Indirect Multi-Energy Transactions of Energy Internet With Deep Reinforcement Learning Approach

Lingxiao Yang, *Student Member, IEEE*, Qiuye Sun, *Senior Member, IEEE*, Ning Zhang, *Student Member, IEEE*, and Yushuai Li, *Member, IEEE*

***Abstract*—With the new feature of multi-energy coupling and the advancement of the energy market, Energy Internet (EI) has higher requirements for the efficiency and applicability of integrated energy response. This paper proposes an indirect multi-energy transaction (IMET) to promote multi-energy collaborative optimization in local energy market (LEM) and improve energy utilization through personalized responses from We-Energies (WEs). Firstly, an indirect customer-to-customer multi-energy transaction is modeled for local multi-energy coupling market which can satisfy privacy, preference and autonomy of users. The efficiency of energy matching can be promoted through the participation of conversion devices. In addition, multi-time scale hybrid trading mechanism is constructed with the consideration of the transmission speed of different energy sources. Meanwhile, energy transaction process is built as a Markov decision process (MDP) with deep reinforcement learning algorithm so that the system modeling error can be successfully avoided. Furthermore, a distributed training structure is utilized to obtain more experience for a wider range of scenarios. The results of numerical simulations demonstrate the performance of the proposed method.**

***Index Terms*—Indirect multi-energy transactions, market mechanism, deep reinforcement learning, energy internet.**

NOMENCLATURE

p	Bidding information with energy price
e	Bidding information with energy quantity
s	Seller in the market
b	Buyer in the market
e	Type of energy: electricity
g	Type of energy: natural gas
i	Index for electricity seller
j	Index for electricity buyer
k	Index for gas seller
l	Index for gas buyer
t	Index for time during operation
p_-	Price of energy purchased by utility company
p_+	Energy service price of utility company
Δt	Time interval of transaction
n	Number of transactions
W	Time interval of participator to stay in market
w	Record of staying status in market
E	Type of energy includes electricity and gas
L	Energy outputs of conversion devices
P	Energy inputs of conversion devices
$\tilde{C}$	Conversion matrix of corresponding devices
$\tilde{p}$	Bidding energy price with conversion
$\tilde{e}$	Amount of trade available for WE
x	Total trading volume received by buyer
y	Total trading volume sent by seller
dir	Energy that are traded directly
con	Energy converted by the conversion device
c	Conversion cost of corresponding device
N	Set of participator in market
ψ	Penalty coefficient for no transaction condition
z	Energy amount which is not transacted
$\hat{x}$	Energy that seller actually provides to buyer
ξ	Device conversion rate
C	Capacity of conversion device
$\hat{l}$	Distance of transmission channel between two places
cd	Conversion device
ν	Transmission speed of corresponding energy
$\hat{L}_\pi$	Linepack in the pipe
Z	Compressibility factor of pipe
$\hat{p}$	Value of pressure
NTP	Normal temperature pressure
f	Value of gas flow
$\pi1, \pi2$	Start and end of pipe

Manuscript received 16 August 2021; revised 6 December 2021; accepted 7 January 2022. Date of publication 13 January 2022; date of current version 19 August 2022. This work was supported in part by the National Key R&D Program of China under Grant 2018YFA0702200 and in part by the National Natural Science Foundation of China under Grant 62073065. Paper no. TPWRS-01314-2021. *(Corresponding authors: Qiuye Sun; Yushuai Li.)*

Lingxiao Yang is with the School of Artificial Intelligence, Anhui University, Hefei 230601, China (e-mail: ylxiao66@163.com).

Qiuye Sun is with the State Key Laboratory of Synthetical Automation for Processes Industries and the School of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: sunqiuye@ise.neu.edu.cn).

Ning Zhang is with the School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China (e-mail: hbuzhangning@163.com).

Yushuai Li is with the School of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: liyushuai@ise.neu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2022.3142969.

Digital Object Identifier 10.1109/TPWRS.2022.3142969

I. INTRODUCTION

DRIVEN by the contradiction between energy utilization and environmental pollution, the researches on high proportion of renewable energy accommodation have been developed [1]-[4]. Meanwhile, the high quality development of economy promotes the research of Energy Internet (EI) to become a focus, which is coupled with smart distribution network,

0885-8950 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

natural gas, heating and cooling systems [5]-[7]. Unlike the conventional user in the energy system for pure energy load, energy producers and consumers have diverse functions in EI which derives distributed energy trading with user as the center [8]-[10]. The characteristics of multi-energy integration and the further development of the energy market make the EI put forward higher requirements for the efficiency and applicability of user response. Therefore, the study of energy transactions of EI becomes more significant to promote the coordinated and optimized utilization of multi-energy and the efficient accommodation of distributed energy resources.

It is notable that the energy transactions are a market-based distributed mechanism that coordinates multi agents based on bidirectional communication [11]. It aims to improve energy efficiency and economy by optimizing response through personalized responses from various types of users. Up to now, a lot of energy transaction strategies have been presented in electricity market, which put forward many explorations on the participants, system architecture and trading model of the transaction-driven energy system [12]-[15]. An energy sharing model among peer-to-peer photovoltaics prosumers was proposed in [8] so that greater economics are achieved than the independent operation of prosumers in the microgrid. The work in [14] focused on a localized event-driven market to facilitate indirect customer-to-customer energy trading, in which customers and prosumers have the willingness to directly participate in the retail electricity market. [15] investigated a parent-child matching energy transaction model to manage the peak load demand among multiple homes, which considers photovoltaics, battery storage and electric vehicles.

With the rapidly evolving of EI, on the one hand, the expansion of prosumers improves the ability of users to participate in energy transactions. On the other hand, users have various energy forms such as gas, cold and heat besides electricity [16]-[17]. It is a critical issue in investigating the energy transaction mechanism between energy hub and multi-energy system. For instance, reference [18] described a commercial building system with heating, ventilation and air-conditioning base on thermal and power model. An energy transaction market approach was proposed for peak shaving, load shifting, while having little impact on the building indoor temperature. Reference [19] focused on the complementarity of multiple energy resources in energy hubs to solve possible distribution network congestion. In [20], a transactive energy market was used as a platform to realize end-to-end bilateral transactions between local energy producers and consumers, so as to maximize the interests of all parties. The noncooperative game framework was investigated in [21]-[22] and the energy demands of each We-Energy (WE) were allocated to balance the profits of each WE. However, the existing researches on the retail side energy market of EI are still in the preliminary stage, and several critical problems need to explore: 1) With the development of local energy market (LEM) in EI and intelligent users, the user participation mode is more flexible, and the traditional transaction mode cannot fully satisfy the privacy, preference and autonomy of users. 2) Considering that natural gas and other energy sources have great inertia, and there are significant differences in transmission speed and storage capacity, LEM of EI has different clearance cycles for different types of energy. Once there is coupling between energies, the trading mechanism is more complex. 3) Considering the complexity of the system model, deep reinforcement learning with model-free feature is an effective way to solve the sequential decision problem [23]-[26]. However, some reinforcement learning approaches have limits on continuous action set in EI and so on.

Therefore, to address these challenges, this paper focuses on building an indirect multi-energy transaction (IMET) scheme for LEM to achieve the economic and energy efficiency of EI and ensure the supply and demand balance at the local level. Table I shows the difference between relevant works in the references and the method proposed in the paper. The major contribution of this paper can be expressed as follows:

1) An indirect multi-energy transaction model of LEM is established. With this effort, the privacy, preference and autonomy of users can be satisfied. The market efficiency and local-level energy balance are improved for the EI.
2) Based on the physical model of market, the paper proposes a multi-energy multi-time scale hybrid trading mechanism, and constructs an electricity-gas joint bilateral bidding energy market model. The inertia difference of different energy sources is fully taken into account.
3) Combined with the deep reinforcement learning algorithm, the Markov multi-energy interaction model is established with distributed structure, and the problem of continuous action in the model is solved, and finally the energy profit of the LEM in EI is maximized.

The rest of this paper is organized as follows. Section II describes the IMET framework of the EI. Section III presents transaction mechanism. The MDP and modified distributed proximal policy optimization (PPO) algorithm is presented in Section IV. In Section V, simulation results are presented to show the performance of the proposed algorithm whilst the conclusion is drawn in Section VI.

