Data-driven Peer-to-peer Energy Trading Based on Prosumer-driven Carbon-aware Distribution Locational Marginal Price

Xingyu Liu, Yunting Yao, Tianran Li, Yening Lai, Qi Wang, and Zhenya Ji

***Abstract*—Peer-to-peer (P2P) energy trading enables an efficient regulation of distributed renewable energy among prosumers, implicitly promoting low-carbon operation. This study proposes a novel P2P energy trading scheme with coupled electricity-carbon (E/C) market that co-optimizes both power and carbon emission flows. To facilitate the low-carbon operations in the market, we introduce a prosumer-driven carbon-aware distribution locational marginal price (PDC-DLMP) to serve as a pricing signal for the distribution system operator (DSO). To efficiently determine the optimal trading solutions, we adopt a two-layer data-driven approach. The first layer employs a reinforcement learning algorithm named multi-agent twin-delayed deep deterministic policy gradient (MATD3); the second layer uses a deep neural network (DNN) driven surrogate model, which is designed to map the PDC-DLMP signals, thereby eliminating the need for direct DSO intervention during market operation. This approach protects the physical model parameters of the distribution network and ensures multi-level privacy protection. Simulation results validate the effectiveness of the proposed P2P energy trading scheme with coupled E/C market, demonstrating its ability to achieve both reduced carbon emissions and lower operational costs for microgrid prosumers.**

***Index Terms*—Peer-to-peer (P2P) energy trading, electricity market, carbon market, distribution locational marginal price, reinforcement learning, deep neural network (DNN), surrogate model.**

Nomenclature

A. Sets and Indices

$\mathcal{L}_n^*$	Set of all lines connected to node n
Φ_L	Set of load nodes
Φ_{MG}^n	Set of network nodes where microgrid (MG) attaches
g, Φ_G	Index and set of thermal generators
$g(n)$	Index of nodes where generator g attaches
i, Φ_{MG}	Index and set of MG prosumers
k	Index of neurons in deep neural network (DNN)
l	Index of power network lines
l^+, l^-	Indices of originating- and receiving-end nodes of line l
L_n^+, L_n^-	Set of power network lines that flow into and out of nodes
n, N	Index and set of power network nodes
s	Index of samples in DNN
t, T	Index and set of time periods

B. Variables

$\tilde{\epsilon}$	Clipped Gaussian noise
λ^C	Nodal carbon emission marginal price
λ^E	Distribution locational marginal price (DLMP)
$\lambda_{s,t}^{E/C*}$	Predicted price signal of sample s in DNN at time t
$\lambda_{t,+}^{\text{grid}}, \lambda_{t,-}^{\text{grid}}$	Electricity purchasing and selling prices in peer-to-grid (P2G) market at time t
λ_t^{CT}	External carbon emission price at time t
λ_t^{BP}	Electricity purchasing price from grid at time t
$\lambda_{t,+}^{P2P}, \lambda_{t,-}^{P2P}$	Electricity purchasing and selling prices in peer-to-peer (P2P) market at time t
λ^{NUC}	Network usage charge
$\lambda_{n,t}^{PDC\text{-}DLMP}$	Prosumer-driven carbon-aware distribution locational marginal price (PDC-DLMP) of node n at time t
ξ	Set of decision variables for distributed resources in power network
$a_{l,t}$, $v_{n,t}$	Squared current flow of line l and squared nodal voltage magnitude of node n at time t
$a_{i,t}^{Bess}$	Charging and discharging action space of battery of MG i at time t
A_i	Set of individual actions in partially observable Markov game (POMG) of MG i
c	Set of decision variables related to carbon flow
$C_{i,t}^{\text{cost}}$	Operational cost of MG i at time t
$C_{i,t}^{PG}$	Generation cost of distributed diesel generator

Manuscript received: May 25, 2024; revised: July 2, 2024; accepted: December 2, 2024. Date of CrossCheck: December 2, 2024. Date of online publication: February 28, 2025.

This work was supported in part by the National Natural Science Foundation of China (No. 52107100) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. SJCX24-0657).

X. Liu, Y. Yao, T. Li (corresponding author), Q. Wang, and Z. Ji are with the NARI School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China (e-mail: 1962270695@qq.com; fionaforreal@163.com; litianran@njnu.edu.cn; wangqi@njnu.edu.cn; 61214@njnu.edu.cn).

Y. Lai is with NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing, China (e-mail: laiyening@sgepri.sgcc.com.cn).

DOI: 10.35833/MPCE.2024.000548

	in MG i at time t
$C_{i,t}^{P2G}$	Transaction cost in P2G market of MG i at time t
$C_{i,t}^{\text{carbon}}$	Carbon emission cost of MG i at time t
$C_{i,t}^{P2P}$	Transaction cost in P2P market of MG i at time t
e_t^{BAC}	Carbon emission intensity of battery at time t
e^N, e^L	Carbon emission intensities of node and line
f	Set of decision variables related to power flow in power network
h_s	Input layer activation for sample s in DNN
O_i	Set of individual observations in POMG
O_{global}	State of global observation
$P_t^{BAC,CH}$, $P_t^{BAC,DC}$	Charging and discharging power of battery at time t
P_t^G	Output of distributed diesel generators at time t
P_t^{RES}	Output of renewable energy generators at time t
$P_{l,t}^B$, $Q_{l,t}^B$	Active and reactive power flows of line l at time t
P_t^{Buy}	Electricity purchased from power network at time t
$P_{n,t}^{in}$	Total power flow inflow of node n at time t
$P_{n,t}^{inj}$	Net load of node n at time t
P_t^{Load}	Output of loads at time t
$\hat{R}_{n,t}$	Total carbon emission of node n at time t
$R_{l,t}^B$	Total carbon emission accompanying active power flow of line l at time t
$R_{n,t}^{in}$	Total carbon emission inflow of node n at time t
R^G, R^L	Total carbon emissions of generator and load
R^{BAC}	Total carbon emission of battery
R_i	Set of individual rewards in POMG
SOC_t^{BAC}	Energy storage state of battery at time t
W	Target value in critic network
$x_t^{i,j}$	Cleared transaction energy between MG i and MG j at time t
$\boldsymbol{x}_t^i$	Vector of cleared transaction energies of MG i at time t
$\boldsymbol{y}_t^i$	Vector of trading energies of MG i at time t
z_s	Output layer activation for sample s in DNN

C. Parameters

α	Step size of updates for Q-function
γ	Discount factor in POMG
Δt	Length of each time slot
θ_{DNN}	DNN parameter for surrogate model
τ	Soft update parameter in target network
ω_{jk}	Linear relationship coefficient assigned to neuron k for feature j in DNN
a_0, b_0	Cost coefficients of distributed diesel generators
b_k	Bias coefficient assigned to neuron k in DNN
B	Size of mini-batch used for updating policy network
B_n, G_n	Susceptance and conductance of node n
e^G	Carbon emission intensity of generators
M	Episode of training process repeated
N_S	Number of training samples in DNN
P^{POMG}	Probability of state transition in POMG
$\overline{P}_i^{ev}$	The maximum charging/discharging value of battery i
r_l, x_l	Resistance and reactance of line l
S_l	Complex power of line l
$\overline{(\cdot)}$, $\underline{(\cdot)}$	Upper and lower bounds

D. Functions

$E(\cdot)$	Function of electricity cost
$C(\cdot)$	Function of carbon emission cost
$C_{Con}(\cdot)$	Function of carbon emission flow (CEF) constraints
$C_{Eq}(\cdot)$	Function of CEF balance equations
$P_{Con}(\cdot)$	Function of power flow constraints
$P_{Eq}(\cdot)$	Function of power flow balance equations

I. Introduction

EXCESSIVE carbon emissions have led to increasingly obvious climate change. Currently, 40% of the global carbon emissions originate from fossil fuel combustion during power production [1]. To mitigate these environmental challenges, decarbonizing the energy supply side has become an urgent task. Policymakers are actively encouraging the integration of distributed energy resources (DERs) into distribution systems and promoting their participation in electricity markets [2]. Advancements in smart grid technologies have enabled third-party entities to produce and store electricity locally using DERs and batteries, giving rise to the concept of prosumers, i.e., consumers who also produce electricity [3]. Among these, microgrids (MGs) have emerged as key players, which can aggregate renewable energy and participate in electricity markets as prosumers [4]. To achieve the low-carbon operation of power system, it is crucial to incorporate carbon policies into decision-making and enhance the utilization of DERs during power dispatch optimization. These efforts support the transition to a more sustainable and efficient energy ecosystem.

The market mechanism for coupled electricity-carbon (E/C) operations is an effective approach for integrating carbon policies and promoting decarbonization among prosumers. Reference [3] developed a blockchain-based trading framework that integrates energy and carbon markets, facilitating the exchange of energy and carbon quotas among MGs. Reference [5] proposed an energy trading mechanism for E/C co-optimization based on carbon quotas, specifically targeting building communities. Reference [6] introduced a combined operating mechanism for distributed photovoltaics (PVs) and carbon markets, leveraging blockchain technology

[6]. However, the conventional studies have primarily focused on "observed" emissions, mostly considering the interactions between the electricity and carbon markets from the generation side [7], [8].

