probability from current state $(s^i \in S^i, \forall i \in \boldsymbol{N})$ to the next state $(s^{\cdot i} \in S^i, \forall i \in \boldsymbol{N})$, by taking certain joint actions $(a^i \in A^i, \forall i \in \boldsymbol{N})$. Each agent i obtains rewards as a function of the state and agent's action $r_i: S^i \times A^i \rightarrow R^i$. Each agent i's objective is to maximize its own cumulative expected return $R^i: R^i = \sum_{t=0}^{Tmax} \gamma_t r_t^i$, where γ is a discount factor and T_{max} is the time horizon. More details about the state and the action of agents are elaborated in the following subsections.

4.1.2. State and observation

The state (s) refers to the system's status. In a multi-agent environment, each agent may have a partial view of the system state (s), primarily due to privacy concerns. Agent i's observation is represented as o^i. In this study, agent i's observation (o^i) comprises both public and private components, as illustrated in Eq. (1). The public component, denoted as $\boldsymbol{I}_P^i$, encompasses information shared among all agents, such as time, upstream electricity prices, gas prices, heating prices, and meteorological information. On the other hand, $\boldsymbol{I}_p^i$, represents agent i's private state, which includes EH i's operational status, including renewable energy generation profiles, electrical demands, thermal responses, and the state of energy storage units (SoC) within EH i. $\boldsymbol{x}$ denotes the set of all agents' observation in Eq. (2).

$$o^i = \left(\boldsymbol{I}_P^i, \boldsymbol{I}_p^i\right) \quad \forall i \in \boldsymbol{N} \tag{1}$$

$$\boldsymbol{x} = \{o^1, o^2, ..., o^N\} \tag{2}$$

4.1.3. Action

The action represents the energy-related behaviors that agents take in response to a given state and scenario. Agent i's action, denoted as a^i is defined in Eq. (3) and consists of two components. The former, $\boldsymbol{a}_{P2P}^i$, refers to the P2P electricity sharing behaviors of EH i, encompassing the price and quantity of electricity exchanged in the local P2P market. The second component, $\boldsymbol{a}_{EH}^i$, encompasses the multi-energy flow management of EH i, which includes the following energy behaviors: 1) importing or exporting multi-energy from or to the upstream grid, 2) converting energy within the EH, such as energy flows to combined heat

and power, heat pumps, boilers, and chillers, etc., and 3) managing the charge/discharge rates of multi-energy storage units, including electricity, heating, and cooling. The full action space A^i of agent i consists of all possible combinations of a^i within its feasible space. Typically, the feasible space is regulated by physical operational constraints.

$$a^i = \left(\boldsymbol{a}_{P2P}^i, \boldsymbol{a}_{EH}^i\right) \quad \forall i \in \boldsymbol{N} \tag{3}$$

4.1.4. State transition

The state transition process is governed by the transition function (T), as depicted in Eq. (4), and this function is emulated through CPSs. It can be observed that the state of agent i in the subsequent time step is related to the collective action set comprising the P2P energy trading and multi-energy dispatch actions of all agents($i \in \boldsymbol{N}, -i \in \boldsymbol{N}, i \neq -i$), as well as the disturbance of Gaussian noiseω.

$$s^{\cdot i} = T(s^i, a^i, a^{-i}, \omega) \quad \forall i \in \boldsymbol{N} \tag{4}$$

4.1.5. Reward

Likewise, the reward (r) of each agent earned by taking specific actions is simulated by the CPS. In this case, an integrated reward function consisting of energy cost ($Cost^i$) and penalty term ($Penalty^i$) is employed, as shown in Eq. (4). It means, the objective of each agent is to minimize the energy cost of its corresponding EH and meet its end-users' demand.

$$r^i = -\left(Cost^i + Penalty^i\right) \quad \forall i \in \boldsymbol{N} \tag{5}$$

4.1.6. Network update

An actor-critic structure is used by MADDPG. As to each agent i, the actor is a fully connected neural network that maps the local observation (o^i) to action (a^i) following the policy ($\boldsymbol{\mu}_{\varphi^i}$). Its objective is to obtain the largest maximum expected cumulative reward ($Q_{\boldsymbol{\mu}_{\varphi^i}}(\boldsymbol{x}, a^1, a^2, ..., a^N)$). This objective can be achieved by directly adjusting the parameter (φ^i) of the policy. The gradient of the policy can be written as Eq. (6), and the parameter (φ^i) will be updated by taking steps in the direction ($\nabla_{\varphi^i} J(\boldsymbol{\mu}_{\varphi^i})$), as shown in Eq. (7):

$$\nabla_{\varphi^i} J(\boldsymbol{\mu}_{\varphi^i}) = E_{\boldsymbol{x}, a \sim \boldsymbol{D}} \left[\nabla_{\varphi^i} \boldsymbol{\mu}_{\varphi^i}(a^i | o^i) \nabla_{a^i} Q_{\boldsymbol{\mu}_{\varphi^i}}(\boldsymbol{x}, a^1, a^2, ..., a^N) | a^i = \boldsymbol{\mu}_{\varphi^i}(o^i) \right] \tag{6}$$

$$\varphi^i \leftarrow \varphi^i + \eta_a \nabla_{\varphi^i} J(\boldsymbol{\mu}_{\varphi^i}) \tag{7}$$

where η_a denotes the learning rate of the actor network, and E denotes expectation. $\boldsymbol{D}$ denotes the historical dataset provided by the CPS, used for the experience replay buffer. Specifically, operational data tuples $[\boldsymbol{x}, \underbrace{a^1, a^2, ..., a^N}_{a}, \boldsymbol{x}^{\cdot}, r^i]$ are stored in $\boldsymbol{D}$ and then sampled in batches for the actor and critic network training.

$Q_{\boldsymbol{\mu}_{\varphi^i}}(\boldsymbol{x}, a^1, a^2, ..., a^N)$ denotes the action-value indicating the cumulative reward obtained by taking policy ($\boldsymbol{\mu}_{\varphi^i}$). The action-value

Table 1
The actor-critic structure of MADDPG.

	Network	Formulation	Input	Output
Actor	Actor	$a^i = \boldsymbol{\mu}_{\varphi^i}(o^i)$	1) Observation (o^i)	Action(a^i),to calculate $Q_{\boldsymbol{\mu}_{\varphi^i}}$ in the critic network
	Target Actor	$a^{\cdot i} = \boldsymbol{\mu}_{\varphi^{\cdot i}}(o^{\cdot i})$ $\varphi^{\cdot i} = \tau\varphi^i + (1-\tau)\varphi^{\cdot i}$	1) The next observation ($o^{\cdot i}$)	The next action ($a^{\cdot i}$), to calculate$Q_{\boldsymbol{\mu}_{\varphi^{\cdot i}}}$ in the target critic for the critic network update
Critic	Critic	$Q_{\boldsymbol{\mu}_{\varphi^i}}(\boldsymbol{x}, a^1, a^2, ..., a^N)$	1) Observation 2) Action 3) Output of actor	Action-value ($Q_{\boldsymbol{\mu}_{\varphi^i}}$), for parameter update of the actor (Eq. (6)(7)) and critic network (Eq. (8)(9))
	Target Critic	$Q_{\boldsymbol{\mu}_{\varphi^i}}(\boldsymbol{x}^{\cdot}, a^{\cdot 1}, a^{\cdot 2}, ..., a^{\cdot N})$ $\theta^{\cdot i} = \tau\theta^i + (1-\tau)\theta^{\cdot i}$	1) The next observation 2) Output of target actor	The next action-value estimation ($Q_{\boldsymbol{\mu}_{\varphi^i}}$), for the critic network update (Eq. (8)(9))

$(Q_{\boldsymbol{\mu}_{\varphi^i}}(\boldsymbol{x}, a^1, a^2, ..., a^N))$ is approximated by the critic network to assess the action output of the actor network. This value is used to guide the update of the policy network. The objective of the value network is to achieve an unbiased estimation of $Q_{\boldsymbol{\mu}_{\varphi^i}}(\boldsymbol{x}, a^1, a^2, ..., a^N)$. The data sets for training are sampled from $\boldsymbol{D}$, and the loss function (L) is formulated as Eq. (8) and (9). Data training processes are conducted to adjust parameters (θ) of the deep neural network of the critic network and minimize the training loss.

$$L(\varphi^i) = \mathbb{E}_{\boldsymbol{x}, a, r, \boldsymbol{x}'} \left[\left(Q_{\boldsymbol{\mu}_{\varphi^i}}(\boldsymbol{x}, a^1, a^2, ..., a^N) - y \right)^2 \right] \tag{8}$$

$$y = r^i + \gamma Q_{\boldsymbol{\mu}'_{\varphi'^i}}\left(\boldsymbol{x}', a'^1, a'^2, ..., a'^N\right) \Big| a'^i = \boldsymbol{\mu}'_{\varphi'^i}(o'^i) \tag{9}$$

where $-i$ denotes all agents except agent i, and $\boldsymbol{\mu}' = \{\boldsymbol{\mu}_{\varphi'^i}, \boldsymbol{\mu}_{\varphi'^{-i}}\}$ $(i \in \boldsymbol{N}, -i \in \boldsymbol{N}, i \neq -i)$ is the set of target policies with delayed parameters (φ'). It is updated periodically in a soft update way, as presented in Eq. (10). τ is the soft update coefficient. Similarly, the soft update of the delayed parameter (θ') of the target critic network can be conducted according to Eq. (11).

$$\varphi'^i = \tau \varphi^i + (1-\tau)\varphi'^i \tag{10}$$

$$\theta'^i = \tau \theta^i + (1-\tau)\theta'^i \tag{11}$$

It can be observed from Eq. (9) that the training of the critic network involves the policies of all agents. However, due to privacy concerns or communication limitations, this global assumption is impractical. To mitigate this assumption, a decentralized policy approximation is introduced. As outlined in Eq. (12), each agent i undertakes an additional approximation $\widehat{\mu}_{\vartheta_j^i}$ to estimate the true policy of agent j. This approximation process employs an entropy regularizer to maximize the logarithmic probability of agent j's actions, allowing agent iii to learn this approximate policy.

$$L\left(\vartheta_j^i\right) = -\mathbb{E}_{o^j, a^j}\left[\log \widehat{\mu}_{\vartheta_j^i}\left(a^j \middle| o^j\right) + \lambda H\left(\widehat{\mu}_{\vartheta_j^i}\right) \right] \tag{12}$$

where ϑ refers to the approximation parameter. H is the entropy of the policy distribution and λ is its factor.

The structure of the abovementioned networks is summarized in Table 1.

4.2. Modeling energy behavior of multiple energy hubs using MADDPG

The pseudocode of this Markov game of multi-agent environment is presented in below Table and the schematic diagram of MADDPG-based energy behavior modeling is presented in Fig. 3.

The pseudocode of using MADDPG to model the energy behavior of MEH.

	Decision-making process modeling of MEHs using MADDPG
1	Initialize the state of each EH and stochastic process (N)
	Initialize parameters of actor and critic network of each agent i, φ^i and θ^j, respectively
2	**For** episode $ep = 1$ to max_episode:
3	Each agent $i(i \in \boldsymbol{N})$ would interact with CPS $i(i \in \boldsymbol{N})$ to get local observation (o^i)
	Update the noise according to the decay factor (ξ: $\omega^{ep} = \omega^{ep^*}\xi$)
4	**For** time_step $t = 1$ to T_{max}:
5	**For** each agent i:
6	Select the action based on policy $(\boldsymbol{\mu}_{\varphi^i})$ and Gaussian noise (ω_{ep}): $a_t^i \leftarrow N(\boldsymbol{\mu}_{\varphi^i}(o_t^i|\varphi^i), \omega^{ep})$
7	Send the P2P energy sharing action $(a_{t,P2P}^i)$ and multi-energy flow management action $(a_{t,EH}^i)$ to the local energy market and CPS environment respectively for execution

	Decision-making process modeling of MEHs using MADDPG
8	Governed by the state-action pair (s_t^i, a_t^i), according to the state translation function (Eq. (4), calculated by virtual digital models of CPS), the state of CPS environment translates to s_{t+1}^i
9	Update the observation (o_{t+1}^i) and reward (r_t^i) (Eq. (5))
10	Communicate with other agents, and store the tuple $([\boldsymbol{x}_t, \boldsymbol{a}_t, \boldsymbol{x}_t', r_t^i])$ in the replay buffer database $(\boldsymbol{D}^i)$
11	**End for**
12	**End for**
13	**If** len $(\boldsymbol{D}^i) >$ batch_size **then:**
14	**For** each agent i:
15	Sample $([\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{x}', r^i])$ from the replay buffer $(\boldsymbol{D}^i)$ to form a mini batch (M)
16	Update the parameter (θ^i) of the critic network by minimizing the loss function $(L(\varphi^i) = \frac{1}{M}\sum_i \left(Q_{\boldsymbol{\mu}_{\varphi^i}}(\boldsymbol{x}, a^1, a^2, ..., a^N) - y \right)^2)$ using the Adam optimizer (Eq. (8)(9))
17	Update the parameter (φ^i) of actor-network following the max cumulative reward principle using the Adam optimizer (Eq. (6)(7))
18	**End for**
19	Conduct soft parameter updates for the target actor and target critic network according to Eq. (10)(11)
20	**End if**
21	**End for**

5. Scenarios

A digital twin establishes a cyber-physical environment capable of running simulations across various hypothetical scenarios. It enables the digital testing of novel energy system techniques, providing decision-makers with insights into the operational patterns of energy systems under different scenarios. This digital twin-enabled multi-scenario testing approach has the potential to advance energy system development by providing valuable insights to decision-makers before practical implementation. The scenarios explored in this study are illustrated in Fig. 4.