II. Indirect Multi-Energy Transaction Structure

A. Market Participant Model

With the development of information and communication technology, energy market is further opened in the EI, and intelligence user equipment are further improved. Through the energy transaction strategy, the optimal response of users can be better achieved. The EI consists of a number of We-Energies (WEs) that an interconnection network connecting them. WE, as a smart energy unit coupled by multiple energy sources, can operate independently or in connection with utility grid. The main body of WE will be the individual or company with distributed generation equipment [27]. To be specific, WE can be a buyer or a seller. Here, all the WEs can participate in the energy market as sellers or buyers based on the forecasting of energy deficit or surplus for each WE. The IMET will realize multi-energy matching between WEs through market operator to improve the overall energy efficiency of the EI.

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

TABLE I
THE COMPARISON OF RELEVANT WORKS PROPOSED IN THE REFERENCES AND THE METHOD PROPOSED IN THIS PAPER

Reference	Mechanism	Energy type	Features of paper
[15]	Market matching	Electricity	Exploring the emerging energy brokers in an event-driven market
[20]	System optimization	Multi-Energy	Solving distribution network congestions for energy hubs
[22]	Noncooperative game	Multi-Energy	Modeling a game between hubs and utility companies
[23]	Noncooperative game	Multi-Energy	Modeling a game between WEs in the market
Proposed	Market matching	Multi-Energy	Considering the impact of multi-time scale on trading

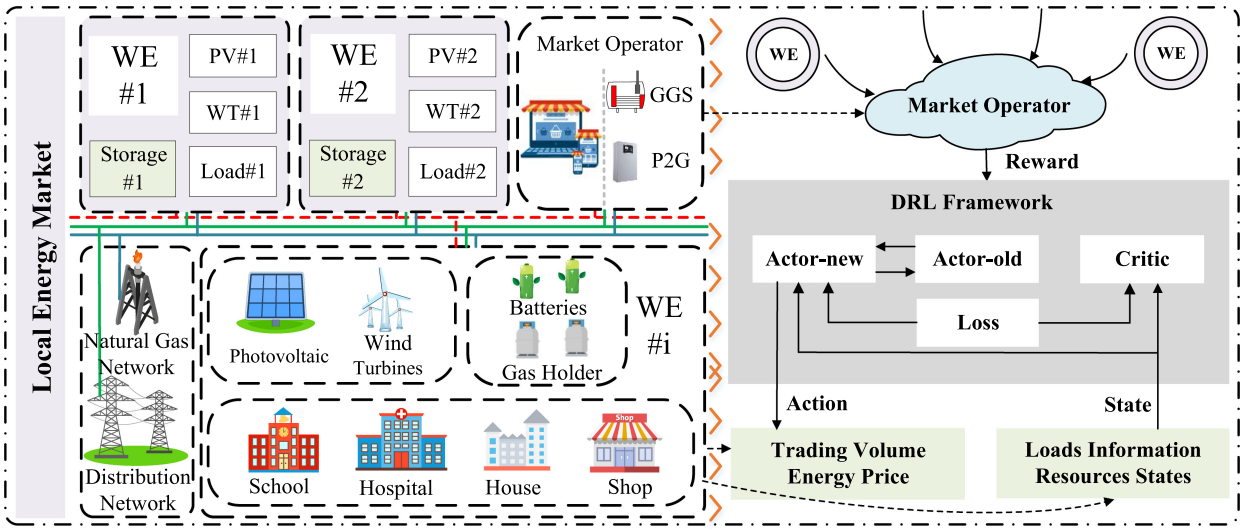

Fig. 1. Electricity-gas coupled transaction structure.

TABLE II
ALGORITHM FLOW

Algorithm 1 Modified distributed PPO Algorithm (chief)

Input number of workers W, number of workers needed to update global network parameters D
Step-1: **for** iteration $\in \{1, \ldots, N\}$ **do**
Step-2: **for** actor $\in \{1, \ldots, M\}$ **do**
Step-3: Wait until at least $W - D$ gradients wrt. θ are available
Step-4: Average gradients and update global θ and critic network parameters
Step-5: **end for**

B. Indirect Electricity-Gas Transaction Structure

The energy transactive framework of this paper is designed for a LEM which contains electricity-gas coupled transactions. The IMET structure is presented by Fig. 1.

The LEM has predefined boundaries and market operator with communal conversion devices is added to participate in operation of the local energy trading platform. The energy conversion devices include power to gas (P2G) device and gas generator set (GGS). Thus, electricity and gas can be converted to each other to participate in natural gas and electricity market transactions. The LEM is designed following a day-trading process. The indirect customer-to-customer multi-energy transaction means that trade involves not only WEs as buyers and sellers, but also market operator as intermediaries in such a LEM. WE formulates market bidding strategies based on equipment operating status. The market operator will only receive the price, amount of energy and the demand of time for each WE. The load and the equipment information are not needed. Then, market operator will summarize user bidding information and choose different actions according to the supply and demand situation at each time interval. Eventually, energy shortage or energy surplus will be traded with utility company to realize power balance. From this, the additional options for WEs who have the willingness to directly participate in the LEM occasionally are added. In addition, the market efficiency and local-level energy balance can be improved, while taking into account the privacy, preference and autonomy of WEs.

In the whole transaction process, the amount of energy that users participate in the market, the time of demand, and the bid for energy prices are knowable only to market operator. The user can only get information about whether the match is finally completed, the information of the rest of the market participants is not needed and completely unknowable. Meanwhile, because the energy supplied or demanded by different users does not have particularity in the energy market, part of the matching is realized by market operator through energy conversion device. Therefore, even after completing a market transaction, users who complete the transaction do not need to know the information of both parties, which ensures the privacy of users. In addition, users can choose and decide the bid for the energy prices, the amount of energy and the demand of time for participating in the market according to their own preferences. In order to ensure the autonomy of users, market operator can only match the information provided by users, and market operator has no

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

right to change the decisions made by users based on their own preferences. Therefore, the proposed energy transaction can operate optimally with privacy, preference and autonomy of users.

III. Electricity-Gas Bilateral Energy Transaction Mechanism

A. Bidding Strategy of WE

WE can be sellers or buyers in the LEM. The highest and lowest market bidding prices of sellers can be set as:

$$p_{-}^{E} \leq p^{s^E} \leq p_{+}^{E} - \varepsilon^{E}, \tag{1}$$

where the bidding prices of sellers are higher than the direct trading with utility company so that sellers will be more inclined to participate in market transactions. Meanwhile, the highest bidding prices of sellers should be lower than the energy service price of utility company to attract buyers to enter the market and trade with sellers. ε^E is set to avoid the bidding prices of sellers being equal to the service price and enhance pairing chance of sellers and buyers.

The bidding price of buyers should also be guaranteed that

$$p_{-}^{E} + \varepsilon^{E} \leq p^{b^E} \leq p_{+}^{E}, \tag{2}$$

where p^{b^E} is designed to be lower than the energy sevice price of unility company to encourage buyers to enter the energy market. And it also need to leave enough price margin ε^E to attract sellers.

We set the time interval of transaction is equal to 24/nh. Sellers and buyers will participate in market transaction according to their own situations.

Meanwhile, we use $w_{i,t}^{s^e} = W^e$, $w_{j,t}^{b^e} = W^e$, $w_{k,t}^{s^g} = W^g$, $w_{l,t}^{b^g} = W^g$ to record the staying status in energy market. As the time evolves: $t \to t+1$, the remaining time record will descrease after each time: $w_{i,t}^{s^e} \to w_{i,t}^{s^e} - 1$, $w_{j,t}^{b^e} \to w_{j,t}^{b^e} - 1$, $w_{k,t}^{s^g} \to w_{k,t}^{s^g} - 1$, $w_{l,t}^{b^g} \to w_{l,t}^{b^g} - 1$ until buyers and sellers withdraw from LEM.

The market operator, which has the communal energy conversion devices, can achieve energy conversion with a certain conversion efficiency which can be expressed as:

$$L = \tilde{C}P. \tag{3}$$

The marginal cost of devices that provides energy conversion is designed based on different device operating costs. The market operator will perform appropriate conversion and configuration based on the energy information provided by WE to achieve the optimal match of LEM.