Existing solutions that focus solely on the generation side fail to adequately capture the dual effects of carbon emissions from both the supply and demand sides [9]. This often results in an increased economic burden on the supply side due to marginal carbon emissions and leads to misaligned incentives [8], [10]. To address this, [11] introduced the concept of carbon emission flow (CEF) to clarify the carbon emission responsibilities of both sides. Building on the CEF model, [8] designed a coupled E/C market that incorporates the demand side. In addition, [7] developed an integrated pricing model for carbon emissions in multi-energy systems, and [12] employed the carbon-integrated distribution locational marginal price (DLMP) at the distribution network level to guide low-carbon operations of electric ships [12]. However, these studies do not account for the application of the CEF model to prosumers or provide dispatch price signals tailored to their specific E/C coupling.

Although the CEF-based coupled E/C market captures carbon emission responsibilities more effectively and promotes low-carbon operation within the energy system, the increasing penetration of distributed renewable energy into distribution networks necessitates more flexible trading methods to enhance its utilization [13]. Peer-to-peer (P2P) energy trading, as an emerging energy management approach, offers greater scheduling flexibility for DERs among prosumers [14]. Conventional optimization methods for P2P energy trading can be categorized into four types: game theory [15], [16], auction theory [17], constrained optimization [18], and blockchain [19]. As model-based methods, they face two main challenges: ① some optimization methods, such as mixed-integer linear programming (MILP), assume linear relationships among variables [20], whereas the alternating direction method of multipliers (ADMM) assumes regularization and convexity, leading to slow consensus convergence [21]; and ② due to the stochastic nature of renewable energy generation, load consumption, and distribution network constraints [22], along with the prosumer trading and the introduction of CEF coupling concept [10], the nonlinearity and non-convexity of P2P energy trading problems are significantly heightened, rendering them intractable.

Deep reinforcement learning (DRL) is an alternative approach for addressing these challenges [23]. By combining deep learning with reinforcement learning, DRL offers strong exploration capabilities and model-free characteristics, making it an increasingly powerful tool for solving P2P energy trading problems [24], [25]. In [26], the deep Q-learning is employed to tackle electricity market trading, whereas in [27] and [28], the multi-agent deep deterministic policy gradient (MADDPG) algorithm is used to demonstrate its potential for approaching the optimal decision-making in highly dynamic P2P energy trading markets, where trading strategies remain undisclosed and privacy is preserved. However, both the deep Q-learning and DDPG suffer from two key limitations: ① they struggle to perform effectively in environments with a large number of actions in a continuous action space; and ② they tend to overestimate Q-values, which might lead to sub-optimal decision-making and unstable performance in real-world applications [29]. In addition, in multi-agent energy trading environments, these methods are typically implemented within a centralized training and decentralized execution (CTDE) framework. Although this framework improves the training stability even in the face of continuous and dynamic actions from the agents, it requires access to a precise physical model of the distribution network during the training phase. This introduces several issues: ① exposure of the sensitive physical model parameters of the distribution network; ② increased computational burden for the distribution system operator (DSO) during training; and ③ deviation from the original intent of model-free problem-solving [30].

In summary, this paper narrows three critical research gaps that require urgent attention by: ① designing an economic incentive price signal based on the CEF theory that accommodates prosumers, which could effectively guide low-carbon operations; ② developing an appropriate P2P energy trading scheme with coupled E/C market and employing more suitable techniques to clear coupled E/C trading between prosumers; and ③ addressing the issue of exposing actual parameters during the training of distribution network models. Table I summarizes the key differences of technical features between this paper and previous research.

TABLE I
COMPARATIVE TECHNICAL FEATURES OF DIFFERENT STUDIES

Reference	Network constraint	P2P energy trading	Carbon tracing for prosumers	Optimization method
[3]	√	√	×	Blockchain
[7], [8]	√	×	√	Fixed-point iteration
[13]	√	√	×	Distributed optimization
[15]	√	√	×	Game theory
[5], [22]	×	√	×	Multi-agent DRL (MADRL)
[12]	√	×	√	Distributed optimization
[31]	√	×	√	MADRL
[10]	√	√	√	Distributed optimization
This paper	√	√	√	MADRL + surrogate model

Note: "√" denotes consideration of the factor; "×" denotes non-consideration of the factor.

The main contributions of this study are as follows.

1) We first propose a carbon-aware optimal power flow (C-OPF) model integrated with CEF theory. Using this model, we introduce a decomposable prosumer-driven carbon-aware distribution locational marginal price (PDC-DLMP) at the DSO level. The DSO utilizes this price signal to guide the low-carbon operation of MG prosumers.

2) We develop a P2P energy trading scheme with coupled E/C market based on the PDC-DLMP signal, which effectively clarifies the economic responsibility of network usage and excessive carbon emissions, providing a more flexible dispatching space for DERs in distribution networks. To address these highly non-convex and nonlinear challenges, we employ a multi-agent twin-delayed deep deterministic policy gradient (MATD3) algorithm from MADRL.

3) To protect the model parameters of distribution network during the centralized training process of DRL, we propose a two-layer data-driven approach. The first layer employs the MATD3 algorithm for the optimal P2P energy trading, and the second layer utilizes a deep neural network (DNN) based surrogate model to map the PDC-DLMP signal. This approach ensures that the model parameters of distribution network remain private, thereby facilitating secure interactions with third-party agents.

The remainder of this paper is organized as follows. Section II provides an overview of the system framework. Section III presents the proposed system model, which introduces the C-OPF model to derive the PDC-DLMP signal for the DSO and develops a P2P energy trading scheme with coupled E/C market based on this signal. Section IV introduces the two-layer data-driven approach. Section V presents case simulation studies. Section VI concludes this study.

II. System Framework Establishment

As shown in Fig. 1, the framework of the proposed P2P energy trading scheme with coupled E/C market consists of three layers, which are briefly described as follows.

1) Physical layer: this layer is composed of multiple MG prosumers and the point of common coupling (PCC) connecting the MGs. Each MG prosumer includes distributed renewable energy generators, distributed diesel generators, batteries, and loads. In addition, CEFs interact between each MG prosumer and the distribution network via the PCC.

2) Market layer: this layer includes the P2P energy trading market with coupled E/C market and the peer-to-grid (P2G) market. MG prosumers first engage in the P2P market, where surplus energy is cleared in the P2G market. The DSO reflects the information of coupled E/C market and network operation through price signals with E/C coupling.

3) Coordination strategy layer: coordination strategies are implemented using data-driven DNN and MATD3 algorithms. The DSO pretrains a surrogate model, which is then distributed to local agents to generate PDC-DLMP signals. Each MG prosumer is modeled as an RL agent and trained using an embedded surrogate model to achieve the optimal trading solutions.

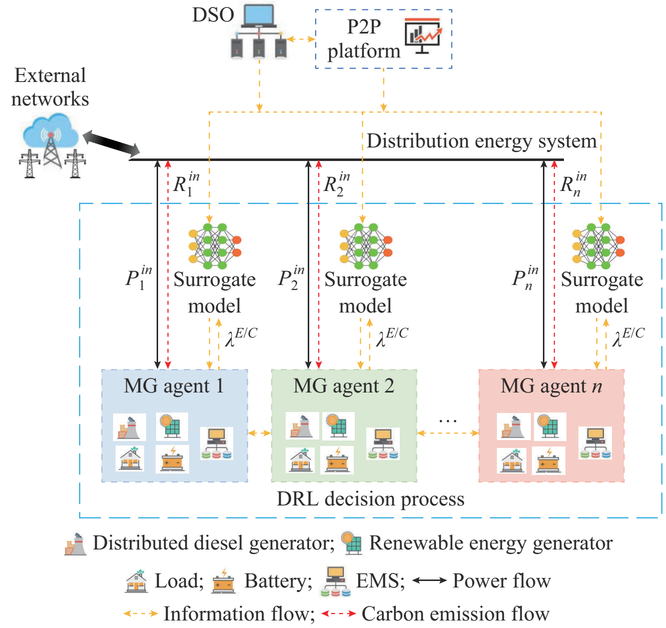

Fig. 1. Framework of proposed P2P energy trading scheme with coupled E/C market.

III. System Modeling

A. CEF Model

Under the concept of CEF, the carbon emissions from power generators are treated as "virtual" emission flow, which runs parallel to the physical power flow, as shown in Fig. 2. These "virtual" emission flows are injected into the power network at the generator node and transported to end users through the power system [32]. To quantify and analyze this type of "virtual" emission flow, a CEF model [11] based on the conservation law of carbon mass (CLCM) and the proportional sharing principle (PSP) is used, and thus the distribution of carbon emissions across the network can be tracked. The DSO can then charge third-party entities for their carbon emissions based on the cleared carbon emissions. In the CEF model, the carbon emission intensity is a critical indicator for quantitatively analyzing carbon emissions in a network, describing carbon emissions coupled with each unit of power flow [33].

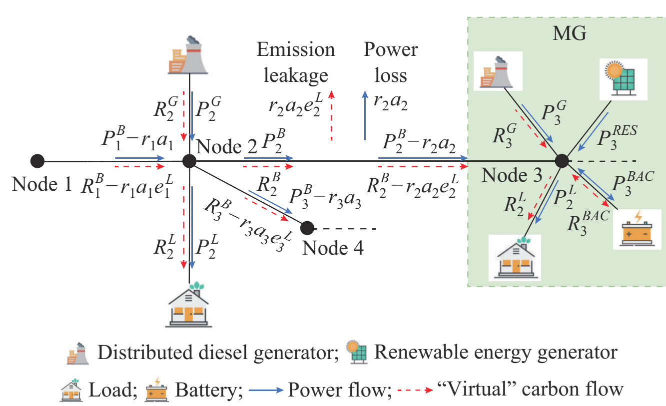

Fig. 2. Illustration of CEF and physical power flow.