These scenarios involve decision-making strategies employed by autonomous agents and P2P energy sharing strategies regulating energy matching, as elaborated in Sections 5.1 and 5.2, respectively. Additionally, global optimum scenarios are considered as reference cases in this research, accounting for both scenarios with and without P2P sharing. These reference cases are derived using a mixed-integer nonlinear programming (MINLP) solver for a central optimization model. The economic and technical performance of energy systems in these scenarios will be assessed using a comprehensive evaluation system (see Section 6).

5.1. Decision-making strategies of autonomous agents

The interaction between energy hubs (EHs) can take one of two forms: cooperative or competitive. Typically, EHs owned by different stakeholders prioritize their individual benefits and may not share all information with other EHs, leading to a non-cooperative game. Conversely, EHs under the same ownership, like government agencies, are more likely to share all information and engage in cooperative operations.

Consequently, this work examines two categories of decision-making strategies. For competitive EHs, each agent i optimizes its policy to maximize its local reward (r^i), as indicated in Eq. (5), based on its limited knowledge of other agents. The non-cooperative game reaches a Nash equilibrium when no agent can improve its benefits by altering its action policy.

In contrast, for cooperative EHs, all agents engage in bidirectional communication, sharing information including observations (o), actions (a), and policies $(\boldsymbol{\mu})$. They collaborate to achieve the maximum collective total reward R: $R = \sum_{i=1}^{N} r^i$.

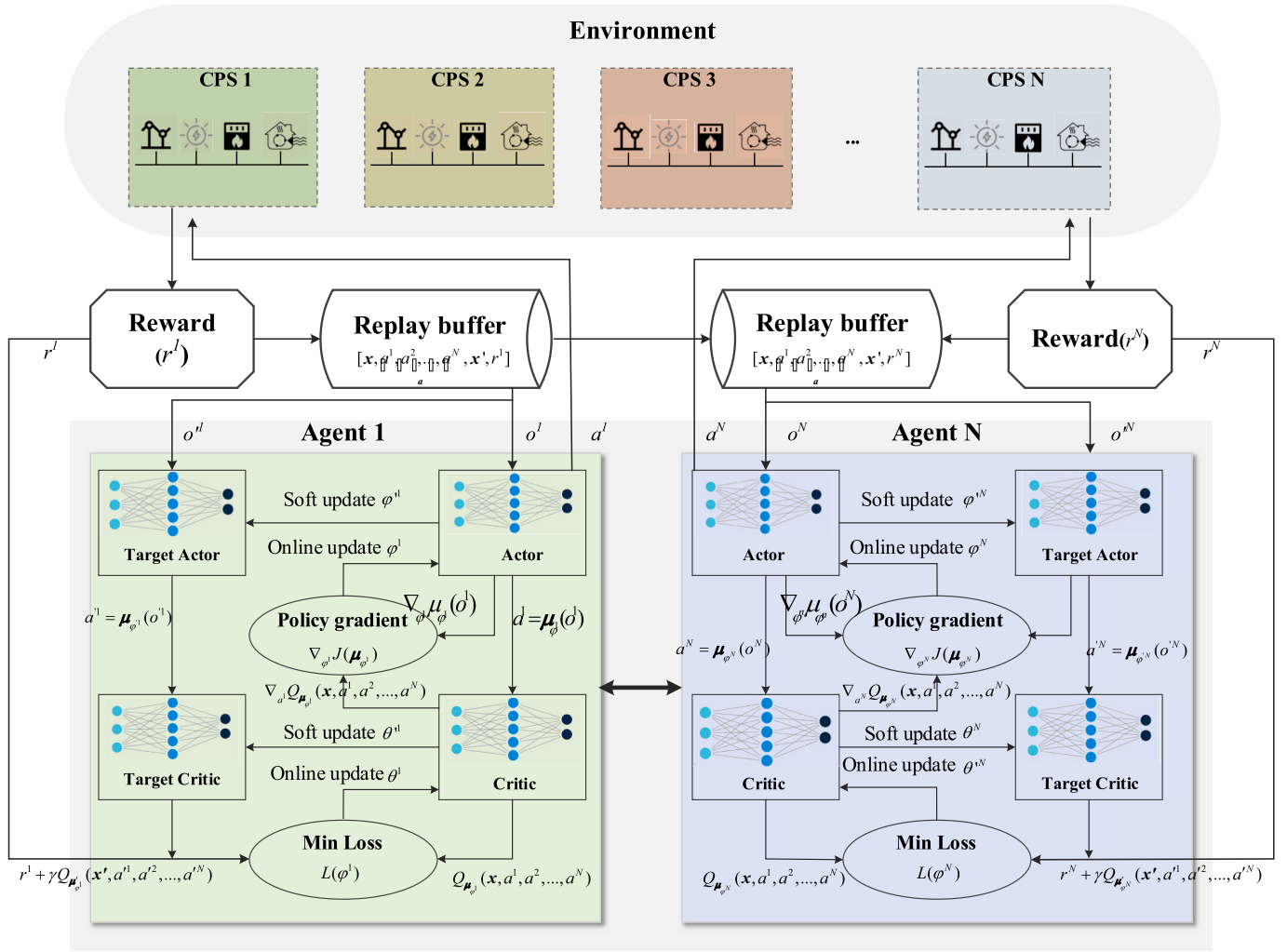Additionally, as a baseline case in this study, a random decision-

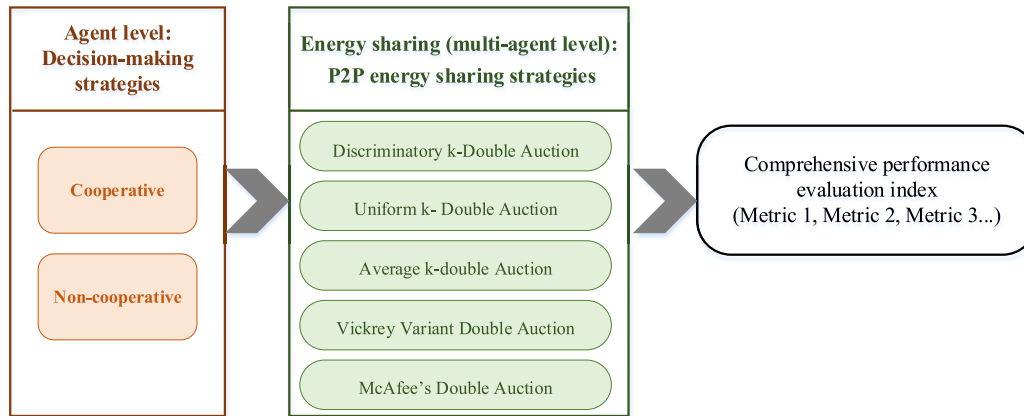**Fig. 3.** The schematic diagram of the MADDPG method [66].

Fig. 4. Hypothetical decision-making and P2P sharing scenarios

making strategy is employed, wherein each agent selects an action randomly from its action space.

5.2. P2P energy sharing strategies

In contrast to conventional energy trading, where EHs can only purchase energy from the upstream grid at a predetermined price (P_{grid}),

and sell their generated energy back at a lower price (P_{feed_in}), the concept of a local energy market (LEM) offers advantages for EHs by allowing them to establish energy prices within the range of (P_{feed_in}, P_{grid}). The effectiveness of the LEM depends on the specific P2P energy sharing strategies employed. A well-designed energy sharing strategy can provide EHs with the incentives to engage in the LEM, resulting in efficient energy trading and fair allocation of payoffs.

In this work, several P2P energy sharing strategies reported in the literature are examined. They include Discriminatory k-Double Auction (k-DA), Uniform k-Double Auction (U-k-DA), Average k-Double Auction (A-k-DA), Vickrey Variant Double Auction (VV-DA), and McAfee's Double Auction (M-DA). The supply and demand curves of P2P energy trading are provided in Fig. 5. Specifically, the energy trading price and quantity are determined based on the information presented in Fig. 5. The mathematical formulations for different strategies are listed in Table 2. For more comprehensive details regarding these P2P sharing strategies, please refer to references [24, 67].

where S denotes the set of all sellers participating in the P2P energy trading; B denotes the set of buyers; $Q*$ is the critical intersection point where aggregate demand and supply meet; Around the intersection point, the lowest cleared bids (including price and quantity) from buyers are denoted as (p_{bL}, q_{bL}), while the second-lowest bids are denoted as (p_{bL-1}, q_{bL-1}); Likewise, highest cleared asks from sellers are denoted as (p_{sH}, q_{sH}), and the second-highest asks are (p_{sH-1}, q_{sH-1}). Here, k is the predetermined constant in the closed interval [0,1].

6. Performance evaluation index

An index system, originally introduced in reference [68], has been modified and utilized in this study to assess the multi-dimensional performance of energy systems, particularly with regard to P2P energy trading. In particular, the economic and technical indices, along with their respective calculation methods, are detailed in Table 3.

The detailed meanings of the parameters in Table 3 are elaborated in the Appendix.

Among these metrics, *VI*, *PI*, and *EI* represent the economic metrics for assessing energy system operation. They provide insights into the extent of cost-saving potential achieved, the additional benefits gained through the adoption of specific P2P energy sharing strategies, and the fairness of payoff allocation among participants based on their contributions.

On the other hand, *EBI*, *PFI*, and *SSI* function as technical metrics, offering insights into the energy balance, peak power shifting, and the reliance on the external bulk power grid when employing specific P2P energy sharing strategies.

All these metrics are normalized within the [0, 1] range to ensure consistent and comparable evaluation. Detailed explanations of these indices can be found in [68].

Notably, The equality index (*EI*) is defined as the Euclidean distance (norm 2) between the profit allocation values of participants and the Sharpe values under a specific P2P trading strategy. This index serves as a metric for assessing the fairness of profit allocation among participants in P2P transactions.

The Shapley value ($income_i^{Shapley}$) is a method used to distribute total gains among participants in a cooperative game. In the context of a cooperative game composed of *n* players (with the set of players denoted as N), let *v* be the characteristic function defined as follows: Given an alliance of players denoted as S, $v(S)$ represents the value of alliance S, indicating the expected sum of rewards that members within alliance S can obtain through cooperation. The function v maps subsets of players to the real number set $v : 2^n \rightarrow R$ and complies with the specific condition $v(\varnothing) = 0$. According to the Shapley value, in a given coalition game (*v*, N), the payoff allocated to player *i* is represented as shown in Eq. (13).

$$\varphi_i(v) = \sum_{S \in N/\{i\}} \frac{|S|!(n - |S| - 1)}{n!} (v(S) - v(S/\{i\}))$$

$$= \sum_{S \in N/\{i\}} \binom{n}{1, |S|, n - |S| - 1}^{-1} (v(S) - v(S/\{i\}))$$

(13)

where, $S \in N/\{i\}$ represents the set of players in alliance S excluding player *i*, and $|S|$ is the number of members within alliance S. This formula can be explained as follows: Each coalition S consists of multiple players, and player *i*'s marginal contribution to coalition S is denoted as $v(S) - v(S/\{i\})$. Player *i*'s total contribution is calculated as the weighted average of their marginal contributions in various possible alliances.

7. Case study design

An energy system that includes 4 energy hubs (4-EH) in China is used as the case study. The physical structure of this system is shown in Fig. 6
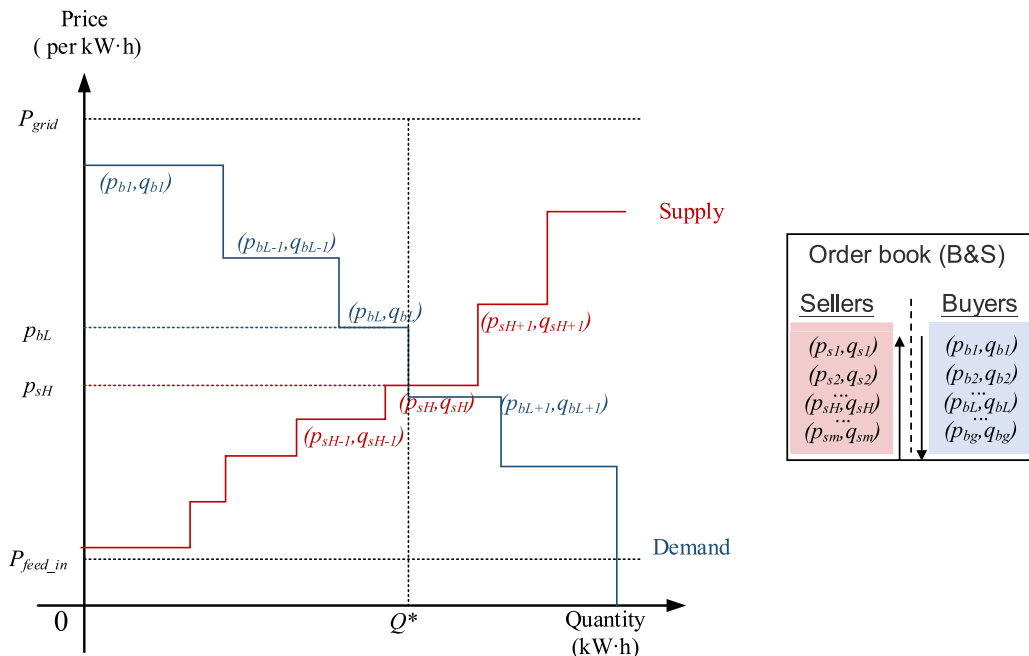

Fig. 5. The supply and demand curves of various P2P sharing strategies [24].