B. Market Model With Electricity-Gas Bilateral Bidding

Local electricity-gas bilateral bidding market is an important trading platform for local sellers and buyers to participate. Market operator is responsible for facilitating local energy transactions between sellers and buyers in an indirect way. Sellers and buyers provide bidding information to market operator and market operator will choose different actions based on the current market conditions to achieve the optimal configuration of energy production and consumption. Due to the greater inertia of natural gas, the daily market of EI has different transmission time for different types of energy, which may lead to a transmission time delay between participants in the transaction. Therefore, the effect of transmission time on transactions is explored in the local clearing model (4) shown at the bottom of this page.

The objective function of LEM is designed as eq. (4) and it aims to maximize the profit of LEM. In (4), the first term of first line $p_j^{b^e}(t)x_j^{b^e}(t)$ means the total income of market operator given by electricity buyers. The sum of other two terms $p_i^{s^e}(t)y_i^{s^e,dir}(t)$ and $\tilde{p}_i^{s^e}(t)y_i^{s^e,con}(t)$ means the income for electricity sellers given by market operator. The purpose of splitting the income of electricity sellers into two parts is to show that the energy sold by electricity sellers can be converted in the market. Similarly, the first term of second line $p_l^{b^g}(t)x_l^{b^g}$ means the total income of market operator given by gas buyers. The sum of other two terms $p_k^{s^g}(t)y_k^{s^g,dir}(t)$ and $\tilde{p}_k^{s^g}(t)y_k^{s^g,con}(t)$ means the income for gas sellers given by market operator. The third line of the equation represents the cost of energy conversion through conversion device. ψz is a penalty term for failure to match energy sources and z can be expressed as:

$$\begin{aligned} z = &\sum_{i\in N_t^{s^e}} \left(\tilde{e}_i^{s^e}(t) - y_i^{s^e}(t)\right) + \sum_{j\in N_t^{b^e}} \left(\tilde{e}_j^{b^e}(t) - x_j^{b^e}(t)\right) \\ &+ \sum_{k\in N_t^{s^g}} \left(\tilde{e}_k^{s^g}(t) - y_k^{s^g}(t)\right) + \sum_{l\in N_t^{b^g}} \left(\tilde{e}_l^{b^g}(t) - x_l^{b^g}(t)\right). \end{aligned} \tag{5}$$

In the equation, $\tilde{e}_i^{s^e}(t) - y_i^{s^e}(t)$, $\tilde{e}_j^{b^e}(t) - x_j^{b^e}(t)$, $\tilde{e}_k^{s^g}(t) - y_k^{s^g}(t)$ and $(\tilde{e}_l^{b^g}(t) - x_l^{b^g}(t)$ denote the mismatch of electricity sellers, electricity buyers, gas sellers and gas buyers, respectively. Equality constraints and inequality constraints are shown in (6)-(21).

S.t.

$$\sum_{j\in N_t^{b^e}} x_j^{b^e}(t) - \sum_{i\in N_t^{s^e}} y_i^{s^e}(t) = 0, \tag{6}$$

$$F = \left\{ \begin{array}{l} \sum_{t=1}^{T} \sum_{j\in N_t^{b^e}} p_j^{b^e}(t)x_j^{b^e}(t) - \sum_{t=1}^{T} \sum_{i\in N_t^{s^e}} p_i^{s^e}(t)y_i^{s^e,dir}(t) - \sum_{t=1}^{T} \sum_{i\in N_t^{s^e}} \tilde{p}_i^{s^e}(t)y_i^{s^e,con}(t) \\ + \sum_{t=1}^{T} \sum_{l\in N_t^{b^g}} p_l^{b^g}(t)x_l^{b^g}(t) - \sum_{t=1}^{T} \sum_{k\in N_t^{s^g}} p_k^{s^g}(t)y_k^{s^g,dir}(t) - \sum_{t=1}^{T} \sum_{k\in N_t^{s^g}} \tilde{p}_k^{s^g}(t)y_k^{s^g,con}(t) \\ - \sum_{t=1}^{T} \sum_{i\in N_t^{s^e}} \alpha_i c^{P2G} y_i^{s^e,con}(t) - \sum_{t=1}^{T} \sum_{k\in N_t^{s^g}} \beta_k c^{GGS} y_k^{s^g,con}(t) \\ -\psi z \end{array} \right\}. \tag{4}$$

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

$$y_i^{s^e}(t) = y_i^{s^e,dir}(t) + y_i^{s^e,con}(t), \tag{7}$$

$$\sum_{l \in N_t^{b^g}} x_l^{b^g}(t) - \sum_{k \in N_t^{s^g}} y_k^{s^g}(t) = 0, \tag{8}$$

$$y_k^{s^g}(t) = y_k^{s^g,dir}(t) + y_k^{s^g,con}(t), \tag{9}$$

$$y_i^{s^e,dir}(t) = \hat{x}_i^{b^e,dir}(t), \tag{10}$$

$$y_k^{s^g,con}(t)\xi^{GGS} = \hat{x}_k^{b^e,con}\left(t + \Delta\tau_3\right), \tag{11}$$

$$\sum_i \hat{x}_i^{b^e,dir}(t) + \sum_k \hat{x}_k^{b^e,con}(t) = \sum_j x_j^{b^e}(t), \tag{12}$$

$$y_k^{s^g,dir} = \hat{x}_k^{b^g,dir}\left(t + \Delta\tau_1\right), \tag{13}$$

$$y_i^{s^e,con}\xi^{P2G} = \hat{x}_i^{b^g,con}\left(t + \Delta\tau_2\right), \tag{14}$$

$$\sum_k \hat{x}_k^{b^g,dir}(t) + \sum_i \hat{x}_i^{b^g,con}(t) = \sum_l x_l^{b^g}(t), \tag{15}$$

$$0 \le y_i^{s^e}(t) \le \tilde{e}_i^{s^e}(t), \forall i \in N_t^{s^e}, \tag{16}$$

$$0 \le x_j^{b^e}(t) \le \tilde{e}_j^{b^e}(t), \forall j \in N_t^{b^e}, \tag{17}$$

$$0 \le y_k^{s^g}(t) \le \tilde{e}_k^{s^g}(t), \forall k \in N_t^{s^g}, \tag{18}$$

$$0 \le x_l^{b^g}(t) \le \tilde{e}_l^{b^g}(t), \forall l \in N_t^{b^g}, \tag{19}$$

$$0 \le y_i^{s^e,con}(t)\xi^{GGS} \le C^{GGS}, \forall i \in N_t^{s^g}, \tag{20}$$

$$0 \le y_k^{s^g,con}(t)\xi^{P2G} \le C^{P2G}, \forall k \in N_t^{s^g}. \tag{21}$$

In the equations, $\tilde{p}^{s^e}$ and $\tilde{p}^{s^g}$ also need to satisfy constraints $p_-^E \le \tilde{p}^{s^E} \le p_+^E - \tilde{\varepsilon}^E$. $\tilde{\varepsilon}^E$ is intended to facilitate the transaction considering the energy conversion. α_i, β_k are binary variables, the value 1 indicates that the device is selected for energy conversion, and the value 0 indicates that the device does not participate in the conversion.

It should be remarked that the model designed in this paper is local-level. The realization of energy transaction between users in the region can effectively reduce the loss of long-distance energy transmission and ensure social welfare. Furthermore, since users can choose to trade with utility company or enter the market according to their own preferences. Although the objective function of the market operator is its profit, due to the autonomy of the participating market users, the market operator has no right to make any changes to the demand of users. So as long as the matching is achieved, the user's requirements will be satisfied, and there is no possibility for the market operator to reduce the user welfare for its own profit. Meanwhile, due to the penalty term z, the market operator will not deliberately refuse to match suitable users because of its own profit. Therefore, the above mechanism guarantees the interests and social welfare of users in the market first. On this basis, market operator can maximize its own profit through more reasonable matching strategies to maintain its own operation as well as equipment maintenance and other needs.