According to [11], the carbon emission intensity of any node in the distribution network depends on two factors: ① the incoming CEF from the connected lines; and ② the power flow injected into the node. The carbon emission intensity of line in the network is equal to that of the node from which the line originates.

The carbon emission intensity of node can be expressed as:

$$e_{n,t}^{N}=\frac{R_{n,t}^{in}}{P_{n,t}^{in}}=\frac{\sum_{l\in L_n^+}(R_{l,t}^{B}-r_l a_{l,t}e_{l,t}^{L})+\sum_{g(n)\in\Phi_G}R_{g(n),t}^{G}}{\sum_{l\in L_n^+}(P_{l,t}^{B}-r_l a_{l,t})+\sum_{g(n)\in\Phi_G}P_{g(n),t}^{G}}\quad\forall n,\forall t\tag{1}$$

The carbon emission intensity of line can be expressed as:

$$e_{l,t}^{L}=e_{n,t}^{N}\quad\forall n,\forall t,\forall l\in L_n^-\tag{2}$$

Different operating states of energy storage devices like battery result in varying carbon emissions. In the discharging state, the batteries are treated as generators and their carbon emission intensity remains constant. In the charging state, they are treated as loads that absorb power from the distribution network and their carbon emission intensity increases, which is continuously updated based on the carbon intensity of the absorbed electricity [31]. The carbon emission intensity of battery $e_{i,t+1}^{BAC}$ can be updated by:

$$e_{i,t+1}^{BAC}=\frac{e_{i,t}^{BAC}\cdot SOC_{i,t}^{BAC}+e_{i,t}^{N}P_{i,t}^{BAC,CH}\Delta t}{SOC_{i,t}^{BAC}+P_{i,t}^{BAC,CH}\Delta t}\quad\forall n,\forall t,\forall i\in\Phi_{MG}^{n}\tag{3}$$

Based on the effects of batteries on MG prosumers, the carbon emission intensity of node in CEF model can be reformulated as:

$$e_{n,t}^{N}=\frac{\sum_{l\in L_n^+}(P_{l,t}^{B}-r_l a_{l,t})e_{l,t}^{L}+P_{i,t}^{G}e_i^{G}+P_{i,t}^{BAC,DC}e_{i,t}^{BAC}}{\sum_{l\in L_n^+}(P_{l,t}^{B}-r_l a_{l,t})+P_{i,t}^{RES}+P_{i,t}^{G}+P_{i,t}^{BAC,DC}}\quad\forall n,\forall t,\forall i\in\Phi_{MG}^{n}\tag{4}$$

B. C-OPF Model

We then construct the C-OPF model as a market clearing strategy based on the CEF theory. The general compact form of the C-OPF model is given by [32]:

$$\min\left(E(\xi,f,c)+C(\xi,f,c)\right)\tag{5a}$$

s.t.

$$P_{Eq}(\xi,f)=0\tag{5b}$$

$$P_{Con}(\xi,f)\le 0\tag{5c}$$

$$C_{Eq}(\xi,f,c)=0\tag{5d}$$

$$C_{Con}(\xi,f,c)\le 0\tag{5e}$$

In a distribution network, $E(\cdot)$ can be expressed as the cost of network losses or that of purchasing electricity from a higher-level grid. In this paper, the objective function of the C-OPF model in (6a) includes the costs of carbon emission charged by the DSO to each node as well as the cost of purchasing electricity from the higher-level grid. The equations and constraints for the power flow in a distribution network are similar to those in [34]. The nonlinear problems for power flow calculations in the distribution network are addressed using second-order cone relaxation, as shown in (6b)-(6l). The constraints for carbon flow are expressed as (6m) and (6n).

$$\min\left(\sum_{n\in N}\lambda_t^{CT}\hat{R}_{n,t}\Delta t+\lambda_t^{BP}P_t^{Buy}\Delta t\right)\quad\forall n,\forall t\tag{6a}$$

$$\sum_{l\in L_n^-}P_{l,t}^{B}-\sum_{l\in L_n^+}(P_{l,t}^{B}-r_l a_{l,t})-P_{n,t}^{inj}+G_n v_{n,t}=0:\lambda_{n,t}\quad\forall n,\forall t\tag{6b}$$

$$\sum_{l\in L_n^-}Q_{l,t}^{B}-\sum_{l\in L_n^+}(Q_{l,t}^{B}-x_l a_{l,t})+B_n v_{n,t}=0:\mu_{n,t}\quad\forall n,\forall t\tag{6c}$$

$$(P_{l,t}^{B})^2+(Q_{l,t}^{B})^2\le S_l^2:\eta_{l,t}^{+}\quad\forall l,\forall t\tag{6d}$$

$$(P_{l,t}^{B}-r_l a_{l,t})^2+(Q_{l,t}^{B}-x_l a_{l,t})^2\le S_l^2:\eta_{l,t}^{-}\quad\forall l,\forall t\tag{6e}$$

$$v_{l^+,t}-2(r_l P_{l,t}^{B}+x_l Q_{l,t}^{B})+a_{l,t}(r_l^2+x_l^2)=v_{l^-,t}:\alpha_{l,t}\quad\forall l,\forall t\tag{6f}$$

$$\frac{(P_{l,t}^{B})^2+(Q_{l,t}^{B})^2}{a_{l,t}}\le v_{l^+,t}:\beta_{l,t}\quad\forall l,\forall t\tag{6g}$$

$$\underline{V}_n\le v_{n,t}:\gamma_{n,t}^{-}\quad\forall n,\forall t\tag{6h}$$

$$v_{n,t}\le\bar{V}_n:\gamma_{n,t}^{+}\quad\forall n,\forall t\tag{6i}$$

$$\underline{Q}_{n,t}\le Q_{n,t}\le\bar{Q}_{n,t}:\nu_{n,t}\quad\forall n,\forall t\tag{6j}$$

$$P_{n,t}^{inj}=P_{n,t}^{Load}\quad\forall n\in\Phi_L,\forall t\tag{6k}$$

$$P_{n,t}^{inj}=P_{n,t}^{Load}+P_{n,t}^{BAC,CH}-P_{n,t}^{G}-P_{n,t}^{RES}-P_{n,t}^{BAC,DC}\quad\forall n\in\Phi_{MG},\forall t\tag{6l}$$

$$\frac{1}{2}\sum_{l\in\mathcal{L}_n^*}e_{l,t}^{L}a_{l,t}r_l+\frac{1}{2}P_{n,t}^{Load}e_{n,t}^{N}\Delta t\le\hat{R}_{n,t}:\varphi_{n,t}^{Load}\quad\forall n\in\Phi_L,\forall t\tag{6m}$$

$$\frac{1}{2}\sum_{l\in\mathcal{L}_n^*}e_{l,t}^{L}a_{l,t}r_l+\frac{1}{2}(P_{n,t}^{Load}+P_{n,t}^{BAC,CH})+\frac{1}{2}(P_{n,t}^{G}e_n^{G}+P_{n,t}^{BAC,DC}e_n^{BAC})\Delta t\le\hat{R}_{n,t}:\varphi_{n,t}^{MG}\quad\forall n\in\Phi_{MG},\forall t\tag{6n}$$

where the dual variables $\lambda_{n,t}$, $\mu_{n,t}$, $\eta_{l,t}^{+}$, $\eta_{l,t}^{-}$, $\alpha_{l,t}$, $\beta_{l,t}$, $\gamma_{n,t}^{-}$, $\gamma_{n,t}^{+}$, $\nu_{n,t}$, $\varphi_{n,t}^{\text{Load}}$, and $\varphi_{n,t}^{MG}$ are associated with constraints (6b)-(6j), (6m), and (6n), respectively. The left-hand sides of (6m) and (6n) represent the equivalent carbon emissions of nodes, which include the carbon emissions from power generation and load consumption at the nodes as well as the cumulative carbon emissions from network losses on all adjacent lines.

The CEF model assumptions state that the power emitted by all DERs in the distribution network along with the carbon emissions is transmitted through the network and ultimately absorbed by the loads. Therefore, this study assumes that both generators and consumers share equal responsibilities for carbon emissions by multiplying the equivalent carbon emissions from power generation and load consumption at all nodes by a factor of 1/2. Notably, for regular load nodes, only the equivalent carbon emissions from load absorption must be calculated; whereas for MG prosumers, all carbon emissions resulting from both power generation and load consumption must be considered.