Table 2
P2P energy sharing strategies [24,67].

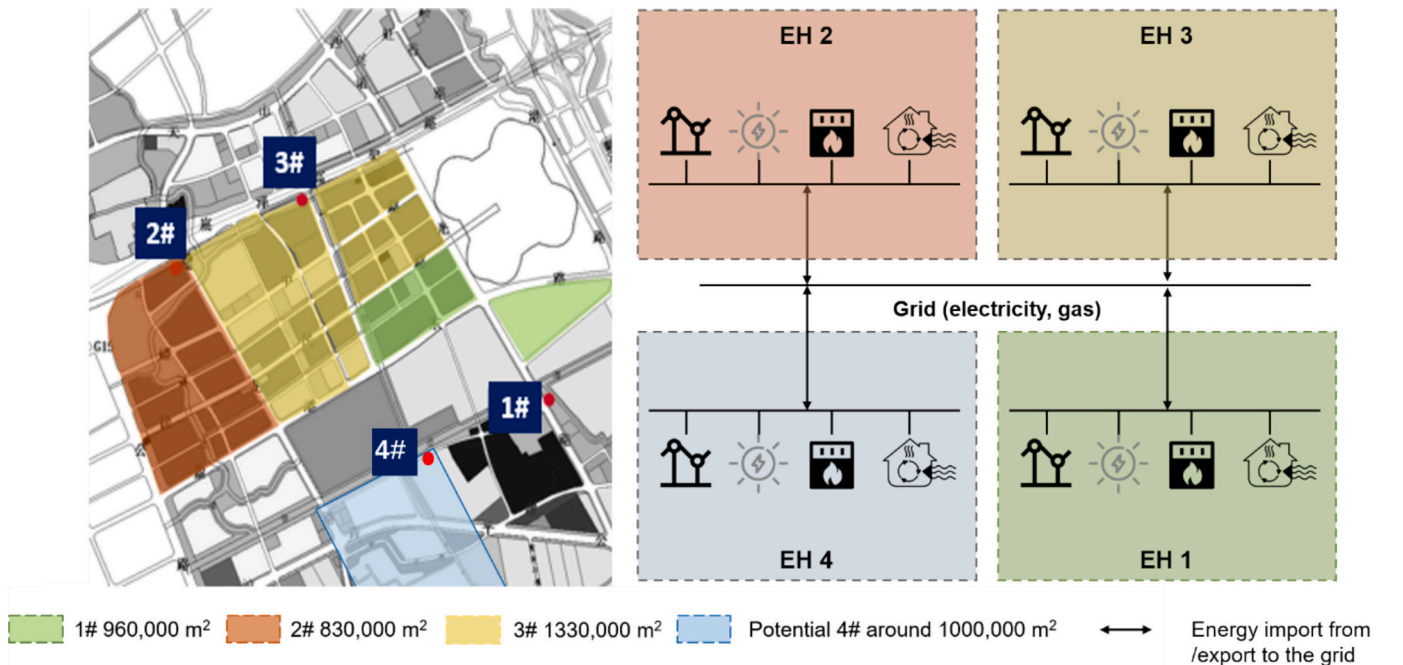| P2P energy sharing strategy	Buying price	Selling price	Trading quantity
Discriminatory k-Double Auction (k-DA)	If $p_{bi} > p_{sj}: p_{i,j} = kp_{bi} + (1-k)p_{sj}$ $(i \in B, j \in S)$		Q^*
Uniform k- Double Auction (U-k-DA)	$kp_{bL} + (1-k)p_{sH}$		Q^*
Average k-Double Auction (A-DA)	$k\sum_{i=1}^{L} p_{bi} + (1-k)\sum_{j=1}^{H} p_{sj}$ $(i \in \boldsymbol{B}, j \in \boldsymbol{S})$		Q^*
Vickrey Variant Double Auction (VV-DA)	p_{bL}	p_{sH}	$\begin{cases}\sum_{j=1}^{H-1} q_{sj}, if \sum_{i=1}^{L-1} q_{bi} \geq \sum_{j=1}^{H-1} q_{sj} \\ \sum_{i=1}^{L-1} q_{bi}, if \sum_{i=1}^{L-1} q_{bi} \leq \sum_{j=1}^{H-1} q_{sj}\end{cases}$
McAfee's Double Auction (M-DA)	$p_0 = \frac{1}{2}(p_{sH+1} + p_{bL+1})$ $\begin{cases} p_0 \in [p_{sH}, p_{bL}] \\ \text{works the same as } U-k-DA \\ p_0 \notin [p_{sH}, p_{bL}] \\ \text{works the same as } VV-DA \end{cases}$		$\begin{cases} p_0 \in [p_{sH}, p_{bL}] \\ \text{works the same as } U-k-DA \\ p_0 \notin [p_{sH}, p_{bL}] \\ \text{works the same as } VV-DA \end{cases}$

Table 3
The economic and technical index system.

Index	Calculation
Value tapping index (*VI*)	$VI = \dfrac{value_{test} - value_{ref}}{value_{max} - value_{ref}}$
Participation willingness index (*PI*)	$PI = 1 - \dfrac{cost_{P2P} - cost_{min}}{cost_{noP2P} - cost_{min}}$
Equality index (*EI*)	$EI = 1 - \sqrt{\sum_{i=1}^{N}\left(income_i - income_i^{Shapley}\right)^2}$
Energy balance index (*EBI*)	$EBI = 1 - \dfrac{\sum_{t\in T}\left\|\sum_{n\in\Psi}\left(\sum_{j\in L}l_{n,j,t} + \sum_{j\in G}g_{n,j,t}\right)\right\|}{\sum_{t\in T}\sum_{n\in\Psi}\sum_{j\in L}l_{n,j,t} + \sum_{t\in T}\sum_{n\in\Psi}\sum_{j\in G}\left\|g_{n,j,t}\right\|}$
Power flatness index (*PFI*)	$1 - \dfrac{\sum_{t\in P\subset T}\left\|\sum_{n\in\Psi}\left(\sum_{j\in L}l_{n,j,t} + \sum_{j\in G}g_{n,j,t}\right)\right\|}{\sum_{t\in T}\left\|\sum_{n\in\Psi}\left(\sum_{j\in L}l_{n,j,t} + \sum_{j\in G}g_{n,j,t}\right)\right\|}$
Self-sufficiency index (*SSI*)	$SSI = 1 - \dfrac{\sum_{t\in T^+}\sum_{n\in\Psi}\left(\sum_{j\in L}l_{n,j,t} + \sum_{j\in G}g_{n,j,t}\right)}{\sum_{t\in T}\sum_{n\in\Psi}\sum_{j\in L}l_{n,j,t}}$ $T^+ = \left\{t\Big|\sum_{n\in\Psi}\left(\sum_{j\in L}l_{n,j,t} + \sum_{j\in G}g_{n,j,t}\right) > 0, t\in T\right\}$

7.1. The cyber-physical system: 4-EH system

This 4-EH system comprises four energy hubs {EH 1, EH 2, EH 3, EH 4} designed to operate autonomously. These hubs are interconnected by P2P energy sharing and distribution networks such as the grid. The internal balance of multi-energy supply and demand within each hub is achieved through 1) purchasing or selling energy from/to the upstream grid, and 2) P2P energy sharing with other interconnected EHs.

P2P energy trading occurs directly between EHs, allowing each EH to sell its surplus energy to or purchase energy to cover deficits from other EHs. This enables all EHs to benefit from local P2P energy prices. Local energy prices are typically lower than the purchase price and higher than the feed-in tariffs offered by the upstream grid. Additionally, achieving local energy balance through P2P sharing can reduce the EHs' dependence on the grid, facilitating peak load reduction and energy balancing.

Given the fact that heating and cooling distribution networks are rarely shared between independent EHs, and the long-distance

1# 960,000 m²   2# 830,000 m²   3# 1330,000 m²   Potential 4# around 1,000,000 m²   ⟷ Energy import from /export to the grid

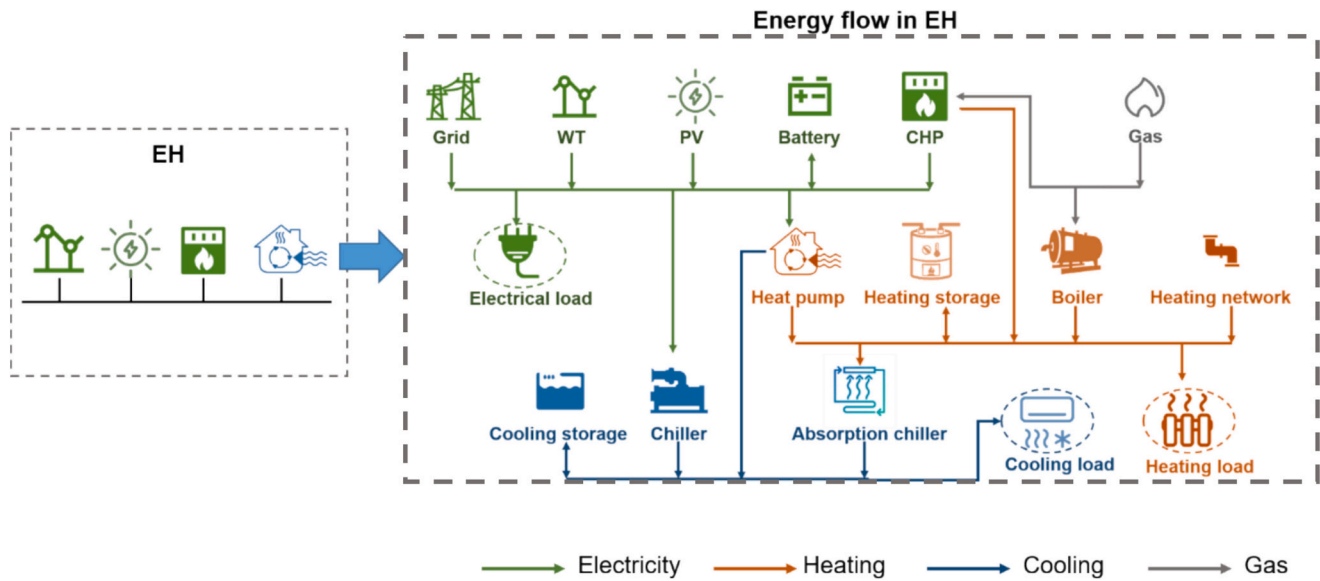**Fig. 6.** The structure of the studied 4-EH system.

Fig. 7. Multi-energy flows of a typical energy hub.

transmission of hot/chilled water results in significant losses, this study focuses solely on P2P electricity trading.

Specifically, the multi-energy structure of a representative EH studied in this work is depicted in Fig. 7. This EH has access to various energy sources, including the upstream grid, which provides electricity, gas, and space heating, the local P2P electricity market, and renewable generation devices such as photovoltaic (PV) panels and wind turbines (WT).

After acquiring multi-energy inputs, these inputs are either converted into other forms or dispatched to end-users to fulfill their multi-energy demands. Several energy conversion devices are considered in this study, including:

1) Combined heat and power (CHP): CHPs generate both electricity and heat by consuming natural gas, reducing energy costs due to the typically lower price of natural gas compared to electricity. CHPs can directly supply electricity, provide space heating through heat exchangers, and provide cooling by coupling with absorption chillers to serve end-users.
2) Absorption chiller (AC): ACs can utilize the heat produced by CHP to generate cooling, contributing to efficient cascade utilization of energy.
3) Electrical chiller (EC): ECs are used for cooling.
4) Heat pump (HP): HPs are employed for heating and cooling applications.
5) Gas boiler (GB): GBs are used for heating by burning natural gas.

In addition to these energy conversion devices, energy storage (ES) units such as batteries and thermal tanks are also integrated into the EH system. These units play a crucial role in addressing temporal and spatial mismatches between energy demand and supply.

Different forms of energy are distributed to end-users to meet their electrical and thermal (heating and cooling) demands. End-users in this study include office, commercial, and hotel buildings. This study also explores thermal demand response measures for these buildings.

Taking EH1 as an example, detailed information about its energy conversion devices, renewable energy generation and energy storage are provided in Table 4, Table 5, and Table 6 respectively.

To formulate the multi-energy input/output and energy conversion of each connected EH, this work employs a coupling matrix-based approach for energy hubs, as proposed in [69]. To determine the thermal state of different building types with varying occupant schedules

Table 4
Energy conversion devices in EH1.

No.	Device	Capacity (kW)	Number	Total capacity (kW)	Rated efficiency
1	CHP	4624	4	18,496	1) 0.5 for power generation; 2) 0.35 for heat generation
2	LiBr absorption chiller	4324	3	12,972	1.2
3	Electrical chiller	6997	8	55,976	5.5
4	Heat pump	Cooling: 1052 Heating: 1094	8	Cooling: 8416 Heating: 8752	Cooling: 3 Heating: 3.5
5	Gas boiler	6000	2	12,000	0.95

Table 5
Renewable energy generation of EH1.