To be specific, in the current day-trading optimization, the transaction time t of electricity and natural gas buyers is the time they receive the energy. Transaction time t of electricity and natural gas sellers is the time they send the energy. Therefore, the delays caused by gas transmission need to be taken into account as $\Delta\tau_1$, $\Delta\tau_2$, $\Delta\tau_3$ in (11), (13)-(14). Electricity sellers can directly trade with electricity buyers during the transaction period t due to the faster transmission speed of electricity i.e. $t^{s^e} \to t^{b^e}$, $w_{i,t}^{s^e} \to w_{j,t}^{b^e}$. Because the transmission speed of natural gas is slower than that of electricity, its clearing cycle is also longer than that of electricity. Therefore, natural gas sellers will directly trade with natural gas buyers during the transaction period and the transmission time between the buyers and sellers of natural gas $\Delta\tau_1$ should be considered as:

$$t^{s^g} \to t^{b^g} + \Delta\tau_1, \quad w_{k,t^{s^g}} \to w_{l,t^{b^g}}, \tag{22}$$

$$\Delta\tau_1 = round\left(\frac{\hat{l}_{s^g,b^g}}{\nu_g \Delta t}\right), \tag{23}$$

where the transmission speed of natural gas can be calculated based on [28]. It should be noticed that different energy systems have different inertia differences since the natural gas system has the linepack, which impacts the energy transmission of natural gas network. The linepack in natural gas network can be expressed as [29]:

$$\hat{L}_\pi = \frac{ZV}{\hat{p}_{\mathrm{NTP}}} \frac{2}{3}\left(\hat{p}_{\pi 1} + \hat{p}_{\pi 2} - \frac{\hat{p}_{\pi 1} p_{\pi 2}}{\hat{p}_{\pi 1} + \hat{p}_{\pi 2}}\right). \tag{24}$$

The market is settled in different periods of time to complete the matching as shown above. So the energy transmission model of the market is essentially a steady state model of end-to-end energy transmission, and the purpose of the market is to ensure the accuracy of end-to-end energy transmission. The relationship among linepack, inflow gas and outflow gas can be expressed as [30]:

$$\left(f_{\pi 1} - f_{\pi 2}\right)\Delta t + \hat{L}_{\pi,t-1} = \hat{L}_{\pi,t}. \tag{25}$$

According to the above equation, the amount of energy transferred in the process of energy transaction in the market should be ensured i.e. $f_{\pi 1} = f_{\pi 2}$. Therefore, linepack in the pipeline should remain constant before and after transaction, and the constraint should be met during operation:

$$\begin{aligned} &Z_{t-1}V_{t-1}\left(\hat{p}_{\pi 1,t-1} + \hat{p}_{\pi 2,t-1} - \frac{\hat{p}_{\pi 1,t-1}\hat{p}_{\pi 2,t-1}}{\hat{p}_{\pi 1,t-1} + \hat{p}_{\pi 2,t-1}}\right) \\ &= Z_t V_t\left(\hat{p}_{\pi 1,t} + \hat{p}_{\pi 2,t} - \frac{\hat{p}_{\pi 1,t}\hat{p}_{\pi 2,t}}{\hat{p}_{\pi 1,t} + \hat{p}_{\pi 2,t}}\right). \end{aligned} \tag{26}$$

Due to the conversion devices existence in LEM, it is possible to convert between electricity and gas since the energy conversion devices are controlled by the market operator in LEM. However, due to the multi-time scale characteristics of electricity and natural gas, when the electricity and gas are traded, the limitation of natural gas transmission speed needs to be considered. If electricity sellers can enter the gas market for transactions through P2G, it should consider the transmission time of natural gas from the conversion device to the buyer of natural gas. The transmission time from the seller of electricity to the conversion device can be ignored due to the fast transmission and conversion speed of electricity.

$$t^{s^e} \to t^{b^g} + \Delta\tau_2, \quad w_{i,t^{s^e}} \to w_{l,t^{b^g}}, \tag{27}$$

$$\Delta\tau_2 = round\left(\frac{\hat{l}_{cd,b^g}}{\nu_g \Delta t}\right), \tag{28}$$

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

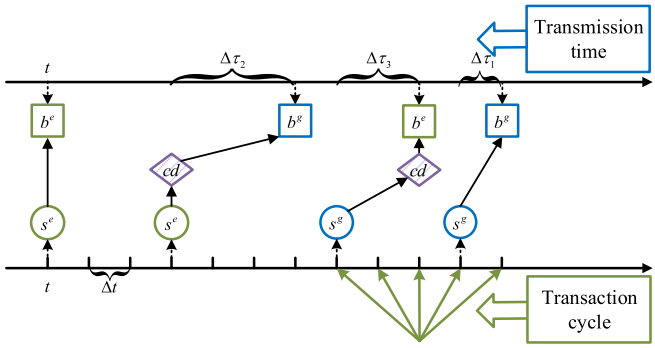

Fig. 2. Multi-time scale transaction process of electricity-gas market.

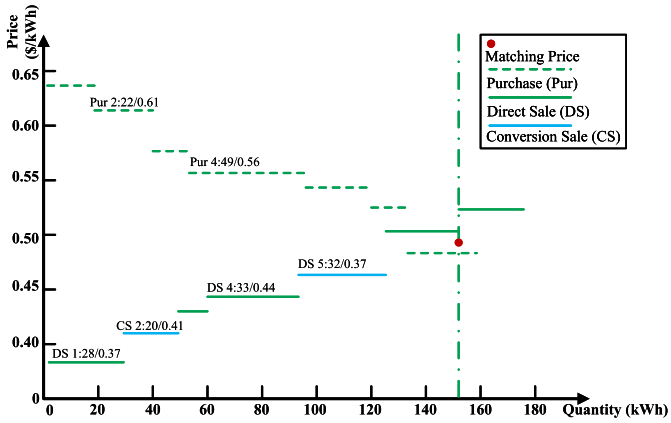

Fig. 3. The market clearance of electricity in LEM.

where $\hat{l}_{cd,b^g}$ is the distance of transmission channel between conversion devices and buyers of natural gas.

However, when gas sellers enter the electricity market for transactions through conversion, the transmission time of natural gas from the seller of natural gas to the conversion devices need to consider and the transmission of electricity is ignored. The corresponding clearing cycle is set to:

$$t^{s^g} \to t^{b^e} + \Delta\tau_3, \quad w_{k,t^{s^g}} \to w_{j,t^{b^e}}, \tag{29}$$

$$\Delta\tau_3 = round\left(\frac{\hat{l}_{s^g,cd}}{\nu_g \Delta t}\right), \tag{30}$$

where $\hat{l}_{s^g,cd}$ is the distance of transmission channel between conversion devices and sellers of natural gas.

The multi-time scale transaction process of electricity-gas market is depicted by Fig. 2. As shown in the figure, due to the rapid transmission of electricity, the trading between the electricity buyers and electricity sellers can ignore the time delay of energy transaction. In addition, transactions involving natural gas have time delays due to its slow transmission speed.

Remark 1: Compared with the traditional market transaction model, multi-time scale mechanism is designed into the transaction model which can be embodied in (10)-(15). Due to the transmission inertia of different energy sources, their trading cycles are biased. With the conversion devices in the LEM model, there is a coupling conversion between multi-energy, which will lead to the delay of the transaction mode. Under the design of this trading mechanism, energy interaction is more complex but more practical.

The actions of market operator can be designed to accept all the bids and offers and realize optimal price-quantity pairs matching through conversion, and the time record of chosen sellers/buyers is set to 0 to quit LEM. The rest of the customers that fail to be selected continue to stay in the market for future chances with their remaining time counters updated by $w \to w - 1$. The market operator will make further conversions and adjustments based on the information provided by buyers and sellers, and the market clearance of electricity is shown in Fig. 3. The market clearance of natural gas is similar to electricity.

Above all, the IMET in EI can be divided into two parts:

1) In the day trading, WE sends the transaction information to the market operator through the intelligent control system. Market operator makes the optimal matching scheme according to the energy supply and demand relationship and publishes it to the control system. The control system realizes local WE interaction according to the matching relation.
2) When WE does not complete the energy transaction within the time limit, it will trade with the utility company to fill the demand-generation gap so that a local trade balance can be achieved.

Remark 2: In this paper, the proposed LEM is suitable for some local energy scenarios. The utility company generally is the managers of large energy networks. For the utility company, the LEM is similar to a major customer. The utility company would trade energy with the LEM but would not normally be directly involved in the management of a LEM. Therefore, the market operator and utility company are usually not interchangeable.

Remark 3: There are several mechanisms in the market to ensure the equilibrium: Firstly, the penalty term in objective function can ensure that the market operator will not deliberately refuse to match the appropriate users. Secondly, the market allows users to enter and participate in matching repeatedly to increase the success rate of matching. Thirdly, the user can finally trade with the utility company to avoid the user's requirements can not be met. However, compared with matching in the market, trading with the utility company will lose part of the economic benefits. Under the mechanisms, different matching situations will only affect the profit of market operator, but not affect the normal operation of the market.