C. Pricing for P2P Energy Trading with Coupled E/C Market Based on PDC-DLMP

In the proposed market framework, DSOs are tasked not only with ensuring the safe and stable operation of the power network but also with developing appropriate price signals to facilitate market transactions in the coupled E/C market. Inspired by [10] and [35], this study introduces the concept of PDC-DLMP signals based on the C-OPF model as an incentive for MG prosumers with both carbon and network awareness. In addition, the coupled E/C market enables a reasonable collection of network usage charge (NUC) and carbon emission costs through PDC-DLMP signals. The PDC-DLMP $\lambda_{n,t}^{PDC\text{-}DLMP}$ is expressed as:

$$\lambda_{n,t}^{PDC\text{-}DLMP}=\lambda_{n,t}^{E}+\lambda_{n,t}^{C} \quad \forall n \in N, \forall t \tag{7}$$

$$\lambda_{n,t}^{C}=\frac{1}{2} e_{n,t}^{N} \varphi_{n,t}^{MG} \Delta t \quad \forall n \in \Phi_{MG}, \forall t \tag{8}$$

1) NUC

NUC is a price signal provided by the DSO to guide buyers and sellers to comply with the constraints of a distribution network [36]. From (7) and (8), it is evident that the PDC-DLMP is an extension of the DLMP. Therefore, the NUC between buyer agent i and seller agent j derived from PDC-DLMP is expressed as:

$$\lambda_{ij}^{NUC}=\frac{\left|\lambda_{i}^{E}-\lambda_{j}^{E}\right|}{2} \quad \forall i,j \in \Phi_{MG}^{n} \tag{9}$$

2) P2P and P2G Energy Trading Prices

In the P2G market, MGs typically pay higher prices when purchasing electricity from distribution networks than those that MGs receive when selling electricity back to the network. The initial settings for the trading prices of P2P energy and wholesale electricity market must satisfy the following constraint:

$$\lambda_{t,-}^{grid}<\lambda_{t,-}^{P2P} \approx \lambda_{t,+}^{P2P}<\lambda_{t,+}^{grid}=\lambda_{n,t}^{E} \tag{10}$$

This constraint ensures that energy trading in the P2P market is more advantageous than that in the wholesale electricity market, thus prioritizing the clearing of P2P energy trading. Note that the NUC can be utilized to facilitate P2P energy trading while ensuring the distribution network security and reflecting the trading preferences of neighboring nodes. Therefore, (10) can be modified as:

$$\lambda_{t,-}^{grid}<\lambda_{t,-}^{P2P}-\lambda^{NUC}<\lambda_{t,+}^{P2P}+\lambda^{NUC}<\lambda_{t,+}^{grid} \tag{11}$$

Under the joint incentives of market trading prices, NUC, and nodal carbon emission marginal prices $\lambda_{n,t}^{C}$, MG operators can adjust their energy dispatch strategies to increase the share of cheaper and cleaner energy purchased in the P2P market. This not only reduces operating costs but also helps meet the DSO requirements for grid stability, safe operation, and carbon reduction goals.

D. Modeling for Batteries

The modeling of battery in MG prosumers is expressed as:

$$SOC_{i,t+1}^{BAC}=SOC_{i,t}^{BAC}+P_{i,t}^{BAC,CH}\Delta t-P_{i,t}^{BAC,DC}\Delta t \tag{12a}$$

$$\underline{SOC}_{i}^{BAC} \leq SOC_{i,t}^{BAC} \leq \overline{SOC}_{i}^{BAC} \tag{12b}$$

$$0 \leq P_{i,t}^{BAC,CH} \leq \bar{P}_{i,t}^{BAC,CH} \tag{12c}$$

$$0 \leq P_{i,t}^{BAC,DC} \leq \bar{P}_{i,t}^{BAC,DC} \tag{12d}$$

Equation (12a) is used to update the state of charge (SOC) of battery. Constraint (12b) indicates that the remaining SOC after charging/discharging should be within the limits of SOC, and (12c) and (12d) restrict the charging and discharging power of battery, respectively.

E. Peer Matching Method for Proposed P2P Energy Trading Scheme with Coupled E/C Market

Prior to initiating a transaction, all MGs in the distribution network will predefine their trading strategies to identify potential transactions in the P2P market. Define $y_t^{i,j}$ as initial trading energy between MG i and MG j: when $y_t^{i,j}>0$, MG i intends to buy electricity from MG j, and when $y_t^{i,j}<0$, MG i intends to sell electricity to MG j. The vector of trading energies for MG i with other MGs is defined as:

$$\boldsymbol{y}_t^{i}=[y_t^{i,1},y_t^{i,2},\ldots,y_t^{i,i-1},y_t^{i,i+1},\ldots,y_t^{i,j},\ldots,y_t^{i,I}] \quad \forall i,j \in \Phi_{MG}, i \neq j \tag{13}$$

Since initial trading intentions of each market participant (i.e., MG) are self-determined, the challenges arise when decisions involve another MG, especially when both are potential buyers or sellers. This makes it difficult to match transactions. The cleared trading strategy is then calculated as:

$$\begin{cases}\left|y_t^{i,j}\right| \leq \bar{y}^{i,j} \\ \bar{y}^{i,j}=\bar{y}^{j,i}\end{cases} \quad i \neq j \tag{14}$$

$$x_t^{i,j}=\begin{cases}0 & y_t^{i,j} y_t^{i,j} \geq 0 \\ \dfrac{y_t^{i,j}}{\left|y_t^{i,j}\right|} \min\left(\left|y_t^{i,j}\right|,\left|y_t^{i,j}\right|\right) & y_t^{i,j} y_t^{i,j}<0\end{cases} \tag{15}$$

F. Power Balancing

At time t, each MG must balance its power generation and power consumption. We define $y_t^{i,i}$ and $x_t^{i,i}$ as the expected and cleared trading energies of MG i in the P2G at time t, respectively, where positive and negative values indicate the purchasing electricity from and selling electricity to the distribution network, respectively. The vectors of expected and cleared trading energies of MG i are then updated as:

$$\boldsymbol{y}_t^{i}=[y_t^{i,1},y_t^{i,2},\ldots,y_t^{i,i-1},y_t^{i,i},y_t^{i,i+1},\ldots,y_t^{i,I}] \tag{16}$$

$$\boldsymbol{x}_t^{i}=[x_t^{i,1},x_t^{i,2},\ldots,x_t^{i,i-1},x_t^{i,i},x_t^{i,i+1},\ldots,x_t^{i,I}] \tag{17}$$

The trading strategies should satisfy the power balance as:

$$P_{i,t}^{G}+P_{i,t}^{RES}+P_{i,t}^{BAC,DC}+\sum x^{i,j}(t)=P_{i,t}^{Load}+\frac{1}{2}P_{i,t}^{BAC,CH} \quad \forall i,j \in \Phi_{MG}, i \neq j \tag{18}$$

At time t, each MG independently determines its trading list based on (18). The P2P energy trading mechanism updates the trading volume according to the cleared trading strategy in (14) and (15). Finally, the variable $x_t^{i,i}$ is used to adjust the trading power between the MG and distribution network.

G. Objective Function

The operational cost of each MG $C_{i,t}^{\text{cost}}$ should be minimized over a dispatch cycle (24 hours), which includes the power generation cost of MG $C_{i,t}^{PG}$, market transaction costs of MG in the P2G market $C_{i,t}^{P2G}$ and P2P market $C_{i,t}^{P2P}$, and carbon emission cost $C_{i,t}^{\text{carbon}}$ depending on the power flow:

$$C_{i,t}^{\text{cost}}=C_{i,t}^{PG}+C_{i,t}^{P2G}+C_{i,t}^{P2P}+C_{i,t}^{\text{carbon}} \quad \forall i \in \Phi_{MG}, \forall t \tag{19a}$$

$$C_{i,t}^{PG}=a_0(P_{i,t}^G)^2+b_0P_{i,t}^G \quad \forall i\in\Phi_{MG},\forall t \tag{19b}$$

$$C_{i,t}^{P2G}=x_t^{i,i}(I_{(x_t^{i,i}>0)}\lambda_{t,+}^{grid}-I_{(x_t^{i,i}\le 0)}\lambda_{t,-}^{grid}) \quad \forall i\in\Phi_{MG},\forall t \tag{19c}$$

$$C_{i,t}^{P2P}=\sum x_t^{i,i}[I_{(x_t^{i,j}>0)}(\lambda_{t,+}^{P2P}+\lambda_{ij}^{NUC})-I_{(x_t^{i,j}\le 0)}(\lambda_{t,-}^{P2P}-\lambda_{ij}^{NUC})] \quad \forall i,j\in\Phi_{MG},i\ne j,\forall t \tag{19d}$$

$$C_{i,t}^{\text{carbon}}=(P_{n,t}^{Load}+P_{n,t}^{BAC,CH}+P_{n,t}^{G}+P_{n,t}^{BAC,DC})\lambda_{n,t}^{C} \quad \forall n\in N,\forall i\in\Phi_{MG}^n,\forall t \tag{19e}$$

IV. Proposed Two-layer Data-driven Approach

To address the highly nonlinear and nonconvex nature of the model proposed in (1)-(19) and to efficiently solve the optimization problem, we employ a two-layer data-driven approach: DRL algorithm embedded with a DNN-based surrogate model.

A. DNN-based Surrogate Model

A DNN-based surrogate model is used to map the price signals governing the trading in coupled E/C market, specifically addressing the optimization problem represented by (1)-(9). Using a DNN to construct the surrogate model offers two main advantages.

1) A DNN demonstrates stronger generalization capabilities and higher computational efficiency.

2) When the DNN is trained with a large amount of randomly generated energy interaction data and price signals calculated from the C-OPF model, the DSO can protect its power network information, as the parameters of distribution network model remain undisclosed.

1) Framework of DNNs

DNNs are artificial neural network (NN) models consisting of multiple hidden layers. Compared with conventional shallow NNs, DNNs have more hidden layers between the input and output layers. The input and output layers are related by the following mapping:

$$z_{sk}^{(l)}=\sum_{j=1}h_{sj}^{(l)}\omega_{jk}^{(l)}+b_k^{(l)} \tag{20}$$

Equation (20) represents a linear relationship within the l^{th} hidden layer. The input to each layer in the DNN is obtained by weighting and summing the output from the previous layer according to the mapping relationship.

2) DNN Training Algorithm

In a DNN, both the linear relationship and bias coefficients are network parameters that must be optimized. A back propagation algorithm is then applied to optimize these parameters by minimizing the loss function, using the mean square error (MSE):

$$L(\theta_{DNN})=\frac{1}{N_S}\sum_{s=1}^{N_S}(\lambda_{s,t}^{E/C*}-\lambda_{s,t}^{E/C})^2 \tag{21}$$

B. MADRL Algorithm

We model the proposed P2P energy trading scheme with coupled E/C market discussed in Section III as a partially observable Markov game (POMG). The MATD3 algorithm is then applied to explore the optimal strategies for each participant and solve the proposed problem.