No.	Device	Rated capacity (kW·h)
1	PV	20,000
2	WT	30,000

Table 6
Energy energy storage of EH1.

No.	Device	Rated capacity (kW·h)	Rated efficiency
1	Battery	1280	Charge:0.92; Discharge: 0.85
2	Heating/ Cooling storage	20,000	Charge:0.90; Discharge: 0.85

(such as office and commercial buildings), aggregated building thermal dynamic models reported in [70] are developed using a data-driven approach. The historical data required for these models can be provided by Cyber-Physical Systems (CPSs).

Additionally, it is essential to ensure that the operational constraints of EHs are met. For instance, constraints related to the state of energy

storage units, including their lower and upper limits. The state of these energy storage units can be calculated using the model presented in reference [71].

EHs located within the same geographic area generally share a similar energy flow structure, as illustrated in Fig. 7, but may exhibit slight variations in device capacities, renewable energy generation, and demand-side flexibility. To account for these differences and capture the interactions among EHs with varying characteristics, the following distinctions in supply, demand sides, and storage units, are considered:

1) EH 1: This EH has higher thermal tolerance, allowing greater flexibility in responding to thermal demand fluctuations.
2) EH 2: EH 2 does not have an energy storage unit available.
3) EH 3: EH 3 lacks renewable generation (e.g., PV or WT).
4) EH 4: EH 4 is equipped with 40,000 kW·h of PV generation capacity, enhancing its renewable energy generation capability.

These distinctions allow for a more comprehensive analysis of interactions and operations of heterogeneous EHs within the same geographical area.

The natural gas price is set as ¥3.3/Nm^3 and that for space heating is ¥0.36/kW·h. The electricity selling price is ¥0.135/kW·h and purchasing prices are depicted in Fig. 8.

7.2. The multi-agent reinforcement learning-based energy behavior modeling:4-agent system

As stated in Section 4.2, for agent i, its observation (o_t^i) at time t (see Eq. (14)) includes 1) public information ($I_{P,t}^i$): ambient environment temperature ($T_{a,t}$), solar radiation ($Solar_t$), time ($hour_t$), time of use energy tariff $\left(p_{grid\_e,t}, p_{grid\_h,t}, p_{grid\_g,t}\right)$, and 2) private information ($I_{p,t}^i$): PV generation (PV_t^i), WT generation (WT_t^i), electrical load ($L_{e,t}^i$) at time t, aggregated indoor air temperature ($T_{in,t-1}^i$), $L_{e,t}^i$ the state of battery ($S_{e,t-1}^i$), the state of cooling storage ($S_{c,t-1}^i$), the state of heating storage ($S_{h,t-1}^i$) at the end of time t-1.

Subsequently, the energy behavior (a_t^i) of agent i is generated by the actor network according to its observation (o_t^i). The quality of observation whether it is perfect or imperfect, depends on the cooperative or non-cooperative interaction among EHs. Action (a_t^i) indicting energy behaviors of agent i encompasses both its P2P energy sharing behaviors, including price and quantity $\left(p_{P2P,t}^i, q_{P2P,t}^i\right)$, and multi-energy management behaviors within the corresponding EH. The latter includes the multi-energy exchanges with the upstream grid ($P_{e,t}, P_{g,t}, P_{h,t}$), the generation of CHP ($G_{CHP,e,t}^i, G_{CHP,h,t}^i$), heat pump ($G_{HP,t}^i$), electrical chiller ($G_{EC,t}^i$), gas boiler ($G_{GB,t}^i$), and absorption chiller ($G_{AC,t}^i$) as well as the charge/discharge rates of the battery ($C_{e,t}^i$), cooling storage ($C_{c,t}^i$), and heating storage ($C_{h,t}^i$). The a_t^i of agent i is formulated in Eq. (15).

In summary:

$$o_t^i = \left\{ \underbrace{T_{a,t}, Solar_t, hour_t, p_{grid\_e,t}, p_{grid\_g,t}, p_{grid\_h,t}}_{\text{Public information}}, \underbrace{PV_t^i, WT_t^i, L_{e,t}^i, T_{in,t-1}^i, S_{e,t-1}^i, S_{c,t-1}^i, S_{h,t-1}^i}_{\text{Private information of agent } i} \right\} \tag{14}$$

$$a_t^i = \left\{ \underbrace{p_{P2P,t}^i, q_{P2P,t}^i}_{\text{P2P energy sharing among EHs}}, \underbrace{P_{e,t}^i, P_{h,t}^i, P_{g,t}^i, G_{CHP,e,t}^i, G_{CHP,h,t}^i, G_{EC,t}^i, G_{GB,t}^i, G_{AC,t}^i, C_{e,t}^i, C_{c,t}^i, C_{h,t}^i}_{\text{Multi-energy management within EH } i} \right\} \tag{15}$$

Aligned with the 4-EH physical system, this work develops a reinforcement learning-enabled 4-agent system to mirror the decision-making processes of each EH in various scenarios (see Section 5).

Actions including both the P2P energy sharing (a_{P2P}^i) and internal energy management (a_{EH}^i) are executed in the cyber-physical system (CPS). The state transition and rewards for the state-action pair (s_t, a_t)
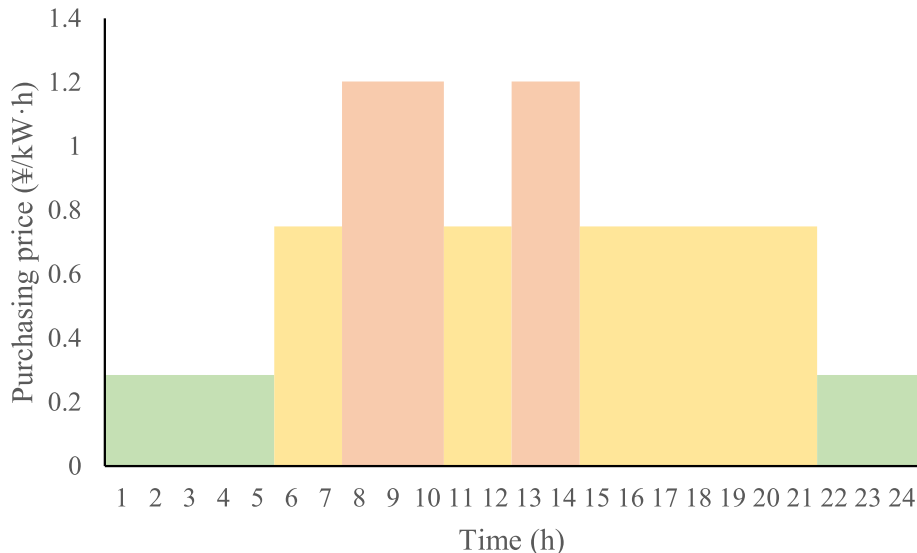

Fig. 8. Purchasing price of electricity (time of use).

Table 7

hyperparameters in multi-agent reinforcement learning.

Hyperparameters		
Actor/Critic network	Number of layers	5
	Number of neurons in hidden layers	64
	Activation function	Tanh
	Learning rate	5e-4
Time step		1 h
Episode length		24 h
Discount factor		0.99
Batch size		32
Training interval		1
Noise decay		0.99

are calculated by CPS. The reward (r_t^i) received by agent i as a result of the chosen state-action pair (s_t^i, a_t^i) is calculated based on the energy cost ($Cost_t^i$) of EH i and the deviation ($T_{exceed,t}^i$) of the indoor air temperature ($T_{in,t}^i$) from the upper limit ($\overline{T}_{tolerance}^i$) and lower limit ($\underline{T}_{tolerance}^i$) of the temperature tolerance. This deviation serves as a penalty (*thermal_factor* is the penalty factor), as outlined in Eq. (16).

$$r_t^i = -\left(Cost_t^i + thermal\_factor \times T_{exceed,t}^i\right)T_{exceed,t}^i$$
$$= \begin{cases} T_{in,t}^i - \overline{T}_{tolerance}^i \text{ cooling season} \\ \underline{T}_{tolerance}^i - T_{in,t}^i \text{ heating season} \end{cases} \quad (16)$$

The virtual operational trajectory ($s_t^i, a_t^i, s_{t+1}^i, r_t^i$) of CPS is stored as an independent digital database for each agent. This database serves as a source of data for training through a data-driven approach, enhancing the decision-making capabilities and reasoning of individual agents. This experience replay mechanism is a well-established technique in reinforcement learning and is explained in detail in Section 4.

The hyperparameters used in this 4-agent system for reinforcement learning are listed in Table 7.

It is notable that, as the Tanh activation function is used, the action output of the actor network is in the range $(-1,1)$. To meet the physical constraints of devices within the EH, such as operational boundaries, the action output is scaled to the operational range (lower limit, upper limit) of each device, such as CHP, heat pumps, and boilers. In addition to the operational boundaries (charging and discharging rate constraints), there is a capacity limit for energy storage units. Therefore, after the scaling process, an additional clipping process is necessary to ensure that the capacity of each storage unit remains within its minimum and maximum limits, preventing overcharging. As mentioned in Section 7.1, the operational boundaries are provided by CPSs.

Table 8

a) The cases on a typical summer day. b) The cases on a typical winter day.

a)

		P2P energy sharing strategies				
		k-DA	U-k-DA	A-DA	VV-DA	M-DA
Decision-making strategy	Cooperative	S-C	S-C-U	S-C-A	S-C-VV	S-C-M
	Non-cooperative	S-N	S-N-U	S-N-A	S-N-VV	S-N-M
	Random	S-R	S-R-U	S-R-A	S-R-VV	S-R-M

b)

Decision-making strategy	Cooperative	W-C	W-C-U	W-C-A	W-C-VV	W-C-M
	Non-cooperative	W-N	W-N-U	W-N-A	W-N-VV	W-N-M
	Random	W-R	W-R-U	W-R-A	W-R-VV	W-R-M

7.3. Multi-scenario case design

A series of scenarios are conducted using the proposed extensive digital twin to simulate the operation of the 4-EH system under various decision-making and P2P sharing strategies during the typical summer and winter day. These scenarios are detailed in Table 8, each associated with a unique identifier.

To account for the stochastic nature of autonomous participants in the local energy market, simulation-based tests are repeated 100 times. This iterative approach allows for the collection of statistical information that reflects the system's energy behaviors, enabling a comprehensive analysis of system performance and outcomes.

In Table 8, k-DA, U-k-DA, A-DA, VV-DA and M-DA refer to different P2P energy sharing strategies discussed in this work: Discriminatory k-Double Auction, Uniform k-Double Auction, Average k-Double Auction, Vickrey Variant Double Auction, and McAfee's Double Auction, respectively. The definition and calculation of these strategies can be found in Section 5.2 and references [24, 67].

To evaluate the performance of the cases and measure the economic and technical advantages resulting from P2P local energy sharing, two reference cases were simulated using the digital twin-based approach:

1) Centralized optimization model considering P2P energy sharing (COP): This reference case incorporates P2P energy sharing into the centralized optimization.
2) Centralized optimization model without P2P energy sharing (CO): This reference case does not consider P2P energy sharing in the centralized optimization model.

By comparing the results of the test cases with these reference cases, the impact and benefits of P2P local energy sharing on both economic and technical aspects can be assessed.

8. Simulation results and discussion

The test cases of the proposed extensive digital twin were programmed in Python using TensorFlow on a laptop with an Intel i7-9750H CPU and 8 GB RAM. The reference cases were modeled as a nonlinear problem in Python and solved using Gurobi, which is a commercial optimization solver widely used for linear, nonlinear, and mixed-integer programming problems.

As the actor-critic structure of MADDPG is adopted in this work, the decision-making and training processes are independent. Decision-making occurs at every time step (1 h in this study) using a well-trained neural network, with the average time to compute an action being 6 s. The training process is scheduled at set intervals, specifically after each episode (24 h in this study), and each training session takes 17 min on average across repeated simulations. Notably, the training and decision-making processes are conducted separately, allowing the actor to use parameters updated from the training without impacting the calculation time for decision-making.

For the centralized reference cases, where temporal coupling is primarily introduced by energy storage, the optimization of an entire episode is formulated as a large-scale complex MINLP problem. The computation time per episode using Gurobi is 95 min, with a tolerance gap set to 0.0001. This time may vary based on the specific solver settings and tolerance conditions applied.