IV. Markov Decision Process and Modified Distributed Proximal Policy Optimization Algorithm

The energy market proposed above aims to achieve the day-trading optimal energy transaction based on objective function. Considering the uncertainty of the market, it is hard to obtain the accurate environment model for forecasting the user behaviors. Since the DRL is based on model-free environment and the updating rule of DRL adopts long-term accumulation

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

reward [31]-[32], the problem described above is transformed to MDP and PPO algorithm is improved to solve the problem through learning and selection with the dynamic environment. To the specific, the improved PPO can better handle the model with continuous action space.

A. Formulation of Markov Decision Process

In the proposed multi-energy interaction model, the trading information is entered into the market before, and the state at each time slot will be adjusted to be judged as the new state at the next time slot. Therefore, the previous state information is not needed, and the current time slot contains all the information needed for decision-making, having nothing to do with the previous period. Hence, the multi-energy interaction model can be formulated as a MDP problem.

The MDP modeling has four fundamental elements $(\hat{s}, a, \widehat{p}, r)$, where $\hat{s}$ stands for an array of limited environment states, a expresses an array of limited actions, $\widehat{p} : \hat{s} \times a \times \hat{s}' \rightarrow R$ is the state transition probability from state $\hat{s}$ to state $\hat{s}'$, r is the immediate reward function obtained from environment.

In LEM, the actions are set for the decision variables which denote the transaction price and volume of LEM participants, as well as the cost of energy conversion through conversion devices.

$$a_{LEM} = \left\{ p_j^{b^e}, x_j^{b^e}, p_i^{s^e}, y_i^{s^e,dir}, \tilde{p}_i^{s^e}, y_i^{s^e,con}, p_l^{b^g}, x_l^{b^g}, p_k^{s^g}, y_k^{s^g,dir}, \tilde{p}_k^{s^g}, y_k^{s^g,con}, \alpha_i, x_i^{s^e,con}, \beta_k, y_k^{s^g,con} \right\}. \tag{31}$$

The LEM states contains four parameter vector, including power mismatch, customers' bidding parameters, the timestamp and the states of the remaining time counters associated with current customers.

The reward r is designed to express the reward gets from system. The more reward will feedback under the condition of satisfying the constraints.

$$r_{LEM} = \begin{cases} 0, & if\ constraints\ are\ violated \\ F, & otherwise \end{cases} \tag{32}$$

where r_{LEM} stands for the profit obtained in LEM.

B. Modified Distributed Proximal Policy Optimazation Algorithm

As stated in Section IV A, considering that the states and actions of the system contain a large number of continuous variables, the PPO algorithm with actor-critic architecture is put into use and further extended. The PPO directly samples states, actions, and rewards [33]. The rule of iteration is to maximize the expectation of rewards $\overline{R}_\theta$.

$$\begin{aligned} \overline{R}_\theta &= \sum_\tau R(\tau)\widehat{p}_\theta(\tau) \\ &= \hat{E}_{\tau\sim\widehat{p}_\theta(\tau)}\left[R(\tau)\right] \end{aligned}, \tag{33}$$

where τ is policy trajectory, θ is a parameter for policy update, $\widehat{p}_\theta(\tau)$ is probability of trajectory occurrence. Then, calculate the gradient for the reward:

$$\begin{aligned} \nabla\overline{R}_\theta &= \sum_\tau R(\tau)\widehat{p}_\theta(\tau)\nabla\log\widehat{p}_\theta(\tau) \\ &= \hat{E}_{\tau\sim\widehat{p}_\theta(\tau)}\left[R(\tau)\nabla\log\widehat{p}_\theta(\tau)\right] \end{aligned}. \tag{34}$$

When sampled sufficiently, the expectation of the reward can be approximated to the average value of the reward for $\hat{N}$ rounds:

$$\begin{aligned} \nabla\overline{R}_\theta &\approx \frac{1}{\hat{N}}\sum_{n=1}^{\hat{N}} R(\tau^n)\nabla\log\widehat{p}_\theta(\tau^n) \\ &= \frac{1}{\hat{N}}\sum_{n=1}^{\hat{N}}\sum_{t=1}^{T_n} R(\tau^n)\nabla\log\widehat{p}_\theta(a_t^n|\hat{s}_t^n) \end{aligned}. \tag{35}$$

The advantage function $A^\theta(\hat{s}_t, a_t)$ is set as follow to evaluate performance with action a in state $\hat{s}$.

$$A^\theta(\hat{s}_t, a_t) = R^\theta(\hat{s}_t, a_t) - \hat{V}(\hat{s}_t). \tag{36}$$

In order to improve the speed and efficiency of training, off-policy has been adopted in place of on-policy. Therefore, the sample is used from $\hat{\pi}_{\theta'}$ to train θ.

$$J^{\theta'}(\theta) = \hat{E}_{(\hat{s}_t,a_t)\sim\hat{\pi}_{\theta'}}\left[\frac{\widehat{p}_\theta(a_t|\hat{s}_t)}{\widehat{p}_{\theta'}(a_t|\hat{s}_t)}A^{\theta'}(\hat{s}_t, a_t)\right]. \tag{37}$$

Because θ' is fixed, the sample data from θ' can be reused. So we can sample the data from θ' and use the data to train θ many times.

$$J(\theta) = \frac{1}{\hat{N}}\sum_{n=1}^{\hat{N}}\sum_{t=1}^{T_n}\frac{\widehat{p}_\theta(a_t|\hat{s}_t)}{\widehat{p}_{\theta'}(a_t|\hat{s}_t)}A^\theta(a_t|\hat{s}_t)\nabla\log\widehat{p}(a_t|\hat{s}_t,\theta). \tag{38}$$

If θ and θ' differ greatly and not enough samples are taken, their variances will differ greatly and its training effect will be affected. So a objective with clipped probability ratios is established, which forms a lower bound of the performance of the policy:

$$clip\left(\frac{\widehat{p}_\theta(a_t|\hat{s}_t)}{\widehat{p}_{\theta'}(a_t|\hat{s}_t)}, 1-\hat{\varepsilon}, 1+\hat{\varepsilon}\right), \tag{39}$$

where $\hat{\varepsilon}$ is a hyperparameter that is less than 1. In this way, the algorithm can be improved as:

$$J^{\theta^K}(\theta) \approx \sum_{(\hat{s}_t,a_t)}\min\left(\begin{array}{l}\frac{\widehat{p}_\theta(a_t|\hat{s}_t)}{\widehat{p}_{\theta^K}(a_t|\hat{s}_t)}A^{\theta^K}(\hat{s}_t,a_t), \\ clip\left(\frac{\widehat{p}_\theta(a_t|\hat{s}_t)}{\widehat{p}_{\theta^K}(a_t|\hat{s}_t)}, 1-\hat{\varepsilon}, 1+\hat{\varepsilon}\right)A^{\theta^K}(\hat{s}_t,a_t)\end{array}\right). \tag{40}$$

The PPO will dynamically select tasks which further affects the overall system performance optimization. However, Note that the transactions in LEM are complex and random. In order to guarantee the better performance in various random scenes, the distributed PPO is adopted in the paper with distributed feature which places actor-critic network in multiple workers for synchronous training. From this, the experience of multiple workers will be synchronously uploaded to the chief learner, and then each worker will get the latest running policy from the chief learner. In this way, each worker can receive feedback from the chief learning to apply better experience to its own scenario. The chief learning can also aggregate the training experience of all