1) Formulation of POMG

The POMG is a natural extension of the partially observable Markov decision process (POMDP). POMG is defined as $POMG(N,O_{global},A_i,O_i,P^{POMG},R_i,\gamma)$. The goal of each agent i is to maximize the cumulative discounted reward.

The components of POMG are described as follows.

1) Observation

The observation of each agent i is expressed as:

$$o_{i,t}=\{\hat{P}_{i,t}^{RES},\hat{P}_{i,t}^{Load},SOC_{i,t},P_{i,t}^{inj},\lambda_{n,t}^{E},\lambda_{n,t}^{C}\} \quad \forall n,\forall i\in\Phi_{MG}^n \tag{22}$$

As no MG can accurately obtain load fluctuation curves for renewable energy, to achieve better generalization capabilities and handle uncertainties, the symbol “^” in the observation space denotes Gaussian noise, which serves as a means of incorporating uncertainty-aware training, helping the model better generalize and adapt to the inherent unpredictability of renewable energy sources.

2) Action

The action of each agent i at time t is expressed as:

$$a_{i,t}=\{P_{i,t}^G,a_{i,t}^{Bess},\boldsymbol{y}_t^i\} \tag{23}$$

This study constructs a safety model in the environment to constrain actions $a_{i,t}^{Bess}$. This modeling approach replaces the use of penalty functions to punish actions that exceed capacity limits, allowing the agent to make better choices during exploration and thus reducing operational costs. We reformulate the charging and discharging power of batteries $P_{i,t}^{BAC,CH}$ and $P_{i,t}^{BAC,DC}$ as:

$$P_{i,t}^{BAC,CH}=\left[\min\left(a_{i,t}^{Bess}\bar{P}_i^{ev},\frac{\overline{SOC}_i-SOC_{i,t}}{\Delta t}\right)\right]^+ \tag{24}$$

$$P_{i,t}^{BAC,DC}=\left[\max\left(a_{i,t}^{Bess}\bar{P}_i^{ev},\frac{\underline{SOC}_i-SOC_{i,t}}{\Delta t}\right)\right]^- \tag{25}$$

The intended charging/discharging output of the battery $a_{i,t}^{Bess}\overline{P}_i^{ev}$ may cause the SOC to violate its capacity limits when $SOC_{i,t-1}+a_{i,t}^{Bess}\overline{P}_i^{ev}<\underline{SOC}_i$ in the discharging state or $SOC_{i,t-1}+a_{i,t}^{Bess}\overline{P}_i^{ev}>\overline{SOC}_i$ in the charging state. To address this issue, we compare $a_{i,t}^{Bess}\overline{P}_i^{ev}$ with the remaining current capacity. If $a_{i,t}^{Bess}\overline{P}_i^{ev}$ causes the SOC to violate its capacity limits, we directly replace it with the remaining SOC. Under this clip restriction technique, the behavior of battery is safe for maintaining a feasible SOC.

3) Reward

The reward of each agent i at time t is expressed as:

$$r_{i,t}=-C_{i,t}^{\text{cost}} \tag{26}$$

2) TD3

TD3 is a model-free and off-policy actor-critic algorithm that leverages DNNs to learn policies in high-dimensional continuous state-action spaces [37]. It is an extension of the DDPG, which improves the training stability using twin critic networks and a drift correction mechanism, optimizing the overall performance through appropriate parameter updates.

DDPG is a reinforcement learning algorithm designed to solve MDP problems with continuous action spaces [38]. The chief purpose of the DDPG is to parameterize the policy and value functions by approximating them using NNs. The critic network is represented by the Q-function, of which the update formula is expressed as:

$$Q(s_t,a_t)\leftarrow Q(s_t,a_t)+\alpha(r_{t+1}+\gamma\max Q(s_{t+1},a_{t+1})-Q(s_t,a_t)) \tag{27}$$

The update formula for the value function network is expressed as:

$$L_{\text{critic}}=\frac{1}{|B|}\sum_{i=1}^{B}(Q_\theta(s_t,a_t)-W^t)^2 \tag{28}$$

The parameters of the policy network ϕ are updated as follows. First, we compute the policy gradient, and then we update the policy network using gradient ascent.

$$\nabla_\phi J\approx\frac{1}{|B|}\sum_{i=1}^{B}(\nabla_a Q(s,a)|_{s=o_i,a=\pi_{\phi_i}(s)}\nabla_\phi\pi_\phi(s)|_{s=o_i}) \tag{29}$$

$$W_t=r_t+\gamma Q'(s_{t+1},a_{t+1}) \tag{30}$$

However, the DDPG often suffers from Q-value overestimation, which can result in suboptimal decisions and unstable performance in real-world applications. In TD3, the critic consists of two Q-networks ($Q_{\theta 1}$ and $Q_{\theta 2}$) and their target networks ($Q'_{\theta 1}$ and $Q'_{\theta 2}$), while the actor is formed by a deterministic policy network π_ϕ and its target network $\pi_{\phi'}$. The TD3 addresses this issue by using twin critic networks to evaluate the performance of the current policy and one of the critic networks is taken as the target for computing action gradients, thereby reducing the estimation error. In TD3, the target values are updated using two critic networks:

$$W_t=r_t+\gamma\min_{i=1,2}Q'_{\theta i}(s_{t+1},a_{t+1}) \tag{31}$$

3) MATD3 Algorithm

Based on the proposed framework, the coupled E/C market involving MG prosumers is modeled as a multi-agent system. To facilitate convergence to a stable policy while safeguarding the privacy of agents, we employ a CTDE approach. The centralized Q-value function is updated by:

$$\mathcal{L}(\theta_i)=\mathbb{E}_{o_t,a_t,r_t,o_{t+1}}((Q_{\theta_i}(o_{1,t},\dots,o_{N,t},a_{1,t},\dots,a_{N,t})-y_i^t)^2) \tag{32}$$

The target function is updated by:

$$W_i=r_{i,t}+\gamma\min_{j\in\{1,2\}}Q_{\theta'_{ij}}(o_{1,t},\dots,o_{N,t},\tilde{a}_{1,t+1},\dots,\tilde{a}_{N,t+1}) \tag{33}$$

The action function is updated by:

$$\tilde{a}_{i,t+1}=\pi_{\phi'_i}(o_{i,t})+\tilde{\epsilon}_i\quad \tilde{\epsilon}_i\sim clip(\mathcal{N}(0,\tilde{\sigma}_i^2),-c_i,c_i) \tag{34}$$

The gradient of policy network can be written as:

$$\nabla_{\phi_i}J\approx\mathbb{E}_{o_t,a_t}[\nabla_{a_i}Q_{\theta_i}(o_{1,t},\dots,o_{N,t},a_{1,t},\dots,a_{N,t})|_{o_i=o_{i,t},a_i=\pi_{\phi_i}(o_{i,t})}\cdot \nabla_{\phi_i}\pi_{\phi_i}(o_i)|_{o_i=o_{i,t}}] \tag{35}$$

4) Training and Implementation of Two-layer Data-driven Approach

Figure 3 illustrates the information flow of the proposed two-layer data-driven approach.

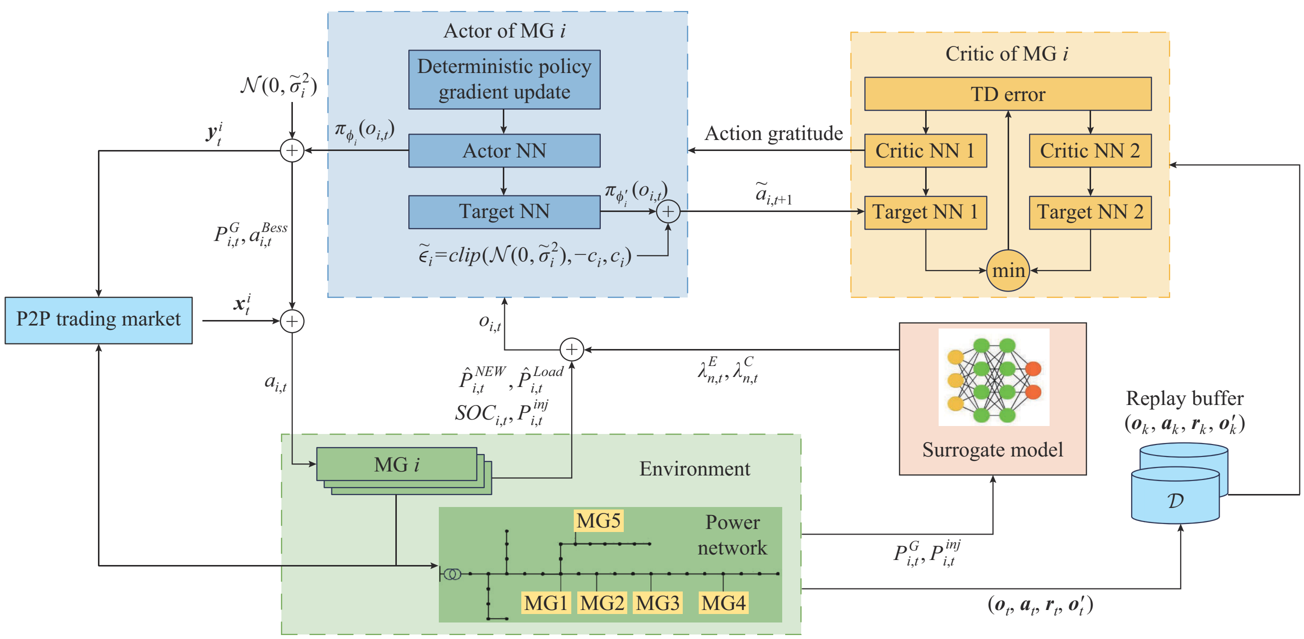

Fig. 3. Information flow of proposed two-layer data-driven approach.