8.1. Economic and technical benefits of P2P local energy trading

In Fig. 9 and Fig. 10, COP and CO represent the centralized optimization cases with and without P2P energy sharing, respectively. These two cases serve as reference cases under ideal assumptions, where full access to all information and a central operator is available. This comparison illustrates the economic and technical benefits of P2P energy sharing. In these figures, the financial benefits are primarily identified
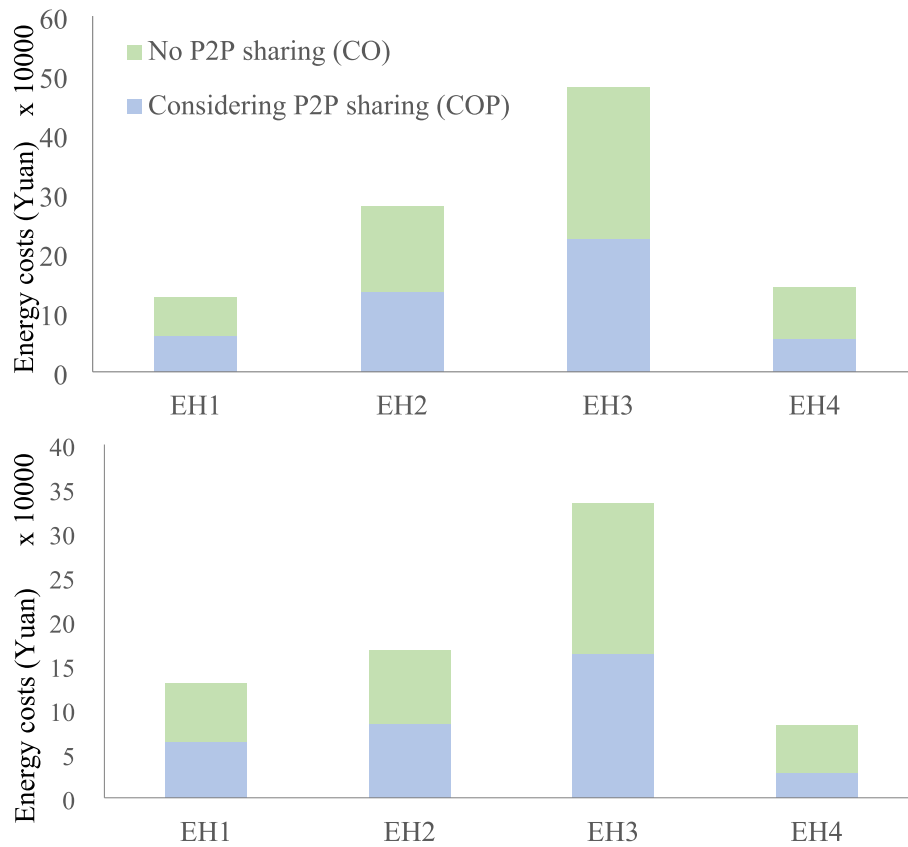
Fig. 9. a) Costs of EHs in CO and COP cases (typical summer day).
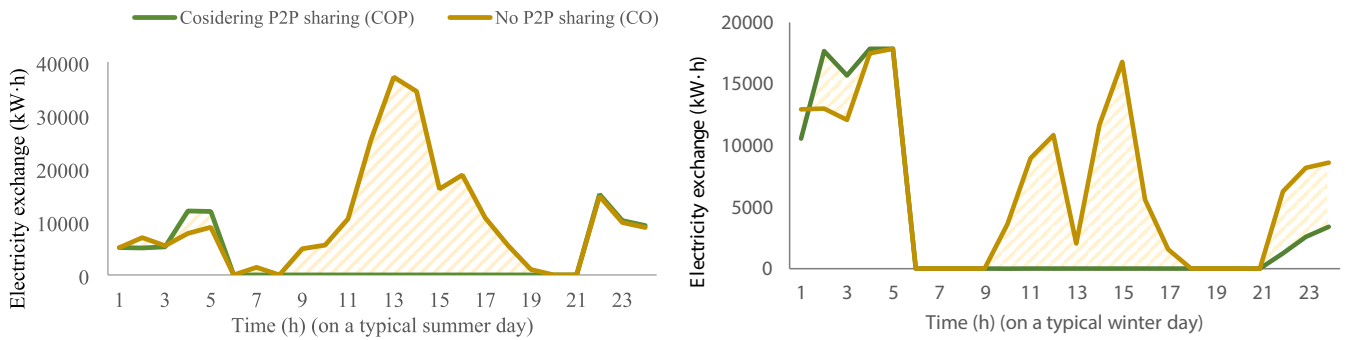b) Costs of EHs in CO and COP cases (typical winter day).

Fig. 10. a) Energy exchange between 4-EH system and the grid.
b) Energy exchange between 4-EH system and the grid.

through energy cost savings achieved by participating in P2P trading, while the technical benefits are realized by reducing electricity exchanges between the 4-EH system and the upstream grid, thus enhancing system-wide self-sufficiency.

Among these EHs, EH 4 gains the most significant benefits, achieving energy bill savings of 25.7 % on a typical summer day and 29.6 % on a typical winter day, primarily due to its larger PV capacity. EH 4 can sell surplus PV generation to other EHs at a relatively high price, resulting in a profit. EH 3 follows as the second-largest beneficiary, with energy bill savings of 12.0 % on a typical summer day and 7.2 % on a typical winter day. Since EH 3 does not possess renewable generations, it benefits from purchasing electricity from other EHs at a relatively low price through P2P sharing.

Overall, the total energy bill savings generated by P2P sharing of this 4-EH system amount to ¥72,538 (15.2 %) on a typical summer day and ¥33,647 (9.9 %) on a typical winter day.

Apart from the economic benefits, from a technical perspective, the P2P local energy market facilitates reducing the electricity exchange between the 4-EH energy system and the upstream grid, particularly during peak periods. This reduction enables more extensive peak shaving and valley filling, as depicted in Fig. 10. In the scenario considering local P2P trading (COP), each EH prioritizes purchasing or selling energy to meet local load deficits or surplus through P2P trading with other EHs. This reduces the need to procure or sell electricity from/ to the upstream grid, thereby enhancing the system's self-sufficiency and resilience.

8.2. Multi-energy management of each EH

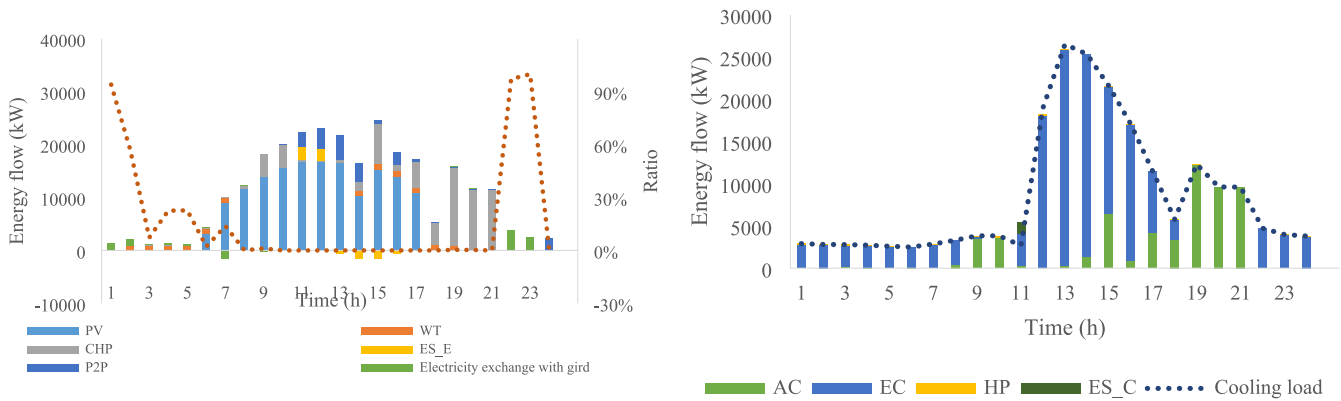In addition to the P2P energy sharing discussed in Section 8.1, agents

Fig. 11. a) Electricity flow dispatch in EH1 (typical summer day).
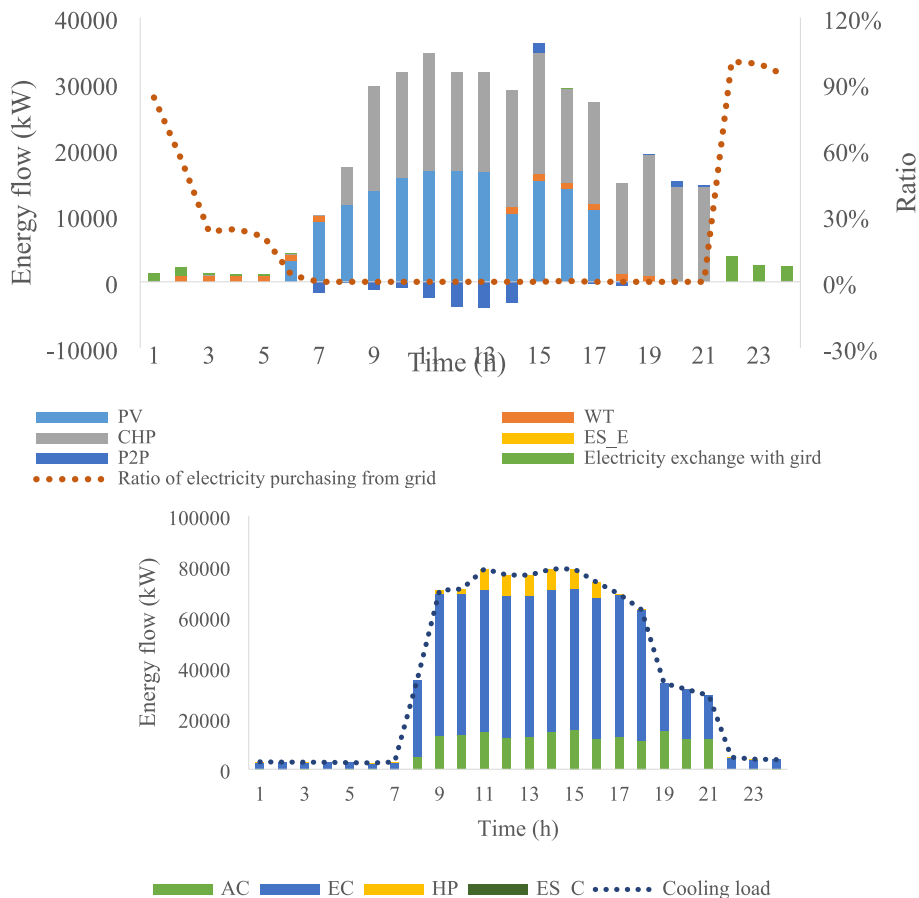b) Cooling flow dispatch in EH1 (typical summer day).

Fig. 12. a) Electricity flow dispatch in EH2 (typical summer day).
b) Cooling flow dispatch in EH2 (typical summer day).

also determine the inner multi-energy flow dispatch within each EH. This includes the coordination of energy input, energy storage, and energy conversion devices to meet power, cooling, and heating demands. For example, the power and cooling loads of each EH on a typical summer day—supplied through input energy, energy conversion devices, and energy storage unit are depicted in Figs. 11–14. These Figures aim to illustrate how electricity and cooling flows are managed or allocated within each EH.

The simulation results indicate that during valley electricity price periods (such as 1:00–6:00 and 22:00–24:00) when both renewable energy production and the cooling and heating demands are at their lowest, EHs tend to purchase electricity from the upstream grid at a reduced price to meet their energy needs. Conversely, during peak and flat electricity price periods (7:00–22:00), as power and thermal demands rise, reliance on the upstream grid decreases. During these times, EHs are more inclined to utilize increasing renewable energy generation and employ the CHP-AC system to produce electricity while recovering waste heat from the electricity generation process. This approach effectively and economically meets the multi-energy demands of end-users.

In cases of electricity surplus or deficit, excess energy generation is either stored in energy storage units (ES) or shared with other
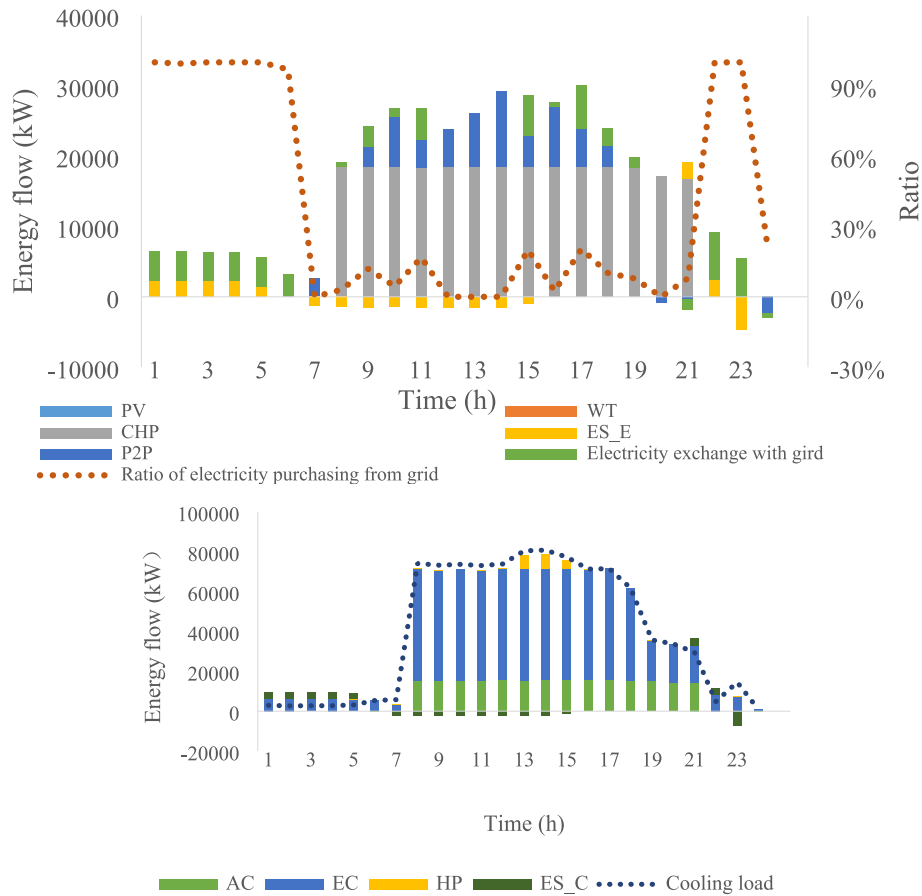
Fig. 13. a) Electricity flow dispatch in EH3 (typical summer day).
b) Cooling flow dispatch in EH3 (typical summer day).

interconnected EHs via the local P2P market. Any remaining surplus or unmet demand that cannot be addressed through storage or P2P trading is subsequently managed by exporting to or importing from the upstream grid. For example, EH 4, which possesses significant renewable generation capacity, manages peak production periods (such as at 13:00) by first meeting its energy requirements, then storing excess energy or selling it to other EHs through P2P sharing. If additional electricity remains, it is exported to the upstream grid. (see Fig. 14).