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

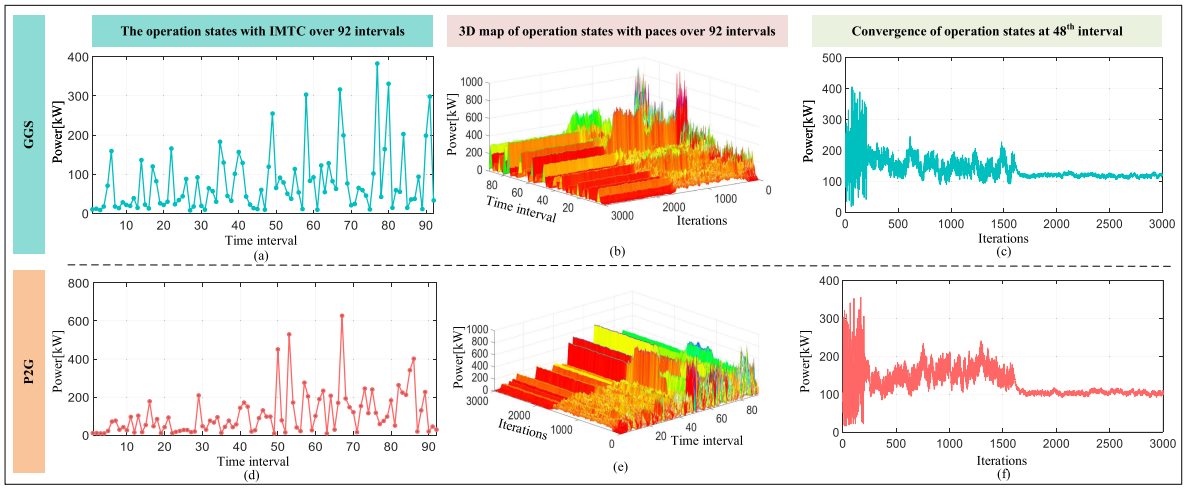

Fig. 4. Operating status of GGS and P2G over 92 time intervals. (a), (d) are the input power of GGS and P2G over 92 time intervals respectively. (b), (e) are the 3D map of operating status of GGS and P2G with paces over 92 time intervals respectively. (c), (f) are the operating status of GGS and P2G with paces at the 48th interval respectively.

workers, thus avoiding the continuity of single-threaded updates and making it possible to obtain good updates without the need for a memory bank like Deep Q-learning (DQN), or Deep Deterministic Policy Gradient (DDPG). The algorithm flow of the distributed PPO is shown in Algorithm 1 and Algorithm 2.

Remark 4: The distributed PPO algorithm contains two parts. Algorithm 1 corresponds to the chief and Algorithm 2 is for workers. Workers collect data in parallel in different environments and share a global PPO. Workers send the collected data to global PPO i.e. chief and then the chief updates the parameters based on the batch of datum collected from workers. With this effort, the efficiency of data collection is improved and the distributed PPO algorithm can gain wider and better experience for the better performance.

V. Case Studies

In this section, we focus on evaluating the performance regarding the proposed IMET model and verifying the effectiveness of our market learning algorithm. The timeline, one day, consists of 96 time intervals, each of which lasts for 15 min. We set 23th to 24th hour to stop trading, so the time intervals are 92. The number of customers is 384 which includes sellers and buyers. The conversion efficiency of GGS and P2G is 0.5 and 0.4 respectively.

Four cases are represented to show the effectiveness of the proposed IMET with the distributed PPO. Case 1 studies the performance of the IMET under normal circumstances. Case 2 investigates on the performance of the IMET without the consideration of multi-time scale. And case 3 focuses on the efficiency of the IMET without conversion devices. Case 4 shows the comparison of different methods for IMET to show the performance of the proposed method.

A. Performance Test Under Normal Circumstances

In the first case study, we run the IMET with the distributed PPO approach to complete the multi-energy transaction of WEs under normal circumstances. For the proposed distributed PPO above, the discount is 0.9, the adam update rate for actor and critic is 0.0001 and 0.0005 respectively. The minibatch size is 24 and the clipping parameters is 0.2.

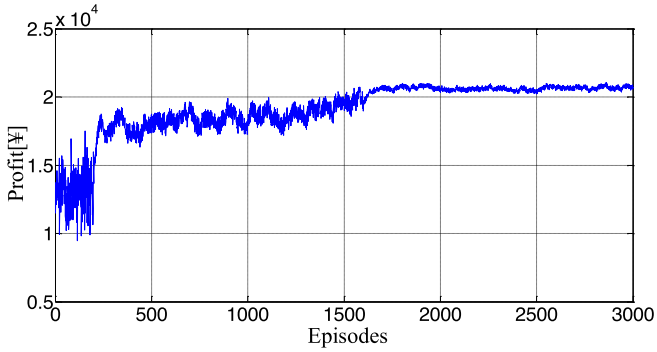

Fig. 5. The profit for each episode with the distributed PPO.

From Fig. 4, it can be clearly seen the operating status of conversion devices with IMET over 92 time intervals. Fig. 4(a), (d) describe the input power of GGS and P2G over 92 time intervals. Fig. 4(b), (e) depict the 3D map of operating status of conversion devices with paces over 92 time intervals. The depth of the color reflects the input power. The convergence of 92 time intervals operation results of devices are reached after around 1700 steps. Fig. 4(c), (f) are the operating status of GGS and P2G with paces at the 48th interval. The operating results in Fig. 4(a), (d) are depicted from the y0z-axis of Fig. 4(b), (e). Fig. 4(c), (f) demonstrate the convergence results through running more episodes that correspond to the x0z section of Fig. 4(b), (e) at the 48th interval. The profit for each episode is shown in Fig. 5. It can be seen that there is no obvious improvement for the episode around 1700 steps which means the improved PPO tends to be stable. And the optimal transaction strategy is obtained. Fig. 6 shows the performance of the conventional PPO algorithm in the proposed model. It is clear that both methods converge to similar profit value. However, the convergence speed of the proposed

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

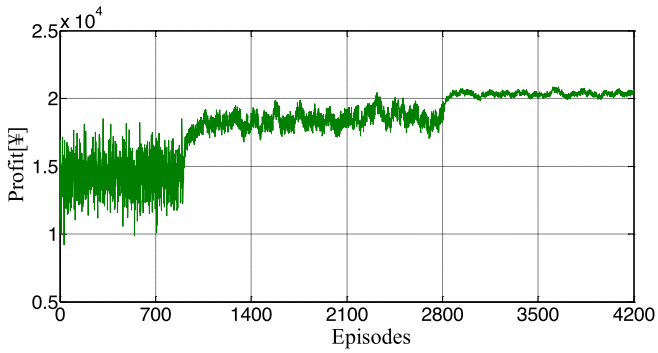

Fig. 6. The profit for each episode with the conventional PPO.

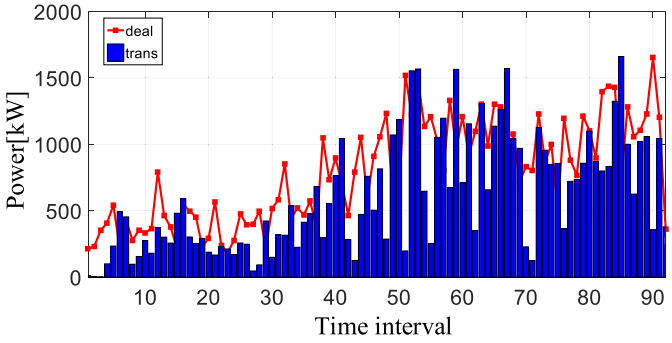

Fig. 7. The total volume of energy transactions and transmissions in the EI based on the proposed model.

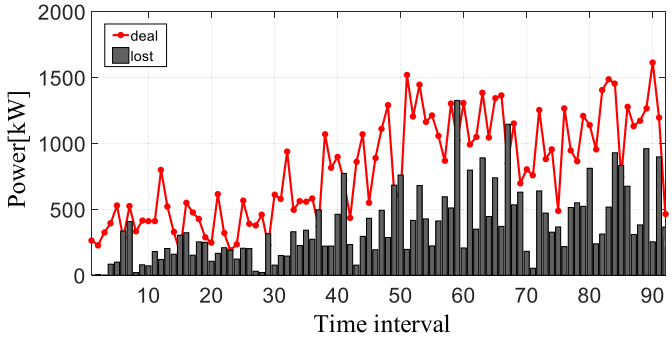

Fig. 8. The total volume of energy transactions and losses without considering multi-time scale.

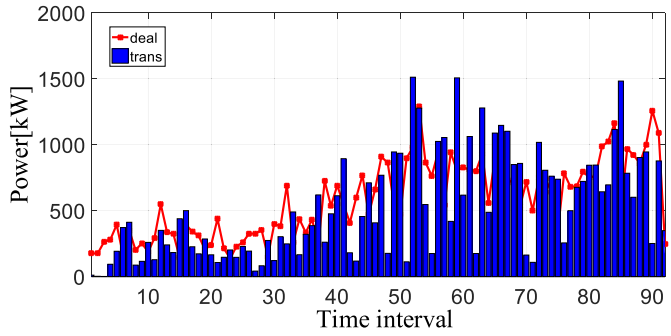

Fig. 9. The total volume of energy transactions and transmissions in the EI without conversion devices.

method is significantly faster than that of the conventional PPO algorithm. The results show that the multiple workers for synchronous training structure which is introduced in the proposed method can improve the efficiency of the algorithm.