First, the parameters $\omega_{jk}^{(l)}$ and $b_k^{(l)}$ of DNN-driven surrogate model are initialized, and then they are updated via (21). The reinforcement learning parameters and the replay buffer of the NN are also initialized. For each MG, it receives its observation $o_{i,t}$, including price signals $\lambda_{n,t}^E$ and $\lambda_{n,t}^C$ provided by the surrogate model, and then selects generator actions $P_{i,t}^G$, battery actions $a_{i,t}^{Bess}$, and P2P energy trading actions $\boldsymbol{y}_t^i$ based on its observation and policy π_{ϕ_i}. In the P2P energy trading stage, MG i must negotiate with other MGs in the environment to obtain the cleared energy transaction $\boldsymbol{x}_t^i$. Then, $\boldsymbol{x}_t^i$ is used to operate MG i in the energy conversion stage. MG i receives the reward r_i^t, and the next state observation $o_{i,t+1}$ is calculated based on the surrogate model. Finally, all observations, actions, rewards, and transitions to the next stage of each MG $(\boldsymbol{o}_t,\boldsymbol{a}_t,\boldsymbol{r}_t,\boldsymbol{o}_t')$ are stored in the replay buffer $\mathcal{D}$. For centralized training, each MG will randomly sample a small batch of data $(\boldsymbol{o}_k,\boldsymbol{a}_k,\boldsymbol{r}_k,\boldsymbol{o}_k')$ from $\mathcal{D}$ for parameter updates. The parameters of the critic network θ_i are updated by minimizing the loss function via (32)-(34), and the actor network parameter is updated using the policy gradient via (35). The target network of MG i is then updated using the following formulas:

$$\theta'_{i1} \leftarrow \tau\theta_{i1} + (1-\tau)\theta'_{i1} \tag{36}$$

$$\theta'_{i2} \leftarrow \tau\theta_{i2} + (1-\tau)\theta'_{i2} \tag{37}$$

$$\phi'_i \leftarrow \tau\phi_i + (1-\tau)\phi'_i \tag{38}$$

Each episode contains T time steps, and the training process is repeated M times to ensure the algorithm convergence.

V. Experiment and Simulation Results

A. System Parameterization and Experimental Setup

To demonstrate the effectiveness of the proposed P2P energy trading scheme with coupled E/C market and validate its performance, simulations are conducted on an IEEE 33-bus distribution system. The test distribution system consists of five MG prosumers connected to nodes 6, 8, 11, 15, and 28. The DER composition of each MG is detailed in Table II, and the key input data for the distribution system, i.e., forecasts of typical load, PV output, and wind turbine (WT) output are presented in Fig. 4. The study is based on multiple test days, with averages calculated using well-trained data. According to the parameters presented in Table III, the decision space for internal resource dispatch within the MGs is defined as follows. The decision space for batteries is set to be $[-100, 100]$kW, whereas that for distributed diesel generators is set to be $[0, 400]$kW. The external carbon emission price λ_t^{CT}, which reflects government penalties for carbon emissions, is assumed to vary with carbon futures prices. We set λ_t^{CT} at 90 \$/tCO$_2$ based on relevant data from the European carbon market. The datasets used for training and testing the surrogate model include randomly generated net load, output of distributed diesel generator, and the corresponding PDC-DLMP, all of which are calculated using the C-OPF model in the IEEE 33-bus distribution system. The clearing period is assumed to be 1 hour during a 24-hour cycle [10]. An analysis of a larger-scale IEEE 123-bus system [39] is presented in Supplementary Material A.

TABLE II
DER Composition of Each MG

MG	Location node	DER composition
MG 1	6	PV, load type 1, battery, distributed diesel generator
MG 2	8	WT, load type 1, battery, distributed diesel generator
MG 3	11	PV, load type 2, battery, distributed diesel generator
MG 4	15	WT, load type 1, battery, distributed diesel generator
MG 5	28	PV, load type 1, battery, distributed diesel generator

The MATD3 algorithm is trained for 3000 episodes. The batch size for each time step is 128. The discount factor is 0.97. The target smoothing coefficient is set to be 0.005. The target update frequency is 2.

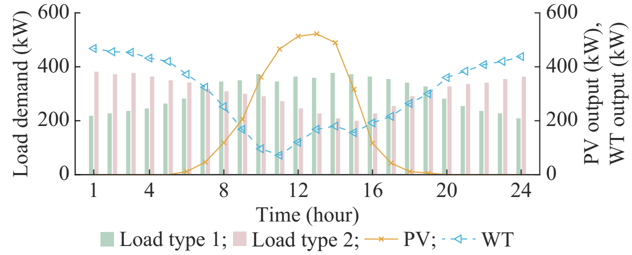

Fig. 4. Forecasts of typical loads, PV output, and WT output.

TABLE III
System Parameter Settings

DER	Parameter	Value
Distributed diesel generator	a_0 (\$/kWh^2)	0.012
	b_0 (\$/kWh)	0.02
	e^G (tCO$_2$/MWh)	0.9
Battery	SOC (kWh)	30-270
	$\dot{P}_i^{ev}$ (kW)	100

The size of experience pool is 2×10^5. The learning rate for both the actor and critic NNs is 0.005.

B. Case Settings

To demonstrate the superiority of the proposed P2P energy trading scheme with coupled E/C market based on PDC-DLMP, three comparative cases are established.

Case 1: the coupled E/C market is excluded, and energy trading is limited to P2P trading between MGs.

Case 2: the coupled E/C market is incorporated with PDC-DLMP signals but energy trading is limited to the P2G market for MGs.

Case 3 (this study): the coupled E/C market is incorporated with PDC-DLMP signals while considering both P2P and P2G energy trading among the MGs.

From the average reward in Case 3 shown in Fig. 5, it is evident that after 3000 training episodes, the proposed two-layer data-driven approach achieves stable training with good convergence.

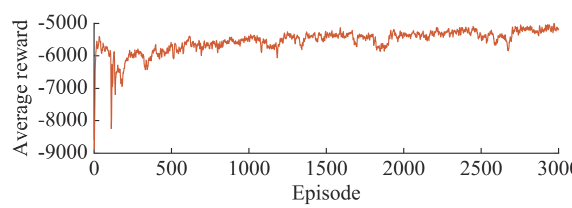

Fig. 5. Average reward in Case 3.

Table IV presents the key performance indicators in three cases. The total carbon emissions are calculated by multiplying the output of distributed diesel generators by their respective carbon emission intensities. This is based on the assumption that the carbon emission intensity of the root bus and the initial carbon emission intensity of the batteries are set to be zero, meaning that all carbon emissions in the distribution system are attributed to distributed diesel generators.

TABLE IV
COMPARISON OF KEY PERFORMANCE INDICATORS IN THREE CASES

Case	Output of distributed diesel generator (MWh)	Total carbon emission (t)	Total cost ($)
Case 1	16.785	15.106	5113.83
Case 2	15.052	13.546	4342.14
Case 3	14.413	12.971	4218.51

From Table IV, it is evident that in Case 1, the MGs predominantly rely on high-carbon-emission distributed diesel generators to meet their energy needs. In this case, the total output of distributed diesel generators reaches 16.785 MWh, resulting in 15.106 t of carbon emissions. By contrast, Cases 2 and 3, which incorporate a coupled E/C market, demonstrate a significant reduction in carbon emissions. This reduction is due to the clearing strategy in coupled E/C market, which imposes additional charges on carbon emissions. In addition, both cases experience a marked decrease in total costs, which is attributed to the PDC-DLMP signals based on the CEF theory, which allocates carbon emission costs between the generation and demand sides. Furthermore, Case 3 offers additional trading flexibility to reduce costs. This results in a win-win scenario that balances economic benefits with environmental sustainability.

C. Operational Effects of Considering Coupled E/C Market with PDC-DLMP Signals on MG Prosumers

By comparing the simulation results of Cases 1 and 3, we can analyze the effects of the coupled E/C market on MG operations. In Fig. 6, we present the output of distributed diesel generator and nodal carbon marginal prices across different cases. Table V summarizes the outputs and operational costs of the distributed diesel generators for the MGs in each case.

From the data in Fig. 6, we can observe that for MGs 1, 3, and 5, which utilize PV for renewable energy generation, the output of distributed diesel generators is higher at night and lower during the daytime. By contrast, MGs 2 and 4, which rely on WT for renewable energy generation, exhibit higher power generation during the daytime and lower at night. When incorporating the coupled E/C market, we find that nearly all MGs reduce their reliance on distributed diesel generators during peak hours, while increasing the share of surplus renewable energy purchased from the P2P market. This shift demonstrates the efficiency of the coupled E/C market in optimizing energy usage and promoting low-carbon operations.