Regarding conversion devices, priority is given to those with higher energy efficiency to address cooling and heating needs. In terms of cooling, absorption chillers (ACs) are used to meet part of the cooling demand, while high-efficiency electrical chillers (ECs) handle the remaining cooling needs. Heat pumps (HPs) complement the system during peak cooling periods. For heating, waste heat from CHP is primarily recycled using heat exchangers, with additional heating requirements met by HPs. Gas boilers (GBs) are employed during peak heating periods.

As for thermal demand response, load profiles (such as those shown in Fig. 11b–14b) demonstrate that employing demand response measures (such as a broader indoor air temperature comfort zone) enables EH 1 to significantly reduce its thermal demands and achieve lower energy costs compared to other EHs.

8.3. Comparative analysis of 4-EH system across various scenarios

8.3.1. Multi-dimensional performance evaluation of various P2P sharing strategies

Appropriate P2P energy sharing can incentivize EHs to participate in the local P2P energy market and allocate payoffs fairly based on their contributions. However, due to privacy concerns, non-cooperative

decision-making strategies are more common among EHs, even though they may lead to performance degradation owing to imperfect information.

In non-cooperative scenarios, EHs have different roles in the P2P local energy market based on their characteristics: EH4, with its larger renewable generation capacity, functions as a producer, selling surplus PV generation to other EHs; EH3, which lacks access to PV or WT generation, purchases electricity from other EHs; EH2, without a storage unit, tends to sell all extra renewable generation through the P2P market; and, due to thermal demand response measures, EH1 has the lowest thermal load to meet, resulting in less frequent trading with other EHs.

Different P2P sharing strategies employ different trading mechanisms, leading to variations in hourly power exchanges in the P2P local market: Average k-Double Auction (A-DA) promotes P2P energy sharing and achieves the highest transaction frequency and volume, followed by Discriminatory k-Double Auction (k-DA) and Uniform k-Double Auction (U-k-DA); In contrast, McAfee's Double Auction (M-DA) and the Vickrey Variant Double Auction (VV-DA) show limited effectiveness in encouraging P2P sharing, possibly due to their more complex mechanisms.

It is noteworthy that M-DA and VV-DA result in the highest energy bills on both typical summer and winter days, suggesting that simpler P2P trading rules, like A-DA, can provide greater financial benefits through energy sharing. The financial benefits realized by each EH through participation in P2P energy sharing on typical summer and winter days are shown in Table 9.

Overall, the daily energy cost of the 4-EH system (see Section 8.3.2) corresponds with each P2P strategy's effectiveness in stimulating energy sharing, highlighting the financial advantages of P2P energy sharing in the local energy market.

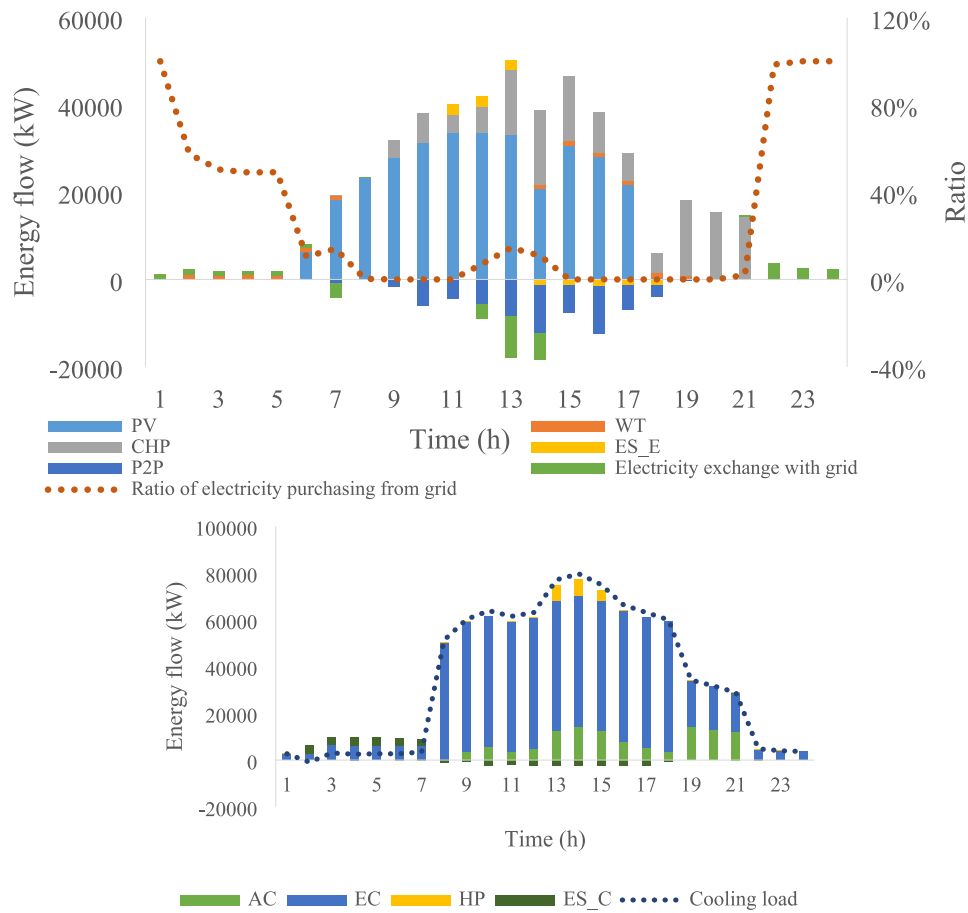In addition to the financial benefits associated with different P2P

Fig. 14. a) Electricity flow dispatch in EH4 (typical summer day).
b) Cooling flow dispatch in EH4 (typical summer day).

Table 9
Financial benefits of each EH resulting from P2P energy sharing.

Cost saving	EH 1	EH 2	EH 3	EH 4
Typical summer day	¥ 9857.6 (14.8 %)	¥ 5369.0 (3.7 %)	¥ 31,412.8 (12.0 %)	¥ 21,309.1 (25.7 %)
Typical winter day	¥ 7707.1 (10.8 %)	¥ 2003.4 (2.4 %)	¥ 13,260.6 (7.2 %)	¥ 16,019.4 (29.6 %)

sharing strategies discussed earlier, a comprehensive evaluation is conducted, taking into account various dimensions, including economic and technical impacts. Detailed information about this multi-dimensional evaluation index system can be found in Section 6. The simulation results in Fig. 15, present assessments for both a typical summer day (left) and a typical winter day (right). These assessments provide a more holistic view of how different P2P energy sharing strategies discussed in Section 5.2), influence various aspects of the energy system.

When using different P2P trading strategies, the 4-EH energy system exhibits similar performance in terms of energy balance (*EBI*) and self-sufficiency (*SSI*), with no significant differences observed. However, notable variations appear in other indicators, particularly the value tapping index (*VI*), participation willingness index (*PI*), and power flatness index (*PFI*).
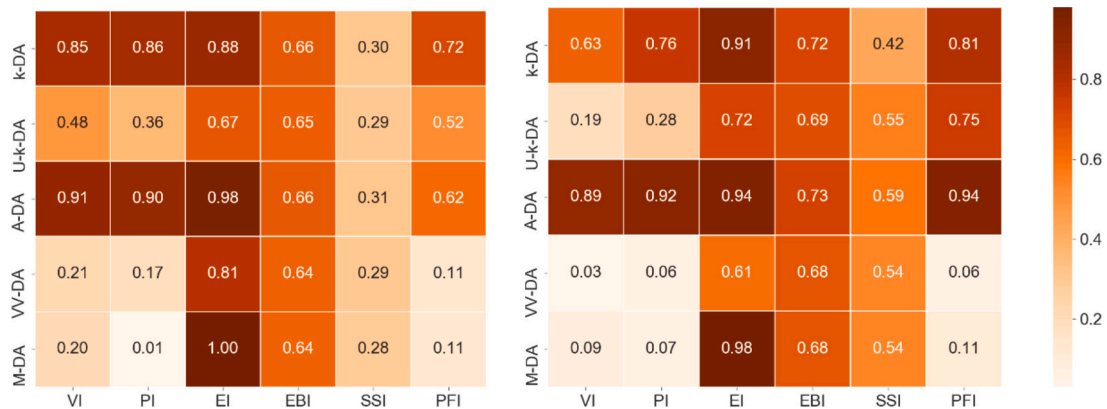

Fig. 15. Multi-dimensional performance evaluation of various P2P sharing strategies. x-axis includes evaluation indices such as *VI*, *PI*, and other metrics. The y-axis covers the six P2P energy sharing strategies. A larger value indicates better performance.

The A-DA strategy performs well across all evaluation indicators due to its straightforward supply and demand matching mechanism and fair price determination. Following A-DA is the k-DA strategy. The U-k-DA strategy, a variant of k-DA that uses uniform transaction prices for all EHs, fails to differentiate EHs' contributions to P2P sharing based on price. As a result, the equality index (*EI*) is lower than that of the k-DA strategy, and it does not effectively incentivize EHs to participate in P2P trading.

As previously mentioned, M-DA and VV-DA, being the two most complex P2P trading strategies in terms of trading mechanisms, do not adequately encourage EHs' participation in P2P trading. Consequently, these strategies lead to lower transaction frequency and volume and perform worse than the other three P2P trading strategies across various indicators, except for the equality index (*EI*).

8.3.2. Energy cost of 4-EH system using various decision-making and P2P sharing strategies

Given the stochastic nature of the decision-making processes in the studied 4-EH system, repeated simulation experiments were conducted in this work. Statistical distributions of daily energy costs were obtained for different strategies that incorporate both decision-making and P2P energy sharing strategies, as illustrated in Fig. 17.

The results indicate that when connected EHs employ non-cooperative strategies with partial information sharing, the resulting energy costs fall between those of cooperative cases and random decision-making scenarios. This suggests that the virtual data-driven learning process for agents is effective and provides better economic performance compared to random energy behavior selection (as seen in the random cases). The random strategy presents a much wider distribution of costs on both typical summer and winter days and results in higher energy costs.

The cooperative strategy achieves the lowest energy costs and exhibits a narrower cost distribution. For example, under the k-DA strategy on a typical summer day, the cooperative decision-making approach achieves an average energy cost reduction of 1.85 %, while under the A-DA strategy, it achieves a 1.32 % reduction compared to non-cooperative cases. This indicates that cooperation among EHs, including information sharing and a common optimization goal, can mitigate the stochasticity in the multi-EH decision-making process. The cooperative strategy not only regulates energy sharing behavior in the

P2P local energy market but also optimizes multi-energy dispatching within each EH, thereby accessing greater financial benefits.

Regarding the P2P energy sharing strategies, a similar trend is observed on both typical summer and winter days. Specifically, A-DA delivers the best economic performance, followed by k-DA, which is widely used in existing studies on P2P trading. In contrast, the energy bills under M-DA and VV-DA sharing strategies are the highest, with potential reasons for this analyzed in Section 8.3.1.

When compared with the reference case (COP), nearly optimal results are achieved when cooperative EHs engage in P2P energy sharing using the A-DA and k-DA strategies, with deviations listed in Table 10. This outcome demonstrates the effectiveness of the proposed multi-agent reinforcement learning-based energy behavior prediction method. It also suggests that the self-interested decision-making behavior of individual EHs, aimed at maximizing their own benefits while keeping private information unshared, can lead to a slight degradation in the overall economic performance of the system.

In this table, S refers to a typical summer day, and W denotes a typical winter day. C represents cooperative scenarios, while N stands for non-cooperative scenarios. A indicates the A-DA strategy, the. The specific identifier of cases can be found in Section 7.3. COP refers to the centralized optimization model that incorporates P2P energy sharing.

9. Conclusion and future work

The multiple energy hub systems (MEH) with local P2P energy sharing present a promising paradigm for future smart grid. MEHs enable direct energy trading among multiple EHs, promoting the efficient utilization of locally generated renewable energy and enhancing overall system flexibility.

To provide energy system operators with valuable insights into the operation of MEHs under uncertain future scenarios, we proposed an extensive digital twin (EXDT). This approach goes beyond conventional digital twins, which typically focus on the cyber-physical system (CPS) and are traditionally limited to historical data analysis and real-time monitoring in the energy sector. The EXDT integrates multiple modules, including a data-driven energy behavior prediction model, hypothetical scenarios, evaluation metrics, and the CPS, all interconnected through data exchange.