The total volume of energy transactions and transmissions in the EI is shown in Fig. 7. Noted that the input power of P2G and GGS increase from the 50th to the 70th interval in Fig. 4(a), (d). Correspondingly, the trading volume in Fig. 7 is high between the 50th to the 70th interval. Conversion devices may have contributed to the increase in energy trading volume. Meanwhile, from Fig. 7, it can be seen that there is a deviation between the volume of energy trading and the volume of energy transmission, which is caused by the delay in the transmission speed of gas. The energy accumulation caused by the delay of energy transmission makes the volume of energy transmission greater than the volume of energy trading at some moments like the 52th and 53th interval.

B. Performance Test Without Considering Multi-Time Scale

In case 2, we aim at verifying the effectiveness of IMET with or without considering multi-time scale. The energy trading volume and losses are shown in Fig. 8 without considering multi-time scale. The loss means that the participants have entered into a transaction in the energy market, but due to the transmission delay of gas, the energy loss is caused by the inability to transmit in time within the trading time. It can be observed from Fig. 8 that the most of the time the energy loss is exist. However, compared with Fig. 7, it is interesting to find that the energy trading volume of IMET without considering multi-time scale is larger than the volume of IMET with considering multi-time scale sometimes like the 60th, 63th and 83th interval. Meanwhile, the total volume of transaction in case 2 is 74728 kW and the total volume of transaction in case 1 is 72658 kW. It is because energy is traded in case 2 regardless of multi-time scale. Some transactions with considering the multi-time scale could not match were also matched in case 2 without considering the multi-time scale which reduce some constraints in IMET, so the overall volume increases slightly. Therefore, the multi-time scale characteristic has a great impact on multi-energy trading and the consideration of the characteristic is necessary.

C. Performance Test Without Conversion Devices

In this case, the focus is on testing the performance of IMET without conversion devices. Fig. 9 demonstrates the total volume of energy transactions and transmissions in the EI without conversion devices. From Fig. 9, its maximum energy transmission volume is about 1500 kW at 52th time interval and its maximum energy trading volume is about 1600 at 53th time interval, while the maximum energy transmission and trading volumes are about 1700 kW and 1700 kW respectively in Fig. 7 which are higher than the IMET without conversion devices. Meanwhile, it can be clearly seen that compared with Case 1, the energy transaction and transmission volumes are generally lower. It implies that energy conversion devices plays a great role in promoting energy transaction in IMET.

In addition, we further compare the reward of the proposed IMET with conversion and without conversion for energy transaction. Fig. 10 represents the comparison results of the reward for the two methods. Compared with the IMET without conversion, the reward of the proposed IMET with conversion is larger than

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

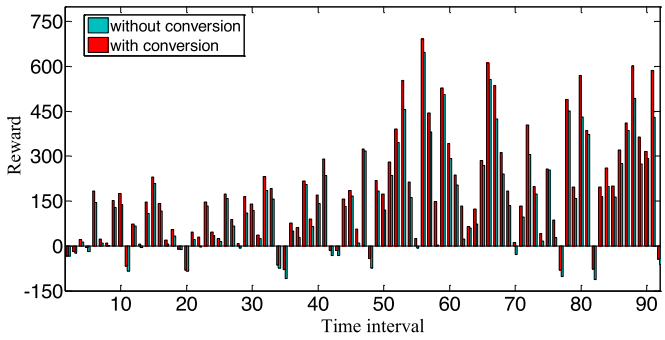

Fig. 10. Comparison results of the reward with or without conversion devices.

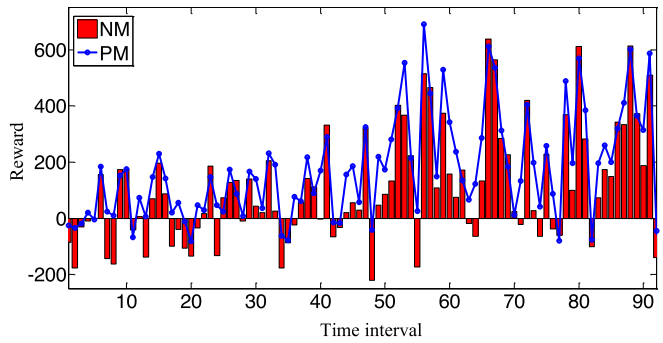

Fig. 11. Comparison results of the reward for different methods.

it most of the time. Meanwhile, there are more negative values like the 28th, 35th and 70th interval. So the transaction volume is lower when there is no conversion device. In addition, the profits of the proposed IMET with conversion and the IMET without conversion are RMB 21107, RMB 12301, respectively. Those results show that the proposed IMET with conversion works well and has the better performance.

D. Comparison of the Methods

In case 4, a comparison simulation of the conventional linear programming method and the DRL approach is provided to compare the difference between the two methods more clearly. The reward of the conventional linear programming method (NM) and the improved PPO approach (PM) for energy transaction are shown in Fig. 11. Because the conventional method can only solve the optimal strategy according to the given information, using conventional method means the timely matching transaction mechanism that trading cannot wait and buying or selling is done immediately. Table IV shows the comparison results of the two methods. Compared with NM, the reward of PM is larger than the conventional method most of the time. It is because the PM in this paper can make better use of waiting mechanism. Although there are times when the reward of NM is higher than PM like the 66th, 80th interval, there are a lot of negative situations of NM like the 24th, 48th interval which mean many WEs cannot match in the LEM and have to trade with utility company. In addition, as shown in the table, the PM obtains better profit than NM. Those results show that compared with conventional optimization method, the DRL approach has better performance for the proposed market model.

TABLE III
ALGORITHM FLOW

Algorithm 2 Modified distributed PPO Algorithm (worker)
Step-1: **for** iteration $\in \{1, \ldots, N\}$ **do**
Step-2: **for** actor $\in \{1, \ldots, M\}$ **do**
Step-3: Run policy $\pi_{\theta'}$ in environment for T timesteps, collecting $\{\hat{s}_t, a_t, r_t\}$
Step-4: Compute advantage estimates $A_1, \ldots, A_T$
Step-5: Store partial trajectory information
Step-6: **end for**
Step-7: Optimize surrogate L wrt. θ, with K epochs
Step-8: $\theta' \leftarrow \theta$
Step-9: **for** m $\in \{1, \ldots, M\}$ **do**
Step-10: Compute $\nabla_\theta J_{clip}$ by (35)
Step-11: Send gradient wrt. θ to chief
Step-12: Wait until gradient accepted or dropped, update parameters
Step-12: **end for**
Step-14: **end for**

TABLE IV
THE COMPARISON OF TWO METHODS

Method	Total reward	Number of reward under 0	Number of better reward
PM	21107	15	72
NM	9250.1	30	20

VI. CONCLUSION

This paper investigates the adaptive and indirect energy transaction problem for EI. An IMET model is proposed that can drive each WE to flexibly participate in coupled energy markets according to its own preference. To solve the impact of different energy transmission times, the multi-time scale hybrid trading mechanism is also formulated. In addition, the improved PPO algorithm is employed to learn the optimal control policy which can adapt to the model-free fashion and day-trading optimal operation performance of EI. The proposed method is demonstrated by four example cases. The consequences of the cases are analyzed in view of the volume of energy transactions and transmissions. Moreover, the comparison of the proposed method with another conventional optimal method is also discussed in the paper. As the results shown in the cases, the innovative IMET model based on the improved PPO can realize the multi-energy collaborative optimization and the efficiency of energy utilization can be improved.

REFERENCES

[1] N. Zhang, Q. Sun, J. Wang, and L. Yang, "Distributed adaptive dual control via consensus algorithm in the energy internet," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 4848–4860, Jul. 2021.
[2] B. Huang, L. Liu, H. Zhang, Y. Li, and Q. Sun, "Distributed optimal economic dispatch for microgrids considering communication delays," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 49, no. 8, pp. 1634–1642, Aug. 2019.
[3] Y. Li, D. W. Gao, W. Gao, H. Zhang, and J. Zhou, "Double-mode energy management for multi-energy system via distributed dynamic event-triggered newton-raphson algorithm," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5339–5356, Nov. 2020.
[4] Y. Li, H. Zhang, X. Liang, and B. Huang, "Event-triggered based distributed cooperative energy management for multienergy systems," *Trans. Ind. Inform.*, vol. 15, no. 14, pp. 2008–2022, Apr. 2019.
[5] Y. Mu *et al.*, "A double-layer planning method for integrated community energy systems with varying energy conversion efficiencies," *Appl. Energy*, vol. 279, 2020, Art. no. 115700.