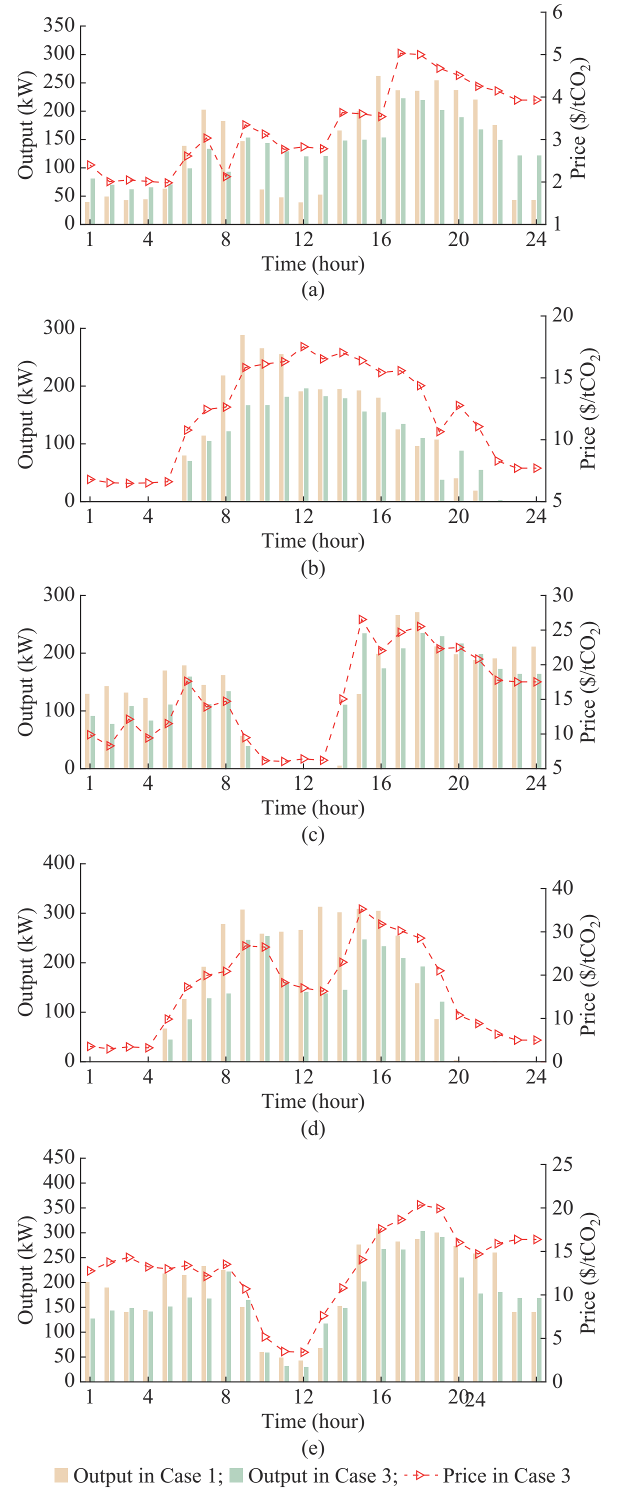

Fig. 6. Output of distributed diesel generator and nodal carbon emission prices in Cases 1 and 3. (a) MG 1. (b) MG 2. (c) MG 3. (d) MG 4. (e) MG 5.

TABLE V
OUTPUTS AND OPERATIONAL COSTS OF DISTRIBUTED DIESEL GENERATORS FOR MGS

Case	Output (kW)					Operational cost ($)				
	MG 1	MG 2	MG 3	MG 4	MG 5	MG 1	MG 2	MG 3	MG 4	MG 5
1	3145.98	2606.69	3069.17	3145.98	2606.69	1049.01	858.14	809.73	1049.01	858.14
2	2563.50	2191.31	2109.43	2563.50	2191.31	647.32	494.97	490.36	647.32	494.97
3	3111.19	2700.97	2860.99	3111.19	2700.97	1112.80	996.01	994.75	1112.80	996.01

Simultaneously, most MGs increase their generation during off-peak hours for distributed diesel generators, taking advantage of the lower DLMPs and nodal carbon emission marginal prices during these periods. In the P2P market, the high demand for electricity enables MGs to profitably trade low-cost electricity. These generation decisions not only enhance their economic benefits but also shift power generation away from peak hours, thereby stabilizing the operation of distributed diesel generators across the MGs.

The nodal carbon emission marginal price varies in tandem with the output of distributed diesel generators, generally increasing with the quantity of generation. Specifically, the nodal carbon emission marginal prices for MGs 1-4 tend to increase over time. This is because in a distribution network, the power flow and CEF typically diverge from the initial node toward the terminal nodes. As (2) shows, the carbon emission intensity of node is often greater than or equal to that of line flowing into the nodes. Consequently, carbon emissions accumulate more significantly at terminal nodes, resulting in higher carbon emission prices. As the nodal carbon emission marginal price increases, the reduction in power generation by the MGs increases from 76 to 1005 kW. MG operators facing higher nodal carbon emission marginal prices, such as MGs 4 and 5, tend to reduce the outputs of distributed diesel generator to lower their operating costs. By contrast, MG operators experiencing lower nodal carbon emission marginal prices, such as MG 1, are more likely to increase their power generation during low-demand periods and sell low-carbon and low-cost electricity in the P2P market, thus improving their economic benefits.

Figure 7 illustrates the energy trading results for MG 1 in the P2P market in Cases 1 and 3. The P2P energy trading data are predominantly negative, indicating that MG 1 frequently sells electricity via P2P energy trading over many time intervals, as driven by its lower DLMP. In addition, the P2P energy trading in Case 3 involves more transactions than that in Case 1. This is because with the introduction of the coupled E/C market and the application of carbon emission marginal prices, purchasing electricity from the P2P market becomes more cost-effective.

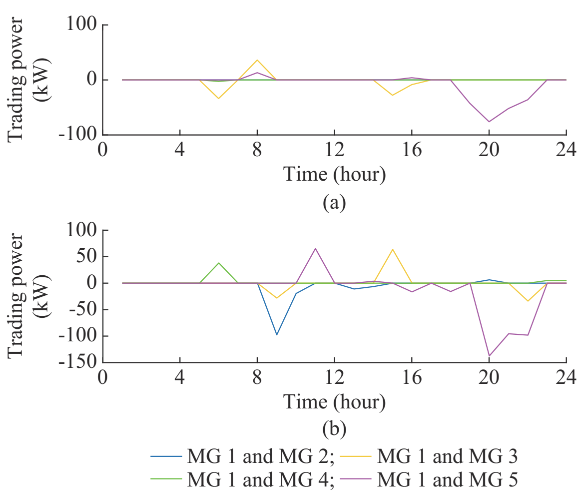

Fig. 7. Energy trading results for MG 1 in P2P market in Cases 1 and 3. (a) Case 1. (b) Case 3.

D. Operational Effects of P2P Energy Trading on MG Prosumers

A comparison of the simulation results in Cases 2 and 3 enables us to investigate the effects of the P2P energy trading on the operation of each MG. The output of distributed diesel generators in each MG as well as the nodal carbon emission marginal prices in both cases are presented in Supplementary Material B.

Figure 8 illustrates the battery operation status of MG3 in Cases 2 and 3. During the peak hours of renewable energy generation (10:00-13:00), the batteries in both cases choose to charge, allowing them to store energy for later use and supply electricity during the next high-demand period. In Case 2, the battery also charges during the low DLMP period (01:00-04:00), because this case relies solely on the P2G market for energy trading, making it more susceptible to the purchasing and selling prices set by the DSO.

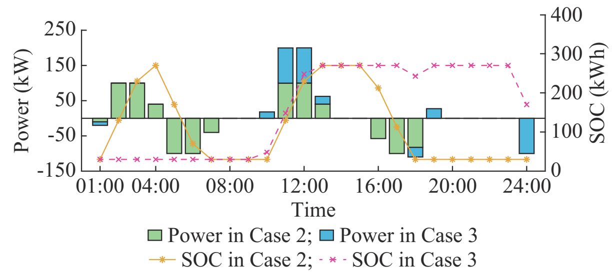

Fig. 8. Comparison of battery operation status of MG3 in Cases 2 and 3.

With the introduction of the P2P market in Case 3, the operational statuses of the batteries change. During periods of power shortage, MG3 tends to prefer trading electricity through the more affordable P2P market, thus reducing its reliance on batteries for storage.

In summary, without the P2P market, each MG can purchase electricity only from the P2G market or use battery storage to store the excess renewable energy generated during peak production periods. However, with the introduction of P2P energy trading, the purchasing price in the P2P market is typically lower than that in the P2G market, whereas the selling price is higher. Consequently, each MG is more inclined to trade its surplus renewable energy in the P2P market. Thus, the P2P market reduces and increases power generation at nodes with high and low carbon emission marginal prices, respectively, thereby implicitly promoting low-carbon operations.

VI. Conclusion

This study establishes a novel P2P energy trading scheme with coupled E/C market for prosumers. It leverages PDC-DLMP signals to guide low-carbon operations. The MATD3 algorithm, embedded with a surrogate model, constitutes a two-layer data-driven approach and demonstrates excellent performance within the framework. Numerical simulations show that the coupled E/C market not only reduces total carbon emissions within the distribution network but also significantly lowers the operational costs for MG prosumers. The introduction of P2P energy trading transfers power generation demands away from nodes with high carbon emissions, promoting the use of surplus renewable energy and further reducing operational costs for MG prosumers.

In future studies, we will focus on developing more effective P2P matching methods and pricing mechanisms to prioritize low-carbon energy trading in the novel P2P energy trading market with coupled E/C market. We also aim to investigate the training of more reliable surrogate models at the distribution network level, with further exploration of the deployment of these models onto edge devices for real-time execution. Finally, we plan to investigate the coordination of large-scale P2P trading and assess the scalability of DRL algorithms in these contexts.