Specifically, the EXDT incorporates an array of hypothetical
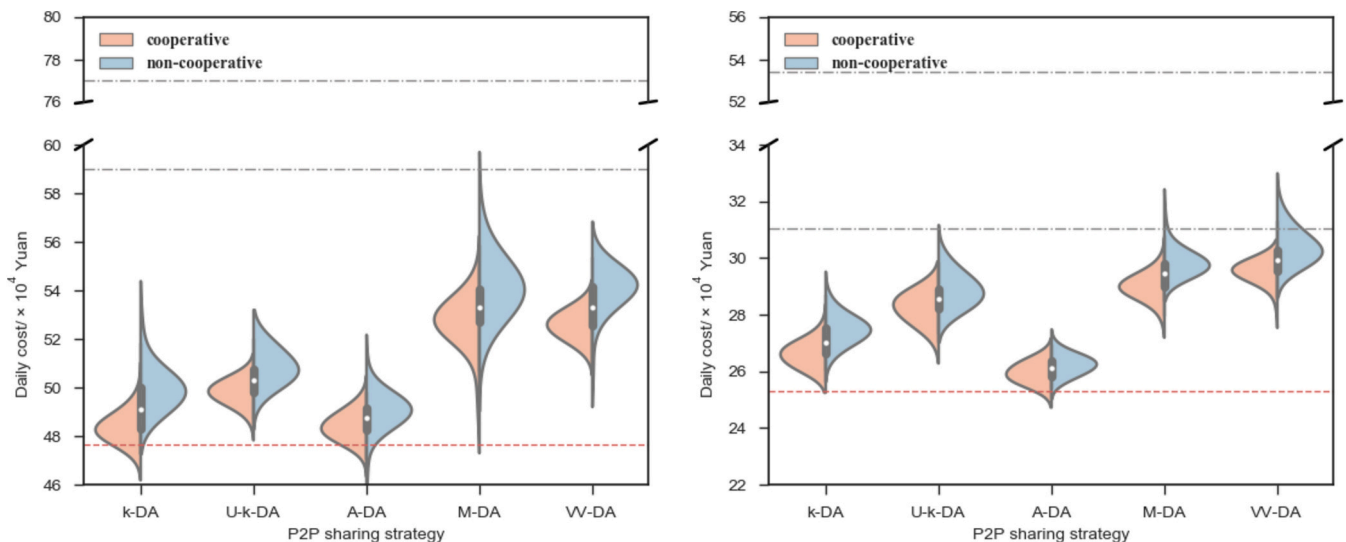
Fig. 16. a) Energy cost of 4-EH system in various scenarios on a typical summer day.
b) Energy cost of 4-EH system in various scenarios on a typical winter day. The red line represents the energy cost obtained from the centralized optimization model that considers P2P energy sharing (COP). The lower grey line indicates the lower quartile of energy costs when using a random strategy, while the upper grey line denotes the upper quartile of energy costs for the random strategy. The violin plot illustrates the distribution of the 4-EH system's energy costs across different P2P energy sharing strategies.

Table 10

The deviation between test cases from the global optimum.

Test cases	S-C	S-C-N	S-C-A	S-N-A	W-C	W-C-N	W-C-A	W-N-A
Deviation from the reference case (COP)	0.45 %	2.29 %	0.08 %	1.40 %	2.87 %	4.66 %	0.21 %	1.40 %

scenarios that encompass a range of decision-making and P2P sharing strategies. These scenarios provide system operators with various operational options, enabling them to explore different pathways. We then conducted "what-if" simulations to assess the potential outcomes associated with these scenarios. During this simulation process, we employed a multi-agent reinforcement learning framework, MADDPG, to predict and model the stochastic energy behaviors of multiple EHs, which may be owned by different stakeholders, thereby adding complexity to the decision-making process.

Compared to conventional deterministic approaches involving a central operator or model-based assumptions, the MADDPG-based energy behavior model yields superior performance in handling the nonlinearity and high-dimensional nature of the decision-making process for interconnected EHs. Additionally, it captures the stochastic characteristics of autonomous EHs with privacy concerns through a multi-agent structure. Finally, we utilized a comprehensive multi-dimensional evaluation index system to assess the economic and technical performance of different P2P energy sharing strategies.

This EXDT enables the prediction and modeling of interconnected EHs' behaviors under diverse scenarios, providing insights into the MEH system's performance from both economic and technical perspectives. Ultimately, this approach empowers energy system operators to make more informed decisions regarding MEHs.

An illustrative 4-EH multi-energy system in China was employed as a case study to evaluate the economic and technical advantages of P2P sharing. The system was also used to conduct a comparative analysis of various scenarios, providing valuable insights for energy system operators. The key findings from the simulation results are as follows:

Financial and technical benefits of P2P energy sharing for MEH systems:

1) Economic benefits: by engaging in local P2P energy sharing, interconnected EHs can benefit from reasonable local energy prices. For instance, EH4, which has a large PV generation, can earn profits by selling its surplus energy through direct P2P sharing. The financial benefits gained by individual EHs contribute to significant energy bill savings for the entire system. In this study, the 4-EH system achieves cost reductions of ¥72,538 (15.2 %) on a typical summer day and ¥33,647 (9.9 %) on a typical winter day.

2) Technical benefits: from a technical perspective, combining P2P energy sharing with energy storage and demand response measures significantly enhances self-sufficiency and peak load management. During peak hours, driven by higher energy tariffs, EHs tend to share their energy and manage their flexibility through energy storage and demand response measures. Consequently, only a small proportion of demand (20 % for EH3 and 15 % for EH4) is met by the upstream grid, allowing locally produced generation to be widely distributed among EHs. This approach not only enhances local energy utilization but also contributes to CO2 emission reductions and decreased transmission losses over the long term.

In terms of future scenarios and their impact on system performance:

1) Decision-making strategies: cooperative operational strategies show a slight advantage over non-cooperative approaches due to the access to perfect information and shared optimization goals. For example, on a typical summer day, non-cooperative cases using k-DA and A-DA P2P strategies experience average economic performance degradations of 1.85 % and 1.32 %, respectively. Given that EHs are often managed by different stakeholders with privacy concerns,

accessing complete information can be challenging, making non-cooperative strategies more practical in real-world scenarios.

2) P2P energy sharing strategies: among P2P sharing strategies, A-DA and k-DA stand out, with A-DA being particularly effective in delivering better economic and technical performance. These strategies improve EHs' willingness to engage in the P2P local energy market and ensure a fair allocation of profits from P2P energy sharing. Therefore, they are recommended for future P2P energy markets. Conversely, more complex sharing mechanisms like M-DA and VV-DA contribute minimally to P2P energy sharing and produce less satisfactory system performance.

These findings highlight the significant benefits of P2P local energy sharing and provide valuable insights into choosing effective operational and sharing strategies for energy system operators in future energy markets.

In this work, we validated the effectiveness and reasonableness of the MADDPG-based approach for modeling energy behavior. When compared to a centralized optimization model, the proposed method closely approximates global optimal solutions, with slight deviations observed, such as 0.08 % for S-C-A cases and 0.45 % for S-C cases.

Apart from energy behavior modeling, MADDPG can also be applied to real-time operation optimization of MEHs and will be explored in the upcoming work. As discussed in Section 8, the benefits of MADDPG arise from the decoupled nature of the training and decision-making processes for the actor and critic networks. The well-trained actor network can independently perform decision-making by outputting actions without relying on concurrent training. Training is carried out periodically using historical operation data, ensuring that decision-making computations are not disrupted. In this study, the average decision-making process took an average of 6 s, demonstrating the method's ability to meet real-time management requirements while significantly reducing computational time compared to traditional physical model-based approaches.

Additionally, our future research agenda will explore advanced technologies such as electric vehicles (EVs) and power-to-gas (P2G) systems, integrate additional P2P sharing strategies, and develop robust operational strategies capable of addressing uncertainties in renewable generation and load forecasting. To simplify the model, this study utilizes only the virtual part of the digital twin, with the connection between the virtual and physical systems omitted. In future work, a complete DT will be considered, incorporating both virtual and physical components to fully leverage the capabilities of the cyber-physical system.

CRediT authorship contribution statement

Shiyao Li: Writing – original draft, Validation, Methodology, Investigation. **Yue Zhou:** Writing – review & editing, Supervision, Methodology. **Jianzhong Wu:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Yiqun Pan:** Supervision, Project administration, Investigation. **Zhizhong Huang:** Supervision, Investigation. **Nan Zhou:** Writing – review & editing.

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

Information about the data used in the case study of this paper including how to access them, can be found in the Cardiff University data catalogue at 10.17035/cardiff.27900204.

Acknowledgement

Appendix A. Appendix

(1) Parameters in Table 3

$value_{test}$	The total income of participants under the test P2P energy sharing strategy
$value_{ref}$	The total income of participants without P2P energy sharing
$value_{max}$	The potential maximum income
$cost_{P2P}$	The total energy cost under test P2P energy sharing strategy
$cost_{min}$	The minimum energy cost of all participants
$cost_{noP2P}$	The total energy cost of all participants without P2P energy sharing
$income_i$	The income of participants i under the test P2P energy sharing strategy
$income_i^{Shapley}$	The Shapley value showing the reasonable payoff allocation among participants of the local energy market
T	The time horizon
Ψ	The set of participants in the local P2P energy market
L	The set of devices that consume energy
G	The set of devices that produce energy
l	The energy consumed
g	The energy generated
P	The set of peak hours

(2) Nomenclature

Nomenclature

Abbreviation

ADMM	Alternating direction multiplier method	k-DA	k-Double Auction
CPS	Cyber-physical system	U-k-DA	Uniform k- Double Auction
COP	Centralized optimization model considering P2P energy sharing	A-k-DA	Average k-Double Auction
CO	Centralized optimization model without P2P energy sharing	VV-DA	Vickrey Variant Double Auction
DT	Digital twin	M-DA	McAfee's Double Auction
EXDT	Extensive digital twin	VI	Value tapping index
EH	Energy hub	PI	Participation willingness index
ESS	Energy storage systems	EI	Equality index
GHG	Greenhouse gas	EBI	Energy balance index
MINLP	Mixed-integer nonlinear programming	PFI	Power flatness index
MADDPG	Deep deterministic policy gradient	SSI	Self-sufficiency index
MES	Multi-energy system	PV	Photovoltaic panel
MEH	Multi-energy hub system	WT	Wind turbine
MARL	Multi-agent reinforcement learning	CHP	Combined heat and power
MG	Markov Game	AC	Absorption chiller
P2P	Peer-to-peer	EC	Electrical chiller
PLR	Partial load rate	HP	Heat pump
PSO	Particle swarm optimization	GB	Gas boiler
RES	Renewable energy source	ES	Energy storage

Sets

D	Virtual database	a_{P2P}^i	P2P electricity sharing actions of EH i
N	The set of agents	a_{EH}^i	The multi-energy optimization actions of EH i
S	State space	μ_{φ^i}	The set of stochastic policies of EH i
A	Action space	I_P^i	The public information obtained by EH i
x	The set of all EH' observation	$I_{p,}^i$	The private information obtained by EH i

Variables and parameters

Multi-agent reinforcement learning

T_{max}	Time horizon	φ, θ	Parameters of the deep neural network of the actor and critic network respectively
V_{set}	Parameters set by end users	$L(\varphi^i)$	Loss function when taking policy μ_{φ^i}
a	Action	$\widehat{\mu}_{\vartheta_j^i}$	The estimation of the true policy of agent j
r	Reward	τ	Soft update coefficient for target networks
s'	The next step state	ϑ_j^i	The approximation parameter of agent j's policy estimated by agent i
o	Observation	H	The entropy of the policy distribution
γ	Discount factor	M	A mini batch
ω	Gaussian noise	η_a	The learning rate of the actor network
$Cost^i$	Energy cost of EH i	$Q_{\mu_{\varphi^i}}$	The action-value indicates the cumulative reward obtained by taking policy μ_{φ^i}
$Penalty^j$	Penalty term for end-users' thermal comfort		

(continued on next page)

(continued)

Nomenclature			
Case study			
$p_{grid,t}$	The purchase price of the upstream grid at time t, Yuan/kW·h or Yuan/ Nm^3	$p_{P2P,t}^i$	The price of P2P energy sharing of EH i at time t, Yuan/kW·h
P_{feed_in}	The feed-in price of the upstream grid, Yuan/kW·h	$q_{P2P,t}^i$	The quantity of P2P energy sharing of EH i at time t, Yuan/kW·h
$Q*$	The critical intersection point where the aggregate demand and supply meet	$G_{CHP,t}^i$	The generation of CHP of EH i at time t, kW
$income_i^{Shapley}$	The Shapley value is a method used to distribute total gains among participants in a cooperative game	$G_{HP,t}^i$	The generation of HP of EH i at time t, kW
$\varphi_i(v)$	in a given coalition game (v, $\boldsymbol{N}$), the payoff allocated to player i	$G_{EC,t}^i$	The generation of EC of EH i at time t, kW
$T_{a,t}$	Ambient environment temperature at time t,°C	$G_{GB,t}^i$	The generation of GB of EH i at time t, kW
$T_{in,t}^i$	Aggregated indoor air temperature at time t,°C	$G_{AC,t}^i$	The generation of AC of EH i at time t, kW
$Solar_t$	Solar radiation	C_t^i	The charge/discharge rates of EH i at time t, kW
$hour_t$	Time	$\overline{T}_{tolerance}^i$	The upper limit of the indoor temperature tolerance,°C
PV_t^i	PV generation at time t of EH i, kW·h	$\underline{T}_{tolerance}^i$	The lower limit of the indoor temperature tolerance,°C
WT_t^i	WT generation at time t of EH i, kW·h	$thermal\_factor$	Penalty factor for violating thermal comfort constraints
$L_{e,t}^i$	Electrical load of at time t of EH i, kW	$T_{exceed,t}^i$	The deviation of the indoor air temperature from the upper limit and lower limit of the temperature tolerance, °C
S_t^i	The state of storage device at time t of EH i, kW·h		
Superscript/ subscript			
i	Agent series number	'	The next step
$-i$	Other components except for component i	t	Time step
e	Electricity	g	Natural gas
h	Heating	c	Cooling

References

[1] Bouckaert S, Pales AF, Mcglade C, et al. Net zero by 2050: A roadmap for the global energy sector. 2021.