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.

[6] C. Chen, T. Xiao, T. Qiu, N. Lv, and Q. Pei, "Smart-contract-based economical platooning in blockchain-enabled urban internet of vehicles," *IEEE Trans. Ind. Inform.*, vol. 16, no. 6, pp. 4122–4133, Jun. 2020.
[7] L. Xing, B. Jiao, Y. Du, X. Tan, and R. Wang, "Intelligent energy-saving supervision system of urban buildings based on the Internet of Things: A case study," *IEEE Syst. J.*, vol. 14, no. 3, pp. 4252–4261, Sep. 2020.
[8] N. Liu, X. Yu, C. Wang, C. Li, L. Ma, and J. Lei, "Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3569–3583, Sep. 2017.
[9] N. Liu *et al.*, "Online energy sharing for nanogrid clusters: A Lyapunov optimization approach," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4624–4636, Sep. 2018.
[10] Y. Li, D. W. Gao, W. Gao, H. Zhang, and J. Zhou, "A distributed double-Newton descent algorithm for cooperative energy management of multiple energy bodies in energy internet," *IEEE Trans. Ind. Inform.*, vol. 17, no. 9, pp. 5993–6003, Sep. 2021.
[11] "GAC. The GridWise transactive energy framework version 1.0 [EB/OL]," Jan. 2015. [Online]. Available: http://www.gridwiseac.org/pdfs/te framework report pnnl-22946.pdf
[12] L. Kristov, P. De Martini, and J. D. Taft, "A tale of two visions: Designing a decentralized transactive electric system," *IEEE Power Energy Mag.*, vol. 14, no. 3, pp. 63–69, May 2016.
[13] F. Rahimi, A. Ipakchi, and F. Fletcher, "The changing electrical landscape: End-to-end power system operation under the transactive energy paradigm," *IEEE Power Energy Mag.*, vol. 14, no. 3, pp. 52–62, May 2016.
[14] T. Chen and W. Su, "Indirect customer-to-customer energy trading with reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4338–4348, Jul. 2019.
[15] K. Mahmud, M. S. H. Nizami, J. Ravishankar, M. J. Hossain, and P. Siano, "Multiple home-to-home energy transactions for peak load shaving," *IEEE Trans. Ind. Appl.*, vol. 56, no. 2, pp. 1074–1085, Mar. 2020.
[16] N. Zhang, Q. Sun, L. Yang, and Y. Li, "Event-triggered distributed hybrid control scheme for the integrated energy system," *IEEE Trans. Ind. Inform.*, vol. 18, no. 2, pp. 835–846, Feb. 2022.
[17] F. Rahimi and A. Ipakchi, "Using a transactive energy framework: Providing grid services from smart buildings," *IEEE Electrific. Mag.*, vol. 4, no. 4, pp. 23–29, Dec. 2016.
[18] H. Hao, C. D. Corbin, K. Kalsi, and R. G. Pratt, "Transactive control of commercial buildings for demand response," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 774–783, Jan. 2017.
[19] J. Hu, X. Liu, M. Shahidehpour, and S. Xia, "Optimal operation of energy hubs with large-scale distributed energy resources for distribution network congestion management," *IEEE Trans. Sustain. Energy*, vol. 12, no. 3, pp. 1755–1765, Jul. 2021.
[20] J. Lian, H. Ren, Y. Sun, and D. J. Hammerstrom, "Performance evaluation for transactive energy systems using double-auction market," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 4128–4137, Sep. 2019.
[21] S. Bahrami and A. Sheikhi, "From demand response in smart grid toward integrated demand response in smart energy hub," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 650–658, Mar. 2016.
[22] L. Yang, Q. Sun, D. Ma, and Q. Wei, "Nash Q-learning based equilibrium transfer for integrated energy management game with We-energy," *Neurocomputing*, vol. 396, pp. 216–223, 2020.
[23] L. Yang, Q. Sun, N. Zhang, and Z. Liu, "Optimal energy operation strategy for we-energy of energy internet based on hybrid reinforcement learning with human-in-the-loop," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 52, no. 1, pp. 32–42, Jan. 2022.
[24] T. Yu, H. Z. Wang, B. Zhou, K. W. Chan, and J. Tang, "Multi-agent correlated equilibrium Q(λ) learning for coordinated smart generation control of interconnected power grids," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1669–1679, Jul. 2015.
[25] W. Liu, P. Zhuang, H. Liang, J. Peng, and Z. Huang, "Distributed economic dispatch in microgrids based on cooperative reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2192–2203, Jun. 2018.
[26] Hua. H. *et al.*, "Optimal energy management strategies for energy internet via deep reinforcement learning approach," *Appl. Energy*, vol. 239, pp. 598–609, Apr. 2019.
[27] Q. Sun, R. Fan, Y. Li, B. Huang, and D. Ma, "A distributed double-consensus algorithm for residential we-energy," *IEEE Trans. Ind. Inform.*, vol. 15, no. 8, pp. 4830–4842, Aug. 2019.
[28] D. Wolf and Y. Smeers, "The gas transmission problem solved by an extension of the simplex algorithm," *Manage. Sci.*, vol. 49, pp. 1454–1465, Nov. 2000.
[29] S. Clegg and P. Mancarella, "Integrated electrical and gas network flexibility assessment in low-carbon multi-energy systems," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 718–731, Apr. 2016.
[30] J. Mei, Z. Wei, and Y. Zhang, "Dynamic optimal dispatch with multiple time scale in integrated power and gas energy systems," *Automat. Electric Power Syst.*, vol. 42, no. 13, pp. 36–42. 2018.
[31] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: Reinforcement learning framework," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 427–435, Feb. 2004.
[32] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, vol. 1. Cambridge, MA, USA: MIT Press, 1998.
[33] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv: 170706347*.

Lingxiao Yang (Student Member, IEEE) received the B.S. degree in electrical engineering and automation from Henan Normal University, Xinxiang, China, in 2015, the M.S. degree in power system and automation, and the Ph.D. degree in power electronics and power transmission from Northeastern University, Shenyang, China, in 2017 and 2022, respectively.

She is currently a Postdoctoral Research Scholar with the School of Artificial Intelligence, Anhui University, Hefei, China. Her main research interests include machine learning and its various applications in energy internet, microgrids, and power distribution network.

Qiuye Sun (Senior Member, IEEE) received the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2007.

He is currently a Full Professor with Northeastern University. He has authored or coauthored more than 200 papers and more than ten books or textbooks. His current research interests include the stability analysis and stabilization control of novel power system with electric vehicles, distributed network control and equipment manufacturing of energy internet, integrated energy systems, and microgrids.

He is a IET Fellow. He is also a Member of the Editorial Board of four international journals and several Chinese journals, such as the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IET Cyber-Physical Systems: Theory Applications, Acta Automatica Sinica, and Proceedings of the CSEE.

Ning Zhang (Student Member, IEEE) received the B.S. degree in electrical engineering and automation from Hebei University, Baoding, China, in 2015, the M.S. degree in power system and automation, and the Ph.D. degree in power electronics and power transmission from Northeastern University, Shenyang, China, in 2017 and 2022, respectively.

He is currently a Lecturer with the School of Electrical Engineering and Automation, Anhui University, Hefei, China. His current research interests include scheduling strategy of energy hub, energy trading, security operation of energy internet, integrated energy systems, and microgrids.

Yushuai Li (Member, IEEE) received the B.S. degree in electrical engineering and automation and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2014 and 2019, respectively.

From 2019 to 2021, he was a Postdoctoral Research Scholar with the Department of Electrical and Computer Engineering, University of Denver, Denver, CO, USA. He is currently a Lecturer with the School of Information Science and Engineering, Northeastern University. His main research interests include distributed modelling, control, energy management and optimization of energy internet and multi-energy systems, and distributed machine learning algorithm with applications in microgrids.

Authorized licensed use limited to: Northeastern University. Downloaded on June 26,2025 at 23:32:05 UTC from IEEE Xplore. Restrictions apply.