References

[1] Department of Trade and Industry. (2003, Feb.). Energy White Paper: our energy future – creating a low carbon economy. [Online]. Available: https://m.360docs.net/doc/c318256787.html/files/file10719.pdf

[2] T. Zhou, C. Kang, X. Chen *et al*., "Evaluating low-carbon effects of demand response from smart distribution grid," in *Proceedings of 2012 3rd IEEE PES Innovative Smart Grid Technologies Europe* (*ISGT Europe*), Berlin, Germany, Oct. 2012, pp. 1-6.

[3] W. Hua, J. Jiang, H. Sun *et al*., "A blockchain based peer-to-peer trading framework integrating energy and carbon markets," *Applied Energy*, vol. 279, p. 115539, Dec. 2020.

[4] J. Wang and N. Liu, "Distributed optimal scheduling for multi-prosumers with CHP and heat storage based on ADMM," in *Proceedings of IECON 2017 – 43rd Annual Conference of the IEEE Industrial Electronics Society*, Beijing, China, Oct. 2017, pp. 2455-2460.

[5] D. Qiu, J. Xue, T. Zhang *et al*., "Federated reinforcement learning for smart building joint peer-to-peer energy and carbon allowance trading," *Applied Energy*, vol. 333, p. 120526, Mar. 2023.

[6] H. He, Z. Luo, Q. Wang *et al*., "Joint operation mechanism of distributed photovoltaic power generation market and carbon market based on cross-chain trading technology," *IEEE Access*, vol. 8, pp. 66116-66130, Apr. 2020.

[7] Y. Cheng, N. Zhang, B. Zhang *et al*., "Low-carbon operation of multiple energy systems based on energy-carbon integrated prices," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1307-1318, Mar. 2020.

[8] Y. Wang, J. Qiu, Y. Tao *et al*., "Carbon-oriented operational planning in coupled electricity and emission trading markets," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3145-3157, Jul. 2020.

[9] L. Sang, Y. Xu, and H. Sun, "Encoding carbon emission flow in energy management: a compact constraint learning approach," *IEEE Transactions on Sustainable Energy*, vol. 15, no. 1, pp. 123-135, Jan. 2024.

[10] Z. Lu, L. Bai, J. Wang *et al*., "Peer-to-peer joint electricity and carbon trading based on carbon-aware distribution locational marginal pricing," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 835-852, Jan. 2023.

[11] C. Kang, T. Zhou, Q. Chen *et al*., "Carbon emission flow from generation to demand: a network-based model," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2386-2394, Sept. 2015.

[12] J. Qiu, Y. Tao, S. Lai *et al*., "Pricing strategy of cold ironing services for all-electric ships based on carbon integrated electricity price," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 3, pp. 1553-1565, Jul. 2022.

[13] H. Sheng, C. Wang, X. Dong *et al*., "Incorporating P2P trading into DSO's decision-making: a DSO-prosumers cooperated scheduling framework for transactive distribution system," *IEEE Transactions on Power Systems*, vol. 38, no. 3, pp. 2362-2375, May 2023.

[14] M. Khorasany, A. Paudel, R. Razzaghi *et al*., "A new method for peer matching and negotiation of prosumers in peer-to-peer energy markets," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2472-2483, May 2021.

[15] M. Yan, M. Shahidehpour, A. Paaso *et al*., "Distribution network-constrained optimization of peer-to-peer transactive energy trading among multi-microgrids," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1033-1047, Mar. 2021.

[16] X. Yan, C. Gao, H. Ming *et al*., "Optimal scheduling strategy and benefit allocation of multiple virtual power plants based on general Nash bargaining theory," *International Journal of Electrical Power & Energy Systems*, vol. 152, p. 109218, Oct. 2023.

[17] Z. Li and T. Ma, "Peer-to-peer electricity trading in grid-connected residential communities with household distributed photovoltaic," *Applied Energy*, vol. 278, p. 115670, Nov. 2020.

[18] A. Lüth, J. M. Zepter, P. C. del Granado *et al*., "Local electricity market designs for peer-to-peer trading: the role of battery flexibility," *Applied Energy*, vol. 229, pp. 1233-1243, Nov. 2018.

[19] L. Thomas, Y. Zhou, C. Long *et al*., "A general form of smart contract for decentralized energy systems management," *Nature Energy*, vol. 4, no. 2, pp. 140-149, Jan. 2019.

[20] C. A. Floudas and X. Lin, "Mixed integer linear programming in process scheduling: modeling, algorithms, and applications," *Annals of Operations Research*, vol. 139, no. 1, pp. 131-162, Oct. 2005.

[21] S. Boyd, N. Parikh, E. Chu *et al*., "Distributed optimization and statistical learning *via* the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1-122, Jan. 2011

[22] C. Samende, J. Cao, and Z. Fan, "Multi-agent deep deterministic policy gradient algorithm for peer-to-peer energy trading considering distribution network constraints," *Applied Energy*, vol. 317, p. 119123, Jul. 2022.

[23] X. Liu, C. Zhang, S. Li *et al*., "Data-driven VVC for unbalanced networks with tuning-friendly deep reinforcement learning," in *Proceedings of 2024 IEEE PES General Meeting*, Seattle, USA, Jul. 2024, pp. 1-5.

[24] T. Chen, S. Bu, X. Liu *et al*., "Peer-to-peer energy trading and energy conversion in interconnected multi-energy microgrids using multi-agent deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 715-727, Jan. 2022.

[25] T. Xu, T. Chen, C. Gao *et al*., "Intelligent home energy management strategy with internal pricing mechanism based on multiagent artificial intelligence-of-things," *IEEE Systems Journal*, vol. 17, no. 4, pp. 6045-6056, Dec. 2023.

[26] Y. Ye, Y. Tang, H. Wang *et al*., "A scalable privacy-preserving multi-agent deep reinforcement learning approach for large-scale peer-to-peer transactive energy trading," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5185-5200, Nov. 2021.

[27] J. Wang, D. K. Mishra, L. Li *et al*., "Demand side management and peer-to-peer energy trading for industrial users using two-level multi-agent reinforcement learning," *IEEE Transactions on Energy Markets, Policy and Regulation*, vol. 1, no. 1, pp. 23-36, Mar. 2023.

[28] T. Chen and W. Su, "Indirect customer-to-customer energy trading with reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4338-4348, Jul. 2019.

[29] X. Liu, S. Li, and J. Zhu, "Optimal coordination for multiple network-constrained VPPs via multi-agent deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 14, no. 4, pp. 3016-3031, Jul. 2023.

[30] D. Cao, J. Zhao, W. Hu *et al*., "Data-driven multi-agent deep reinforcement learning for distribution system decentralized voltage control with high penetration of PVs," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4137-4150, Sept. 2021.

[31] X. Shi, Y. Xu, G. Chen *et al*., "An augmented Lagrangian-based safe reinforcement learning algorithm for carbon-oriented optimal scheduling of EV aggregators," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 795-809, Jan. 2024.

[32] X. Chen, A. Sun, W. Shi *et al*. (2023, Aug.). Carbon-aware optimal power flow. [Online]. Available: https://arxiv.org/abs/2308.03240

[33] B. Li, Y. Song, and Z. Hu, "Carbon flow tracing method for assessment of demand side carbon emissions obligation," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 1100-1107, Oct. 2013.

[34] L. Bai, J. Wang, C. Wang *et al*., "Distribution locational marginal pricing (DLMP) for congestion management and voltage support," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4061-4073, Jul. 2018.

[35] A. Papavasiliou, "Analysis of distribution locational marginal prices," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4872-4882, Sept. 2018.

[36] J. Kim and Y. Dvorkin, "A P2P-dominant distribution system architecture," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2716-2725, Jul. 2020.

[37] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research*, Stockholm, Sweden, Jul. 2018, pp. 1587-1596.

[38] X. Meng, Y. Jia, C. Ren *et al*., "DDPG-based backstepping controller for DC solid state transformer in DC microgrid with constant power loads," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4269-4283, Nov. 2022.

[39] J. Xu, H. Gao, R. Wang *et al*., "Real-time operation optimization in active distribution networks based on multi-agent deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 3, pp. 886-899, May 2024.

Xingyu Liu received the B.E. degree in electrical engineering from Nanjing Institute of Technology, Nanjing, China, in 2022. He is currently pursuing the M.Sc. degree in Nanjing Normal University, Nanjing, China. His research interests include operation optimization of distribution network, deep reinforcement learning, and energy market.

Yunting Yao received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2022. She is currently a Lecturer at Nanjing Normal University, Nanjing, China. From January 2023 to August 2023, she was a Research Associate with the Cardiff University, Cardiff, UK. Her research interests include multi-energy trading strategy and market mechanism, and analysis on uncertainty of renewable energy output.

Tianran Li received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2011. He is currently an Associate Professor with the School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China. His research interests include electricity market, and integrated multi-energy system modelling and management.

Yening Lai received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2006. He is currently a Professor of Engineering with NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing, China. His research interests include safety analysis and control of power system and multi-energy system.

Qi Wang received the Ph.D. degrees in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2008. From September 2010 to January 2013, she was a Postdoctoral with Hohai University, Nanjing, China. Since 1998, she has been with the School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China, where she is currently a Professor. Her current research interests include optimization of power system, renewable energy generation, and application of power electronics in power system.

Zhenya Ji received the B.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2009 and 2018, respectively. From 2012 to 2013, she was a Visiting Student with Rheinisch-Westfälische Technische Hochschule, Aachen, Germany. She is currently an Associate Professor with Nanjing Normal University, Nanjing, China. Her research interests include demand-side management and integrated energy system.