[2] Alabi TM, Agbajor FD, Yang Z, et al. Strategic potential of multi-energy system towards carbon neutrality: a forward-looking overview. Energy Built Environ 2023; 4(6):689–708.

[3] Cheng Y, Zhang N, Lu Z, et al. Planning multiple energy systems toward low-carbon society: a decentralized approach. IEEE Trans Smart Grid 2018;10(5):4859–69.

[4] Jasinski M, Najafi A, Homaee O, et al. Operation and planning of energy hubs under uncertainty—a review of mathematical optimization approaches. IEEE Access 2023;11:7208–28.

[5] Mohammadi M, Noorollahi Y, Mohammadi-Ivatloo B, et al. Energy hub: from a model to a concept–a review. Renew Sustain Energy Rev 2017;80:1512–27.

[6] Aljabery AAM, Mehrjerdi H, Mahdavi S, et al. Multi carrier energy systems and energy hubs: comprehensive review, survey and recommendations. Int J Hydrog Energy 2021;46(46):23795–814.

[7] Wang L, Hou C, Ye B, et al. Optimal operation analysis of integrated community energy system considering the uncertainty of demand response. IEEE Trans Power Syst 2021;36(4):3681–91.

[8] Raheli E, Wu Q, Zhang M, et al. Optimal coordinated operation of integrated natural gas and electric power systems: a review of modeling and solution methods. Renew Sustain Energy Rev 2021;145:111134.

[9] Zhang M, Wu Q, Wen J, et al. Optimal operation of integrated electricity and heat system: a review of modeling and solution methods. Renew Sustain Energy Rev 2021;135:110098.

[10] Xi Y, Fang J, Chen Z, et al. Optimal coordination of flexible resources in the gas-heat-electricity integrated energy system. Energy 2021;223:119729.

[11] Abedinia O, Shorki A, Nurmanova V, et al. Synergizing efficient optimal energy hub Design for Multiple Smart Energy System Players and Electric Vehicles. IEEE Access 2023;11:116650–64.

[12] Jogunola O, Ajagun AS, Tushar W, et al. Peer-to-peer local energy market: opportunities, barriers, security and implementation options. IEEE. Access 2024; 12:37873–90.

[13] Kim HJ, Chung YS, Kim SJ, et al. Pricing mechanisms for peer-to-peer energy trading: towards an integrated understanding of energy and network service pricing mechanisms. Renew Sust Energy Rev 2023;183:113435.

[14] Gan W, Yan M, Yao W, et al. Peer to peer transactive energy for multiple energy hub with the penetration of high-level renewable energy. Appl Energy 2021;295: 117027.

[15] Junhui LI, Pan Y, Mu G, et al. A hierarchical demand assessment methodology of peaking resources in multi-areas interconnected systems with a high percentage of renewables. Appl Energy 2024;367:123371.

[16] Zhang C, Wu J, Zhou Y, et al. Peer-to-peer energy trading in a microgrid. Appl Energy 2018;220:1–12.

[17] Li S, Pan Y, Xu P, et al. A decentralized peer-to-peer control scheme for heating and cooling trading in distributed energy systems. J Clean Prod 2021;285:124817.

[18] Lyu C, Jia Y, Xu Z. Fully decentralized peer-to-peer energy sharing framework for smart buildings with local battery system and aggregated electric vehicles. Appl Energy 2021;299:117243.

[19] Gökçek T, Turan MT, Ateş Y. A new decentralized multi-agent system for peer-to-peer energy market considering variable prosumer penetration with privacy protection. Sustain Energy, Grids Networks 2024;38:101328.

[20] Aznavi S, Fajri P, Shadmand MB, et al. Peer-to-peer operation strategy of PV equipped office buildings and charging stations considering electric vehicle energy pricing. IEEE Trans Ind Appl 2020;56(5):5848–57.

[21] Ding Y, Sun X, Ruan J, et al. Customized decentralized autonomous organization based optimal energy management for smart buildings. Appl Energy 2024;376: 124223.

[22] Mohamed MA, Hajjiah A, Alnowibet KA, et al. A secured advanced management architecture in peer-to-peer energy trading for multi-microgrid in the stochastic environment. IEEE access 2021;9:92083–100.

[23] Yao Y, Yang J, Chen S, et al. Design of distributed power trading mechanism based on P2P contract. In: Proceedings of the 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC). IEEE; 2020. p. 1–5.

[24] Zhao Z, Feng C, Lu AL. Auction design through multi-agent learning in peer-to-peer energy trading. arXiv preprint. 2021. p. 10714.

[25] Khorasany M, Paudel A, Razzaghi R, et al. A new method for peer matching and negotiation of prosumers in peer-to-peer energy markets. IEEE Trans Smart Grid 2020;12(3):2472–83.

[26] Siqin Z, Niu D, Li M, et al. Distributionally robust dispatching of multi-community integrated energy system considering energy sharing and profit allocation. Appl Energy 2022;321:119202.

[27] Tushar W, Saha TK, Yuen C, et al. A coalition formation game framework for peer-to-peer energy trading. Appl Energy 2020;261:114436.

[28] Azim MI, Alam MR, Tushar W, et al. A cooperative P2P trading framework: developed and validated through hardware-in-loop. IEEE Trans Smart Grid 2023; 14(4):2999–3015.

[29] Liu Y, Zuo K, Liu XA, et al. Dynamic pricing for decentralized energy trading in micro-grids. Appl Energy 2018;228:689–99.

[30] Görgülü H, Topçuoğlu Y, Yaldız A, et al. Peer-to-peer energy trading among smart homes considering responsive demand and interactive visual interface for monitoring. Sustain Energy, Grids Networks 2022;29:100584.

[31] Hashemipour N, Crespo Del Granado P, Aghaei J. Dynamic allocation of peer-to-peer clusters in virtual local electricity markets: a marketplace for EV flexibility. Energy 2021;236:121428.

[32] Qiu D, Ye Y, Papadaskalopoulos D, et al. Scalable coordinated management of peer-to-peer energy trading: a multi-cluster deep reinforcement learning approach. Appl Energy 2021;292:116940.

[33] Wang J, Zhong H, Wu C, et al. Incentivizing distributed energy resource aggregation in energy and capacity markets: an energy sharing scheme and mechanism design. Appl Energy 2019;252:113471.

[34] Ullah MH, Park JD. Peer-to-peer energy trading in Transactive markets considering physical network constraints. IEEE Trans Smart Grid 2021;12(4):3390–403.

[35] Li W, Qian T, Zhao W, et al. Decentralized optimization for integrated electricity–heat systems with data center based energy hub considering communication packet loss. Appl Energy 2023;350:121586.

[36] Moseley P. EU support for innovation and market uptake in smart buildings under the horizon 2020 framework programme. Buildings 2017;7(4):105.

[37] Tiwari S, Singh JG. Optimal energy management of multi-carrier networked energy hubs considering efficient integration of demand response and electrical vehicles: a cooperative energy management framework. J Energy Storage 2022;51:104479.

[38] Qamar N, Malik TN, Qamar F, et al. Chapter 15 - energy hub: Modeling, control, and optimization. In: Azar AT, Kamal NA, editors. Renewable energy systems. Academic Press; 2021. p. 339–62.

[39] Huang Y, Zhang W, Yang K, et al. An optimal scheduling method for multi-energy hub systems using game theory. Energies 2019;12(12):2270.

[40] Ha T, Xue Y, Lin K, et al. Optimal operation of energy hub based micro-energy network with integration of renewables and energy storages. J Mod Power Syst Clean Energy 2020;10(1):100–8.

[41] Nasiri N, Yazdankhah AS, Mirzaei MA, et al. A bi-level market-clearing for coordinated regional-local multi-carrier systems in presence of energy storage technologies. Sustain Cities Soc 2020;63:102439.

[42] Chen T, Bu S, Liu X, et al. Peer-to-peer energy trading and energy conversion in interconnected multi-energy microgrids using multi-agent deep reinforcement learning. IEEE Trans Smart Grid 2022;13(1):715–27.

[43] Thomas L, Zhou Y, Long C, et al. A general form of smart contract for decentralized energy systems management. Nat Energy 2019;4(2):140–9.

[44] Azim MI, Lankeshwara G, Tushar W, et al. Dynamic operating envelope-enabled P2P trading to maximize financial returns of prosumers. IEEE Trans Smart Grid 2024;15(2):1978–90.

[45] Xie S, Hu Z, Wang J, et al. The optimal planning of smart multi-energy systems incorporating transportation, natural gas and active distribution networks. Appl Energy 2020;269:115006.

[46] Yang W, Guo J, Vartosh A. Optimal economic-emission planning of multi-energy systems integrated electric vehicles with modified group search optimization. Appl Energy 2022;311:118634.

[47] Zhang B, Hu W, Cao D, et al. Novel data-driven decentralized coordination model for electric vehicle aggregator and energy hub entities in multi-energy system using an improved multi-agent DRL approach. Appl Energy 2023;339:120902.

[48] Glaessgen E, Stargel D. The digital twin paradigm for future NASA and US Air Force vehicles 2012. 1818.

[49] Boschert S, Rosen R. Digital twin—The simulation aspect. In: Mechatronic futures: Challenges and solutions for mechatronic systems and their designers; 2016. p. 59–74.

[50] Vatankhah Barenji A, Liu X, Guo H, et al. A digital twin-driven approach towards smart manufacturing: reduced energy consumption for a robotic cell. Int J Comput Integr Manuf 2021;34(7–8):844–59.

[51] White G, Zink A, Codecá L, et al. A digital twin smart city for citizen feedback. Cities 2021;110:103064.

[52] Saad A, Faddel S, Mohammed O. IoT-based digital twin for energy cyber-physical systems: design and implementation. Energies 2020;13(18):4762.

[53] Sleiti AK, Kapat JS, Vesely L. Digital twin in energy industry: proposed robust digital twin for power plant and other complex capital-intensive large engineering systems. Energy Rep 2022;8:3704–26.

[54] Gitelman LD, Kozhevnikov MV, Kaplin DD. Asset management in grid companies using integrated diagnostic devices. 2019.

[55] Merkle L, Pöthig M, Schmid F. Estimate e-golf battery state using diagnostic data and a digital twin. Batteries 2021;7(1):15.

[56] Dong L, Lin H, Qiao J, et al. A coordinated active and reactive power optimization approach for multi-microgrids connected to distribution networks with multi-actor-attention-critic deep reinforcement learning. Appl Energy 2024;373:123870.

[57] Silver D, Huang A, Maddison CJ, et al. Mastering the game of go with deep neural networks and tree search. Nature 2016;529(7587):484–9.

[58] Silver D, Hubert T, Schrittwieser J, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. Science 2018;362 (6419):1140–4.

[59] Vinyals O, Babuschkin I, Czarnecki WM, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature 2019;575(7782):350–4.

[60] Berner C, Brockman G, Chan B, et al. Dota 2 with large scale deep reinforcement learning. arXiv preprint arXiv:191206680. 2019.

[61] Brown N, Sandholm T. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. Science 2018;359(6374):418–24.

[62] Gu S, Lillicrap T, Sutskever I, et al. Continuous deep q-learning with model-based acceleration. 2016. PMLR. 2829-2838.

[63] Shalev-Shwartz S, Shammah S, Shashua A. Safe, multi-agent, reinforcement learning for autonomous driving. arXiv preprint arXiv:161003295. 2016.

[64] Lowe R, Wu YI, Tamar A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments. Adv Neural Inf Proces Syst 2017;30.

[65] Littman ML. Markov games as a framework for multi-agent reinforcement learning. In: Machine learning proceedings 1994. Elsevier; 1994. p. 157–63.

[66] Zhang G, Hu W, Cao D, et al. 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 2022;255:115340.

[67] Lin J, Pipattanasomporn M, Rahman S. Comparative analysis of auction mechanisms and bidding strategies for P2P solar transactive energy markets. Appl Energy 2019;255:113687.

[68] Zhou Y, Wu J, Long C. Evaluation of peer-to-peer energy sharing mechanisms based on a multiagent simulation framework. Appl Energy 2018;222:993–1022.

[69] Liu T, Zhang D, Wang S, et al. Standardized modelling and economic optimization of multi-carrier energy systems considering energy storage and demand response. Energy Convers Manag 2019;182:126–42.

[70] Guo Z, Coffman AR, Munk J, et al. Aggregation and data driven identification of building thermal dynamic model and unmeasured disturbance. Energ Build 2021; 231:110500.

[71] Wu Y, Liu Z, Liu J, et al. Optimal battery capacity of grid-connected PV-battery systems considering battery degradation. Renew Energy 2022;181:10–23.